



# 論文集

SAS、SASを構成するプロダクト群は、SAS Institute Inc.の登録商標です。  
その他、本論文集に記載されている会社名、製品名は、一般にそれぞれ各社の商標または登録商標です。  
本論文集の一部または全部を無断転載することは、著作権法上の例外を除き、禁止されています。  
本論文集の内容を実際に運用した結果の影響については、責任を負いかねます。

# 目次

## CDISC 企画セッション

- 医薬品製造販売承認申請時の CDISC 標準のバリデーションにおける考察** ..... 3  
高浪 洋平(武田薬品工業株式会社)  
山崎 文寛、山田 佳代子
- Multiple Imputation を実施した際の電子データ提出方法の検討** ..... 4  
大内 喜海(協和発酵キリン株式会社、CJUG-ADaM)  
吉崎 正浩(小野薬品工業株式会社、CJUG-ADaM)、浅見 由美子(第一三共株式会社、CJUG-ADaM)
- Therapeutic Area Data Standards User Guide for Breast Cancer の  
解析データ使用報告** ..... 5  
坂上 拓(株式会社中外臨床研究センター)
- 日米同時申請する際の電子データ提出に関連する問題点と解決方法の検討** ..... 6  
鈴木 正人(MSD株式会社)
- 医薬品承認申請時の電子データ提出開始に向けて** ..... 7  
安藤 友紀(独立行政法人医薬品医療機器総合機構)

## オープンデータ

- 家計調査に基づく収入・支出の統計解析** ..... 11  
武田 凌波(東京理科大学大学院)

## 金融・経済・経営関連

- 国勢調査マイクロデータから全自動で構築した SAS データベースに基づく  
新世帯類型の設計と開発** ..... 15  
安井 浩子(公益財団法人 統計情報研究開発センター)  
中川 雅義(公益財団法人 統計情報研究開発センター)、周防 節雄(兵庫県立大学)
- 新世帯類型による国勢調査マイクロデータの分析結果の概要** ..... 23  
新井 郁子(公益財団法人 統計情報研究開発センター)  
米澤 香、中川 雅義
- 4回の学生調査に基づく学生の文化・芸術に対する意識の現状と変化** ..... 33  
有馬 昌宏(兵庫県立大学)
- 全国消費実態調査の匿名データから見た食料品目別消費傾向による  
食生活の変化について** ..... 43  
池田 瑞穂(京都大学大学院)
- 匿名データを用いた夫婦共働き世帯における  
子あり世帯と夫婦のみ世帯の比較:横断研究** ..... 52  
井上 貴博(ノバルティス ファーマ株式会社)  
矢田 徹(イーピーエス株式会社)
- 匿名データを用いたコメ離れと食事づくりの簡素化についての検証** ..... 62  
中松 建(株式会社タクミンフォメーションテクノロジー)

## SAS システム

DS2言語によるデータハンドリングについて	73
森岡 裕	
デフォルトと複数格付の同時推計によるロジットモデル構築手法の提案	85
木村 和央(株式会社金融工学研究所)	
ガンマ過程を用いた比例ハザードモデルによる解析	100
矢田 真城(東京理科大学大学院、エイツーヘルスケア株式会社) 浜田 知久馬(東京理科大学)	
落語協会の有カ一門の勢力分布をD3.jsで「分かりやすく」ビジュアル化する -JSONプロシジャを用いて-	110
坂部 裕美子(公益財団法人 統計情報研究開発センター)	

## SAS システム/その他関連分野

SASによる迷路の自動生成	113
松沢 享(株式会社 インターサイエンティフィックリサーチ)	

## 生物統計・医薬関連

Server版SASを基盤にした「生物統計解析システム」の紹介	117
栃澤 欣之(大塚製薬株式会社) 角元 慶二、佐々木 和典、菱垣 晴次	
SASで作成した統計解析帳票の柔軟なQC体制構築に向けた取り組みの事例紹介	122
仲川 慎太郎(中外製薬株式会社) 吉本 拓矢、吉田 寛輝、久力 洋、河田 祐一	
「区間打ち切りを考慮したノンパラメトリックな生存関数推定法の 推定性能を向上させる診断スケジュールの検討」	123
野村 怜史(東京理科大学) 寒水 孝司、浜田 知久馬	
PROC SGPLOTによるSwimmer Plot -見やすさを求めて-	124
福田 裕章(MSD株式会社) 吉田 秀幸(株式会社タクミンフォメーションテクノロジー)、韓 士栄(MSD株式会社)	
<b>企画セッション</b> 欠測のあるデータの解析のためのSASプログラムの紹介 ~データ発生・DIAマクロとプロシジャの進展~ <b>(1)セッションの概要</b>	125
日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4 欠測のあるデータの解析検討チーム 土居 正明(東レ株式会社)	
<b>企画セッション</b> 欠測のあるデータの解析のためのSASプログラムの紹介 ~データ発生・DIAマクロとプロシジャの進展~ <b>(2)シミュレーションデータ発生方法</b>	126
日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4 欠測のあるデータの解析検討チーム 横山 雄一(持田製薬株式会社) 大浦 智紀(日本イーライリリー株式会社)、土居 正明(東レ株式会社)	

<b>企画セッション</b> 欠測のあるデータの解析のためのSASプログラムの紹介～データ発生・DIAマクロとプロシジャの進展～ <b>(3) MNARの仮定の下での Selection Modelに基づく解析</b>	— 127
日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4 欠測のあるデータの解析検討チーム 黒田 晋吾(武田薬品工業株式会社)	
<b>企画セッション</b> 欠測のあるデータの解析のためのSASプログラムの紹介～データ発生・DIAマクロとプロシジャの進展～ <b>(4) Pattern Mixture ModelとMultiple Imputationに基づく解析1 (Delta adjustmentとTipping Point Analysis)</b>	— 128
日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4 欠測のあるデータの解析検討チーム 大浦 智紀(日本イーライリリー株式会社)	
<b>企画セッション</b> 欠測のあるデータの解析のためのSASプログラムの紹介～データ発生・DIAマクロとプロシジャの進展～ <b>(5) Pattern Mixture ModelとMultiple Imputationに基づく解析2 (Reference-based imputation)</b>	— 129
日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4 欠測のあるデータの解析検討チーム 藤原 正和(塩野義製薬株式会社)	
<b>企画セッション</b> 欠測のあるデータの解析のためのSASプログラムの紹介～データ発生・DIAマクロとプロシジャの進展～ <b>(6) MI プロシジャで実行可能な Pattern Mixture ModelとMultiple Imputationに基づく解析</b>	— 130
日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4 欠測のあるデータの解析検討チーム 渡邊 大丞(サノフィ株式会社)	
<b>企画セッション</b> 欠測のあるデータの解析のためのSASプログラムの紹介～データ発生・DIAマクロとプロシジャの進展～ <b>(7) セッションのまとめと質疑応答</b>	— 131
日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4 欠測のあるデータの解析検討チーム 大江 基貴(株式会社大塚製薬工場) 土居 正明(東レ株式会社)	
<b>SASでのNetwork Meta-Analysisの実施例</b> ～ 頻度論に基づくアプローチ ～	— 132
舟尾 暢男(武田薬品工業株式会社) 黒田 晋吾	
<b>ナニワデータサイエンス研究会</b> ～第1回ダイジェスト版～	— 154
高浪 洋平(ナニワデータサイエンス研究会) 北西 由武(塩野義製薬株式会社)、藤原 正和(塩野義製薬株式会社)、坂井 絵理(塩野義製薬株式会社)、 黒田 晋吾(武田薬品工業株式会社)、舟尾 暢男(武田薬品工業株式会社)	
<b>Sales Analyticsの可能性</b> — 医薬品営業を例として —	— 155
武藤 猛(MarkeTech Consulting)	
<b>JMP Clinicalを用いたセントラルモニタリングの経験</b> ～RBMを推進するために～	— 165
福升 悠一(エイソーヘルズケア株式会社) 近藤 秀宣、樋口 天裕、小幡 智憲、畑山 知慶、林 行和	
<b>SAS XML Mapperを用いたdefine.xmlのメタ情報の有効活用及び</b> <b>業務効率化の取り組み</b>	— 166
林 三男(大日本住友製薬株式会社) 山本 繁	
<b>申請時電子データのLengthにまつわる実務上の問題点と対処法</b>	— 167
片山 雅仁(塩野義製薬株式会社) 平野 勝也、鈴木 ゆら	

<b>欠測を含む順序カテゴリカル経時データの解析 -MI プロシジャの有用性-</b>	168
藤原 正和(塩野義製薬株式会社) 駒寄 弘(マルホ株式会社)	
<b>ADaM/TLF 経験者がSDTMを作成できるまでのサクセスストーリー</b>	169
小林 千鶴(塩野義製薬株式会社) 神谷 亜香里(塩野義製薬株式会社)、坂本 佳代子(株式会社エス・シー・イー)、勘場 大(塩野義製薬株式会社)、 江川 広明(塩野義製薬株式会社)、角谷 伸一(塩野義製薬株式会社)	
<b>欠測を含む順序カテゴリカル経時データの解析 -GEE プロシジャの有用性-</b>	170
駒寄 弘(マルホ株式会社) 藤原 正和(塩野義製薬株式会社)	
<b>Multi-Regional Clinical Trialの生存時間解析における 地域ごとのイベント発現数の予測</b>	171
淀 康秀(大日本住友製薬株式会社) 浜田 知久馬(東京理科大学)	
<b>用量設定試験におけるMCP-Modの3つの実行例 ～ 多重比較と用量反応関係のモデリング ～</b>	172
加藤 雅章(武田薬品工業株式会社) 清水 康平、佐々木 英磨、舟尾 暢男	
<b>指数分布に従うデータへの層別ログランク検定の検出力の検討</b>	193
小松 邦岳(臨床評価研究会)	
<b>PPK解析時の潜在的なデータ不備が推定値に与える影響の検討</b>	204
川崎 勝己(株式会社ヘル・メディカルソリューションズ) 碓 新里、新城 博子	
<b>ポアソン分布に従うと仮定された総ケース数が固定された条件の下での rate ratioの条件付き信頼区間</b>	205
伊藤 要二(アストラゼネカ株式会社)	
<b>JMPスクリプト言語 (JSL) を用いた前臨床における割付け手法のアドイン開発</b>	215
原茂 恵美子(株式会社タクミンフォメーションテクノロジー) 佐藤 拓哉、篠津 和夫、佐藤 耕一	
<b>Kaplan-Meier プロット・Forestプロット作成の応用: グラフ出力範囲内・範囲外への数値出力</b>	216
魚住 龍史(京都大学) 吉田 早織(日本化薬株式会社)、平井 隆幸(日本化薬株式会社)、浜田 知久馬(東京理科大学)	
<b>臨床試験におけるSG プロシジャを用いたGraphic Designの考察</b>	233
山崎 文寛(武田薬品工業株式会社) 清水 康平、高浪 洋平	
<b>SAS プロシジャを用いた生存時間データに対する例数設計の変革</b>	250
魚住 龍史(京都大学) 矢田 真城(エイツーヘルスケア株式会社)、浜田 知久馬(東京理科大学)	

<b>生存時間データに対するベイズ流例数設計</b> .....	268
矢田 真城(エイソーヘルズケア株式会社)	
魚住 龍史(京都大学)、浜田 知久馬(東京理科大学)	

## 生物統計・医薬関連 / SAS システム

<b>SASにおける文芸的プログラミングへの試み</b> .....	281
吹谷 芳博(株式会社 エスアールティ)	
<b>医療・健診データに対するSAS Visual Analytics/ Visual Statisticsの適用可能性の検討</b> .....	294
大津 洋(国立国際医療研究センター 臨床研究センター)	
大橋 靖雄(中央大学)	

## チュートリアル

<b>SASによる生存時間分布の予測「Death Noteの統計学」</b> .....	297
浜田 知久馬(東京理科大学)	
魚住 龍史(京都大学)	
<b>データハンドリングにおけるSQLプロシジャの利活用 —PROC SQL入門—</b> .....	298
古隅 弘樹(兵庫県立大学)	
<b>JMPによる第4回Let'sデータ分析の規定課題の解析</b> .....	313
高橋 行雄(BioStat研究所株式会社)	
<b>SASユーザー総会2016における「Let'sデータ分析 第4回マイクロデータ分析コンテスト」の 規定課題のSASプログラム解説</b> .....	330
周防 節雄(兵庫県立大学)	

## 統計理論

<b>局外トレンドのある回帰モデルの推測</b> .....	351
藤原 滉(県立広島大学)	
丸川 真友子、富田 哲治	
<b>ベイジアンモデルによる地域人口予測モデルの可能性について</b> .....	352
片桐 智志(ネイチャーインサイト株式会社)	
山下 論史	

## Let's データ分析

<b>参加カテゴリーB   ラドラー</b>	
自由課題レポート: 貧困・中流・富裕世帯における教育費の実態調査 .....	365
規定課題レポート .....	374
<b>参加カテゴリーA   チーム ベイ</b>	
自由課題レポート: 消費税率の引上げ及び軽減税率制度の導入は低所得者にどのような影響を与えるか .....	380
規定課題レポート .....	389

<b>参加カテゴリーC</b>	<b>3-girls from China</b>	
自由課題レポート:	シングルマザー世帯の家計の実態 ー片働き世帯と共働き世帯との比較からー	393
規定課題レポート		403
<b>参加カテゴリーB</b>	<b>田中 健一郎</b>	
自由課題レポート:	世代別及び経常収入別に見た官民の平均貯蓄率の差異に関する考察	409
規定課題レポート		419
<b>参加カテゴリーB</b>	<b>株式会社コネクティボ</b>	
自由課題レポート:	一戸建て・共同住居住まい世帯における傾向の比較	423
規定課題レポート		428
<b>参加カテゴリーC</b>	<b>三度の飯より分析</b>	
自由課題レポート:	収入と医療費の関係から見える「医療格差」に関して	430
規定課題レポート		434
<b>参加カテゴリーB</b>	<b>FEG Freshers</b>	
自由課題レポート:	業種で見る中小企業の収入 ～大企業との比較～	449
規定課題レポート		456
<b>参加カテゴリーB</b>	<b>筑波銀行営業企画部ダイレクトマーケティンググループ</b>	
自由課題レポート:	マイナス金利下におけるローン事業の発展性と疑似マイクロデータに見るローン利用者の行動特性	464
規定課題レポート		474



CDISC企画セッション

## 医薬品製造販売承認申請時のCDISC標準のバリデーションにおける考察

○高浪 洋平<sup>1</sup>、山崎 文寛<sup>1</sup>、山田 佳代子<sup>1</sup>

(<sup>1</sup>武田薬品工業株式会社)

2015年4月に「承認申請時の電子データ提出に関する実務的事項について」（実務的通知）及び「承認申請時の電子データ提出等に関する技術的ガイドについて」（技術的ガイド）が発出され、いよいよ本邦において2016年10月より医薬品製造販売承認申請時電子データ提出が開始される。さらに、2015年11月にPMDAより電子データ提出対象となるCDISC標準のSDTM, ADaM及びDefine.xmlについて「バリデーションルール一覧」が公開され、電子データ作成時の具体的な要件が定められた。一方、FDAでは既に電子データの受け付けを開始しており、SENDとSDTMについてバリデーションルールが公開されているが、グローバルに展開する製薬企業は、それら両方の要件を満たす電子データの準備が必要となる。本発表では、日本で医薬品製造販売承認申請を実施する申請者側の視点から、PMDAのバリデーションルールの概要を整理するとともに、実務上、電子データを準備するにあたって留意すべきポイントを報告する。

## Multiple Imputationを実施した際の電子データ提出方法の検討

○大内喜海<sup>1</sup>、吉崎正浩<sup>2</sup>、浅見由美子<sup>3</sup>

(<sup>1</sup>協和発酵キリン株式会社及びCJUG-ADaM、<sup>2</sup>小野薬品工業株式会社及びCJUG-ADaM、<sup>3</sup>第一三共株式会社及びCJUG-ADaM)

臨床試験は様々な理由により欠測値が発生することがあり、臨床試験の成績は欠測値を含むデータから評価される。欠測値は、ICH E9において「臨床試験において偏りを起こし得る代表的な原因」とされ、現在検討されているICH E9 (R1)のEstimandにも関連するため、臨床試験の成績を評価するうえで適切に取扱う必要がある。FDAの依頼によりNational Research Councilが2010年に公表した「The Prevention and Treatment of Missing Data in Clinical Trials」、EMAが2010年に公表した「Guideline on Missing Data in Confirmatory Clinical Trials」により、欠測値の予防や欠測値を含んだデータに対する解析手法について提言がなされたことをきっかけに、国内においても欠測値を含む臨床試験データの解析手法が活発に議論されるようになった。そのような状況下で、規制当局より平成26年6月20日付で「承認申請時の電子データ提出に関する基本的考え方について」、平成27年4月27日付で「承認申請時の電子データ提出に関する実務的事項について／技術的ガイドについて」が発出されたことにより、平成28年10月1日以降（経過措置期間あり）の承認申請品目は、CDISC標準（SDTM/ADaM）に基づくデータセット、Define-XML（Analysis results metadataを含む）、Data Guide（Reviewer's Guide）及びADaM/解析帳票作成プログラム等の提出が必要となった。

これまで、欠測値に対する解析手法に関して活発な議論がなされているものの、それらの電子データ提出方法に関する議論は活発になされていなかった。そこでCDISC、規制当局、FDAがそれぞれ公表している「ADaM Implementation Guide」、「承認申請時の電子データ提出に関する実務的事項について／技術的ガイドについて」、「STUDY DATA TECHNICAL CONFORMANCE GUIDE」等の関連資料と照らし合わせて考察したが、Multiple Imputation（以下、MI）を実施するうえで以下のような点で電子データ提出方法が明確ではないと考えられた。

- ・MIプロシジャがADaMのBDS型で実行できないため、実行前後でCDISC非準拠のデータセットが発生するが、それを電子データとして提出する必要性があるのかが明確ではない

- ・MIによる解析結果を得る過程で中間データセットが発生するが、それらも電子データとして提出する必要性があるのかが明確ではない

- ・上記2点のようにMIを実施することでCDISC非準拠のデータセットや中間データセットが発生するが、どのようにTraceabilityを残せばよいのかが明確ではない

本発表では上記の点を踏まえてMIを実施した際の電子データ提出方法を検討し、その検討結果を報告する。

## Therapeutic Area Data Standards User Guide for Breast Cancerの解析データ使用報告

○坂上 拓

(株式会社 中外臨床研究センター バイオメトリクス部 データサイエンスグループ)

近年、ADaM標準のバリエーションや、安全性解析に関わるADaMのサンプルが増えてきたことから、これらを元にADaMを作成し、安全性に係る解析帳票の開発を効率的に進める事が出来るようになりつつある。このため、安全性解析で作成するようなADaMに関しては、各社で共通の認識を得ることができるようになり、問題点の共有や情報交換もスムーズになってきている一方で、有効性の解析は、疾患領域特有の要件が多く、公開されているサンプルも少ないことから、ADaM標準をベースに各社各様の解析仕様やデータ構成で開発を進めているのが現状である。

CFASTが主導して開発している、TAUG (Therapeutic Area Data Standard User Guide)は、特性疾患領域の臨床試験に対する、CDISC標準の活用方法を多くの例と共に示した文書で、有効性解析を目的としたADaMの標準についても定義されていることから、このような問題を解決し、CDISC標準を更に活用する上で重要な役割を果たす文書となっている。

現在リリースされているTAUGの一つである、TAUG Breast Cancerには、腫瘍縮小効果判定結果と生存時間解析を実施するために、4つのADaM (ADSL, ADEVENT, ADTTE, ADRESP)のメタデータが提案されている。その中でも特に興味深いのが、Time-to-event情報が格納されているADTTEを作成するための中間データセットであるADEVENTで、本データセットを作成することにより、ADTTEが効率的に作成することができ、且つ、シンプルにトレーサビリティを保持することができるようになった。また、ADRESPは、腫瘍縮小効果判定結果の解析を目的として、非常にシンプルな構造で最良総合効果を格納している。

しかし当然ながら、TAUGで示されているメタデータや例示だけでは、実際に実施されるすべての解析を網羅することはできない。このため、TAUGに定義されているメタデータや例示を元に、各社の解析方針、ケースカードの収集項目やデザイン、SDTMのマッピング方針を考慮しながら応用していく必要がある。

実際の業務で使用した経験を元に、OS (Overall Survival)の解析の打ち切りデータの格納方法や、OSやPFS (Progression Free Survival)とTime-to-eventを算出するための起点日が異なる、DOR (Duration of Response)のADEVENTやADTTEへの応用について解説し、こういったチャレンジが議論を呼び、更なる標準化の発展につながることを期待している。

## 日米同時申請する際の電子データ提出に関連する問題点と解決方法の検討

○鈴木正人<sup>1</sup>

(<sup>1</sup>MSD株式会社)

FDAから発出されているConformance Guide等の規制要件とPMDAから発出されている各種通知と技術的ガイドの間に違いがあることから、国際共同試験であっても、FDA申請とは別にPMDA申請用にデータガイドや定義ファイルを追加・修正する必要があり、いわゆるDouble standardが発生している。また、legacy試験のデータを提出する場合、例えばSDTM IG3.1.1であっても、米国においては、猶予期間中はFDAと事前合意することでUp-versioningをすることなく電子データ提出をすることができる。一方、日本においてはSDTM IG3.1.2以上にUp-versioningすることが実務者通知で求められており、その際にPMDAに提出するUp-versioningしたSDTM及びADaMから作成した解析結果とSDTM IG3.1.1から米国本社が作成したCTDにおける解析結果が一貫したものになるようにTraceabilityを担保する必要がある。特にOncology領域においては、SDTM IG3.1.3からOncology domainが追加されたことから、生存時間解析用のADaMデータセット、例えばADTTEの作成に際し、米国本社が準備したSASプログラムを複製して再生することが簡単にできなくなった。また、中間解析で有効中止になった場合、データカットオフに関連する諸問題が発生する。10月1日以降の電子データ提出の準備に際し、国際共同試験を使って日米同時申請する場合の規制要件や承認申請プロセスの違いによる悩ましい問題点を紹介する。

## 医薬品承認申請時の電子データ提出開始に向けて

○安藤友紀

(医薬品医療機器総合機構)

医薬品医療機器総合機構（PMDA）では、承認申請データを一層活用した承認審査や相談の質の向上を図るため、次世代審査・相談体制の構築として、医薬品承認申請時に添付される臨床試験データ利用のための具体的検討を行ってきている。実際に、本年10月1日以降の承認申請品目より、医薬品承認申請時の臨床試験の電子データ提出が開始される。申請時電子データ提出に関しては、これまで、「承認申請時の電子データ提出に関する基本的考え方について」（平成26年6月20日薬食審査発0620第6号）、「承認申請時の電子データ提出に関する実務的事項について」（平成27年4月27日薬食審査発0427第1号）、技術的詳細に関する「承認申請時の電子データ提出に関する技術的ガイドについて」（平成27年4月27日薬機次発第0427001号）が発出され、その方針が示されてきている。さらに、電子データ提出において用いるべき各種標準とそのバージョンを示した「申請電子データ提出に際して利用可能な規格一覧」（平成27年7月30日）、CDISC標準に準拠した電子データに関する「バリデーションルール一覧」（平成27年11月18日）、PMDAにおける経験や業界からの問い合わせ等に基づきまとめられた「申請電子データに関するFAQ」が公表され、提出すべきデータの詳細が明らかにされてきている。

PMDAにおける提出された電子データの利用に関しては、平成25年よりパイロットとして4回にわたり、試行的に提供されたCDISC標準に準拠した臨床試験データ、及び臨床薬理試験データを用いた検討が行われてきた。また、平成27年5月からは、治験相談の新たな相談枠として、電子データの提出内容（仕様等、定義ファイルやデータセット作成用プログラムを含む）等に関して助言を行う申請電子データ提出確認相談が設置され、具体的な提出内容に関する相談に応じている。これらパイロット及び治験相談において経験した内容の一部は、これまで発出された通知、ガイド及びFAQに反映されてきており、さらに、本年10月1日のデータ提出開始に向けて、技術的ガイドの更新やFAQの追加等により、今後も情報が提供されていく予定である。

本発表では、PMDAにおける次世代審査・相談体制の構築状況、及びこれまでに実施されてきたパイロットや申請電子データ提出確認相談の経験を踏まえた、電子データ提出に際して留意すべき事項等の最新の状況について紹介する。

オープンデータ

## 家計調査に基づく収入・支出の統計解析

○武田 凌波, 寒水孝司, 浜田知久馬

(東京理科大学大学院工学研究科経営工学専攻)

総務省統計局は、国民生活における家計収支の実態を把握し、国の経済対策・社会政策の立案のための基礎資料を提供することを目的として、家計調査を行っている<sup>[1]</sup>。家計調査は毎年行われ、世帯や世帯主の年齢、職業などを調査し、経常収入や消費支出などと言った、収入・支出の詳細まで調査する。本研究では、経常収入と消費支出についての解析を行う。経常収入とは、家計の消費行動に大きな影響を与える周期性あるいは再現性のある収入のことを表し、消費支出とは、日常生活を過ごすのに必要な商品やサービスを購入して実際に支払った金額のことを表す。

現在、総務省統計局での発表では、平成2年から平成27年までの年齢階級や世帯人員数などと言った条件での収入・支出の結果が掲載されている。先行研究<sup>[2]</sup>では、経常収入・消費支出は年齢に依存すること、世帯人員が多いほど消費支出が高くなることが発表された。しかし、多変量解析のモデルは発表されていない。例えば、年齢が30代で世帯人員が4人のときの収入・支出の結果は不明である。また、世帯主の職業によって収入が異なるため、総務省統計局が発表されたものが、各家庭においての基準になることは難しいと考えられる。

そこで、年齢や世帯人員を含む13項目を利用して、経常収入・消費支出についての統計モデルを検討する。交絡の影響もあるため、複数の項目における収入・支出についての多変量解析を行う。外れ値が存在し、等分散性が成り立たないため、分位点回帰<sup>[4]</sup>を用いて統計モデルを作成することを目的とする。

本研究では、収入・支出の関係がどのようになっているかをそれぞれの分位点ごとに傾向を表し、格差がどれだけ広がるかどうか、ということを明らかにする。それに基づき、現在の総務省統計局での発表方法は適しているかどうかを述べる。

### 参考文献

- [1] 大橋靖雄, 浜田知久馬. 生存時間解析-SASによる生物統計. 東京大学出版会. 1995.
- [2] 総務省統計局. 日本の統計2016. 総務省. 2016.
- [3] 阿部修人. 近年の日本家計消費の動向——家計調査の結果を中心として——. 一橋大学機関リポジトリ. 2010.
- [4] Roger Koenker, Kevin F. Hallock. "Quantile Regression", *Journal of Economic Perspectives*, Volume 15 Number 4, Fall 2001. 143-156, 2001.



金融・経済・経営関連

# 国勢調査マイクロデータから全自動で構築した SAS データベースに基づく新世帯類型の設計と開発

安井浩子 (公財)統計情報研究開発センター研究開発部

中川雅義 (公財)統計情報研究開発センター研究開発部

周防節雄 兵庫県立大学・名誉教授

Development and Design of the New Household Category Invented from the  
SAS Database Full-automatically Constructed from the Microdata of the Recent Japan  
Population Census.

Hiroko Yasui, Masayoshi Nakagawa

Statistical Information Institute for Consulting and Analysis R&D Division

Setsuo Suoh

The University of Hyogo Professor Emeritus

## 要旨

国勢調査マイクロデータから SAS データベース構築後、世帯主と直系親族世帯員との続柄の構成で世帯を分類する「新世帯類型分類」を開発、実装した。本報告では、提供された国勢調査のマイクロデータ(CSV ファイル)から全自動で SAS データベースを構築する SAS プログラム、および、新世帯類型分類の開発プロセスとそれを国勢調査の SAS データベースに実装する SAS システムの概略を報告する。

キーワード: 国勢調査、世帯、続柄、新世帯類型分類、公的統計、マイクロデータ、SAS データベース

## 1. はじめに

国勢調査の結果として集計・公表される様々な統計は、国や地方公共団体の各種行政施策の基礎資料として用いられている。民間部門においても、「男女、年齢別人口」など個人属性に関する統計は、消費動機を分析し、製造・販売などの企業活動に広く活用されている極めて重要な情報である一方、個人レベルでも、消費には自分のためだけの消費の他に、家族のため、あるいは家族と一緒に行うための消費もあり、個人が属する世帯の構造も消費動機に少なからず影響を及ぼしている。

国は、世帯の構造別世帯数として、昭和 45 年国勢調査から「家族類型別世帯数」を公表している。我々は、この従来の家族類型が個人の消費に影響を及ぼす世帯の構造を知る分類として適しているのかをつぶさに検討した。その結果、現行の家族類型では、核家族世帯など世帯の構造別分布を大局的に把握するには適しているが、世帯内の最も若い夫婦と他の世帯員との関係によ

って世帯を分類しているので、社会通念上異なる世帯が同じ世帯区分に分類されていることになり、個人の消費に影響を及ぼす世帯の構造を知る分類としては必ずしも最適ではないかもしれないと考えた。

我々は、この点についていろいろ試行錯誤した結果、世帯主と直系親族世帯員との関係から世帯を分類するシステムを開発し、「新世帯類型分類」と名付けた。この新世帯類型分類は、世帯主（または世帯主夫婦）と他の世帯員との続柄に基づいて世帯を分類している<sup>1</sup>。

本報告では、新世帯類型分類に基づく分析結果を開示するのではなく、その新分類体系について、その設計過程、及び実際の国勢調査の調査票情報（マイクロデータ）から構築された SAS データベースへの実装作業の概要を報告するに留める。

## 2. 国勢調査の続柄

本プロジェクトで使用許可を得た国勢調査の調査票情報（マイクロデータ）は、大規模調査の 1980 年、1990 年、2000 年、2010 年調査の 4 回分である。1995 年の簡易調査から「父母」が「世帯主の父母」と「世帯主の配偶者の父母」に分割されたので、表1上側に示すように 1990 年以前と 2000 年以降の調査では続柄のコード値が異なっている。従って、この 4 回次分の調査データを分析する際には、表1下側に示すように 2000 年と 2010 年の「父母」の続柄を分割前の値に変換すると同時に、他の続柄のコードの再編成も行った。

表1 国勢調査の続柄一覧

1980年(昭和55年)		1990年(平成2年)		2000年(平成12年)		2010年(平成22年)	
符号	変更後の符号	符号	変更後の符号	符号	変更後の符号	符号	変更後の符号
01 世帯主又は代表者	01	01 世帯主又は代表者	01	01 世帯主または代表者	01	01 世帯主または代表者	01
02 世帯主の配偶者	02	02 世帯主の配偶者	02	02 世帯主の配偶者	02	02 世帯主の配偶者	02
03 子	03	03 子	03	03 子	03	03 子	03
04 子の配偶者	04	04 子の配偶者	04	04 子の配偶者	04	04 子の配偶者	04
05 孫	05	05 孫	05	05 世帯主の父母	06	05 世帯主の父母	06
06 父母	06	06 父母	06	06 世帯主の配偶者の父母	08	06 世帯主の配偶者の父母	06
07 祖父母	07	07 祖父母	07	07 孫	05	07 孫	05
08 兄弟・姉妹	08	08 兄弟姉妹	08	08 祖父母	07	08 祖父母	07
09 他の親族	09	09 他の親族	09	09 兄弟姉妹	08	09 兄弟姉妹	08
10 住み込みの雇人(家事)	10	10 家事使用人	10	10 他の親族	09	10 他の親族	09
11 住み込みの雇人(営業)	10	11 営業使用人	10	11 住み込みの雇人	10	11 住み込みの雇人	10
12 その他	11	12 その他	11	12 その他	11	12 その他	11
** 対象外	12	VV 不詳	12	VV 不詳(聞き取り調査世帯)	12	VV 不詳(世帯主以外の基本項目 記入不備世帯員)	12

【注】二重線から上が直系親族

↓

新世帯分類に使用した続柄		
01	世帯主又は代表者	直系親族
02	世帯主の配偶者	
03	子	
04	子の配偶者	
05	孫	
06	父母	
07	祖父母	
08	兄弟姉妹	傍系親族
09	他の親族	
10	住み込みの雇人	非親族
11	その他	
12	不詳	

<sup>1</sup> このアルゴリズムの手法を最初に適応したのは、インドネシアのマイクロデータ SUSENAS(2000)である[Suoh, S.(2014)]。今回は、帰宅時両親不在世帯の特定や昼間独居老人世帯の特定などにも応用した。

### 3. 国勢調査の調査票情報の SAS データベース化

利用を許可された 4 回次分の国勢調査の調査票情報は全て SAS データベースに組み込んだ。SAS によるこの処理は、1回次分の「マイクロデータ本体と符号表(エクセルファイル)を自動的に読み込み、SAS データセットを作成する SAS プログラム」を自動生成する SAS マクロプログラムを書くことによって、4回次分が同じプログラムで自動処理されたので、極めて効率的に行うことができた。さらに、第 4 節で解説する作業によって作成されたエクセル表を SAS プログラムで読み込んで、新世帯類型コードを SAS データベースに組み込んだ後、種々の分析作業を行った。詳細な分析結果<sup>2</sup>は伊藤(2016)に収録されている。

SAS プログラムによるマイクロデータの全自動データベース化技法は、昨年の SAS ユーザー総会でを行った報告(周防 2015)において、全国消費実態調査の匿名データで実用化済みであり、本システムでも同様の手法を採用した。その時は、全国消費実態調査の匿名データのレイアウトが複雑で作業が難航したが、今回の国勢調査の匿名データのレイアウトは単純だったので、あまり時間を要せずに作業が完了した。

### 4. 新世帯類型の作成プロセス

新世帯類型の設計には、エクセル表を活用した。この手法は極めて直感的に、かつ、正確に全てのパターンを網羅できる利点がある。

#### 4.1 プロセス1: 世帯主と直系親族の全てのパターンの析出

第 2 節の表 1 下の続柄を使って、直系親族 7 種類の全ての組み合わせパターンを、表 2 の左側にある「0,1,2」のパターンを作成・特定した。単に各続柄の有無だけであれば、「0,1」の組み合わせで済むが、世帯主の親を、「両親」と「ひとり親」に区別もしたかったので、続柄「父母」だけは「0,1,2」として、人数を区別できるようにした。その結果、7 つの続柄<sup>3</sup>の全ての組み合わせパターンは、96 通り(=3×2<sup>5</sup>)になった。

表 2 の「0,1,2」の組み合わせパターンを見ながら、「続柄の構成」欄に日本語で各パターンの世帯構造を表現していき、できあがった「続柄の構成」に基づいて、「同居世代の数」及び「世帯主の世代」を格付けしていった。

#### 4.2 プロセス2: 続柄の構成の整理

表 2 左側の「続き柄の構成」を、以下のルールで、表 2 の右側に示す「新世帯類型用の区分の名称」に変換した。表 2 右側にある黒色セルに白字の箇所がそれに該当する。

- ① 子供夫婦が成立しなかった「子供の配偶者」は、「子供」とした<sup>4</sup>。
- ② 「両親」または「ひとり親」は、4 世代以上の世帯については「親」に統一した<sup>5</sup>。

<sup>2</sup> 分析結果の一部は新井他(2016)にも掲載。

<sup>3</sup> 実際には、「世帯主」は必ずいるので、この続柄の有無は考慮しなくてよい。

<sup>4</sup> 国勢調査報告書で使用されている家族類型に於ける扱いに合わせた。

<sup>5</sup> 親を両親とひとり親に分ける必要性を検討した結果と、区分の数を抑制するために、このように処置した。

### 4.3 プロセス3:新世帯類型番号の付与

プロセス2の処理の結果、当初 96 個あったパターンの中に、同じ「新世帯類型用の区分の名称」が出現した。この点の整理を含め、かつ、世帯内の世代数の多い順番に並べ替えるために、以下の①～⑧のソートキーをエクセル関数を用いて作成後、この順番に 1 回のソートで順序を並べ替えた結果、表3ができた。

- ①「同居世代の数<sup>6</sup>」欄を降順でソート。「同居世代の数」が多い世帯が上位に来るようにした。
- ②「世帯主は何世代目<sup>7</sup>」欄を昇順でソート。世帯主の順位が高い方が上位に来るようにした。
- ③世帯主夫婦のいる世帯を上位に配置するため、「世帯主夫婦の判定」ができるように、表2の左側の「世帯主」と「世帯主の配偶者」の値を加算して、新たな欄を作り、この欄を降順でソート。
- ④「続柄の個数」をカウントする新たな欄を作り、この欄を降順でソート。続柄の数の多い方が上位に来るようにした。
- ⑤子供夫婦の有無については、子供夫婦のいる区分が上位に来るように、「子供夫婦の判定」(=「子」と「子の配偶者」欄の合計)欄を作り、この欄を降順でソート。
- ⑥子供の有無については、子供のいる方が上位に来るように、降順でソート。
- ⑦孫の有無については、孫のいる方が上位に来るように、降順でソート。
- ⑧両親の有無については、両親のいる方が上位に来るように、降順でソート。

このようにできた表3の「新世帯類型用の区分の名称」が同じ行に対して、Excel 関数を使って同じ連番号を 1 番から順番に振り、「小分類符号」を付与、親族世帯の新世帯類型を 56 区分に集約することができた<sup>8</sup>。この他に、親族・非親族同居世帯(1区分)、非親族世帯(1区分)、単独世帯(2区分:世帯主男性、世帯主女性)、特掲(枝番号付与の世帯)した世帯として、核家族世帯(4区分)、兄弟姉妹から成る世帯(1区分)があり、最終的に 65 区分<sup>9</sup>に分類した。

結果的には、エクセル上でのこの一連の手作業は、全てSASプログラムで処理ができる。

## 5. まとめ

今回は、世帯タイプのパターンを特定するのにエクセル表を使って作業を行ったが、これは試行錯誤をしながらパターンを作成するという行程を経たからである。この方式で 8 つのソートキーがエクセル関数を使って自動作成できたという確証を得たので、同様の処理を他のマイクロデータに対して行う際は、最初から SAS で行うことができることが保証された。つまり、エクセルを使って実験を行った結果、成功したので、次に同様の作業を行う際には、最初の段階からSASでこのアルゴリズムを参照しながら新世帯類型を SAS データベースに組み込むことができる。

<sup>6</sup> 定義については付録参照。

<sup>7</sup> 「世帯主は何世代目」とは、年配の世代から数えて何番目の世代かという意味である。例えば、「祖父母と世帯主」だけの世帯の場合、祖父母と世帯主の間に「両親」が抜けているが、世帯主は 3 番目の世代とする。

<sup>8</sup> この分類では核家族世帯の区分はないが、追加処理で、56 区分に枝番号付与した区分(傍系親族「兄弟姉妹、他の親族」)がいる世帯を除外した世帯)を特掲しており、核家族世帯の特定ができる仕組みになっている。

<sup>9</sup> この 65 区分の一覧表は新井他(2016)で表 2 として掲載されている。

エクセル表で数の多いパターンを分類する手法は極めて合理的であり、どの分野のデータでも有効で、かつ、正確性を保証している。仮に、これをSASの if ~then~else 文や、select-when 文で記述した場合、それが正しく機能していることをベリファイすることはかなり厄介である。また、フローチャートを描けたとしても、手間ばかりかかって実用的でない。

### 参考文献

- ①Suoh, S. (2014) The Construction of a Database System for the Micro Data Analysis of Asian Household Surveys: A Case Study of Indonesia, Income Distribution In Asian Regions: Socioeconomic Analyses Based on Micro Data Since the Second Half of the 20th Century (edited by Kamiya, D. et al.), Japan Statistical Association, Inc., pp65-117.
- ②周防節雄(2015) 「全国消費実態調査の匿名データから SAS のリレーショナルデータベースを完全自動構築するシステムの開発」、統計数理研究所 ([http://www.nstac.go.jp/services/pdf/151127\\_3-2.pdf](http://www.nstac.go.jp/services/pdf/151127_3-2.pdf))
- ③伊藤彰彦(2016) 『新たな個人消費を創造する新世帯類型別世帯数・世帯人員』、Sinfonica 研究叢書 No.25、(公財) 統計情報研究開発センター
- ④新井郁子、米澤香、中川雅義(2016) 「新世帯類型による国勢調査マイクロデータの分析結果の概要」、『SAS ユーザー総会 2016 論文集』

### 謝辞

本研究は、独立行政法人日本学術振興会の科学研究費(課題番号 26380280:研究代表:伊藤彰彦、「新世帯類型の構築と世帯構造の変動に係る長期時系列分析」)の助成金で行われた。記して謝意に替えたい。







## 付録 ソートキーの計算方法

ソートキーの作成には、エクセル関数とエクセルの種々の操作を組み合わせた。これら 8 個のキーの作成の詳細な解説は、本稿の目的ではないので、二つのキーを取り上げるに留める。下に示したエクセル表は表 3 の一部を抜粋したものである。

(1) 「同居世代の数」とは、祖父母(第 1 世代)、父母(第 2 世代)、世帯主・世帯主の配偶者(第 3 世代)、子・子の配偶者(第 4 世代)、孫(第 5 世代)のうち、異なる世代の数であるが、単なる数ではない。例えば「世帯主と孫がいる世帯」は、第 3 世代と第 5 世代から成り、第 4 世代がないが、同居世代の数は 2 ではなくて、3 とする。

① 下図の A 列の種類「2」に「世帯主と孫がいる世帯」があるが、列 B～H に「0,1,2」で有無を表現している。この情報から、「世帯主・世帯主の配偶者」(第 3 世代)と「子・子の配偶者」(第 4 世代)の有無をそれぞれひとまとめにして「1,0」で表現した後、第 1 世代～第 5 世代をひとまとめにしたパターン(N 列にある黒いセルに白字)を、エクセルの数式で作る。

$$=B7*10000+ROUND(C7/2,0)*1000+D7*100+ROUND(SUM(F7:G7)/2,0)*10+H7$$

② 計算式で計算した M7 列の(0,1)から成る値を左の L 列のセルに「値」で貼り付ける。

③ その貼り付けたパターンに対して 0 を半角ブランクに一括変換して K 列に保存。その結果、末尾に 0 があった場合は、サプレスされ、桁数が減る。

④ そのサプレスされたパターンの桁数を J 列でエクセルの LEN 関数を使えば、同居世代の数が判る。こうすれば含まれていない世代も半角ブランクで表現されているので、その世代も「同居世代の数」に含まれることになる。

(2) 「世帯主は何世代目」とは、「世帯内で年配の世代から数えて世帯主は何番目」という意味である。これも同じく、「M 列にある黒いセルに白字」に対して「=LEN(M7)-2」を計算すれば求まる。

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
2	有無パターン#	70	60	10	20	30	40	50	続柄の構成	★同居世代の数④	★エクセル計算⑤ 同居世代の数	★エクセル計算⑥ 同居世代の数	★エクセル計算⑦ 同居世代の数	★エクセル計算⑧ 世代か「注」	
3		祖父母	父母	世帯主	世帯主の配偶者	子	子の配偶者	孫		LEN	空白に置換	右のセルの空白を半角に置換	右のセルを値で貼り付け	エクセル	エクセル
4															
5															
6	種類	0	1	2	1	0	1	0							
7	2	0	0	1	0	0	0	1	世帯主と孫がいる世帯	3	1 1	101	101	1	
8	18	0	1	1	0	0	0	1	世帯主と孫とひとり親がいる世帯	4	1 1 1	1101	1101	2	
9	82	1	2	1	0	0	0	1	世帯主と孫と両親と祖父母がいる世帯	5	1 1 1 1	11101	11101	3	
10	83	1	2	1	0	0	1	0	世帯主と子供の配偶者と両親と祖父母がいる世帯	4	1 1 1 1	11110	11110	3	

# 新世帯類型による国勢調査マイクロデータの分析結果の概要

新井 郁子・米澤 香・中川 雅義  
(公財)統計情報研究開発センター

Brief Summary of Statistical Analyses on the Microdata of Japan's Population Censuses Based on the New Household Category.

Yuko Arai, Kaori Yonezawa, Masayoshi Nakagawa  
Statistical Information Institute for Consulting and Analysis

## 要旨

国勢調査のマイクロデータを使って新しい世帯分類「新世帯類型分類」を開発し、SAS データベースに組み込んだ。その新世帯類型分類を用いて種々の分析を行い、研究叢書として出版したので、その一部を紹介する。

キーワード: 国勢調査、新世帯類型、高齢化、少子化、単身世帯、核家族世帯、非親族世帯、マイクロデータ、proc tabulate、proc freq

## 1. はじめに

国勢調査の大規模調査年 1980 年、1990 年、2000 年、2010 年の調査票情報(マイクロデータ)を使って、世帯を分類する新しいシステムを開発した。この新世帯類型分類<sup>1</sup>は、「個人が属する世帯の構造」も個人の消費に少なからず影響を及ぼしていることから、「個人の消費に影響を及ぼす世帯の構造」を把握するために開発した世帯分類である。この分類を用いて国勢調査の 1980 年(昭和 55 年)と 2010 年(平成 22 年)の世帯数を比べてみると、この 30 年間に、世帯構造の分布に大きな変化が起っていたことが判明した。その概要について解説する。なお、集計・分析には SAS を使用した。

表 1 新世帯類型作成用の世帯員の属性

続柄	親族の種類	
① 世帯主または代表者	直系	親族
② 世帯主の配偶者		
③ 子		
④ 子の配偶者		
⑤ 孫		
⑥ 父母		
⑦ 祖父母		
⑧ 兄弟姉妹	傍系	親族
⑨ 他の親族		
⑩ 住み込みの雇人	非親族	非親族
⑪ その他		
⑫ 不詳		

## 2. 新世帯類型の分類区分

国勢調査では世帯員全員について「続柄」を調査しているが、調査年次によって定義が異なるので新世帯類型の開発用に「続柄」を統一した<sup>2</sup>(表1)。

①～⑨までを親族といい、⑩と⑪を非親族という。また、親族のうち①～⑦を直系親族といい、⑧と⑨を傍

<sup>1</sup> 詳細については安井他(2016)を参照。

<sup>2</sup> 続柄の統一作業の詳細についても安井他(2016)を参照。

系親族という。

新世帯類型は国勢調査の一般世帯<sup>3</sup>を対象にしており、2人以上の世帯員から成る世帯のうち、①と②～⑨の続柄から成る世帯を「親族世帯」、2人以上の世帯員から成る親族世帯に⑩または⑪の世帯員が含まれる世帯を「親族・非親族同居世帯」、2人以上の世帯員から成る世帯のうち、①と⑩または⑪の続柄から成る世帯を「非親族世帯」、世帯員が①の世帯主のみの1人世帯を「単独世帯」とした(表2のA～D)。

親族世帯については、直系親族の①と②～⑦の組み合わせにより56区分に分類し、さらに表2に示すとおり、1～5世代世帯に区分される。親族世帯(1)～(55)の世帯区分には、傍系親族が同居

表2 新世帯類型区分

新世帯類型	
<b>A 親族世帯</b>	<b>(3世代世帯・世帯主2世代目)</b>
(5世代世帯)	(33) 世帯主夫婦と子供夫婦と両親がいる世帯
(1) 世帯主夫婦と子供夫婦と孫と親と祖父母がいる世帯	(34) 世帯主夫婦と子供夫婦とひとり親がいる世帯
(2) 世帯主夫婦と子供と孫と親と祖父母がいる世帯	(35) 世帯主夫婦と子供と両親がいる世帯
(3) 世帯主夫婦と子供夫婦と孫と祖父母がいる世帯	(36) 世帯主夫婦と子供とひとり親がいる世帯
(4) 世帯主夫婦と子供と孫と祖父母がいる世帯	(37) 世帯主と子供夫婦と両親がいる世帯
(5) 世帯主夫婦と孫と親と祖父母がいる世帯	(38) 世帯主と子供夫婦とひとり親がいる世帯
(6) 世帯主夫婦と孫と祖父母がいる世帯	(39) 世帯主と子供と両親がいる世帯
(7) 世帯主と子供夫婦と孫と親と祖父母がいる世帯	(40) 世帯主と子供とひとり親がいる世帯
(8) 世帯主と子供と孫と親と祖父母がいる世帯	<b>(3世代世帯・世帯主3世代目)</b>
(9) 世帯主と子供夫婦と孫と祖父母がいる世帯	(41) 世帯主夫婦と両親と祖父母がいる世帯
(10) 世帯主と子供と孫と祖父母がいる世帯	(42) 世帯主夫婦とひとり親と祖父母がいる世帯
(11) 世帯主と孫と親と祖父母がいる世帯	(43) 世帯主夫婦と祖父母がいる世帯
(12) 世帯主と孫と祖父母がいる世帯	(44) 世帯主と両親と祖父母がいる世帯
(4世代世帯)	(45) 世帯主とひとり親と祖父母がいる世帯
(4世代世帯・世帯主2世代目)	(46) 世帯主と祖父母がいる世帯
(13) 世帯主夫婦と子供夫婦と孫と親がいる世帯	<b>(2世代世帯)</b>
(14) 世帯主夫婦と子供と孫と親がいる世帯	<b>(2世代世帯・世帯主1世代目)</b>
(15) 世帯主夫婦と孫と親がいる世帯	(47) 世帯主夫婦と子供夫婦がいる世帯
(16) 世帯主と子供夫婦と孫と親がいる世帯	(48) 世帯主夫婦と子供がいる世帯
(17) 世帯主と子供と孫と親がいる世帯	(49) 世帯主と子供夫婦がいる世帯
(18) 世帯主と孫と親がいる世帯	(50) 世帯主と子供がいる世帯
(4世代世帯・世帯主3世代目)	<b>(2世代世帯・世帯主2世代目)</b>
(19) 世帯主夫婦と子供夫婦と親と祖父母がいる世帯	(51) 世帯主夫婦と両親がいる世帯
(20) 世帯主夫婦と子供夫婦と祖父母がいる世帯	(52) 世帯主夫婦とひとり親がいる世帯
(21) 世帯主夫婦と子供と親と祖父母がいる世帯	(53) 世帯主と両親がいる世帯
(22) 世帯主夫婦と子供と祖父母がいる世帯	(54) 世帯主とひとり親がいる世帯
(23) 世帯主と子供夫婦と親と祖父母がいる世帯	<b>(1世代世帯)</b>
(24) 世帯主と子供夫婦と祖父母がいる世帯	(55) 世帯主夫婦がいる世帯
(25) 世帯主と子供と親と祖父母がいる世帯	(56) 世帯主と傍系親族から成る世帯
(26) 世帯主と子供と祖父母がいる世帯	(56)・2 うち、兄弟姉妹から成る世帯
(3世代世帯)	<b>B 親族・非親族同居世帯</b>
(3世代世帯・世帯主1世代目)	<b>D 単独世帯</b>
(27) 世帯主夫婦と子供夫婦と孫がいる世帯	(57) 世帯主(男性)
(28) 世帯主夫婦と子供と孫がいる世帯	(58) 世帯主(女性)
(29) 世帯主夫婦と孫がいる世帯	<b>(別掲:核家族世帯)</b>
(30) 世帯主と子供夫婦と孫がいる世帯	(48)・2 うち、世帯主夫婦と子供から成る世帯
(31) 世帯主と子供と孫がいる世帯	(50)・2 うち、世帯主(男性)と子供から成る世帯
(32) 世帯主と孫がいる世帯	(50)・3 うち、世帯主(女性)と子供から成る世帯
	(55)・2 うち、世帯主夫婦のみの世帯

<sup>3</sup> 国勢調査では全世帯を「一般世帯」と「施設等の世帯」に分類している。「一般世帯」とは以下の世帯を言う。1) 住居と生計を共にしている人の集まり、又は一戸を構えて住んでいる単身者。ただし、これらの世帯と住居を共にする単身の住み込みの雇人については、人数に関係なく雇主の世帯に含む。2) 前記1)の世帯と住居を共にし、別に世帯を維持している間借り単身者又は下宿屋などに下宿している単身者。3) 会社・団体・商店・官公庁などの寄宿舎、单身寮などに居住している単身者。

ちなみに「施設等の世帯」とは、寮・寄宿舎の学生・生徒、病院・療養所の入院者、社会施設の入所者、自衛隊営舎内居住者、矯正施設の入所者、その他定まった住居を持たない者の世帯を言う。(総務省統計局ホームページ「平成22年国勢調査 ユーザーズガイド」参照)

している世帯も含まれているため、これらの区分の名称の末尾を「がいる世帯」とした。

一方、親族世帯のうち傍系親族が同居していない世帯は、区分名に明示された続柄の世帯員しかいない世帯となるので、これらの世帯区分の名称の末尾は「から成る世帯」とした。表2の末尾の別掲で示した「核家族世帯」は、「から成る世帯」の典型的な例である。

### 3. 新世帯類型別の時系列分析の結果

本節に示す結果の概要は、前節で解説した新世帯類型分類に基づいている。

#### 3.1 新世帯類型(中分類)別世帯数

1980年から2010年の間に世帯数は1.45倍増加している。しかし、3世代以上の世代が同居している世帯は減少し、親子から成る2世代世帯は、1.12倍に留まっている。これに対し、世帯主夫婦のみの世帯が大半を占める1世代世帯(2.24倍)や単独世帯(2.36倍)は大幅に増加している(表3)。

表3 新世帯類型(中分類)別一般世帯数—全国(1980年、1990年、2000年、2010年)

新世帯類型 (中分類)	世帯数				倍率 昭和55年 ～ 平成22年	構成比(%)	
	昭和55年 (1980年)	平成2年 (1990年)	平成12年 (2000年)	平成22年 (2010年)		昭和55年 (1980年)	平成22年 (2010年)
総数	35,821,270	40,670,250	46,771,140	51,843,690	1.45	100.00%	100.00%
A 親族世帯	28,493,320	31,108,050	33,655,510	34,526,310	1.21	79.54%	66.60%
5世代世帯	740	570	760	570	0.77	0.00%	0.00%
4世代世帯	340,430	331,370	250,540	193,500	0.57	0.95%	0.37%
世帯主2世代目	253,680	250,200	194,750	163,260	0.64	0.71%	0.31%
世帯主3世代目	86,750	81,180	55,800	30,230	0.35	0.24%	0.06%
3世代世帯	5,439,040	5,302,390	4,555,430	3,561,990	0.65	15.18%	6.87%
世帯主1世代目	2,110,850	2,137,020	1,940,390	1,639,620	0.78	5.89%	3.16%
世帯主2世代目	3,302,020	3,138,390	2,587,880	1,898,260	0.57	9.22%	3.66%
世帯主3世代目	26,180	26,990	27,160	24,110	0.92	0.07%	0.05%
2世代世帯	17,954,460	18,842,300	19,605,440	20,092,890	1.12	50.12%	38.76%
世帯主1世代目	17,096,480	17,703,180	18,149,370	18,489,200	1.08	47.73%	35.66%
世帯主2世代目	857,980	1,139,120	1,456,070	1,603,700	1.87	2.40%	3.09%
1世代世帯	4,758,650	6,631,420	9,243,340	10,677,370	2.24	13.28%	20.60%
B 親族・非親族同居世帯	169,750	91,980	84,830	106,980	0.63	0.47%	0.21%
C 非親族世帯	61,480	77,440	189,250	345,730	5.62	0.17%	0.67%
D 単独世帯	7,096,720	9,392,780	12,841,550	16,778,530	2.36	19.81%	32.36%
世帯主(男性)	3,950,970	5,251,050	6,920,600	8,807,710	2.23	11.03%	16.99%
世帯主(女性)	3,145,750	4,141,730	5,920,950	7,970,820	2.53	8.78%	15.37%

注) 総数には、新世帯類型不詳を含む。

#### 3.2 世帯の高齢化

世帯主の平均年齢をみると、非親族世帯以外の世帯区分で、世帯主の平均年齢が上がっており世帯の高齢化が進んでいる。全国全世帯平均は、1980年から2010年にかけての年齢差は9歳となっており、中でも、単独世帯の年齢差が12.9歳と大きい(表4)。

表4 新世帯類型(大分類)別世帯主の平均年齢—全国(1980年、1990年、2000年、2010年)

新世帯類型	世帯主の平均年齢				年齢差
	昭和55年 (1980年)	平成2年 (1990年)	平成12年 (2000年)	平成22年 (2010年)	昭和55年 ～ 平成22年
総数	45.17	48.32	51.11	54.19	9.02
A 親族世帯	47.06	50.42	53.22	56.08	9.02
B 親族・非親族同居世帯	51.41	52.88	53.92	54.42	3.01
C 非親族世帯	44.07	41.08	39.99	41.56	-2.51
D 単独世帯	37.47	41.36	45.72	50.38	12.91

世帯主の年齢階級別世帯数をみると2人以上世帯と単独世帯ともに高齢世帯の増加がみられる一方で、2人以上の世帯主の年齢階級「45～54歳」以下の各階級の世帯数は減少し、特に「30～34歳」以下の各階級の減少が大きいことが特徴的である。これは「30～34歳」以下の各階級において既婚者の世帯数が大きく減少したためである(表5)。

### 3.3 核家族世帯の構造別分布

親族世帯に占める核家族世帯の割合は、1980年(74.1%)から2010年(82.2%)にかけて増加しており、核家族化が進行している。

内訳をみると、核家族世帯における主要な世帯である「世帯主夫婦と子供から成る世帯」は1980年当時より僅かながらも減少(0.96倍)し、代わって「世帯主夫婦のみの世帯」(2.31倍)、「世帯主(女性)と子供から成る世帯」(2.23倍)が、それぞれ大幅に増加しており<sup>4</sup>、この30年間に、核家族世帯の構造が大きく変化したことが分かる(表6)。

表5 世帯主の配偶関係(2区分)、世帯主の年齢(5歳階級(45歳以上10歳階級))別、二人以上の一般世帯数—全国(1980年、2010年)

世帯主の配偶関係、 世帯主の年齢	昭和55年 (1980年)	平成22年 (2010年)	倍率
総数	28,724,550	35,065,160	1.22
既婚	28,161,120	33,351,520	1.18
19歳以下	8,050	2,980	0.37
20～24歳	245,270	144,710	0.59
25～29歳	1,571,330	883,290	0.56
30～34歳	3,465,980	1,977,230	0.57
35～39歳	3,784,000	2,924,840	0.77
40～44歳	3,769,480	2,935,310	0.78
45～54歳	7,379,440	5,827,390	0.79
55～64歳	4,501,650	7,650,430	1.70
65～74歳	2,641,430	6,483,560	2.45
75歳以上	794,490	4,521,780	5.69
未婚	563,270	1,318,400	2.34
19歳以下	11,920	8,180	0.69
20～24歳	99,240	84,330	0.85
25～29歳	127,040	141,540	1.11
30～34歳	114,410	129,310	1.13
35～39歳	66,260	148,500	2.24
40～44歳	41,710	151,350	3.63
45～54歳	66,200	300,280	4.54
55～64歳	26,870	249,580	9.29
65～74歳	7,570	71,900	9.49
75歳以上	2,040	33,440	16.39

注)総数には、世帯主の配偶関係不詳及び世帯主の年齢不詳を含む。

<sup>4</sup> 規模としては少ないが、「世帯主(男性)と子供から成る世帯」(2.29倍)も大幅に増加している。

表6 新世帯類型(核家族世帯)別一般世帯数—全国(1980年、1990年、2000年、2010年)

新世帯類型 (核家族世帯)	世帯数(実数)				倍率 昭和55年 ～ 平成22年	親族世帯に対する割合(%)	
	昭和55年 (1980年)	平成2年 (1990年)	平成12年 (2000年)	平成22年 (2010年)		昭和55年 (1980年)	平成22年 (2010年)
総数	21,100,650	23,612,410	26,606,600	28,372,730	1.34	74.05%	82.18%
世帯主夫婦と子供から成る世帯	14,980,160	15,078,360	14,863,160	14,361,230	0.96	52.57%	41.60%
世帯主(男性)と子供から成る世帯	275,720	398,000	504,010	630,450	2.29	0.97%	1.83%
世帯主(女性)と子供から成る世帯	1,403,210	1,853,270	2,407,020	3,127,900	2.23	4.92%	9.06%
世帯主夫婦のみの世帯	4,441,560	6,282,770	8,832,410	10,253,150	2.31	15.59%	29.70%

ここでは、世帯数の増加倍率が最も高い「世帯主夫婦のみの世帯」と、増加倍率が最も低い「世帯主夫婦と子供から成る世帯」に注目した。

### 3.3.1 「世帯主夫婦のみの世帯」

世帯数の増加倍率が高い「世帯主夫婦のみの世帯」について注目すると、1980年から2010年という30年の間に、子供の成長・独立により残されたと考えられる高齢の「世帯主夫婦のみの世帯」(妻の年齢「65～74歳」5.15倍、「75歳以上」15.95倍)が大幅に増加していることと、独立した子供の世帯と考えられる「世帯主夫婦のみの世帯」も、妻の年齢「30～34歳」(2.10倍)、同「35～39歳」(3.01倍)にみられるように増加しており、この2種類の「世帯主夫婦のみの世帯」の併存が「世帯主夫婦のみの世帯」の大幅増加の要因となっていることが分かる(表7)。

表7 妻の年齢(5歳階級(45歳以上10歳階級))別世帯主夫婦のみの世帯数—全国(1980年、2010年)

妻の年齢	昭和55年 (1980年)	平成22年 (2010年)	倍率
総数	4,441,560	10,253,150	2.31
24歳以下	305,640	85,130	0.28
25～29歳	480,000	426,620	0.89
30～34歳	260,550	546,050	2.10
35～39歳	162,270	488,420	3.01
40～44歳	193,690	382,020	1.97
45～54歳	1,081,160	977,440	0.90
55～64歳	1,282,000	2,877,300	2.24
65～74歳	584,990	3,014,950	5.15
75歳以上	91,260	1,455,230	15.95

### 3.3.2 「世帯主夫婦と子供から成る世帯」

「世帯主夫婦と子供から成る世帯」についてみると、30年前と比べて、妻の年齢「55歳～64歳」以上の各階級区分に属する世帯は、大きく増え、「45歳～54歳」以下の各階級区分に属する世帯数は、いずれも減少している。なかでも、妻の年齢「30歳～34歳」以下の各階級区分に属する世帯数は、いずれも半数以下となっている(表8)。

表8 妻の年齢(5歳階級(45歳以上10歳階級))別世帯主夫婦と子供から成る世帯数—全国(1980年、2010年)

妻の年齢	昭和55年 (1980年)	平成22年 (2010年)	倍率
総数	14,980,160	14,361,230	0.96
24歳以下	303,300	141,690	0.47
25～29歳	1,830,310	660,510	0.36
30～34歳	3,075,690	1,545,070	0.50
35～39歳	2,788,550	2,283,620	0.82
40～44歳	2,424,850	2,076,820	0.86
45～54歳	3,318,770	3,277,650	0.99
55～64歳	1,026,500	2,645,190	2.58
65～74歳	192,200	1,303,900	6.78
75歳以上	20,000	426,770	21.34

次に、我が国における人口問題の関

心事といえる少子化の現状について、新世帯類型分類の視点から集計を行った。

親と一緒に暮らしている可能性の高い子供の年齢上限を17歳とし、かつ妻の年齢が44歳以下の「世帯主夫婦と子供から成る世帯」についてみると、子供が2人いる世帯の割合は54.1%から47.5%へと6.7ポイント低下し、子供1人の世帯の割合が28.5%から36.9%へと8.4ポイント上昇している。

さらに多くの事業所が集中している1都3県(東京都、埼玉県、千葉県、神奈川県)と全国を比べてみると、1都3県においては子供1人の世帯の割合は全国平均より4.1ポイント高く、2人、3人以上の世帯は、それぞれ-1.4ポイント、-2.6ポイント全国より低くなっている。東京を中心としたこの地域には、子供を2人以上持つには難しい環境があることが伺われる(表9)。

表9 17歳以下の子供の数(3区分)別世帯主夫婦と子供から成る世帯数(妻の年齢44歳以下)－全国・1都3県(1980年、2010年)

17歳以下の子供の数	昭和55年 (1980年)	平成22年 (2010年)	倍率	構成比 の差
全国	10,422,710	6,707,720	0.64	—
1人	2,968,010	2,474,890	0.83	—
2人	5,640,930	3,183,120	0.56	—
3人以上	1,604,800	961,720	0.60	—
1世帯当たり子供の数	1.89	1.79	—	—
構成比	—	—	—	—
1人	28.48%	36.90%	—	8.42%
2人	54.12%	47.45%	—	-6.67%
3人以上	15.40%	14.34%	—	-1.06%
1都3県の計	3,005,520	2,014,510	0.67	—
1人	896,490	825,520	0.92	—
2人	1,658,630	927,270	0.56	—
3人以上	403,930	235,630	0.58	—
1世帯当たり子供の数	1.85	1.72	—	—
構成比	—	—	—	—
1人	29.83%	40.98%	—	11.15%
2人	55.19%	46.03%	—	-9.16%
3人以上	13.44%	11.70%	—	-1.74%
1都3県の構成比と全国の構成比との差	—	—	—	—
1人	1.40	4.10	—	—
2人	1.10	-1.40	—	—
3人以上	-2.00	-2.60	—	—

表10 男女、配偶関係(4区分)別、単独世帯の増加率・寄与度－全国(1980年～2010年)

男女、 配偶関係	単独世帯数		倍率	増加率・ 寄与度 ※
	昭和55年 (1980年)	平成22年 (2010年)		
総数	7,096,720	16,778,530	2.36	136.43%
男性	3,950,970	8,807,710	2.23	
未婚	3,129,460	5,024,950	1.61	26.71%
有配偶	334,320	725,790	2.17	5.52%
死別	200,290	696,410	3.48	6.99%
離別	227,080	1,051,420	4.63	11.62%
女性	3,145,750	7,970,820	2.53	
未婚	1,606,620	3,118,230	1.94	21.30%
有配偶	117,110	300,340	2.56	2.58%
死別	1,059,440	2,777,460	2.62	24.21%
離別	315,790	982,610	3.11	9.40%

※ 寄与度は増加率に対するもので、寄与度を合計すると増加率になる。ただし、本表では、総数に配偶関係不詳を含んでいるため、寄与度を合計した値と増加率は一致しない。

注1) 増加率・寄与度の欄において、増加率は総数のみ、その他は寄与度である。  
注2) 男性の計、女性の計には、配偶関係不詳を含む。

### 3.4 単独世帯

1980年から2010年の30年の間に単独世帯は2.36倍増加している。男女、配偶関係別にその増加寄与度をみると、男性の未婚(26.7%)及び女性の死別(24.2%)と未婚(21.3%)が突出しており、単独世帯が大増に増加したのは、未婚の男性、未婚の女性、及び死別した女性、がそれぞれ大幅に増えたことによ

るものであることが分かる。

また、男女ともに離別者の増加倍率が高く、離婚も単独世帯の大幅増加の要因の一つとなっている<sup>5</sup>(表 10)。

男女それぞれの増加率に対する寄与度を年齢別にみても、男性は「35～39 歳」「40～44 歳」と「55～59 歳」「60～64 歳」が高く、女性は、「25～29 歳」「60～64 歳」以上の各年齢階級が高くなっている(表 11)。

男性「35～39 歳」「40～44 歳」については、未婚の寄与度が高く、男性「55～59 歳」「60～64 歳」については、未婚の寄与度が高い一方で離別の寄与度も「50～54 歳」に比べて高くなっており、定年や子供の自立を機に離婚を選択した者が増加していることが伺える。

女性「25～29 歳」の寄与度が高いのは、未婚の寄与度が高いためである。だが、未婚の寄与度が高いのは「25～29 歳」に止まらず 30 歳以上の年代の未婚の寄与度も高くなっており、これらの年代の未婚単身者が増加している。一方で、女性「60～64 歳」と「65～69 歳」は、離別の寄与度が高いことが要因となっている(表 12)。

表 11 男女、世帯主の年齢(5歳階級)別、単独世帯の増加率・寄与度—全国(1980年～2010年)

世帯主の年齢	男性				女性			
	単独世帯数		倍率	増加率・寄与度(%)※	単独世帯数		倍率	増加率・寄与度(%)※
	昭和55年(1980年)	平成22年(2010年)			昭和55年(1980年)	平成22年(2010年)		
総数	3,950,970	8,807,710	2.23	122.93	3,145,750	7,970,820	2.53	153.38
24歳以下	1,561,010	1,132,300	0.73	-10.85	864,780	823,620	0.95	-1.31
25歳～29歳	875,880	976,140	1.11	2.54	241,200	617,650	2.56	11.97
30歳～34歳	474,120	799,030	1.69	8.22	163,820	457,990	2.80	9.35
35歳～39歳	231,150	803,520	3.48	14.49	121,710	412,220	3.39	9.23
40歳～44歳	163,970	708,440	4.32	13.78	121,510	339,880	2.80	6.94
45歳～49歳	155,440	626,490	4.03	11.92	166,510	300,620	1.81	4.26
50歳～54歳	139,480	579,940	4.16	11.15	226,060	306,790	1.36	2.57
55歳～59歳	90,870	659,620	7.26	14.40	266,980	415,060	1.55	4.71
60歳～64歳	66,030	685,550	10.38	15.68	281,680	612,050	2.17	10.50
65歳～69歳	65,660	466,750	7.11	10.15	275,020	655,940	2.39	12.11
70歳～74歳	57,630	342,870	5.95	7.22	213,030	729,940	3.43	16.43
75歳以上	69,740	575,870	8.26	12.81	203,460	2,011,770	9.89	57.48

※ 寄与度は増加率に対するもので、寄与度を合計すると増加率になる。ただし、本表では、男性の計、女性の計に年齢不詳を含んでいるため、寄与度を合計した値と増加率は一致しない。

注1) 増加率・寄与度の欄において、増加率は男性の計、女性の計のみ、その他は寄与度である。

注2) 男性の計、女性の計には、年齢不詳を含む。

<sup>5</sup> 単独世帯であるが配偶者のいる者も 6.1%いる。



表 12 年齢(5歳階級)、配偶関係(4区分)別単独世帯の増加率・寄与度—全国(1980年～2010年)

男性単独世帯(30歳～49歳)					男性単独世帯(50歳～69歳)				
年齢、 配偶関係	単独世帯数		倍率	増加率・ 寄与度 ※	年齢、 配偶関係	単独世帯数		倍率	増加率・ 寄与度 ※
	昭和55年 (1980年)	平成22年 (2010年)				昭和55年 (1980年)	平成22年 (2010年)		
男性総数	3,950,970	8,807,710	2.23	122.93%	男性総数	3,950,970	8,807,710	2.23	122.93%
30歳～34歳	474,120	799,030	1.69	8.22%	50歳～54歳	139,480	579,940	4.16	11.15%
未婚	412,040	622,460	1.51	5.33%	未婚	28,470	287,540	10.10	6.56%
有配偶	27,640	32,220	1.17	0.12%	有配偶	62,300	124,680	2.00	1.58%
死別	1,650	700	0.42	-0.02%	死別	13,530	8,780	0.65	-0.12%
離別	23,650	32,210	1.36	0.22%	離別	32,160	109,330	3.40	1.95%
35歳～39歳	231,150	803,520	3.48	14.49%	55歳～59歳	90,870	659,620	7.26	14.40%
未婚	159,850	578,250	3.62	10.59%	未婚	14,190	319,740	22.54	7.73%
有配偶	32,460	49,560	1.53	0.43%	有配偶	33,570	115,660	3.45	2.08%
死別	2,560	1,660	0.65	-0.02%	死別	18,750	24,940	1.33	0.16%
離別	30,400	67,630	2.22	0.94%	離別	22,340	146,150	6.54	3.13%
40歳～44歳	163,970	708,440	4.32	13.78%	60歳～64歳	66,030	685,550	10.38	15.68%
未婚	80,740	453,980	5.62	9.45%	未婚	7,620	303,390	39.82	7.49%
有配偶	42,790	77,930	1.82	0.89%	有配偶	16,560	73,260	4.42	1.44%
死別	3,900	2,560	0.66	-0.03%	死別	24,150	62,570	2.59	0.97%
離別	32,220	87,370	2.71	1.40%	離別	16,080	192,280	11.96	4.46%
45歳～49歳	155,440	626,490	4.03	11.92%	65歳～69歳	65,660	466,750	7.11	10.15%
未婚	47,960	347,270	7.24	7.58%	未婚	5,150	152,450	29.60	3.73%
有配偶	61,370	114,200	1.86	1.34%	有配偶	12,150	36,280	2.99	0.61%
死別	6,920	4,310	0.62	-0.07%	死別	34,690	93,830	2.70	1.50%
離別	35,440	97,800	2.76	1.58%	離別	12,040	143,490	11.92	3.33%

女性単独世帯(25歳～49歳)					女性単独世帯(55歳以上)				
年齢、 配偶関係	単独世帯数		倍率	増加率・ 寄与度 (%)※	年齢、 配偶関係	単独世帯数		倍率	増加率・ 寄与度 (%)※
	昭和55年 (1980年)	平成22年 (2010年)				昭和55年 (1980年)	平成22年 (2010年)		
女性総数	3,145,750	7,970,820	2.53	153.38	女性総数	3,145,750	7,970,820	2.53	153.38
25歳～29歳	241,200	617,650	2.56	11.97	55歳～59歳	266,980	415,060	1.55	4.71
未婚	223,350	524,550	2.35	9.57	未婚	46,810	125,840	2.69	2.51
有配偶	5,790	16,270	2.81	0.33	有配偶	18,940	41,050	2.17	0.70
死別	390	950	2.40	0.02	死別	142,790	77,140	0.54	-2.09
離別	8,510	5,130	0.60	-0.11	離別	53,300	145,520	2.73	2.93
30歳～34歳	163,820	457,990	2.80	9.35	60歳～64歳	281,680	612,050	2.17	10.50
未婚	135,970	368,860	2.71	7.40	未婚	24,840	142,200	5.73	3.73
有配偶	5,470	18,720	3.43	0.42	有配偶	13,290	38,990	2.93	0.82
死別	1,360	1,280	0.94	0.00	死別	203,770	199,770	0.98	-0.13
離別	18,130	14,720	0.81	-0.11	離別	35,020	199,730	5.70	5.24
35歳～39歳	121,710	412,220	3.39	9.23	65歳～69歳	275,020	655,940	2.39	12.11
未婚	86,790	315,590	3.64	7.27	未婚	14,280	109,720	7.68	3.03
有配偶	5,470	15,460	2.83	0.32	有配偶	9,640	25,390	2.63	0.50
死別	3,270	2,950	0.90	-0.01	死別	224,030	339,450	1.52	3.67
離別	23,810	31,070	1.30	0.23	離別	22,900	150,110	6.55	4.04
40歳～44歳	121,510	339,880	2.80	6.94	70歳～74歳	213,030	729,940	3.43	16.43
未婚	70,500	238,820	3.39	5.35	未婚	7,490	86,790	11.59	2.52
有配偶	7,880	15,130	1.92	0.23	有配偶	5,340	21,020	3.93	0.50
死別	9,030	5,490	0.61	-0.11	死別	185,310	488,400	2.64	9.63
離別	31,180	44,750	1.44	0.43	離別	11,380	103,210	9.07	2.92
45歳～49歳	166,510	300,620	1.81	4.26	75歳以上	203,460	2,011,770	9.89	57.48
未婚	73,110	171,010	2.34	3.11	未婚	4,920	155,910	31.71	4.80
有配偶	14,860	23,440	1.58	0.27	有配偶	3,070	44,970	14.65	1.33
死別	29,830	10,820	0.36	-0.60	死別	185,040	1,624,720	8.78	45.77
離別	44,990	66,860	1.49	0.70	離別	7,140	126,650	17.74	3.80

※ 寄与度は増加率に対するもので、寄与度を合計すると増加率になる。ただし、本表では、掲載している年齢階級を限定しているため、各年齢階級の寄与度を合計した値と男性総数または女性総数の増加率は一致しない。

注1) 増加率・寄与度の欄において、増加率は男性総数または女性総数のみ、その他は寄与度である。

注2) 各年齢階級には、配偶関係不詳を含む。

### 3.5 非親族世帯

世帯数の規模は小さいが、世帯主と非親族から成る世帯(非親族世帯)が大幅に増加(5.62倍)しており、これは「世帯主とその他から成る世帯」の増加が影響している(表13)。

「世帯主とその他から成る世帯」を世帯人員別にみると、世帯人員2人の世帯が1980年から7.86倍と著しく増え、その同居形態は異性との同居が91.4%(2010年)を占めており圧倒的に多い。この同居形態は、1980年の時点でも若者世代を中心に幅広い年代層に存在していたが、この30年の間にその厚みを増し、同棲世帯が生活スタイルとして定着しつつあることが分析結果から読み取れる(表14)。

表13 非親族世帯の区分(3区分)別非親族世帯数—全国(1980年、2010年)

非親族世帯の区分	昭和55年 (1980年)	平成22年 (2010年)	倍率
総数	61,480	345,730	5.62
世帯主と住み込みの雇人とその他から成る世帯	390	190	0.49
世帯主と住み込みの雇人から成る世帯	16,350	5,090	0.31
世帯主とその他から成る世帯	44,740	340,450	7.61

表14 世帯人員(4区分)、世帯人員が2人の世帯の同居形態(3区分)、異性同居世帯の世帯主の年齢(7区分)別、世帯主とその他から成る非親族世帯数—全国(1980年、2010年)

世帯人員、世帯人員が2人の世帯の同居形態、異性同居世帯の世帯主の年齢	昭和55年 (1980年)	平成22年 (2010年)	構成比 (2010年)	倍率 昭和55年 ～ 平成22年
総数	44,740	340,450		7.61
世帯人員2人	41,610	327,220	100.00%	7.86
男性同居	370	14,450	4.42%	39.55
女性同居	810	13,600	4.16%	16.84
異性同居	40,430	299,180	91.43%	7.40
24歳以下	11,110	30,650	9.37%	2.76
25歳～34歳	11,300	109,230	33.38%	9.66
35歳～44歳	4,270	54,380	16.62%	12.75
45歳～54歳	4,770	32,250	9.86%	6.76
55歳～64歳	4,200	38,090	11.64%	9.07
65歳～74歳	3,070	24,390	7.45%	7.94
75歳以上	1,710	10,190	3.11%	5.96
平均年齢	38.73	41.19		—
世帯人員3人	2,060	8,980		4.36
世帯人員4人	800	3,100		3.90
世帯人員5人以上	280	1,160		4.16

#### 4. まとめ

以上に示したように、新世帯類型分類を用いたことで少子高齢化の実態やライフスタイルの変化などを量的に示すことができ、1980年と2010年の30年間で世帯の構造別分布に大きな変化が起きていることが分かった。本報告は、Sinfonica 研究叢書『新世帯類型別世帯数・世帯人員』（伊藤彰彦(2016)）に掲載した時系列分析の一部を紹介したものであるが、その他にも、同書には、「高齢者施策用基礎資料」「乳幼児の保育施策用基礎資料」や「消費需要分析用基礎資料（夫婦の数、母親または祖母の数、昼間家にいる者の年齢構成）」「住宅需要分析用基礎資料」等新世帯類型分類と他の項目とのクロス集計の分析結果も収録している。

#### 参考文献

①伊藤彰彦(2016) 『新たな個人消費を創造する新世帯類型別世帯数・世帯人員』、Sinfonica 研究叢書 No.25、(公財) 統計情報研究開発センター。

②安井浩子・中川雅義・周防節雄(2016) 2010年国勢調査マイクロデータから全自動で構築したSAS データベースに基づく新世帯類型の設計と開発、『SAS ユーザー総会 2016 論文集』

#### 謝辞

本研究は、独立行政法人日本学術振興会の科学研究費(課題番号 26380280:研究代表:伊藤彰彦、「新世帯類型の構築と世帯構造の変動に係る長期時系列分析」)の助成金で行われた。記して謝意に替えたい。

# 4回の学生調査に基づく 学生の文化・芸術に対する意識の現状と変化

有馬 昌宏

兵庫県立大学応用情報科学研究科

An Analysis of University Students' Attitude toward Culture and Arts

Masahiro ARIMA

Graduate School of Applied Informatics, University of Hyogo

## 要旨

1986年からほぼ5年間隔で実施されてきた全国の学生を対象とする文化芸術に関する調査の1991年の第2回調査から2008年の第5回調査まで4回の調査の芸術・文化に対する意識に関する10項目の質問への回答から、学生の文化・芸術への意識を対応分析を適用して類型化するとともに、第2回調査から第5回調査にかけて、文化芸術への学生の意識がどのように変化してきているかを考察する。また、意識の相違が実際の文化・芸術のライブとメディアによる鑑賞行動にどのように影響しているかについても考察する。

キーワード：文化資本，実演芸術，文化的オムニボア仮説，学生調査，名義ロジスティック回帰分析

## 1. はじめに

文化や芸術への需要構造をデータに基づいて分析しようとする時、データが十分に整備されていないという問題に直面する。文化・芸術を提供する劇場・ホールや美術館などでは、来館者を対象としたアンケート調査を実施しているが、それらの個票データは一般には公開されていない。広く国民を対象とする調査としては、1976年以降、5年ごとの周期調査として実施されている社会生活基本調査の生活行動調査がある。この調査では、美術鑑賞のほかに、実演芸術に関連する種目としては、演芸・演劇・舞踊鑑賞、音楽会などによるクラシック音楽鑑賞、音楽会などによるポピュラー音楽・歌謡曲鑑賞の3つの種目が取り上げられて過去1年間の鑑賞行動の有無と頻度が調査されている。しかし、文化・芸術に対する意識が鑑賞行動に大きな影響を及ぼすと考えられるにも関わらず、社会生活基本調査はもとより、ほとんどの調査では、行動の有無や行動頻度は調査されているが、文化・芸術に対する意識についての設問はされていないのが現状である。

また、文化芸術への意識を問う設問がされている調査としては、全国規模では1987年、1996年、2003年、2009年に実施された内閣府の「文化に関する世論調査」、1993年と1999年に実施された文化庁の「国民の文化に関する意識調査」、地域的なものでは東京都目黒区で2014年に実施された「文化芸術に関する意識調

査」などの調査があるものの、文化・芸術に対する意識についての継続的な設問はされていないのが現状である。

そこで、本研究では、我々の研究グループが1985年度からほぼ5年間隔で実施している有意抽出による1万人規模の学生を対象とした一連の調査（学生調査）のうち、10の対照的な芸術・文化に対する考え方のどちらに賛成するかを問う同じ質問を設けた1991年度に実施した第2回調査から2008年度に実施した第5回調査までの個票データを用いて、文化・芸術への意識の現状とその変化を考察する。特に、本研究のもとになっている調査は、大学生が調査対象となっていることから、1998年の学習指導要領の改訂（高等学校は1999年）により、「我が国の文化や伝統を尊重する態度を育成する観点などから、小学校から我が国に伝わる楽器を取り上げられるようにするとともに、中学校では3年間を通じて1種類以上の和楽器を体験できるようにする」という指導方針のもとで中学校の音楽の授業で和楽器指導の導入が義務付けられるなどの日本的な文化・芸術を重視する教育の文化・芸術意識への影響がどの程度あるかを、2008年度の第5回調査の結果とそれ以前の第4回調査までの結果と比較して分析することから可能になると考えられる。

以下、第2章では、我々が属する研究グループが実施してきた学生調査について概説し、第3章で、JMPの「2変数の関係」を適用しての文化・芸術への意識の現状と変化を明らかにする。その上で、第4章で対応分析を適用することによって文化・芸術への意識の構造を明らかにする試みの結果を示し、第5章で文化・芸術への意識と実際の鑑賞行動との関係を調べる。最後に、第6章において本研究のまとめを行うとともに、今後に残された課題についても言及する。

## 2. 学生調査の概要

社会生活基本調査は、総務省統計局が所管する基幹統計調査であり、2011年に実施された最新の第8回調査では、層化2段抽出法により、指定する調査区（全国で約6,900調査区）内に居住する世帯のうちから、選定された約8万3千世帯の10歳以上の世帯員約20万人が対象とされている。文化・芸術を対象とした調査でこれだけの大規模な周期調査は他になく、非常に有用な調査であるものの、文化・芸術に関して、分野や種目の細分類による詳細な分析を行いたい場合には、実演芸術は3種目に統合されており、目的を達成することは難しい。また、先述のように、文化・芸術に対する意識に関する設問はされていない。

このような状況を踏まえ、マイクロ統計データ（個票データ）に基づく文化・芸術の需要サイドの分析の重要性を早くから認識していた松田芳郎（一橋大学名誉教授）らの研究グループにより、文化・芸術情報の体系化と統計調査方法の確立を目的として、文化・芸術の需要者側の実態調査『学生の芸術意識と芸術活動に関する調査（以降、「学生調査」と略記）』が1985年以降5回に渡って実施され、調査研究の進行とともにデータの蓄積が図られてきている（加藤・有馬[5]）。学生調査は全国の学生を対象としており、大標本を確保しながら限られた費用で詳細にわたる調査を実施するために、大学教員のネットワークを活用して、全国の大学の地域と専門分野別の分布を考慮する割当法で標本設計を行って調査を実施している。有効回答サンプル数は第1回調査（1985年度実施）で10,570、第2回調査（1991年度実施）で10,770、第3回調査（1996年度実施）で10,061、第4回調査（2002年度実施）で3,767、第5回調査（2008年度実施）では7,475で、第1回調査から第5回調査までのサンプルの構造は、表1に要約して示すとおりである。第5回調査のサンプルの構造では、設立形態別では短期大学からの抽出率が低く、専攻分野（学部）では理工系学部と人文芸術系（文芸系）学部からの抽出率が低くなっている。また、医療福祉系（保健系）学部では、医学部からの抽出率は非常に低く、看護系学部の学生が多くを占めている。地域別では甲信越・北陸・東海と中国・四国

表1 学生調査のサンプルの構造

	第1回調査(1985年度)			第2回調査(1991年度)			第3回調査(1996年度)			
	母集団	有効回収 サンプル	抽出率 (%)	母集団	有効回収 サンプル	抽出率 (%)	母集団	有効回収 サンプル	抽出率 (%)	
設立 形態 別	大学	1,734,080	7,879	0.454	2,052,338	8,312	0.405	2,368,992	8,980	0.379
	短期大学	377,107	1,449	0.384	497,569	1,032	0.207	463,948	733	0.158
	高等専門学校	-----	-----	-----	19,816	180	0.908	21,608	80	0.370
	専修学校	579,274	1,242	0.214	834,713	1,155	0.138	799,551	185	0.023
	大学院	-----	-----	-----	98,650	87	0.088	164,350	83	0.051
性別	男子	1,666,991	5,404	0.324	2,010,715	6,346	0.316	2,136,180	5,386	0.252
	女子	1,023,470	4,949	0.484	1,492,371	4,375	0.293	1,682,269	4,414	0.262
専攻 分野	文芸系学部	879,530	4,757	0.541	1,155,315	3,387	0.293	1,045,345	2,573	0.246
	社会系学部	785,304	3,875	0.493	1,054,629	4,597	0.436	1,157,923	5,497	0.475
	理工系学部	498,968	1,101	0.221	857,863	2,092	0.244	1,076,724	1,280	0.119
	保健系学部	179,226	664	0.370	309,838	548	0.177	372,262	664	0.178
地域 別	北海道・東北	206,564	989	0.479	269,809	792	0.294	314,556	992	0.315
	関東	1,402,891	5,496	0.392	1,420,887	3,445	0.242	1,569,385	3,397	0.216
	甲信越・北陸・東海	76,915	326	0.424	457,024	1,747	0.382	528,336	1,362	0.258
	近畿	553,685	1,751	0.316	652,747	1,724	0.264	747,803	2,197	0.294
	中国・四国	172,743	1,025	0.593	227,008	1,192	0.525	266,181	761	0.286
	九州・沖縄	241,398	983	0.407	309,666	1,867	0.603	370,580	1,352	0.365
全体	2,690,461	10,570	0.393	3,503,086	10,770	0.307	3,818,449	10,061	0.263	

	第4回調査(2002年度)			第5回調査(2008年度)			
	母集団	有効回収 サンプル	抽出率 (%)	母集団	有効回収 サンプル	抽出率 (%)	
設立 形態 別	大学	2,499,147	3,678	0.147	2,520,593	7,009	0.278
	短期大学	267,086	---	0	172,725	160	0.093
	高等専門学校	21,336	---	0	59,446	188	0.316
	専修学校	765,558	41	0.005	---	---	---
	大学院	223,512	44	0.02	262,686	118	0.045
性別	男子	2,085,120	1,804	0.087	2,011,622	3,865	0.192
	女子	1,691,519	1,954	0.116	1,661,331	3,374	0.203
専攻 分野	文芸系学部	1,144,781	2,170	0.19	1,062,559	1,537	0.145
	社会系学部	1,128,365	1,302	0.115	934,495	3,648	0.390
	理工系学部	889,537	162	0.018	725,157	875	0.121
	保健系学部	506,089	129	0.025	295,674	868	0.294
地域 別	北海道・東北	324,816	371	0.114	246,482	415	0.168
	関東	1,579,021	1,066	0.068	1,291,756	2,206	0.171
	甲信越・北陸・東海	523,758	704	0.134	418,189	1,714	0.410
	近畿	757,073	1,124	0.148	620,145	883	0.142
	中国・四国	261,781	178	0.068	211,543	909	0.430
	九州・沖縄	372,227	324	0.087	276,873	475	0.172
全体	3,776,639	3,767	0.1	3,015,450	7,475	0.248	

注1) 母集団の構造は、第1回調査では文部省「昭和59年度学校基本調査」、第2回調査では「平成3年度版学校基本調査」、第3回調査では「平成8年度版学校基本調査」、第4回調査では「平成14年度版学校基本調査」、第5回調査では「平成20年度版学校基本調査」から計算している。

注2) 設立形態別、性別、専攻分野別、地域別の有効回収サンプルには分類が不明のサンプルは除外されているため、合計しても全体の有効回収サンプル数には一致しない。

からの抽出率が高くなっている。

なお、我々の調査は無作為抽出で実施されておらず、表1から明らかのように、各回の学生調査でサンプルに偏りがある。したがって、以下の分析では、割当法でのサンプルの偏りを補正するための復元乗率を用いた調整を行う必要があるが、本研究においては、復元乗率を求めることができていないために、直接的な時系列比較はできないので、注意が必要である。

ところで、学生調査では芸術・文化の需要構造の実態把握に焦点を当てており、調査票の質問は、大別すると、実演芸術7分野44ジャンル(各回調査でジャンルの見直しが行われているので調査回によってジャンル数は異なる)のライブでの鑑賞経験(過去通算と過去1年)と鑑賞希望と実演芸術7分野のメディアによる鑑賞経験(過去1年)を問う質問群、映画鑑賞と美術などの視覚芸術鑑賞の経験と鑑賞場所と同行者を問う質問群、鑑賞行動に影響を与える文化・芸術への意識に関する質問群、稽古事や主体的芸術活動の有無と活動時期に関する質問群、読書(書籍と雑誌)ならびに実演芸術に関する情報の取得源に関する質問群、および鑑賞行動と何らかの関連を持つと考えられる個人的属性(文化資本)に関する事項を問う質問群とから構成されている。文化資本に関する項目は、具体的には、全ての家族構成員について、続柄、年齢、職業、

Q6. あなたは、文学以外の芸術・芸能（映画も含む）について、どのように考えていますか。

以下のAとBの対照的な考え方の組合せのうち、あなたの考え方に近い方の記号に○を付けて下さい。  
 どちらもいえないと思われる場合は、AとBの間のA/Bに○を付けて下さい。  
 A/B以外の意見がある場合は、自由回答欄に記入して下さい。

	A	B	回答欄		
1.	日本的なものを育てていきたい。	外国からいいものを取り入れていけばよい。	A	AB	B
2.	古典的なものを守っていきたい。	現代的な感覚を活かしていきたい。	A	AB	B
3.	テーマや素材には日本のなものを積極的に取り入れるべきだ。	テーマや素材は別に日本のなものにこだわる必要はない。	A	AB	B
4.	芸術復興のために政府・企業の積極的な援助・支援が必要だ。	芸術復興のためには、芸術・芸能家の努力が先決だ。	A	AB	B
5.	芸術・芸能は生（ライブ）でなければ鑑賞したとはいえない。	芸術・芸能は、生でなくても十分に鑑賞できる。	A	AB	B
6.	芸術・芸能は国の違いを超えて普遍だ。	芸術・芸能には、その国産でなければ理解できない面がある。	A	AB	B
7.	余暇の多様化に伴い、芸術・芸能の中で競逐するものも出る。	余暇の時代だからこそ、今後は芸術・芸能が発展していく。	A	AB	B
8.	これからは、芸術・芸能はアマチュアに広がっていく。	これからは、芸術・芸能は専門家中心のものとなっていく。	A	AB	B
9.	芸術・芸能では、教育よりも才能の方が重要である。	芸術・芸能には、才能よりも教育の果たす役割が大きい。	A	AB	B
10.	芸術・芸能は時代とともに変化していくべきだ。	芸術・芸能は時代に迎合する必要はない。	A	AB	B
自由回答欄：芸術・芸能について意見があれば、自由に書いて下さい。					

図1 文化・芸術への意識を捉えるための質問

最終学歴、居住地を問う質問を設けている。また、奨学金の状況、アルバイトの状況、クラブ活動やサークル活動の状況、1ヶ月の小遣いの額なども設問している。

調査方法は、原則として講義などで学生に調査票を配布して、その場で記入させるか、または自宅で記入して次回講義時に回収する方法で実施されている。

第2回調査からの鑑賞経験率（過去1年間に当該分野あるいは当該ジャンルの実演芸術を1回以上鑑賞した学生の比率）の変化については有馬・福永・王[3]で、また稽古事や主体的芸術活動がライブでの実演芸術の鑑賞率に及ぼす影響については福永・有馬[7]で、家族の職業や学歴などの文化資本の影響が鑑賞行動に及ぼす影響については有馬[4]で分析しているが、本研究では、文化・芸術に関する10項目の対照的な意見のどちらに自身の考えが近いかを問う図1に示した質問群の回答データを用いて、文化・芸術への意識の現状と第2回調査からの意識の変化について、クロス集計と対応分析を適用して分析を試みている。

### 3. 文化・芸術に対する意識とその変化

文化・意識を捉えるための質問群は、①日本志向か外国志向か、②古典志向か現代志向か、③日本志向か外国志向か（①とほぼ同じ）、④補助支援か自助努力か、⑤ライブ鑑賞かメディア鑑賞か、⑥国際的普遍性か地域固有性か、⑦衰退か発展か、⑧アマチュア化かプロフェッショナル化か、⑨才能重視か教育重視か、⑩流行か不易か、の10の対照的な意見のどちらに自分自身の考えが近いかを問う形式で構成されている。

表2には、第2回調査から第5回調査までの4回の調査で回答がどのようになっているかを回答者全体と性別で分けた場合の集計結果として示している。

表2より、有意抽出のために統計的検定は行えないものの、学習指導要領の改訂を受けて、1の「日本的なものを育てていきたい」、2の「古典的なものを守っていきたい」、3の「テーマや素材には日本のなものを積極的に取り入れるべきだ」という意見に賛成する比率が第4回調査までと比較して第5回調査では高くなっており、1の「外国からいいものを取り入れていけばよい」、2の「現代的な感覚を活かしていきたい」、3の「テーマや素材は別に日本のなものにこだわる必要はない」、6の「芸術・芸能は国の違いを超えて普遍だ」という意見に賛成する比率は、第4回調査までと比較して第5回調査では低くなっている。

表2 文化芸術への意識の現状とその変化

		全体									
		性別									
		男子学生			女子学生						
	左の意見に賛成	どちらともいえない	右の意見に賛成	左の意見に賛成	どちらともいえない	右の意見に賛成	左の意見に賛成	どちらともいえない	右の意見に賛成		
1	日本的なものを育てていきたい。	16.1	60.1	23.9	18.6	55.7	25.7	12.4	66.2	21.4	外国からいいものを取り入れていけばよい。
		14.4	59.0	26.6	17.0	53.7	29.3	11.3	65.5	23.2	
		18.5	65.3	16.2	21.5	59.8	18.8	15.9	70.3	13.8	
		28.9	59.3	11.9	31.5	54.8	13.7	25.8	64.6	9.7	
2	古典的なものを守っていきたい。	13.5	55.1	31.5	13.8	50.3	36.0	13.0	61.7	25.3	現代的な感覚を活かしていきたい。
		12.3	53.0	34.8	12.9	48.4	38.8	11.5	58.6	29.9	
		14.7	59.4	25.9	13.8	53.5	32.7	15.4	64.7	19.9	
		18.4	56.9	24.8	18.2	52.4	29.4	18.7	61.9	19.5	
3	テーマや素材には日本的なものを積極的に取り入れるべきだ。	14.1	29.7	56.3	15.2	29.5	55.4	12.6	29.9	57.5	テーマや素材は別に日本的なものにこだわる必要はない。
		13.3	31.9	54.7	14.3	32.0	53.7	12.1	32.0	55.9	
		16.4	38.4	45.2	16.0	37.3	46.8	16.6	39.6	43.8	
		21.5	40.3	38.3	22.9	38.4	38.7	19.9	42.2	37.9	
4	芸術振興のために政府・企業の積極的な支援が必要だ。	48.9	35.5	15.5	49.9	33.2	16.9	47.7	38.7	13.6	芸術振興のためには、芸術・芸能家の努力が先決だ。
		40.0	41.1	18.9	39.1	40.1	20.9	41.1	42.3	16.6	
		42.7	40.3	17.0	39.1	41.4	19.5	46.0	39.3	14.7	
		44.7	40.4	14.9	41.3	39.9	18.7	49.1	40.4	10.6	
5	芸術・芸能は生(ライブ)でなければ鑑賞したとはいえない。	26.7	36.3	37.0	27.3	35.1	37.6	25.8	37.9	36.3	芸術・芸能は、生でなくても十分に鑑賞できる。
		22.7	36.8	40.5	23.9	36.9	39.3	21.2	36.9	41.9	
		20.3	36.3	43.4	23.3	35.8	40.9	17.7	36.7	45.6	
		23.2	37.8	39.0	25.4	37.0	37.6	20.9	38.6	40.4	
6	芸術・芸能は国の違いを超えて普遍だ。	42.1	28.0	29.9	36.4	28.0	35.6	50.0	28.0	22.1	芸術・芸能には、その国民でなければ理解できない面がある。
		41.5	29.5	29.0	36.6	29.7	33.7	47.0	29.3	23.7	
		41.5	30.8	27.7	36.6	32.8	30.6	45.9	28.9	25.2	
		35.7	33.5	30.8	32.0	32.6	35.4	39.9	34.5	25.6	
7	余暇の多様化に伴い、芸術・芸能の中で衰退するものも出る。	28.4	33.5	38.1	34.1	33.4	32.6	20.4	33.8	45.8	余暇の時代だからこそ、今後は芸術・芸能が発展していく。
		30.9	35.3	33.7	35.5	35.5	29.0	25.7	35.2	39.2	
		35.1	37.1	27.8	37.1	39.2	23.7	33.2	35.1	31.7	
		36.0	39.8	24.2	40.1	38.8	21.2	31.7	40.8	27.5	
8	これからは、芸術・芸能はアマチュアに広がっていく。	51.3	38.9	9.7	48.4	39.3	12.3	55.4	38.4	6.2	これからは、芸術・芸能は専門家中心のものとなっていく。
		46.3	43.6	10.1	42.2	44.3	13.6	51.0	42.9	6.1	
		46.5	43.7	9.8	40.4	47.1	12.5	52.2	40.6	7.2	
		40.1	47.0	12.9	36.7	46.8	16.5	44.2	47.0	8.8	
9	芸術・芸能では、教育よりも才能の方が重要である。	44.8	42.2	13.0	47.1	38.4	14.5	41.6	47.5	10.9	芸術・芸能には、才能よりも教育の果たす役割が大きい。
		45.0	42.6	12.4	45.5	40.7	13.9	44.3	45.0	10.7	
		35.7	48.6	15.7	37.1	45.1	17.8	34.4	51.8	13.8	
		35.2	48.0	16.8	37.0	45.0	18.0	33.6	51.0	15.3	
10	芸術・芸能は時代とともに変化していくべきだ。	39.8	40.6	19.6	42.9	35.9	21.3	35.6	47.2	17.2	芸術・芸能は時代に迎合する必要はない。
		42.1	40.0	18.0	45.2	35.5	19.3	38.5	45.2	16.3	
		40.0	44.6	15.4	43.5	39.3	17.2	36.7	49.5	13.8	
		39.4	44.0	16.6	41.2	40.5	18.3	37.6	47.9	14.5	
サンプル数	第1段 第2回調査	10,754			第1段 第2回調査	6,108			第1段 第2回調査	4,480	
	第2段 第3回調査	9,119			第2段 第3回調査	4,769			第2段 第3回調査	4,138	
	第3段 第4回調査	3,370			第3段 第4回調査	1,600			第3段 第4回調査	1,759	
	第4段 第5回調査	7,098			第4段 第5回調査	3,673			第4段 第5回調査	3,264	

特に、2002年の第4回調査と2008年の第5回調査との間では、1の「日本的なものを育てていきたい」は10.4ポイントの増加（1996年の第3回調査と比較すると14.5ポイントの増加）に対して「外国からいいものを取り入れていけばよい」は4.3ポイントの減少（1996年の第3回調査と比較すると14.7ポイントの減少）で、賛成率も2002年の第4回調査までと2008年の第5回調査との間では逆転をしている。3の「テーマや素材には日本的なものを積極的に取り入れるべきだ」は「テーマや素材は別に日本的なものにこだわる必要はない」と比較して賛成率はほぼ半分であるが、2002年の第4回調査と2008年の第5回調査との間では5.1ポイントの増加（1996年の第3回調査と比較すると8.2ポイントの増加）に対して、「テーマや素材は別に日本的なものにこだわる必要はない」は6.9ポイントの減少（1996年の第3回調査と比較すると18.0



ポイントの減少)」、6の「芸術・芸能には、その国民でなければ理解できない面がある」は「芸術・芸能は国の違いを超えて普遍だ」を賛成率では下回ってはいるが、2002年の第4回調査と2008年の第5回調査との間で3.1ポイントの増加に対して「芸術・芸能は国の違いを超えて普遍だ」は5.8ポイントの減少を示しており、日本文化や我が国固有の芸術・芸能への回帰の傾向が読み取れる。また、2の「古典的なものを守っていききたい」は、2002年の第4回調査と2008年の第5回調査との比較で3.7ポイントの増加(1996年の第3回調査と比較すると6.1ポイントの増加)に対して「現代的な感覚を活かしていききたい」は1.1ポイントの減少(1996年の第3回調査と比較すると10.0ポイントの減少)を示している。

以上から、今後の詳細な分析が必要ではあるが、「我が国の文化や伝統を尊重する態度を育成する観点などから、小学校から我が国に伝わる楽器を取り上げられるようにするとともに、中学校では3年間を通じて1種類以上の和楽器を体験できるようにする」という指導方針を打ち出した学習指導要領の改訂の影響は、文化・芸術に対する大学生の意識の変化に表れているのではないかと考えられる。

芸術・芸能の発展および広がりについては、「余暇の時代だからこそ、今後は芸術・芸能が発展していく」が1996年の第3回調査までは「余暇の多様化に伴い、芸術・芸能の中で衰退するものも出る」を賛成率で上回っていたが、2002年の第4回調査で逆転しており、「これからは、芸術・芸能はアマチュアに広がっていく」は賛成率で「これからは、芸術・芸能は専門家中心のものとなっていく」を上回ってはいるものの、賛成率は低下してきており、学生は芸術・芸能の発展および広がりについて、バブル期と比較して楽観的には見ていないことが窺える。ただし、芸術・芸能は才能が必要か教育でカバーできるかに関しては、「芸術・芸能では、教育よりも才能の方が重要である」の賛成率が「芸術・芸能には、才能よりも教育の果たす役割が大きい」の賛成率を1996年の第3回調査では32.6ポイントも上回っていたのに対して、2008年の第5回調査では18.4ポイントの差に縮まってきており、教育によって芸術・芸能が広がっていく可能性も意識はされているように考えられる。

インターネットの普及に伴う視聴覚メディアの革新と発展の影響に関しては、2002年の第4回調査の時点では「芸術・芸能は生でなくても十分に鑑賞できる」が「芸術・芸能は生(ライブ)でなければ鑑賞したとはいえない」を23.1ポイント上回っていたが、2008年の第5回調査では15.8ポイントの差に縮まってきている。

文化・芸術に対する意識の男女間の差については、表2から、女子学生の方にどちらともいえないという中庸の意識が高い傾向が見られる。

#### 4. 文化・芸術に対する意識の構造

次ページの図2に示したのは、10の質問の間で作成したモザイク図行列の一部である。図2から類推されるように、10の対照的な意見のペアのどちらかに賛成するかという回答には、項目間で関連性があることが示唆される。

そこで、表側に7,074人の学生、表頭に10の対照的な質問への30の回答選択肢(10の質問に対してAかABかBのどれか)をとって作成した反応行列に対応分析(Correspondence Analysis)を適用し、同じように選択される回答選択肢と10の質問に同じような回答をする学生を分類することを試みた。

主成分分析の適用結果は表3に示す通りであり、3つの主成分スコアのそれぞれで降順に回答選択肢を並べ替えた表から、第1主成分は10の項目についてABの「どちらともいえない」という中庸の選択肢とAまたはBの明確な意見表示の選択肢を弁別する軸であり、第2主成分は1から3までの項目で日本の伝統志

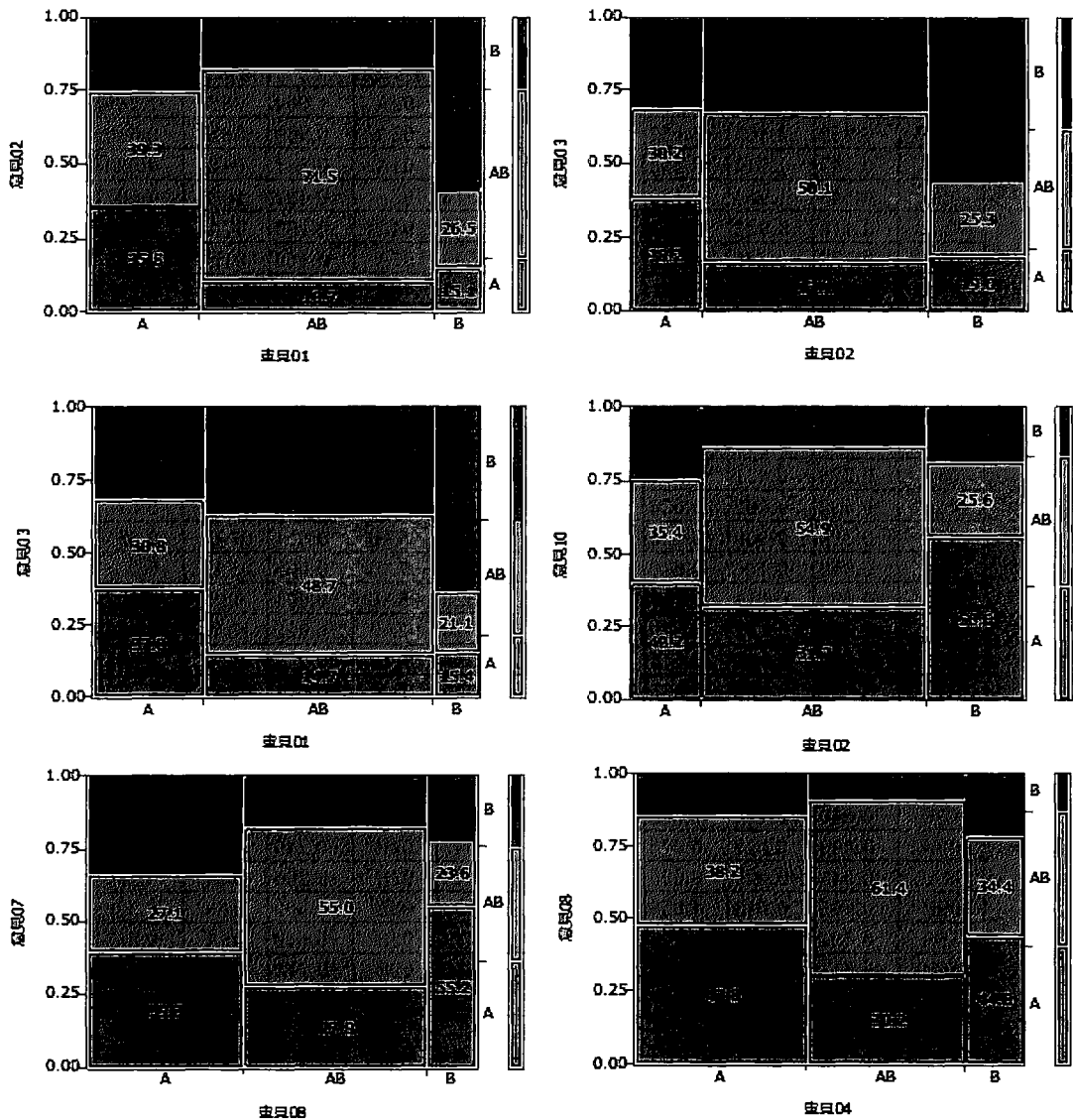


図2 文化・芸術の意識を問う質問間のモザイク図行列の一部

向の A の回答選択肢と外国にも目を向けて現代志向の B の回答選択肢を弁別する軸、第3主成分は8と6と4と10のBの回答選択肢と7のBおよび6と8のAの回答選択肢が両極端に位置しており、「芸術・芸能は時代に迎合することなく専門家中心でその国の国民でなければ理解できない面があって振興も自助努力で対応すべき」という文化・芸術を突き放して見る見方と「芸術・芸能は国の違いを超えて普遍的でアマチュアにも広がっていき発展していく」という楽観的で文化・芸術に寄り添った見方とを弁別する軸であると考えられる。

この対応分析によって得られた第1主成分と第2主成分と第3主成分のサンプルスコアから、性別の各主成分のサンプルスコアの平均を計算すると、男子学生は第1主成分から順に 0.0420, -0.0023, 0.0510 であり、一方の女子学生は-0.0435, 0.0065, -0.0611 であり、第1主成分から女子学生が AB の中庸の回答選択肢を選択する傾向が強いこと、第2主成分からはほとんど差はないが、女子学生は1から3までの項目でAを選択

表3 対応分析の結果

選択肢	c1	c2	c3	選択肢	c1	選択肢	c2	選択肢	c3
01A	0.426	0.841	0.164	01B	0.747	02A	1.064	08B	1.286
01AB	-0.358	-0.167	-0.122	08B	0.680	03A	0.969	06B	0.570
01B	0.747	-1.213	0.213	09B	0.605	01A	0.841	04B	0.523
02A	0.519	1.064	0.130	04B	0.573	05A	0.421	10B	0.507
02AB	-0.416	-0.022	-0.083	02B	0.570	04A	0.168	07A	0.442
02B	0.570	-0.746	0.094	10B	0.549	07A	0.166	01B	0.213
03A	0.443	0.969	0.001	02A	0.519	08B	0.142	04AB	0.176
03AB	-0.589	0.073	-0.040	10A	0.457	06A	0.106	01A	0.164
03B	0.372	-0.619	0.042	07B	0.452	03AB	0.073	06AB	0.149
04A	0.305	0.168	-0.336	09A	0.449	10AB	0.068	02A	0.130
04AB	-0.548	-0.097	0.176	03A	0.443	05AB	0.062	05B	0.125
04B	0.573	-0.249	0.523	01A	0.426	10B	0.054	08AB	0.120
05A	0.328	0.421	-0.296	07A	0.420	09A	0.029	02B	0.094
05AB	-0.532	0.062	0.054	08A	0.420	09B	0.013	09A	0.075
05B	0.320	-0.309	0.125	06B	0.388	06AB	0.011	07AB	0.055
06A	0.318	0.106	-0.630	03B	0.372	07AB	0.006	05AB	0.054
06AB	-0.699	0.011	0.149	05A	0.328	08AB	-0.001	03B	0.042
06B	0.388	-0.132	0.570	05B	0.320	02AB	-0.022	10AB	0.026
07A	0.420	0.166	0.442	06A	0.318	09AB	-0.025	03A	0.001
07AB	-0.655	0.006	0.055	04A	0.305	08A	-0.041	09AB	-0.035
07B	0.452	-0.256	-0.746	01AB	-0.358	04AB	-0.097	03AB	-0.040
08A	0.420	-0.041	-0.555	02AB	-0.416	10A	-0.098	09B	-0.058
08AB	-0.543	-0.001	0.120	05AB	-0.532	06B	-0.132	02AB	-0.083
08B	0.680	0.142	1.286	08AB	-0.543	01AB	-0.167	01AB	-0.122
09A	0.449	0.029	0.075	09AB	-0.543	04B	-0.249	10A	-0.245
09AB	-0.543	-0.025	-0.035	04AB	-0.548	07B	-0.256	05A	-0.296
09B	0.605	0.013	-0.058	03AB	-0.589	05B	-0.309	04A	-0.336
10A	0.457	-0.098	-0.245	10AB	-0.613	03B	-0.619	08A	-0.555
10AB	-0.613	0.068	0.026	07AB	-0.655	02B	-0.746	06A	-0.630
10B	0.549	0.054	0.507	06AB	-0.699	01B	-1.213	07B	-0.746

する可能性が男子学生よりも高く、日本的で伝統的な芸術・芸能を志向する傾向が見られるのに対して、男子学生は現代的なものを活かして素材やテーマは日本的なものにはこだわらず、外国からよいものを取り入れていけばよいという進取の開放的な志向を有している傾向が見られることが示される結果となっている。ただし、第3主成分からは、女子学生は「芸術・芸能は国の違いを超えて普遍的でアマチュアにも広がっていき発展していく」という楽観的で文化・芸術に寄り添った見方をする傾向がある一方で、男子学生は「芸術・芸能は時代に迎合することなく専門家中心でその国の国民でなければ理解できない面があって振興も自助努力で対応すべき」という文化・芸術を突き放して見る見方をする傾向のあることが示唆されている。

## 5. 文化・芸術への意識と鑑賞行動との関係

文化・芸術への意識について、クロス集計結果ならびに対応分析を適用して分析を試みたが、文化・芸術への意識の違いは、実際の文化・芸術の需要行動にも影響を及ぼすはずである。

学生調査では、既述のように実演芸術のライブやメディアによる鑑賞行動や稽古事ならびに主体的芸術活動の経験を設問しているが、例えば、伝統演劇（能、狂言、歌舞伎など）に関しては、項目1のAの回答選

択肢（「日本的なものを育てていきたい」）の回答者群の過去1年の鑑賞経験率は5.22%であるのに対して、項目1のBの回答択肢（「外国からいいものを取り入れていけばよい」）の回答者群の鑑賞経験率は2.22%であるというように、鑑賞経験率に及ぼす影響が認められる。

## 6. まとめと今後の課題

本研究では、1991年から2008年にかけて実施した4回の学生調査のデータに基づき、学生の文化・芸術に対する意識の現状と変化の一端を明らかにした。しかし、有意抽出による調査のために時系列変化の有無を検証できないなどの問題がある。

今後は、性別に加えて、学部別や地域別などで分析することでサンプルの偏りができるだけ生じないグループ内での時系列比較を行うとともに、対応分析による分析をさらに深めて文化・芸術への意識に基づく学生の分類を試み、意識が行動にどのように影響するかを分析するなど、研究をさらに発展させていきたい。

## 謝辞

本研究はこれまで5回にわたって実施されてきた「学生調査」（第4回調査までの経緯は周防[6]を参照）の研究成果に基づいて行われている。これらの調査に関係された諸先生ならびに回答者である学生の皆さんに感謝申し上げます。また、第5回調査のデータ入力にあたっては、科学研究費補助金基盤研究（B）「周期統計調査のマイクロデータによるコーホート分析－文化需要の実証的研究－」（課題番号：23330073、研究代表者：勝浦正樹）の助成を受けており、データの整理にあたっては、兵庫県立大学応用情報科学研究科の博士前期課程の学生であった加藤優希、福永征世、王程、李秋雲の諸氏の協力を得ている。

## 参考文献

- [1] 有馬昌宏、「全国学生調査に基づく実演芸術鑑賞行動の規定要因の分析」、『2008 SAS ユーザー総会 アcademia/テクノロジー&ソリューションセッション 論文集』, pp.93-102, 2008.
- [2] 有馬昌宏、「鑑賞・観覧活動の変化と現状」、『統計』, 第64巻第3号, pp.23-28, 2013.
- [3] 有馬昌宏・福永征世・王程、「学生の実演芸術鑑賞構造とその変化－学生調査と社会生活基本調査の結果から－」、『SAS ユーザー総会Academia/テクノロジー&ソリューションセッション 2013 論文集』, pp.115-124, 2013.
- [4] 有馬昌宏、「学生の芸術鑑賞活動を規定する要因に関する分析」、『2015 SAS ユーザー総会 アcademia/テクノロジー&ソリューションセッション 論文集』, pp.93-102, 2015.
- [5] 加藤優希・有馬昌宏、「5回の学生調査から探る実演芸術鑑賞行動パターンとその規定要因～学生調査データベースの構築と分析を通して～」、『2010 SAS ユーザー総会 アcademia/テクノロジー&ソリューションセッション 論文集』, pp.387-395, 2010.
- [6] 周防節雄（編）、『芸術・文化政策立案のための統計指標の開発と体系化に関する研究』（平成13年度～平成15年度科学研究費補助金(特別研究促進費(1))研究成果報告書）, 2004.

- [7] 福永征世・有馬昌宏,「学生の芸術文化鑑賞活動の現状―第5回学生調査の10%抽出データ分析から―」,  
『SAS ユーザー総会アカデミア・テクノロジー&ソリューションセッション in 神戸 2011 論文集』,  
pp.81-87, 2011.

全国消費実態調査の匿名データから見た  
食料品目別消費傾向による食生活の変化について

○池田 瑞穂<sup>1</sup>, 笹谷 知輝<sup>2</sup>, 大野 嵩護<sup>2</sup>

(<sup>1</sup>京都大学大学院農学研究科, <sup>2</sup>関西学院大学)

Changes in eating Habits by the Food Consumption Tendency  
Using Micro Data of National Survey of Family Income and Expenditure

○Mizuho Ikeda<sup>1</sup>, Tomoki Sasaya<sup>2</sup>, Shugo Ono<sup>2</sup>

(<sup>1</sup>Graduate School of Agriculture, Kyoto University <sup>2</sup>Kwansei Gakuin University)

## 要旨

全国消費実態調査の匿名データを用いて、調査年毎に魚介類支出に対する「若者の魚離れ」の実態や世帯属性の影響に関して検証してきた。本稿では、匿名データを用いたデータハンドリング、データ分析、および、評価に至るまでについて述べる。そして、SASのSQLプロシジャを用いた例を紹介する。

また、等価所得に換算した結果を用いて回帰分析をおこない、これまでの結果と比較した。世帯主の年齢階級の低い世帯ほど魚介類を消費していないなどの結果を同じく得たことを確認した。

キーワード: 匿名データ, 全国消費実態調査, SAS, SQLプロシジャ, GLMプロシジャ

## 1 はじめに

これまで、平成元年、6年、11年、16年の「全国消費実態調査」の匿名データを用いて若者の魚離れの実態を検証してきた [1]。そして、各調査年の世帯主の年齢階級別「魚介類」支出額割合を算出し、他の食料品目や各調査年のデータ間を比較したところ、若いころから消費傾向が決まっていると考えられ、「若者の魚離れ」の対策として若年層からの魚食に関する食育の重要性が高いことを示した。さらに、平成16年二人以上世帯のデータを用いて世帯主の年齢階級や他の世帯属性が魚介類支出に与える影響を調べるため分析をおこなった。等価所得に換算したデータを用い先行研究[2]との回帰分析結果の比較をおこなったところ、等価所得を用いることで「3大都市圏在住世帯」が負の値に転じたが、魚介類支出に対しての考察の影響は殆どないことを確認した。

匿名データを用いたデータ分析にあたり、データハンドリング部分が後に続くデータ分析に大きな影響を与えると考えた。そこで、データ準備の手法とデータハンドリング時に用いたSASのSQLプロシジャのプログラムを紹介する。

## 2 匿名データを用いたデータハンドリング

匿名データは集計用乗率を用いるためデータ量が20万件を超えるほど膨大となり、データ解析の際、時間がかかりかかることが予想された。そこで、コンピュータ環境の整備に要するコストとデータ分析をおこなう時間の削減のため、対象とするデータを絞り込むこととした。

## 2.1 変数の絞り込みと共通変数の作成

匿名データに関する「符号表」が統計センターの Web サイトに公開されている。コンピュータ環境の制約や、解析をおこなう時間などのコスト削減のため、平成 16 年二人以上世帯の符号表を基にあらかじめ今回の研究に必要なと予測した変数を 1,780 変数（約 995MB、集計用乗率計算前）から 170 変数（約 32MB、集計用乗率計算前）に絞り込んだ。

### 2.1.1 共通変数テーブルの作成

分析対象とした全国消費実態調査二人以上世帯 4 年分の変数の番号や項目名は調査年によって異なっていた。そこで平成 16 年の変数をもとに共通変数を作成し、他の調査年と共通する変数のマッピングをおこなった。調査年によっては存在しない項目があるが、本研究では 4 年分同じ変数にマッピングできた共通変数を用いている。表 1 に共通変数とマッピングした変数の一部を示す。平成 16 年のデータの変数は、V0025 の次が V0038 であるように連続していない。そこで、SAS のデータ読み込みの処理の簡便さのためにできるだけ連続する値となるようにソートなどの処理をおこなった。これらの処理により、調査年をマクロ変数に指定するだけで、各調査年に対して同じプログラムを用いてデータの加工、および、検証をおこなうことができた。

### 2.1.2 データの階層構造の変換

匿名データは、データ属性「階層」の 1~4 で構造が表現されている。また、共通変数の項目名は「世帯票\_\_世帯員に関する事項(1)\_\_世帯員番号」のように階層ごとに\_\_（アンダースコア）で名称が区切られており、階層が一目でわかるようになっている。本研究でのデータ分析では最下位階層の名を利用するため、その簡便性から、階層別に項目名を分け共通変数テーブルに追加した（表 1）。

### 2.1.3 新しいテーブルの生成

共通変数テーブルと符号表（データ部のみ）は全て csv ファイルとして保存した。これらのテーブルをデータセットとして取り込んだ後、SQL プロシジャを用いてデータ加工や複数のファイルを結合し、新しいデータセットを作成する方法を用いた。

図 2 のプログラムは本研究で用いる基本のステートメントである。マクロ macro\_nm の引数である dsi はデータの参照元のデータセットであり、dso は新たに作成するデータの出力先のデータセットである。マクロ中ではマクロ変数 &dsi、&dso として扱われる。

図 3 の例に示すように、データセット間の結合をおこなった結果を、新規に作成したデータセットに格納し、データ分析に利用した。変数「共通変数」を含むデータセット &dsi1 に各調査年のデータセット &dsi2 を左外部結合し、その結果をデータセット &dso に格納する例である。ここで di1.cvar は、データセット &dsi1（別名 di1）に存在する変数 cvar という意味である。select 文の直後に、「共通変数」と「変数名」を指定しているのでそれらが最初にデータセット &dsi2 に格納される。

&dsi1 の変数と &dsi2 の変数の等価を条件として左外部結合（left join）を行っている。この場合、&dsi1 に存在する変数「共通変数」に対応する &dsi2 の変数がない場合、&dsi1 のデータはそのまま保存され、対応する &dsi2 の変数欄は空欄となる。左外部結合以外の場合、条件を満足するデータがない場合、抽出する値が変わる。

表1 共通変数と各調査年の変数とのマッピング

1	2	階層別項目名	3	4	共通変数	2004年		1999年	
						変数名	階層	変数名	階層
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
世帯票	世帯人員				w11	V0017	2	V0017	2
世帯票	就業人員				w12	V0018	2	V0018	2
世帯票	世帯員以外の家族	支える人		不在理由	w13	V0020	4	V0020	4
世帯票	世帯員以外の家族	その他の人		入院	w14	V0021	4	V0021	4
世帯票	世帯員以外の家族	その他の人		学業	w15	V0023	4	V0023	4
世帯票	世帯員以外の家族	その他の人		その他	w16	V0025	4	V0025	4
世帯票	単身世帯の形態				w17	V0038	2	V0028	2
世帯票	世帯員に関する事項(1)	世帯員番号			w18	V0039	3	V0043	3
世帯票	世帯員に関する事項(1)	続柄			w19	V0040	3	V0044	3
世帯票	世帯員に関する事項(1)	性別			w20	V0041	3	V0045	3
世帯票	世帯員に関する事項(1)	年齢5歳階級			w21	V0042	3	V0046	3
世帯票	世帯員に関する事項(1)	年齢各歳(15歳未満のみ)			w22	V0043	3	V0047	3
世帯票	世帯員に関する事項(1)	就業・非就業の別			w23	V0044	3	V0048	3
世帯票	世帯員に関する事項(1)	企業区分			w24	V0045	3	V0049	3
世帯票	世帯員に関する事項(1)	企業規模			w25	V0046	3	V0050	3
世帯票	世帯員に関する事項(1)	産業符号			w26	V0047	3	V0052	3
世帯票	世帯員に関する事項(1)	職業符号			w27	V0048	3	V0053	3
世帯票	世帯員に関する事項(1)	国公立・私立の別			w28	V0049	3	V0054	3
世帯票	世帯員に関する事項(1)	学校の種類			w29	V0050	3	V0055	3
世帯票	世帯員に関する事項(1)	専修学校			w30	V0051	3	V0056	3
世帯票	世帯員に関する事項(1)	各種学校・塾など			w31	V0052	3	V0057	3
世帯票	世帯員に関する事項(2)	世帯員番号			w32	V0053	3	V0058	3
世帯票	:	:			:	:	:	:	:
世帯票	:	:			:	:	:	:	:
世帯票	世帯員に関する事項(2)	各種学校・塾など			w45	V0066	3	V0072	3
世帯票	:	:			:	:	:	:	:
世帯票	:	:			:	:	:	:	:
世帯票	世帯員に関する事項(7)	世帯員番号			w102	V0123	3	V0133	3
世帯票	:	:			:	:	:	:	:
世帯票	:	:			:	:	:	:	:
世帯票	世帯員に関する事項(7)	各種学校・塾など			w116	V0136	3	V0147	3
用途	調査月数				w117	V0385	2	V0391	2
:	:	:			:	:	:	:	:
:	:	:			:	:	:	:	:

```
(1) options obs=100;
(2) options firstobs=1,000 obs=1,001,000; /* 対象を1,000~101,000 オブザベーションとする */
```

図1 テスト対象のオブザベーションの指定



```

/*-----
データセット &dsiから取得したデータを
データセット &dsio に格納する.
-----*/

%macro macro_nm(dsi, dso);
proc sql ;
create table &dso as
select
* /* こちらに記述された変数を &dso に格納する. */
/* 格納する変数はコンマ区切りで列挙する. 「*」の場合, すべての変数が対象となる. */
from &dsi as di1 /* データ取得元のデータセット, as 以降は &dsi の別名であり省略してよい. */
quit;
%mend macro_nm;

```

図 2 SQL プロシジャの基本ステートメント

```

/*-----
データセット &dsi1 と &dsi2 を 左外部結合した結果を
データセット &dsio に格納する.
-----*/

%macro add_common_var(dsi1, dsi2, dso);
proc sql ;
create table &dso as
select
di1.cvar, di1.var /* cvar:共通変数, var:変数 */
, * /* その他全ての &dsi1 の変数, その他すべての &dsi2 の変数 */
from &dsi as di1
left join dsi2 as di2 /* 左外部結合は結合時にテーブル &dsi の全データが保持される */
on di1.var=di2.var; /* 条件: 結合するデータセットの変数が等しい */
quit;
%mend add_common_var;

```

図 3 データセット間の結合例

```

/*-----
データセット &dsi の
条件を満たすかどうか判断しその結果を新しい変数に格納する
条件を満たすデータを抽出し
データセット &dsio に格納する.
-----*/

%macro ext_data ( dsi, dso);
proc sql ;
create table &dso as
select
ifn((w399>=4,000,000),1,0) as fig_nensyu /* w399 : 年間収入 */
,ifc((lengthn(w076)=1, 'a', 'b') as fig_syokugyo /* w078 : 職業符号 */
, * /* w399, w078 以外のデータ */
from &dsi
where layer=1; /* 階層が1と等しい */
quit;
%mend ext_data;

```

図 4 データの抽出例

図4はデータセット &dsi に対して条件を満足したデータを抽出するステートメントの例である。select に続くステートメントにおいて、変数 w399 が 4,000,000 以上の場合 1 を、条件を満足しない場合 0 が as で指定している変数 flg\_nensyu に格納される。変数 w076 の文字の長さが 1 の時文字 'a' を、条件を満足しない場合文字 'b' が変数 flg\_syougyo に格納される。ifn の戻り値は数値であり、ifc は文字である。

## 2.2 対象とするオブザベーションの指定

SAS プログラムを実行する際、オブザベーションの調整を図った。

- (1) まず、1 調査年を対象とし、プログラムのテスト用に数 10~10,000 オブザベーションに絞り込み、動作確認をおこなった (図1 (1))。回帰分析等においては約 100,000 オブザベーションを対象にさまざまな分析をおこない傾向を見た。
- (2) データ対象の開始オブザベーションやオブザベーション数を変更し同じ分析をおこなった (図1 (2))。それぞれの結果を対比しずれが大きくないことを確認しつつ、さまざまな分析をおこなった。
- (3) 最後に全オブザベーションを対象にデータ分析をおこなった。

## 3 データ分析の目的と手法

世帯主の年齢階級を 12 階級に分類し、家族形態、在住地域の観点から世帯属性の魚介類支出やその他の支出への影響を探った。家族形態による比較分析として、「年間収入」、「共働きか否か」、「世帯主の子供 (20 歳未満) の有無」、および、「家族形態に 65 歳以上の世帯員の存在の有無」による比較分析をおこなった。

### 3.1 分析の対象データ

匿名データをもとに加工した「世帯主の実子とその親等で構成される世帯」、「未成年のいない世帯」と、両者を併せたものを基に回帰分析をおこなった。本稿では等価所得に換算した値を用いた (式(1))。食料支出は、森[3]に倣い世帯構成の違いを考慮するだけでなく、「エネルギーの食事摂取基準：推定エネルギー必要量」[4]に基づいて、男性の 20 歳から 49 歳 (以下、成人男性) を 1 としたときの各世帯員の重みを作り、世帯ごとの重みの合計値で支出額を割ることで、その世帯の成人男性一人あたりの消費量を算出した。

$$\text{等価所得} = \text{「世帯の年間収入」} / \text{sqrt(「世帯人員」)} \quad (1)$$

### 3.2 食料支出に関する分析

#### 3.2.1 魚介類支出に関して

被説明変数は、成人男性で換算された魚介類支出額とし、世帯属性が魚介類消費に与える影響について GLM プロシジャを用いて回帰分析をおこなった。説明変数の定義を表 2 に示す。使用したモデルは式 (2) のとおりである。

$$\begin{aligned} fish_i = & \beta_1 \text{annual\_income}_i + \beta_2 \text{female\_income}_i + \beta_3 \text{agri\_fish\_income}_i + \beta_4 \text{over65}_i + \beta_5 \text{urban}_i \\ & + \beta_6 \text{age\_head}_i + \beta_7 \text{children}_i + \beta_8 \text{two\_incomes}_i + \beta_9 \text{agri\_fish\_income}_i \times \text{urban}_i \\ & + \beta_{10} \text{age\_head}_i \times \text{urban}_i + \beta_{11} \text{age\_head}_i \times \text{children}_i + \epsilon_i. \end{aligned} \quad (2)$$

agri fish income × urban は「農林漁業収入有世帯」と「3 大都市圏在住世帯」、age head × urban は「世帯主年齢階級」と「3 大都市圏在住世帯」、age head × children は「世帯主年齢階級」と「実子を含む世帯」の交互作用項である。

表 2 変数説明

変数名	定義
annual income (年間収入)	世帯の年間収入(万円)
female income (女性収入)	世帯主が女の収入+配偶者が女の収入(万円)
agri fish income (農林漁業収入有世帯, ダミー変数)	農林漁業収入があれば 1
over65 (65 歳以上を含む世帯, ダミー変数)	65 歳以上の世帯人員がいれば 1
urban (3 大都市圏在住世帯, ダミー変数)	3 大都市圏に住んでいれば 1
age head (世帯主年齢階級)	世帯主年齢階級(12 階級)
children (実子を含む世帯, ダミー変数)	世帯主の子ども(20 歳未満かつ未婚)がいれば 1
two incomes (共働きである世帯, ダミー変数)	世帯主と配偶者の双方が収入を得ていれば 1

※女性収入は月あたりの金額である

表 3 推計結果魚介類支出(平成 16 年)

説明変数	推定値	t 統計量
annual income (年間収入)	2.06	862.84
female income (女性収入)	0.00	-100.39
agri fish income (農林漁業収入有世帯, ダミー変数)	-112.01	-15.56
over65 (65 歳以上を含む世帯, ダミー変数)	117.62	91.21
urban (3 大都市圏在住世帯, ダミー変数)	-18.53	-4.76
age head (世帯主年齢階級)	88.53	1200.2
two incomes (共働きである世帯, ダミー変数)	-220.66	-154.46
children (実子を含む世帯, ダミー変数)	1254.56	220.12
agri fish income × urban	-159.71	-9.08
age head × urban	-1.38	-19.68
age head × children	-45.70	-389.34
Intercept	-1764.05	-396.85
決定係数	0.299	
サンプル数	24,618,205	

### 3.3 分析結果

#### 3.3.1 魚介類支出の分析結果

魚介類支出に対しておこなった回帰分析の推計結果を表 3 に示す。「3 大都市圏在住世帯」以外の変数の t 統計量は十分に大きな値を示しており、統計的に意味のあるものと考えられる。

推計結果によると、「年間収入」、「65 歳以上の世帯員を含む世帯」、「世帯主年齢階級」、「実子を含む世帯」は正の影響を、「農林漁業収入有世帯」、「共働きである世帯」は負の影響を、それぞれ魚介類支出に与えてい

ることがわかる。等価所得を用いることで「3大都市圏在住世帯」が負の値に転じた。それ以外の結果は先行研究[2]と同じである。

表3の推計結果から「世帯主年齢階級」と「3大都市圏在住世帯」の交差項より、在住地域の影響は殆どなく、世帯主の年齢に依存した魚介類支出が行われていることがわかった。年齢の高い世帯員を含む世帯はそうでない世帯よりも多くの支出を行っているが、外食は少なく、また、共働き世帯は魚介類の消費量が少ない。

世帯主年齢階級による、魚介類支出の変化を見るため、実子を含む世帯、成人のみの世帯別に、以下の式(3)、(4)を導出した。age head, children, age head × children 以外は、平均値を代入している。

$$\text{実子を含む世帯 } fish_{children} = -866.39 + 42.14 \times \text{age head} \quad (3)$$

$$\text{成人のみ世帯 } fish_{e \text{ children}0} = -35.80 + 87.85 \times \text{age head} \quad (4)$$

この式のグラフ(図5)において世帯主が20歳代の世帯では、実子の存在は魚介類支出にほとんど影響がなく、実子を含む世帯の方が魚介類支出の増加量が少ないことを示している。

### 3.3.2 他の食料品目を用いた分析結果

肉類、野菜・海藻を被説明変数とした回帰分析の推計結果を表4に示す。肉類において「年間収入」、「3大都市圏在住世帯」、「世帯主年齢階級」、は正の推定値となり、「農林漁業収入有世帯」、「65歳以上の世帯員を含む世帯」、「実子を含む世帯」、「共働きである世帯」の推定値は負の値をとった。

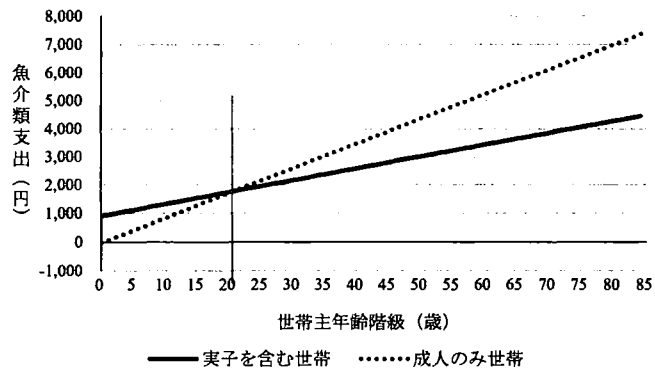


図5 子供の有無別にみる魚介類支出の変化量比較(平成16年)

表4 推計結果(肉類支出, 野菜・海藻支出)(平成16年)

説明変数	肉類		野菜・海藻	
	推定値	t 統計量	推定値	t 統計量
annual income (年間収入)	1.43	871.67	2.12	840.04
female income (女性収入)	0.00	-152.2	0.00	-124.53
agri fish income (農林漁業収入有世帯, ダミー変数)	-57.64	-11.63	-99.66	-13.04
over65 (65歳以上を含む世帯, ダミー変数)	-300.46	-338.59	106.07	77.48

ま	urban (3大都市圏在住世帯, ダミー変数)	-465.64	-173.64	-328.14	-79.32
た,	age head (世帯主年齢階級)	15.31	301.66	95.18	1215.55
野	two incomes (共働きである世帯, ダミー変数)	-134.74	-137.07	-443.00	-292.12
菜	children (実子を含む世帯, ダミー変数)	-122.53	-31.24	1983.75	327.88
・	agri fish income × urban	-219.87	-18.16	68.30	3.66
海	age head × urban	14.66	304.38	13.83	186.05
藻	age head × children	0.95	11.74	-63.46	-509.23
に	intercept	1268.13	414.58	-1336.22	-283.17
お	決定係数	0.0844		0.3331	
	サンプル数	24,618,205			

いて、「年間収入」、「世帯主年齢階級」は正の推定値となり、「65歳以上の世帯員を含む世帯」、「農林漁業収入有世帯」、「実子を含む世帯」、「共働きである世帯」の推定値は負の値をとった。魚介類と同様に、等価所得を用いることで「3大都市圏在住世帯」が負の値に転じた。それ以外の結果は先行研究[2]と同じである。「実子を含む世帯」では、魚介類と同様に、肉類、野菜・海藻支出がその他の世帯よりも少ないことがわかった。

#### 4 まとめ

本稿では、匿名データに対するデータハンドリングの効率の良い手法を紹介した。コンピュータ環境や作業時間のコスト削減のためにおこなったものである。変数抽出作業をおこなうことで、データ分析対象や目的、手法を試行錯誤することができたと考えている。

子どもの食育に影響する世帯属性を検討した。一般的に用いられる等価所得を用いて計算を行ったところ「3大都市圏在住世帯」のみ先行研究[2]では正の値であったが、負の値に転じた。

本分析ではいずれの回帰分析においても決定係数は大きい値を示さなかった。そこで、他の決定要因や経年による世帯属性の影響などをさらに探求する必要がある。ただし、今回利用したのは平成元年、6年、11年、16年の全国消費実態調査であり、最も新しいデータでも平成16年である。平成28年現在に至るまでの12年、社会経済状況はめまぐるしく変化している。平成22年をピークに日本の総人口は減少に転じ、今後ますます減少すると予測されている。また世帯数の増加から核家族化が進行している。また、人口の東京一極集中が加速している。世帯主の高齢化が進行している。人口に関してだけでも大幅な変化が生じている[5]。高齢者対象の宅配、ネットショッピング・電子マネーなどのICT利活用の浸透[6]、コンビニエンスストアの普及など、消費生活のスタイルは大幅に変化している。この間の食品の消費活動を言及するには、現在、平成26年のデータを入手できるオーダーメイド集計を利用することで、より最近の傾向を見ることが可能となると考える。

#### 謝辞

本研究を遂行するにあたり、「匿名データ」をご提供いただいた独立行政法人統計センター様、ご支援をいただいたSASInstitute Japan 株式会社様に感謝いたします。

## 参考文献

- [1] 池田瑞穂, 笹谷知輝, 大野嵩護: 世代別魚介類消費の推移に関する考察, 公的統計のマイクロデータ等を用いた研究の新展開 (平成 26 年度) 報告要旨集, pp.3--9 (2014).
- [2] 大野嵩護, 笹谷知輝, 池田瑞穂: 食生活から見た食料消費傾向に関する考察—匿名データを用いて—, 公的統計のマイクロデータ等を用いた研究の新展開 (平成 27 年度) 報告要旨集, pp.143--162(2015).
- [3] 森宏: 食料消費の年齢・世代効果—文献解題を中心に—: Economic Bulletin of Senshu University, Vol.45, No.3, pp.113--132 (2011).
- [4] 厚生労働省: 食事摂取基準 (2004) : <http://www.mhlw.go.jp/houdou/2004/11/h1122-2a.html> (参照 2015/10/18).
- [5] 国土交通省: 平成 27 年度国土交通白書: <http://www.mlit.go.jp/hakusyo/mlit/h27/hakusho/h28/pdf/np101100.pdf> (参照 2016/06/18).
- [6] 総務省: 平成 27 年度情報通信白書: <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/pdf/index.html> (参照 2016/06/18).

# 匿名データを用いた夫婦共働き世帯における子あり世帯と夫婦のみ 世帯の比較：横断研究

○井上 貴博<sup>1</sup>、矢田 徹<sup>2</sup>

(<sup>1</sup>ノバルティスファーマ株式会社、<sup>2</sup>イーピーエス株式会社)

Comparisons of the dual-career couples with child and without child using Tokumei data

Takahiro Inoue, Toru Yada

Oncology Integrated Science & Operations Department, Novartis Pharma K.K.

Statistics Analysis Department 1, EPS Corporation

## 要旨

日本で少子高齢化が問題となって久しい。そこで全国消費実態調査の匿名データを用いて、夫婦共働きの子あり世帯と夫婦のみ世帯について世帯主年齢、世帯年収、喫煙率などを比較する。

キーワード：匿名データ、全国消費実態調査、少子高齢化、横断研究

## 1 はじめに

平成17年の国民生活白書[1]によると、子どものいない世帯の割合が増加しており、特に30代において、「非親族世帯」と「単独世帯」を除く世帯総数に対する「子どものいない世帯」の割合が、1990年の12.1%から2000年の18.6%と大幅に増加している。少子高齢化の要因としては、晩婚化・非婚化の影響、晩婚化の進展に伴う夫婦の出生力の低下などが指摘されている。夫婦に対する意識調査[2]では、子どもを欲しくない夫婦は少ないことがわかっているが、「子どものいない夫婦」は子育てに必要な年収を高く見込みすぎている、とも指摘されている。子どもを一人育てるのに最低限必要な世帯年収について、子どもなしの男性は714万円であるのに対し、子どもありの男性は591万円と答えており、子どもなしの女性の場合は、594万円であるのに対し、子どもありの女性も567万円と答えている。一方、民間給与実態調査[3]によれば、39歳以下の給与所得者の男性の平均給与は年442万円、女性は年283万円であり、合計すれば700万円を超えている。

全国消費実態調査（平成16年度）から生成された擬似マイクロデータを用いて子どもなしの世帯と子

どもありの世帯ではどのような違いがあるのかを調べた先行研究[4]では、夫婦のみ世帯と一子あり世帯を比較し、子の有無と世帯主年齢、世帯年収、喫煙有無の間に関連が見られた。しかしながら、擬似マイクロデータには含まれないデータもあり、十分な考察が出来なかった。そこで本研究では全国消費実態調査の匿名データを用いて、夫婦共働きで子どもありの世帯と子どもなしの世帯の比較を行った。特に子の有無との関連が強いと考えられる世帯収入については、世帯主（主に男性）とその配偶者（主に女性）別に勤め先収入を比較した。

## 2 方法

### 2.1 全国消費実態調査

全国消費実態調査は、昭和34年以来5年ごとに実施されている国民生活の実態（消費・所得・資産）に係る水準、構造、分布などを明らかにすることを目的とした調査である。平成16年度調査[5]の標本数は約59,400世帯（うち単身世帯約5,000世帯）である。本研究では、二人以上世帯（調査期間9～11月の3か月間）で、世帯票と家計簿（1か月分以上）の両方が揃っている世帯（平成16年：43,861世帯、平成11年：44,540世帯、平成6年：44,687世帯）に対し匿名化措置を施し、乗率を付与した匿名データを用いた。乗率計算前のレコード数が1の場合、収入の中央値を表示しなかった。

集計に際しては、集計用乗率を用いて重み付けした数値を用いた。本研究で対象とする全国消費実態調査（平成16年度）の世帯数は、匿名データ（43,861世帯）に集計用乗率を用いて算出した25,268,124世帯とした。なお、先行研究で用いた擬似マイクロデータ（32,027世帯）についても集計用乗率を用いて495,465世帯として扱った。

### 2.2 抽出条件

表 1 抽出条件

先行研究での抽出条件 (条件 A)		本研究での抽出条件 (条件 B)	
一子＋夫婦世帯 世帯人員 3 名 (有業人員 2 名)	夫婦のみ世帯 世帯人員 2 名 (有業人員 2 名)	一子＋夫婦世帯 就業世帯主 ＋就業配偶者 ＋未婚の子	夫婦のみ世帯 就業世帯主 ＋就業配偶者
介護保険料の支払いなし		--	
世帯主年齢 40 歳未満		夫婦とも 40 歳未満	
公的年金給付なし		--	
男・婦人用被服を共に購入 (被服：洋服、シャツ・セーター類、下着類)		--	

先行研究では、擬似マイクロデータを用いて、40歳未満の夫婦共働き世帯における一子＋夫婦世帯と夫婦のみ世帯の比較を行った。擬似マイクロデータには子どもの有無などの情報が存在しないため、別の情報を用いて抽出条件を設定し、40歳未満の夫婦共働き世帯における一子＋夫婦世帯と夫婦の



み世帯を抽出した。

先行研究での抽出条件（条件A）を用いて、匿名データ（平成16年度）から40歳未満の夫婦共働き世帯における一子+夫婦世帯と夫婦のみ世帯を抽出し、次いで匿名データ（平成16年度）に本研究での抽出条件（条件B）を用いて、40歳未満の夫婦共働き世帯における一子+夫婦世帯と夫婦のみ世帯の抽出を行い、抽出世帯数を比較した。

### 2. 3 先行研究との比較

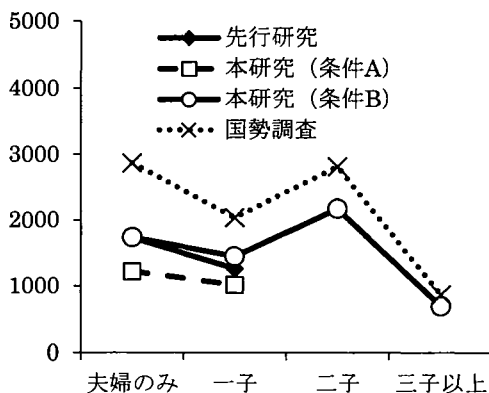
抽出世帯数、世帯主年齢別世帯数、世帯主年齢別・子の年齢別世帯数、世帯主年齢別・喫煙有無別世帯数について、先行研究と本研究を比較した。

先行研究と本研究（条件B）は、対10万世帯比について、ほぼ一致した（表2）。国勢調査（平成17年）対10万世帯比の値に比べて、本研究（条件B）の値は低い傾向が見られたが、子の数に対する推移は似た傾向を示した（図1）。

表2 各抽出世帯数

世帯数 (対10万 世帯比)	先行 研究	本研究 (条件A)	本研究 (条件B)	国勢調査 (H17)
夫婦 のみ 世帯	8589 (1734)	308237 (1220)	439473 (1739)	992131 (2867)
一子+ 夫婦 世帯	6237 (1259)	256094 (1014)	365085 (1445)	703814 (2034)
二子+ 夫婦 世帯	--	--	548564 (2171)	972792 (2811)
三子以 上+夫 婦世帯	--	--	175777 (696)	299866 (867)
全体	495465	25268124	25268124	34605447

図1 対10万世帯比の比較



次に、世帯主年齢ごとに、先行研究と本研究（条件B）の世帯数を比較した（表3）。

表3 世帯主年齢ごとの世帯数

世帯主 年齢	先行研究		本研究（条件B）				合計
	夫婦のみ 世帯	一子+夫婦 世帯	夫婦のみ 世帯	一子+ 夫婦世帯	二子+ 夫婦世帯	三子以上 +夫婦 世帯	
全体	8589	6237	439473	365085	548564	175777	1528899
24歳以下	74 (0.9)	194 (3.1)	15410 (3.5)	12191 (3.3)	4101 (0.8)	270 (0.2)	31971 (2.1)
25～29歳	2825 (32.9)	1129 (18.1)	138757 (31.6)	62656 (17.2)	45154 (8.2)	6499 (3.7)	253067 (16.6)
30～34歳	3584 (41.7)	2891 (46.3)	168539 (38.4)	157704 (43.2)	167244 (30.5)	50699 (28.8)	544186 (35.6)
35～39歳	2106 (24.5)	2024 (32.4)	116766 (26.6)	132535 (36.3)	332065 (60.5)	118309 (67.3)	699675 (45.8)

夫婦のみ世帯と一子+夫婦世帯の世帯主年齢の分布は、先行研究と本研究（条件B）で同様の傾向

が見られた。また、本研究（条件B）と国勢調査（平成17年）は、子の数ごとの世帯主年齢の分布[6]について同様の傾向が見られた。

表3 世帯主年齢ごとの世帯数（続き）

世帯数(%) 世帯主 年齢	国勢調査（平成17年）				
	夫婦のみ 世帯	一子+ 夫婦世帯	二子+ 夫婦世帯	三子以上 +夫婦 世帯	合計
全体	992131	703814	972792	535178	3203915
24歳以下	38076 (3.8)	26665 (3.8)	9272 (1.0)	1106 (0.4)	75119 (2.5)
25～29歳	258527 (26.1)	132149 (18.8)	87996 (9.0)	15691 (5.2)	494363 (16.7)
30～34歳	409518 (41.3)	279345 (39.7)	320709 (33.0)	83134 (27.7)	1092706 (36.8)
35～39歳	286010 (28.8)	265655 (37.7)	554815 (57.0)	199935 (66.7)	1306415 (44.0)

一子+夫婦世帯について、先行研究と本研究（条件A）、本研究（条件B）の子の年齢について比較した（表4）。

先行研究では給食費なし（用途\_学校給食、品目(1)\_学校給食、品目(2)\_学校給食のいずれもが0）が未

就学児に対応すると考えたが、本研究（条件A）の給食費なし、ありの世帯数の比に大きな差が生じた。一方で、本研究（条件A）と本研究（条件B）については、給食費の有無による世帯数の比が0～5歳・6～14歳の世帯数の比と近い値となった。また、本研究（条件B）については、国勢調査（平成17年）の0～5歳・6～14歳の世帯数の比[6]ともよく一致しており、先行研究の抽出条件が擬似マイクロデータに対しては適切ではなかったことが示唆された。

表4 世帯主年齢ごとの子の年齢（給食費の有無）

世帯数(%) 世帯主 年齢	先行研究			本研究（条件A）			本研究（条件B）			
	全体	給食費 なし	給食費 あり	全体	給食費 なし	給食費 あり	全体	0～5歳	6～14 歳	15歳 ～
24歳以下	194	194 (100)	0	7986	7986 (100)	0	12191	12191 (100)	0	0
25～29歳	1129	816 (72.3)	313 (27.7)	42296	39261 (92.8)	3035 (7.2)	62656	60810 (97.1)	1846 (3.0)	0
30～34歳	2891	1309 (45.3)	1581 (54.7)	104036	82926 (79.7)	21110 (20.3)	157704	135656 (86.0)	21663 (13.8)	385 (0.2)
35～39歳	2024	50 (2.5)	1973 (97.5)	101777	50438 (49.6)	51338 (50.4)	132535	69799 (52.7)	61433 (46.4)	1303 (1.0)

表4 世帯主年齢ごとの子の年齢（給食費の有無）（続き）

世帯数(%) 世帯主 年齢	国勢調査（平成17年）			
	全体	0～5歳	6～14 歳	15歳 ～
24歳以下	26665	26149 (98.1)	486 (1.8)	30 (0.1)
25～29歳	132149	123696 (93.6)	8197 (6.2)	256 (0.2)
30～34歳	279345	231029 (82.7)	46853 (16.8)	1463 (0.5)
35～39歳	265655	144226 (54.3)	113458 (42.7)	7971 (3.0)

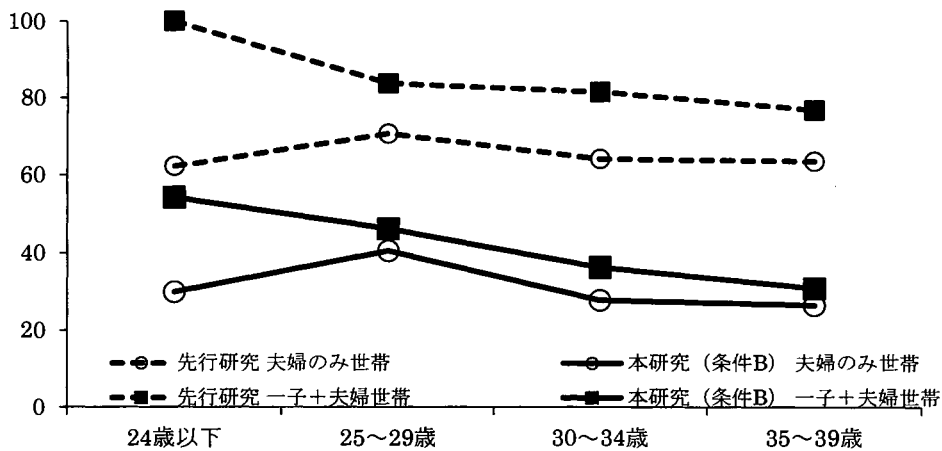
世帯喫煙率（タバコを購入している世帯（用途\_たばこ、品目(1)\_たばこ、品目(2)\_たばこのいずれかに0円超の入力あり）の割合）について、先行研究と本研究（条件B）を比較した（表5）。世帯喫煙率については、数値は先行研究と大きく異なり給食費と同様に先行研究での抽出条件が擬似マイクロデータに対しては適切ではなかったことが示唆された。夫婦のみ世帯の年齢ごとの世

帯喫煙率の推移の形状、一子+夫婦世帯の年齢ごとの世帯喫煙率の推移の形状については、先行研究と本研究（条件B）で似た傾向を示した（図2）。

表5 世帯主年齢ごとの世帯喫煙率

喫煙世帯数 (%)	先行研究		本研究（条件B）			
	夫婦のみ世帯	一子+夫婦世帯	夫婦のみ世帯	一子+夫婦世帯	二子+夫婦世帯	三子以上+夫婦世帯
24歳以下	46(62.2)	194(100)	4605(29.9)	6612(54.2)	3680(89.8)	270(100)
25～29歳	1994(70.6)	945(83.7)	56144(40.5)	28937(46.2)	25706(56.9)	2536(39.0)
30～34歳	2296(64.1)	2355(81.5)	46849(27.8)	57311(36.3)	62504(37.4)	25517(50.3)
35～39歳	1338(63.5)	1554(76.8)	31000(26.5)	40917(30.9)	87888(26.5)	40950(34.6)

図2 世帯主年齢別 世帯喫煙率(%)



### 3 夫婦共働き世帯における子あり世帯と子なし世帯

#### 3.1 世帯年収の比較

表6 世帯年収（万円）

Median (Q1,Q3) n	先行研究		本研究（条件B）			
	夫婦のみ世帯	一子+夫婦世帯	夫婦のみ世帯	一子+夫婦世帯	二子+夫婦世帯	三子以上+夫婦世帯
24歳以下	410.2	438.1	480 (405,590) n=15410	416 (310,701) n=12191	346 (283,346) n=4101	- (-) n=270
25～29歳	563.4	528.3	567 (446,750) n=13757	510 (420,770) n=62656	430 (350,502) n=45154	344 (300,423) n=6499
30～34歳	663.6	622.0	660 (500,888) n=168539	597 (460,770) n=157704	510 (438,651) n=167244	493 (410,667) n=50699
35～39歳	773.3	633.5	760 (560,964) n=116766	648 (500,809) n=132535	633 (510,813) n=332065	597 (482,770) n=118309

匿名データで利用できる変数を使った抽出条件（条件B）を用いて、匿名データから抽出した夫婦

共働きの夫婦のみ世帯と子あり世帯を用いて比較を行った（表6）。

世帯年収については、先行研究と同様に世帯主年齢が上がるほど世帯年収が増える傾向、夫婦のみ世帯に比べて子あり世帯の世帯年収が低い傾向が見られた。世帯主24歳以下の三子以上+夫婦世帯については、対象世帯が270世帯で、乗率を乗じる前のレコード数は1であったため検討から除外した。なお、国勢調査（平成17年）での世帯主24歳以下で夫婦共働きの三子以上の世帯数は、全体の34605447世帯中1106世帯であった。

### 3. 2 世帯主年齢・世帯年収ごとの子の数の比較

世帯主年齢・世帯年収ごとに、子の数の割合を示した（表7）。

世帯主年齢24歳以下の世帯について、世帯年収400～600万円の世帯で子なしの割合が高かった。600万円以上の世帯では一子ありの割合が高くなるが、二子以上ありの世帯は無かった。400万円未満の世帯では子なしの割合が最も低く、一子または二子ありの割合（76.0%）が高かった。世帯主年齢25～29歳の世帯について、世帯年収が多いほど子なしの割合が高くなる傾向が見られた。世帯年収に関わらず子なしの割合が最も高かった。世帯主年齢30～34歳の世帯について、世帯年収600万円未満の世帯で二子ありの割合が最も高く、600～1000万円の世帯では子なしの割合が最も高かった。世帯主年齢35～39歳の世帯について、世帯年収に関わらず二子ありの割合が最も高かった。

表7 世帯主年齢・世帯年収ごとの子の数

世帯数(%)		世帯年収 (万円)					
世帯主年齢	子の数	<400	400= <600	600= <800	800= <1000	1000= <1200	1200=<
24歳以下	全体	12374	13280	6317			
	なし	2970 (24.0)	9774(73.6)	2666(42.2)			
	1人	5303(42.9)	3236(24.4)	3651(57.8)			
	2人	4101(33.1)					
	3人以上		270(2.0)				
25～29歳	全体	57782	102971	50848	29750	11716	
	なし	20885(36.1)	54380(52.8)	34279(67.4)	19323(65.0)	9890(84.4)	
	1人	14821(25.6)	24239(23.5)	12243(24.1)	9527(32.0)	1826(15.6)	
	2人	17850(30.9)	22730(22.1)	3674(7.2)	900(3.0)		
	3人以上	4226(7.3)	1622(1.6)	652(1.3)			
30～34歳	全体	79083	217956	123024	73785	34148	16152
	なし	17772(22.5)	50868(23.3)	42338(34.4)	37348(50.6)	17969(52.6)	2245(13.9)
	1人	22966(29.0)	56152(25.8)	42067(34.2)	18368(24.9)	11392(33.4)	6758(41.8)
	2人	26434(33.4)	88033(40.4)	27560(22.4)	14525(19.7)	4157(12.2)	6497(40.2)
	3人以上	11911(15.1)	22903(10.5)	11059(9.0)	3544(4.8)	630(1.8)	652(4.0)
35～39歳	全体	63218	223250	213135	110909	58970	30194
	なし	7897(12.5)	26695(12.0)	35901(16.8)	22149(20.0)	16668(28.3)	7456(24.7)
	1人	17215(27.2)	42532(19.1)	36985(17.4)	15501(14.0)	15375(26.1)	4927(16.3)
	2人	24516(38.8)	107615(48.2)	110326(51.8)	57358(51.7)	20153(34.2)	12098(40.1)
	3人以上	13590(21.5)	46408(20.8)	29923(14.0)	15901(14.3)	6774(11.5)	5713(18.9)

(網掛け：世帯主年齢・世帯年収ごとに最も割合の高いセル、空欄は対象世帯なし)

### 3. 3 喫煙有無別・世帯主年齢別・世帯年収別の世帯あたりの子の数の比較

喫煙世帯と非喫煙世帯別に、世帯あたりの子の数を計算した（表8）。計算にあたって、子の数が3人以上の世帯について子の数を3人とみなし、各カテゴリ内の子の数の総和を世帯数で除して算出した。

表8 喫煙有無別・世帯主年齢別・世帯年収別の世帯あたりの子の数

子の人数/世帯		世帯年収（万円）					
	世帯主年齢	<400	400= <600	600= <800	800= <1000	1000= <1200	1200=<
喫煙世帯	24歳以下	1.1	0.8	0.8			
	25～29歳	1.2	0.8	0.5	0.4	0.2	
	30～34歳	1.5	1.4	1.2	1.1	0.6	1.9
	35～39歳	1.6	1.8	1.7	1.6	1.2	1.4
非喫煙世帯	24歳以下	1.1	0.1	0.5			
	25～29歳	1.0	0.7	0.4	0.4	0.1	
	30～34歳	1.3	1.4	1.0	0.7	0.6	1.2
	35～39歳	1.7	1.7	1.6	1.6	1.3	1.7
喫煙世帯 －非喫煙世帯	24歳以下	0.0	0.6	0.2			
	25～29歳	0.2	0.2	0.1	0.0	0.1	
	30～34歳	0.2	0.1	0.2	0.4	0.0	0.7
	35～39歳	-0.1	0.1	0.1	0.0	-0.1	-0.3

（喫煙世帯、非喫煙世帯の行の網掛けは、世帯あたりの子の数が1人以上、濃い網掛けは1.5人以上を示す）

全体的に喫煙世帯の方が世帯あたり子の人数が多くなる傾向が見られたが、世帯主年齢が30代後半の層では世帯年収400万円未満と1000万円以上で非喫煙世帯の方が世帯あたり子の人数が多かった。また、世帯年収400万円未満では、喫煙有無・世帯主年齢に関わらず世帯あたりの子の数が1人以上であった。世帯主年齢が30歳代前半の世帯では、世帯年収800万円以上1000万円未満の世帯で、世帯あたりの子の数が少なくなる傾向が見られた。

非喫煙世帯で世帯あたりの子の数が1.5人以上であったのは世帯主年齢が30歳代後半にしか見られなかったのに対し、喫煙世帯では30歳代前半でも世帯あたりの子の数が1.5人以上である層があった。

### 3. 4 世帯主年齢別の世帯主及び世帯主の配偶者の勤め先収入の比較

世帯主と配偶者のそれぞれについて勤め先収入（月収）を集計し、比較した（表9）。

表9 世帯主及び世帯主の配偶者の勤め先収入（月収（千円））

Median (Q1,Q3)	世帯主の勤め先収入				配偶者の勤め先収入			
	夫婦のみ 世帯	一子+夫 婦世帯	二子+夫 婦世帯	三子以上 +夫婦世 帯	夫婦のみ 世帯	一子+夫 婦世帯	二子+夫 婦世帯	三子以上 +夫婦世 帯
24歳以下	271 (208,275)	216 (140,252)	257 (210,257)	- (-)	164 (99,173)	37 (0,102)	91 (23,91)	- (-)
25～29歳	285 (183,328)	264 (225,318)	252 (210,323)	263 (222,279)	146 (63,242)	104 (62,227)	70 (17,90)	49 (33,86)
30～34歳	312 (250,377)	298 (244,360)	304 (233,353)	277 (201,351)	146 (71,260)	93 (28,223)	63 (16,126)	60 (33,97)
35～39歳	358 (283,433)	335 (237,433)	356 (279,427)	343 (249,420)	108 (39,256)	75 (6,182)	73 (24,141)	60 (0,113)

世帯主の勤め先収入は世帯主年齢に依存する傾向が見られ、配偶者の収入は子の数に依存する傾向

が見られた。特に配偶者の収入が、夫婦のみ世帯と子あり世帯で大きく異なっていた。配偶者の就業形態について、配偶者がパートである世帯の割合を計算した（表10）。

表10 配偶者のパートの世帯数（割合）

世帯主年齢	本研究（条件B）			
	夫婦のみ世帯	一子+夫婦世帯	二子+夫婦世帯	三子以上+夫婦世帯
全体	157801 (35.9)	144210 (39.5)	325728 (59.4)	93809 (53.4)
24歳以下	3073 (20.0)	4193 (34.4)	1762 (43.0)	270 (100)
25～29歳	53296 (38.4)	17490 (27.9)	29965 (66.4)	1006 (15.5)
30～34歳	59633 (35.4)	66218 (42.0)	93288 (55.8)	30852 (60.9)
35～39歳	41797 (35.8)	56310 (42.5)	200714 (60.4)	61682 (52.1)

配偶者がパートである世帯の割合は、世帯主年齢の影響よりも子の数による影響が大きかった。特に二子以上の世帯の50%～60%で配偶者がパートだった。次に世帯主の配偶者を対象としてパート以外の就業とパート別に勤め先収入を計算した（表11）。パートの月収については、子の数による影響があまり見られなかった。パート以外の就業

での月収については、子の数に比例して減る傾向が見られた。いずれの場合も、世帯主の月収のように世帯主年齢に比例して月収が増加する傾向はあまり見られなかった。

表11 パート・パート以外別 世帯主の配偶者の勤め先収入（月収（千円））

Median (Q1,Q3) 世帯主年齢	パート以外の就業				パート			
	夫婦のみ世帯	一子+夫婦世帯	二子+夫婦世帯	三子以上+夫婦世帯	夫婦のみ世帯	一子+夫婦世帯	二子+夫婦世帯	三子以上+夫婦世帯
24歳以下	170 (139,204)	0 (0,102)	- (-)		88 (84,166)	49 (37,80)	23 (23,30)	- (-)
25～29歳	212 (152,272)	162 (62,258)	87 (0,197)	49 (33,86)	69 (44,110)	67 (43,88)	66 (25,85)	72 (71,90)
30～34歳	225 (133,283)	199 (0,287)	137 (0,263)	82 (0,191)	81 (53,103)	67 (29,94)	57 (26,83)	58 (37,82)
35～39歳	202 (41,303)	130 (0,256)	150 (0,272)	0 (0,201)	72 (37,101)	55 (28,82)	68 (39,89)	69 (34,87)

表12 調査年度ごとの配偶者の勤め先収入（月収（千円））

世帯主年齢	調査年度	本研究（条件B）			
		夫婦のみ世帯	一子+夫婦世帯	二子+夫婦世帯	三子以上+夫婦世帯
24歳以下	1994年	125 (70,192)	60 (35,139)	- (-)	-
	1999年	108 (77,177)	77 (23,124)	156 (57,160)	- (-)
	2004年	164 (99,173)	37 (0,102)	91 (23,91)	- (-)
25～29歳	1994年	144 (67,223)	68 (17,177)	71 (0,114)	41 (21,64)
	1999年	173 (77,230)	90 (18,210)	53 (21,103)	53 (0,67)
	2004年	146 (63,242)	104 (62,227)	70 (17,90)	49 (33,86)
30～34歳	1994年	144 (71,243)	81 (28,226)	56 (0,120)	56 (0,88)
	1999年	164 (68,253)	99 (32,234)	69 (28,141)	53 (0,87)
	2004年	146 (71,260)	93 (28,223)	63 (16,126)	60 (33,97)
35～39歳	1994年	131 (51,249)	73 (0,177)	59 (0,122)	58 (0,94)
	1999年	107 (49,249)	71 (25,226)	64 (15,127)	59 (0,105)
	2004年	108 (39,256)	75 (6,182)	73 (24,141)	60 (0,113)

配偶者の勤め先収入について、5年前（平成11年（1999年））及び10年前（平成6年（1994年））と比較した（表12）。

世帯主年齢25歳以上の子あり世帯について、1994年から2004年までの間に配偶者の勤め先収入がやや増える傾向が見られた。

## 4 考察

喫煙有無、給食費有無については、擬似マイクロデータを用いた先行研究と異なる結果が見られた。先行研究では、タバコ購入費、給食費が0円であるか否かで有無を判定する方法を用いたが、擬似化の処理によって、実データでは0円であった世帯が擬似マイクロデータ上では0より大きい値をとったことが原因であると考えられる。

世帯主年齢と世帯年収については、先行研究と同様に子の数との関連が見られた。世帯年収に関しては、世帯年収と子の数が比例せず、世帯年収600万～1000万円の層で子の数が少なく、それより世帯年収が低い層と高い層で子の数が増える傾向が見られた。高収入層で子の数が多いのは、育児に対して経済的な懸念が少ないという因果関係が考えられる一方、低収入層で子の数が多いのは、子の数が多いために収入が少ないという高収入層とは逆の因果関係も考えられる。世帯主の勤め先収入を見ると、子の数の影響は少なく世帯主年齢にほぼ比例しているが、配偶者の勤め先収入は子の有無による影響が大きかった。配偶者の就業形態を見ると、子あり世帯でパートである割合が高いことから、出産に伴い配偶者（多くは女性）の就業形態がパートに変わり、その結果として世帯収入を下げているのではないかと考える。また、パート以外の就業であっても配偶者の勤め先収入は世帯主のように年齢に比例して上がっていく傾向があまり見られず、子の数に比例して勤め先収入が減る傾向にあることから、配偶者はパート以外の就業であっても育児のために十分な就業時間が取れず、結果として収入を得られないのではないかと考える。

配偶者の勤め先収入について1994年、1999年、2004年の経時的な変化を見ると、子あり世帯についてはわずかながら改善傾向があるように見えるが、夫婦のみ世帯ではあまり変化が見られない。1994年、1999年については就業形態のデータが取られていないため、この傾向が配偶者の就業環境によるものか、あるいはパートの割合が減ったことによるものかはわからなかった。

## 参考文献

[1] 平成17年版 国民生活白書「子育て世代の意識と生活」

第1章 結婚・出生行動の変化 第1節 最近の出生率低下の要因

<[http://www5.cao.go.jp/seikatsu/whitepaper/h17/01\\_honpen/html/hm01010001.html](http://www5.cao.go.jp/seikatsu/whitepaper/h17/01_honpen/html/hm01010001.html)> (参照)

2016-06-13)

[2] 平成17年版 国民生活白書「子育て世代の意識と生活」

第1章 結婚・出生行動の変化 第3節 子どもを持つという選択

<[http://www5.cao.go.jp/seikatsu/whitepaper/h17/01\\_honpen/html/hm01030002.html](http://www5.cao.go.jp/seikatsu/whitepaper/h17/01_honpen/html/hm01030002.html)> (参照:  
2016-06-13)

<[http://www5.cao.go.jp/seikatsu/whitepaper/h17/01\\_honpen/html/hm01030005.html](http://www5.cao.go.jp/seikatsu/whitepaper/h17/01_honpen/html/hm01030005.html)> (参照:  
2016-06-13)

[3] 民間給与実態統計調査「長期時系列データ」

3-10 1年勤続者の年齢階層別給与所得者数・給与総額・平均給与（男、女、合計）

<[https://www.nta.go.jp/kohyo/tokei/kokuzeicho/jikeiretsu/01\\_02.htm](https://www.nta.go.jp/kohyo/tokei/kokuzeicho/jikeiretsu/01_02.htm)>(参照: 2016-06-16)

[4] 40歳未満の夫婦共働き世帯における一子あり世帯と子どもなし世帯の比較：横断研究

『SASユーザー総会2015 論文集』、pp610-615

[5] 平成16年全国消費実態調査の概要

[6] 平成17年 国勢調査 新分類区分による遡及集計

第1表 世帯の家族類型（22区分）別一般世帯数及び一般世帯人員（3世代世帯，母子世帯及び父子世帯－特掲） — 全国，都道府県，市町村

第10表 夫の年齢（5歳階級），子供の有無・数・年齢（121区分），夫婦の就業・非就業（4区分）別夫婦のいる一般世帯数及び一般世帯人員（雇用者－特掲） — 全国

<[http://www.e-stat.go.jp/SG1/estat/GL08020103.do?\\_toGL08020103\\_&tclassID=000001031076&cycleCode=0&requestSender=search](http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001031076&cycleCode=0&requestSender=search)> (参照: 2016-06-13)

[7] 周防節雄 (2015) 全国消費実態調査の匿名データとその符号表から自動的に SAS のデータセット、変数ラベルと変数フォーマットを作成する SAS プログラム

『SASユーザー総会 2015 論文集』、pp257-278

## 謝辞

全国消費実態調査の匿名データの利用に便宜を図って頂いた（独）統計センター、代理人となり費用を負担して頂いたSAS Institute Japan株式会社に感謝の意を表します。また、本研究に際して、匿名データを自動的にSASデータ化するプログラムを快く提供して頂いた兵庫県立大学 周防節雄先生に感謝の意を表します。



# 匿名データを用いたコメ離れと食事づくりの簡素化についての検証

中松 建

(株式会社タクミインフォメーションテクノロジー、医薬品開発事業1部)

Study of relationships between shift of the preference for rice and faster meal preparation  
with anonymized data

Ken Nakamatsu

Takumi Information Technology Inc.

## 要旨

昨年のマイクロデータ分析コンテストにおいて行った、コメ離れと食事づくりの簡素化についての分析結果を、匿名データを用いて検証した。匿名データに含まれている項目がコメの消費量のうちの程度を占めているかを確認し、各世帯についてコメ離れと考えられるか分類を試みた。コメ離れと考えられる世帯では、外食や清涼飲料等のすぐに飲食が可能な項目が食料費に占める割合が高く、また手間のかかる項目への支出および食料費全体の支出は低くなっていることから、食への関心が低いことが考えられた。昨年の分析結果と匿名データによる分析結果は比較的近かったことから、擬似マイクロデータによる分析は予備的分析として有用であった。

キーワード：マイクロデータ分析 匿名データ コメ離れ

## 1. はじめに

昨年のマイクロデータ分析コンテストにおいて、コメ離れと食事づくりの簡素化について分析を行った。分析結果からはコメへの支出が減少することによって外食や調理済み食品への支出が増えていると考えられた。また世帯主の年齢が低いほど、コメ離れの状態と考えられる世帯の割合は大きくなっていったが、食料費に占める外食および弁当・調理食品等の中食の割合は、コメを一定量以上消費していると考えられる世帯と大きな差はなく、食事づくりの簡素化はコメ離れの強い要因とはなっていないと考えられた。

これまでにコメの消費に関して、農林水産省により web アンケート調査による消費者意向の把握や、大規模家計データからの消費動向の把握が試みられている[1]。この結果から「食事を短時間で済ませたい」「調理や片づけの時間を省きたい」といったことが、パンやめん類を主食に選ぶ際の理由になっていることが報告されている。

しかし、コメを消費している世帯と消費していない世帯の比較や、食料費の他の項目との関係についてなどは着目されておらず、またこれまでの研究・報告は多くないことから昨年の分析結果を実証することは有用であると考えられた。

昨年の分析は平成 16 年全国消費実態調査の結果をもとに生成された、擬似マイクロデータを用いて行った。擬似マイクロデータは平均や相関などの特性をなるべく維持してつくられているが、データ作成時の情報の損

失により、分析結果は実際のデータを用いた結果とは異なることも考えられるため、平成 16 年全国消費実態調査の匿名データ(以下、匿名データ)を用いて分析結果の検証を行った。

## 2. 使用する変数および基準の検討

匿名データには、擬似マイクロデータに含まれない「単身世帯」「2人以上の勤労世帯ではない世帯」が含まれている。本研究では 1. 擬似マイクロデータに結果の検証を行う 2. 食事スタイルの変化が起きにくく、得られる結果が特に有用、という 2 点から対象は 2 人以上の勤労世帯とした。また各集計は集計用乗率を用いて行った。

擬似マイクロデータではコメへの支出金額が 0 であった世帯の割合は 1.6%であったが、上記の対象とした世帯ではおよそ 24.1%であった。そのため現物支出および外食、中食が消費量に占める割合の算出を試み、流通先の割合[2]と比較し、妥当性を検証した。

### 2.1 現物支出

コメの生産者はコメへの支出が 0 になると考えられる。生産者と考えられるのは農業を仕事とする世帯員を含み、食料の現物支出がある世帯である。コメへの支出が 0 であった世帯のうち前の条件に該当したのは 1%程度であったことから、調査対象となった世帯には生産者による消費はほとんど含まれないと考えられた。

贈答用や仕送りによるコメへの支出額は品目分類と用途分類の差により算出が可能であり、品目分類におけるコメへの支出のおよそ 5%であった。

現金の支出の伴わない他の無償譲渡は、食料品への現物支出と交際費のうちの食料費との差が、食料品の無償譲渡と近くなると考えられる。この金額は現物支出のうちおよそ 17%であった。このうちおよそ 2/3 程度をコメと仮定すると贈答等との合計はコメへの支出のおよそ 16%となり、これは流通量からの推測とも近いことから、食料品の現物支出に占めるコメの割合はおよそ 15%であり、コメの消費量のおよそ 10%を占めていると考えられた。

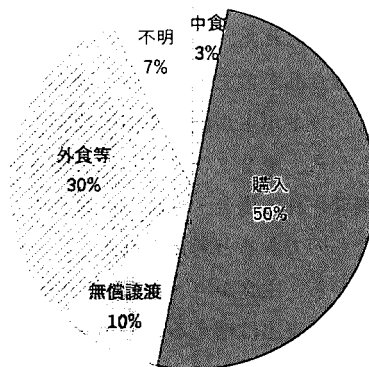


図1 推測されたコメの入手・摂取経路の割合  
各支出は重量に換算  
外食費と加工品による消費量は流通量より推測

### 2.2 外食・中食

昨年の分析ではコメへの支出金額を基準として、各世帯がコメ離れと考えられるかを判定した。しかし世帯主の年齢が低い世帯では中食や外食の利用が多い傾向があり、コメへの支出は少なくともコメは消費されている可能性がある。匿名データでは品目分類の主食的調理食品の項目に、弁当やおにぎりなどコメを主体とする項目があることから、それぞれの単価と 1 個あたりに含まれる精白米の重量から消費量を算出した。単価は平成 19 年全国物価統計調査から、1 個あたりに含まれるコメの重量は農林水産省による報告を参考にし、弁当類は 1 個あたりそれぞれ 498 円・95.5g、おにぎり類は 110 円・45.5g とした。

外食費はコメを主体とする支出金額の項目がなく、支出がこづかいに含まれていることも考えられることから消費量の推測はむずかしい。これらの推測がむずかしい項目は、流通の割合から消費量の 30%を占めて

いると仮定することとした。

### 2.3 コメ離れの基準

野菜では1日あたり350gの摂取が目標とされているが、コメの目標値は設定されていない。そのため1人あたりの消費量の維持を目標とし、コメへの支出と弁当・おにぎり等により消費されたと考えられる量の合計が1500g未満であれば、平均的な消費量の半分に満たないことからコメ離れの可能性が高い世帯とした。世帯員1人あたり1500g以上を消費していると考えられる世帯は便宜上コメ消費世帯とし、1人あたりの消費量が現物支出を合計した場合に1500g以上となる世帯は、どちらかに分類しないこととした。また年齢が低い世帯員は食料品の消費量は少なくなることから、6歳未満の世帯員は0.5人として扱うこととした。

## 3. 結果

斜交主成分クラスター分析により、食料費の各項目は調理や準備の時間に関係すると考えられる3つのクラスターに分割された(図2)。第1クラスターには野菜や魚介類を含んでおり、調理が必要で手間のかかる品目と考えられた。パン、めん類が含まれる第2クラスターには手間のかからない項目として代表的な外食費、調理食品が含まれており、調理や準備が不要・または簡単な手間のかからない項目と考えられた。飲料は缶・ペットボトル飲料や清涼飲料などを含み、菓子類はすぐに手軽に消費でき、食事を代替することもあることから、第2クラスターに分類されると考えられた。第3クラスターは乳卵類と他の穀類であり、これらは調理される場合・されない場合のどちらも考えられることから、第1クラスターと第2クラスターの間的なものであると考えられた。

以上から、主食がコメからパンやめん類となる頻度が増えることによって、調理や準備に手間のかからない項目への支出は増えることが考えられた。

擬似マイクロデータによる結果と比べると説明率は0.469と低くなったが、食料費の各項目はほぼ同様のクラスターに分割された。

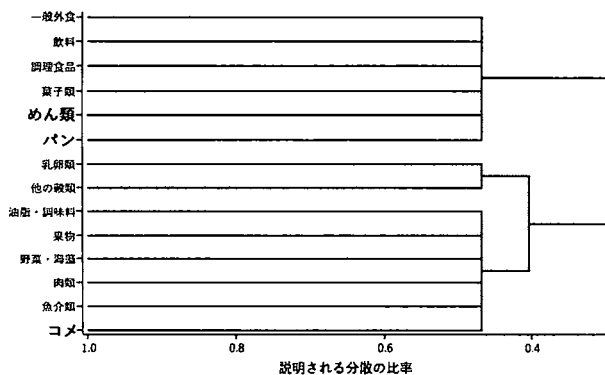


図2 食料費の各項目の分類

上から第2クラスター 第3クラスター  
第1クラスター

コメ離れと考えられる世帯の割合は世帯主の年齢が低いほど大きく、反対に一定量以上コメを消費していると考えられる世帯の割合は小さくなっていった(図3)。また分類に現物支出の内訳が必要な世帯の割合も、世帯主の年齢が低いほど大きくなっていった。年齢が低いほど無償譲渡によってコメを入手する割合が大きいことがこれまでに報告されているが、現物消費のうちコメが占めるのは15%程度であることから、分類していない世帯の多くはコメ離れの状態であると考えられた。擬似マイクロデータの結果と比較すると年齢区分による差は小さくなっているが、世帯主の年齢が低い世帯ほどコメ離れと考えられる世帯が多い傾向は同じであった。

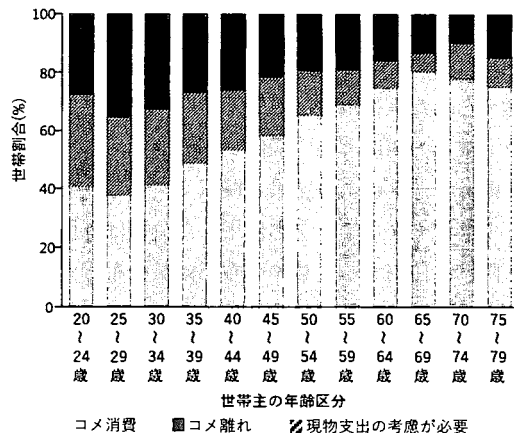


図3 各年齢区分のコメの消費状態による割合

表1 手間のかかる項目に対する1人あたりの支出

年齢区分	コメ離れ	コメ消費	p
20~24歳	6,672	9,981	.0001 **
25~29歳	7,996	9,980	<.0001 ***
30~34歳	8,534	10,756	<.0001 ***
35~39歳	9,070	11,306	<.0001 ***
40~44歳	9,580	12,014	<.0001 ***
45~49歳	10,678	13,455	<.0001 ***
50~54歳	12,807	15,288	<.0001 ***
55~59歳	14,540	17,704	<.0001 ***
60~64歳	16,681	19,916	<.0001 ***
65~69歳	19,405	20,788	0.434
70~74歳	18,998	21,052	0.860
75~79歳	20,523	22,387	0.860

単位：円

穀類(パン・めん類を含む)、魚介類、肉類、乳卵類、野菜類、果物類、油脂・調味料類、飲料(茶葉、コーヒー豆のみ)

p値はT検定(Welch)のあとholmの方法によって修正  
\* 5%で有意 \*\*1%で有意 \*\*\*0.1%で有意

食事づくりの簡素化という点から、1人あたりの消費に手間のかかる項目とにかからない項目への支出の合計を算出し、各年齢区分で比較した。項目の分類には斜交主成分クラスター分析による分類の結果を参考に、手間のかからない項目には外食、中食に加えて菓子類と飲料のうち茶飲料、コーヒー飲料などすぐに消費可能なものとした。手間のかかる項目は穀類(パン・めん類を含む)、魚介類、肉類、乳卵類、野菜類、果物類、油脂・調味料類と、飲料のうち茶葉やコーヒー豆などのすぐに消費できないものを使用した。

その結果、コメ離れと考えられる世帯では手間のかかる項目への支出額が少なくなっており、世帯主の年齢区分が64歳以下の世帯において支出額に有意な差が見られた(表1)。手間のかからない項目への支出は、どちらも同じ程度の金額であり有意な差はなかった。そのためコメ離れと考えられる世帯では食料費への支出が少なく、またそのなかでは手間のかかる項目への支出が比較的高い割合となっていた。そのため、世帯は食に対して時間と費用のどちらもかけない傾向があると考えられた。

### 3. まとめ

匿名データでは各世帯におけるコメの消費量のうちおよそ6割をとらえることができ、食料品の現物支出を加味することにより、コメ離れの状態と考えられる世帯を抽出することができた。

コメ離れと考えられる世帯では、弁当・惣菜などの中食や外食の利用に加えて菓子類や清涼飲料等の調理や準備に手間がかからず、すぐに飲食が可能な項目への支出の割合が比較的高くなっていった。また調理や準備に手間のかかる項目への支出を含む食料費全体の支出ははっきりと少なくなっていた。前回の分析では手

間のかからない項目の占める割合に大きな差はなかったことから、食事づくりの簡素化はコメ離れの強い要因ではないと推察した。しかし、食料費に占める割合ではなく支出額に着目したところ、コメ離れの状態と考えられた世帯は食に対し、時間と費用のどちらもかけない傾向があった。以上より食事づくりの簡素化とコメ離れの間には強い関係があると考えられた。

匿名データによる結果と擬似マイクロデータによって得られた結果はおおむね一致しており、予備的分析として擬似マイクロデータを使用した分析は有用であった。しかし独自の基準により新たにカテゴリーを作成する際に、使用する変数の分布が対数分布と大きく異なる可能性がある場合は結果の差が大きくなることが考えられることから、分析結果の解釈には注意が必要であると考えられた。

匿名データは分析に使用できる項目が多くさまざまな分析が可能な反面、申請にかかる時間や費用から利用までのハードルは低くない。擬似マイクロデータはファイルの読み込みや分析に用いるプログラム作成、ソフトウェアの操作等の準備に用いるだけでなく、分析方法や内容についてもある程度検討することができるため、匿名データの利用環境とともにこうしたデータも整備されることを期待する。

## 謝辞

匿名データの提供依頼申出は一橋大学経済研究所附属社会科学統計情報研究センター・マイクロデータ分析セッションに行い、平成28年度一橋大学経済研究所共同利用・共同研究拠点政府統計匿名データ利用促進プログラムによる助成を頂きました。データの貸与にあたっては多大なお世話をいただいたことを、厚くお礼申し上げます。

## 参考文献

[1] 農林水産省「主食用米消費動向の中期的変化及び要因分析」

[http://www.maff.go.jp/j/seisan/kokumotu/syouhi\\_doukou.html](http://www.maff.go.jp/j/seisan/kokumotu/syouhi_doukou.html)

[2] 農林水産省「米流通をめぐる状況」

[http://www.maff.go.jp/j/study/ryutu\\_system/01/pdf/data8.pdf](http://www.maff.go.jp/j/study/ryutu_system/01/pdf/data8.pdf)

[3] 周防(2015)「全国消費実態調査の匿名データから SAS のリレーショナルデータベースを完全自動構築するシステムの開発」[http://mighty.gk.u-hyogo.ac.jp/confidential/suoh\\_zensho\\_tokumei.html](http://mighty.gk.u-hyogo.ac.jp/confidential/suoh_zensho_tokumei.html)

## \* 匿名データの読み込みについて

匿名データの読み込みは周防(2015)により公開されているプログラム[3]を用い、あわせて動作の検証と利便性を高めるためのプログラムの作成を行った。

### 1 動作の検証

既知の問題である「WARNING: 記号参照 CHKVNAME を展開していません」という警告は以下の変更により解消された。

対象プログラム: transform.sas 104 行目

変更前：programCode="if &chkVname=' ' then delete; run;";  
変更後：programCode="if &chkVname=" " then delete; run;";

また proc import による読み込みがうまくいかない場合、CSV 形式で保存し以下の部分を変更することで、その後の処理が実行可能となった。

対象プログラム auro\_read\_zensho\_tokumei.sas 15 行目

変更前：

```
proc import out=zensho.zensho_fugohyo&year._&F_T  
  datafile="&fugohyo_path%&year%tokumei_zen&H_year._&which..xls"  
  dbms=excel replace;  
  getnames=no;  
  sheet=符号表;  
run;
```

変更後：

```
data zensho.zensho_fugohyo&year._&F_T ;  
  length F1 - F15 $200 ;  
  infile "&fugohyo_path%&year%tokumei_zen&H_year._&which..csv" dsd missover termstr = crlf lrecl  
= 3000 ;  
  input F1 - F15 ;  
run ;
```

CSV ファイルの読み込みの際に、論理レコード長の指定がないためデフォルトの設定が使用されるが、SAS9.3 以前ではデフォルトの長さが短いため、文字切れを防ぐために以下のコードの追記が必要であった。

対象プログラム： autoexec\_zensho\_tokumei または auro\_read\_zensho\_tokumei.sas

追記箇所： auro\_read\_zensho\_tokumei.sas の実行前または先頭に追記

```
options lrecl = 32767;
```

また生成されるプログラムのうち、CSV ファイルの読み込みを行うプログラムはパスにマクロ変数が含まれているため、単体で実行することができない。これは以下の変更によって単体で実行可能になった。

対象プログラム： auro\_read\_zensho\_tokumei.sas 84 行目

変更前：program\_code=' filename inX "&tokumei\_path%&year.zensyou\_&F\_T..csv";';

変更後：program\_code=%unquote(%quote(' filename inX "&tokumei\_path%&year.zensyou\_&F\_T..csv";'));

## 2 追加プログラム

生成されるデータはコード表に沿ったものであるため、コード表およびデータが 0 埋めをされていない変数はコード表の意図したとおりに並ばない。またフォーマットは生成されたプログラムを実行することで作成できるが、データセットへの設定が必要である。そのため作成されたフォーマットより 0 埋めが必要な変

数を判定して0埋めを行い、フォーマットを設定するプログラムを作成した。

```
/*
処理内容：ライブラリ Zensho にある指定された年の各データセットに対して0埋めと
XzenshoXXXX にフォーマットの設定を行う。
パラメータ：YEAR 実行する年
OUT 出力するライブラリ(指定なしの場合は Zensho を上書き)
autoexec_zensho_tokumei.sas 等のプログラムを実行した直後でなければ
下のマクロ変数とライブラリの場所を指定する。
/*let SASPGM = ;*プログラムを置いてある場所のパス;*/
/*libname ZENSHO "D:\Tokumei\%syouhi%\lib" ;*/

%macro FMT_PAD (year = , OUT = ZENSHO ) ;

* フォーマット作成プログラムの存在確認 ;
%if %sysfunc (fileexist("&SASPGM.\%tokumeiFormat&Year..sas")) = 0 %then %do ;
%put &YEAR. 年のフォーマット作成プログラムは存在しません ;
%return;
%end ;

*ライブラリの確認 ;
%if %sysfunc (libref (&OUT.)) ^= 0 %then %do ;
%put 出力先ライブラリ &OUT. は存在しません ;
%return;
%end ;

%if %sysfunc (libref (Zensho)) ^= 0 %then %do ;
%put ライブラリ Zensho は存在しません ;
%return;
%end ;

*Work のフォーマットを削除 ;
%if %sysfunc (cexist("work.FORMATS")) = 1 %then %do ;
proc catalog cat = work.formats kill ;
quit ;
%end ;

*フォーマットの読み込みと内容の長さを取得 ;
%inc "&SASPGM.\%tokumeiFormat&Year..sas" / source2 ;

proc format cntlout = work.FMTDAT ;
run ;

proc sql noprint ;
create table work.ST_LEN as
select FMTNAME ,
max ( lengthn ( START ) ) as MAX_LEN ,
min ( case when notdigit ( strip ( START ) ) = 0 then lengthn ( START ) else 2 end ) as MIN_LEN
from work.FMTDAT
group by FMTNAME
;
quit ;

proc sort data = work.FMTDAT ;
by FMTNAME ;
run ;

* 0埋めが必要なフォーマットへの処理とマクロ変数の作成 ;
data work.FMTDAT ;
merge work.FMTDAT end = EOF
work.ST_LEN
;
by FMTNAME ;
length CODE ZVAR QZVAR $2000 ;
retain CODE ZVAR QZVAR ;
if MAX_LEN = 2 and MIN_LEN = 1 and notdigit ( START ) = 0 then START = put ( input ( START , best. ) , z2. ) ;
if last.FMTNAME = 1 then do ;
CODE = catx ( " ", CODE , scan ( FMTNAME , 1 , "_" ) , cats( ifc( TYPE = "C" , "$" , "" ) , FMTNAME , "." ) ) ;
if MAX_LEN = 2 and MIN_LEN = 1 then do ;
ZVAR = catx ( " ", ZVAR , scan ( FMTNAME , 1 , "_" ) ) ;
QZVAR = catx ( " ", QZVAR , quote ( scan ( FMTNAME , 1 , "_" ) ) ) ;
end ;
end ;
end ;
```

```

if EOF = 1 then do ;
  call symputx ( "CODE", CODE ) ;
  call symputx ( "ZVAR", ZVAR ) ;
  call symputx ( "QZVAR", QZVAR ) ;
end ;
run ;

proc format cntlin = work.FMTDAT lib = &Out. ;
run ;

proc datasets lib = work nolist ;
  delete FMTDAT ;
quit ;

* 0 埋めが必要なデータセットとそれに含まれる変数を特定 ;
data _null_ ;
  length VLIST DLIST $200 ;
  set sashelp.VCOLUMN end = EOF ;
  by MEMNAME ;
  where libname = "ZENSHO" and prxmatch( "/X.*&YEAR./", MEMNAME ) > 0 and upcase( NAME ) in ( &QZVAR. ) ;
  retain DLIST VLIST ;

  if first.MEMNAME then do ;
    DSN + 1 ;
    DLIST = catx ( " ", DLIST, MEMNAME ) ;
    call missing ( VLIST ) ;
  end ;

  VLIST = catx ( " ", VLIST, NAME ) ;

  if last.MEMNAME then do ;
    call symputx ( cats ( "MEMNAME", DSN ), MEMNAME ) ;
    call symputx ( cats ( "VLIST", DSN ), VLIST ) ;
  end ;

  if EOF = 1 then call symputx ( "DSN", DSN ) ;
run ;

* 0 埋めとフォーマットの設定 ;
%do i = 1 %to &DSN. ;
  data &Out. &&MEMNAME&i. ;
  set Zensho.&&MEMNAME&i. ;
  %if %index ( &&MEMNAME&i. , XZENSHO ) > 0 %then %do ;
    format &CODE. ;
  %end ;
  array ZVAR &&VLIST&i. ;
  do over ZVAR ;
    if notdigit ( strip ( ZVAR ) ) = 0 then ZVAR = put ( input ( ZVAR , best. ) , z2. ) ;
  end ;
  run ;
%end ;

%mend ;

/* 2004 年のデータを上書きせずに Work に出力する */
%FMT_PAD ( YEAR = 2004 , OUT = Work ) ;

/* Work 以外に出力した場合、使用時には options fmtsearch で出力先を指定する */
Options fmtsearch = ( Zensho Work ) ;
%FMT_PAD ( YEAR = 2004 ) ; * Zensho の 2004 年のデータに上書きする場合 ;

```



SASシステム

# DS2 言語によるデータハンドリングについて

森岡 裕

Data Handling with DS2

Yutaka Morioka

## 要旨

DS2 言語は SAS9.4 から導入され、従来のデータステップに比べて、より柔軟で拡張された処理を行うことができる機能である。ステートメントや関数名が類似しているが SAS 言語とは基本的に別言語であり、オブジェクト指向を狙って設計されている。処理をメソッド単位に記述することによりコードの再利用性が高く、従来のデータステップよりも構造化したプログラム設計が可能となっている。

また SQL との親和性が高く、データ型も多く用意されているため、データベースとの連携がしやすくなっている。さらに、マルチスレッド処理を簡単に記述することができることや、行列計算、ハッシュテーブルの処理も含めて記述できるなど、SAS における次世代のデータハンドリングの主軸となりえる機能だといえる。

ただし、やはり従来の SAS データステップとは別言語であることから、学習コストは高いと言わざるをえない。実際に DS2 がどのようなものかを知る前に、導入の検討を打ち切っているケースも少なくない。DS2 の詳細まで解説すると一つの言語のリファレンスを作るに等しく、膨大な量になってしまうため、本稿では DS2 の特徴的な部分と、実務に応用しやすい部分について大まかに解説し、感触を掴んでいただくことを目的とする。

## DS2 の文法の基本構造

DS2 言語を理解する上での、必要最低限の基本構造について説明する。詳細な部分についてはリファレンスで補完いただきたい。なお、本稿では区別しやすくするため、従来の SAS データステップのことを旧データステップと呼称することとする。まず、以下の2つのプログラムを比較してみる。

```
/*PG1*/
data OUT1;
  set TEST1 end=eof;
  if _N_=1 then put 'NOTE:処理の開始';
  Y = X + 1;
  if eof then put 'NOTE:処理の終わり';
run;
```

数値変数 X を含む TEST1 というデータセットがあったとして、それを set して X+1 を新規変数 Y に割り当てて、新規データセット OUT1 を作る。なおかつ処理開始時と終了時にメッセージをだすという処理について、旧データステップで記述したのが PG1 で、DS2 で記述したのが PG2 になる。

PG2 について解説する。まず DS2 言語での処理

```
/*PG2*/
proc ds2 ; /*1*/
  data OUT1(overwrite=yes); /*2*/
  declare double Y; /*3*/
  method init(); /*4*/
    put 'NOTE:処理の開始';
  end;
  method run(); /*5*/
    set TEST1; /*6*/
    Y = X + 1;
  end;
  method term(); /*7*/
    put 'NOTE:処理の終わり';
  end;
enddata; /*8*/
run; /*9*/
quit; /*10*/
```

は proc DS2 ~quit;の間に記述する (1 10 の箇所)。(SAS University Edition または SAS on demand 等の環境については 1 の部分に libs=work の記述が追加で必要になる場合がある)

その中の data ~ enddata; run; (2 8 9) が一つの処理ブロックになる。2 の data ステートメントの記述については旧データステップと同様だが、DS2 ではデータセットの自動上書更新はされず、既に同名のデータセットが存在する場合はエラーとなる。これを回避するために overwrite=yes を追加する必要がある(2)。

3 の箇所は declare ステートメントにより、double という数値型 (倍精度浮動小数点型) を使って新規変数 Y を宣言している。

旧データステップには変数の宣言という概念がなく、割り当てステートメントや、length、format ステートメント等による型定義があれば、それによって新規変数が作成されていたが、DS2 では基本的に新規変数には宣言が必要である。

旧データステップでは変数の型には数値型と文字型しか存在しなかったが、DS2 では一般的なデータベースで用いられる多くの型が存在し、型を使い分けた処理やデータベースとの連携が容易になっている。

ただし、SAS データセットは旧データステップで作成しても DS2 で作成しても同じものであるため、date などの日付型を指定してもデータセットを作成すると数値型に変換される。従来の数値型に対応するのが double で、文字型に対応するのが char なので本稿ではそれを説明に用いる。型の詳細についてはリファレンスを参考にしてください。

4 5 7 は全て method XX ~ end という構造をとっており、これをメソッドという。メソッドにはユーザーが自由に定義して使用するユーザー定義メソッドと、システム定義パッケージ (後述) に依存したメソッド、上記以外で予め用意されている特別な働きを持つシステムメソッドに分類される。

4 5 7 の init run term は全てシステムメソッドで、予め決まった働きをそれぞれ持つ。

4 の init メソッドは DS2 の data-enddata ブロック開始時に、最初一度だけ実行される処理を記述する。旧データステップの if \_N\_=1 then に似た働きを持つ。ただし正確には if \_N\_=1 then は 1 オブザベーション読み込んだ時点、init メソッドは 1 オブザベーションも読み込んでいない時点なので完全に同一ではない。

7 の term メソッドは DS2 データステップの終了時に、最後一度だけ実行される処理を記述する。旧データステップの if end 変数 then に似た働きを持つ。ただし、正確には if end 変数 then は最終オブザベーションの読み込み時点、term メソッドは最終オブザベーションが全て読み込み終わった後 (そのため \_N\_ の値はオブザベーション数+1 になっている) となる。

5 の run メソッドは 1 オブザベーション読み込むごとに行いたい処理を記述する。run メソッドの中に 6 のように set ステートメントを記述することで旧データステップと同様に、set で指定したデータセットが 1 オブザベーション読み込まれるごとに以下の run メソッド内の処理が行われる。

## ユーザー定義メソッド

【データセット: TEST2】

ID	SEX	HEIGHT	WEIGHT
1	M	167	66
2	F	155	68
3	M	177	56
4	F	1.6	86
5	M	183	56
6	F	158	46

システムメソッドおよびシステム定義パッケージ付属のメソッドについては、予めメソッド名も機能も決まっているが、ユーザー定義メソッドは自由に名前と機能を定義することができる。

説明用に上図、6 人分の身長・体重のデータセット【TEST2】を使用する。

例として、体表面積を算出する処理を考える。まず体表面積を計算する前に身長に異常値がないかをチェックする必要があるとする。その場合、右図【PG3】のように値をチェックするメソッドを組み込んで書くことができる。

1 の箇所について、データセットを作成する必要がない場合は旧データステップと同様に \_null\_ を指定できる。

```

/*PG3*/
proc ds2 ;
  data _null_ ; /*1*/
  method h_chk(double h); /*2*/
    if . < h < 100 then put
      'WARNIG:測定単位確認' h=;
  end;
  method run();
    set TEST2;
    h_chk(HEIGHT); /*3*/
  end;
enddata;
run;
quit;

```

2でh\_chkという任意の名前でユーザー定義メソッドを定義している。(double h)というのは、double型でhという引数を受け取るメソッドであるということの意味している。

runメソッド内3の箇所呼び出している(HEIGHT)というのはメソッドの引数hにHEIGHTをあてるという意味である。これにより、実行結果として、「WARNING:測定単位確認 h=1.6」というメッセージがログに出力されることになる。

なお、今回はh\_chkメソッドの記述の後にrunメソッドを記述したが、逆にrunメソッドを先に記述して、後にh\_chkメソッドを記述していた場合は、実行するとエラーになる。

コード上、先行するrunメソッド内で、h\_chkメソッドを呼び出しているが、その時点でh\_chkメソッドの定義がまだ読み込まれていないからである。

もし、後ろに記述されているメソッドをそれよりも前で呼び出したい場合は、forwardキーワードを使う。1のdataステートメントの次の行にforward h\_chk;としておけば、先行するrunメソッド内でh\_chkを呼び出している、後で出現する定義部分を参照してくれるのでエラーにはならない。

異常値が混じっていることが確認できたので、次に、HEIGHTが100未満の数字が入った場合、値を100倍して上書き更新するメソッドと、体表面積を算出するメソッドを組み込んだのが下図PG4になる。右図は実行後の出力データセット【OUT2】になる。

```

/*PG4*/
proc ds2 ;
  data OUT2(overwrite=yes);
  forward h_update cal_bsa; /*1*/
  declare double BSA;
  method run();
    set TEST2;
    h_update(HEIGHT); /*2*/
    BSA=cal_bsa(HEIGHT,WEIGHT); /*3*/
  end;
  method h_update(in_out double h); /*4*/
    if . < h < 100 then h=h*100; /*5*/
  end;
  method cal_bsa(double h,double w) returns double; /*6*/
    declare double result; /*7*/
    result=w**0.425 * h**0.725 * 0.007184; /*8*/
    return result; /*9*/
  end;
enddata;
run;
quit;

```

BSA	ID	SEX	HEIGHT	WEIGHT
1.7423762307	1	M	167	66
1.6717567782	2	F	155	68
1.6948327725	3	M	177	56
1.8902321817	4	F	160	86
1.7362940428	5	M	183	56
1.4357036026	6	F	158	46

結果をみるとID=4の1.6であったHEIGHTの値が160となり、体表面積が変数BSAに格納されている。

コードについて説明する。

1はrunメソッド内で呼び出す2つのメソッドについて、runメソッドの後に記述したために必要となるforwardである。逆順に記述する場合は不要となる。

2で、引数の値が100未満の場合、100倍するh\_updateメソッドにHEIGHTを指定して呼び出している。

4から2行先のendまでの部分がh\_updateメソッドの定義部分となる。4で引数指定のdouble hの前にin\_outがついているが、これはキーワードで、引数かつ戻り値が同じ場合につけるものである。5で条件にあてはまった場合、引数を100倍している。

2の部分で引数にHEIGHTをあてているため、つまりHEIGHTが100未満の場合、HEIGHTを100倍した値をHEIGHTに代入、

つまりは上書き更新していることになる。

3の箇所は、身長と体重から体表面積を計算するcal\_bsaメソッドを呼び出し、計算結果の戻り値を変数BSAに代入している箇所となる。

6から4行先のendまでがcal\_bsaメソッドの定義部分となる。7でresultという変数を宣言しているが、【OUT2】にはresultという変数は存在していない。DS2において変数はメソッドの外で宣言されるグローバル変数とメソッド内で宣言されるローカル変数にわけられる。グローバル変数はどのメソッドからも参照でき、出力データセットにも残るが、ローカル変数は宣言されたメソッドの中でのみ参照され、結果のデータセットには残らないため、ローカル変数resultは出力データセット【OUT2】には残っていない。

6の部分で returns double としているのはメソッドの戻り値の型を指定している部分となる。

8が体表面積の計算部分（デュポア式）で、結果をローカル変数の result に代入している。

9のreturnステートメントは指定した値をメソッドの戻り値にする機能がある。例では結果を一旦、ローカル変数に入れてから戻したが、単純な式であるので、789の部分は return w\*\*0.425 \* h\*\*0.725 \* 0.007184;と一行にしても問題ない。

次にユーザー定義メソッドにおけるオーバーロード（多重定義）について説明する。

オーバーロードとはプログラミング用語で、同一名の関数やメソッド等について複数定義し、呼び出される際の文脈に応じてどの定義で実行するかを自動的に機能選択させる仕組みである。

例えば、先の cal\_bsa メソッドは体表面積を計算するために定義したメソッドで、数値型の身長、体重を引数としている。体表面積を計算する場合は、すべてこのメソッドを呼び出すことができれば便利である。

しかし、例えば、使用したいデータセットの身長、体重が文字型で格納されている場合や、他の計算方法で体表面積を出したい場合などに、いちいち新規のメソッドを定義していくのは煩雑である。とはいえ一つのメソッドを、新しい要望がでるたびに分岐を加えて書き換えていくのも管理上問題がある。

```
/*PG5*/
proc ds2 ;
  data OUT3(overwrite=yes);
  forward cal_bsa;
  declare double BSA a b ; /*1*/
  declare char c d ;      /*2*/
  method init();
    a=160; b=50; c='160'; d='50';
    BSA=cal_bsa(a,b);put BSA=; /*3*/
    BSA=cal_bsa(c,d);put BSA=; /*4*/
    BSA=cal_bsa(a,b,'新谷式');put BSA=; /*5*/
    BSA=cal_bsa(a,b,'藤本式');put BSA=; /*6*/
  end;
  method cal_bsa(double h,double w) returns double; /*7*/
    declare double result;
    result=w**0.425 * h**0.725 * 0.007184;
    return result;
  end;
  method cal_bsa(char h,char w) returns double; /*8*/
    declare double result;
    result=inputn(w,'best.32')**0.425 * inputn(h,'best32.')**0.725 * 0.007184;
    return result;
  end;
  method cal_bsa(double h,double w,char cm) returns double; /*9*/
    declare double result;
    if cm ='新谷式' then result=w**0.425 * h**0.725 * 0.007358;
    if cm ='藤本式' then result=w**0.444 * h**0.663 * 0.008883;
    return result;
  end;
enddata;
run;
quit;
```

これは SAS マクロにも同じことが言える。多くの場所で呼び出しているマクロの内容を変えることは危険であり、パラメータもあまり増やしたくない。しかしマイナーチェンジの度に新しいマクロを増やしたくない。

そこで、PG5 をみってみる。

【結果 ログ画面】

```
BSA=1.50112052420814
BSA=1.50112052420814
BSA=1.53747839882009
BSA=1.45960012718282
```

12からabは数値(double)、cdは文字型(char)となっている。

3は引数が数値、4は引数が文字、56は引数が3つで数値、数値、文字のパターンであるが、呼び出しているのはすべて同名の cal\_bsa メソッドである。

789で同じ名前のメソッドを定義しているが引数の数と型指定がそれぞれ異なるため、共存することができる。これによって呼び出し時の引数によって、3の箇所では7のメソッド、4の箇所では8のメソッド、56の箇所では9のメソッドが自動選択される。

このオーバーロードの機能を利用することで同じ働きのものを同じ名前のメソッドでまとめることができるので非常に便利である。

## ユーザー定義パッケージ

DS2 はメソッドの組み合わせで処理を組み立てていく言語であるが、毎回個々のステップの中でメソッドを定義していると大変で、再利用や共有もできない。

メソッドの定義部分と、実行部分を分離して管理できると便利である。DS2 ではメソッドを、パッケージという塊にまとめて保存し、必要に応じてパッケージを呼び出し、そこに含まれるメソッドを使うことができる。パッケージは外部フォルダにも保存できるため、複数のユーザー間で共有でき、開発効率を高めることができる。

PG4 で紹介した `h_update` メソッドと `cal_bsa` メソッドをパッケージ化したい場合、以下 PG6 のように記述できる。

```
/*PG6*/
proc ds2;
  package work.p_bsa/overwrite=yes; /*1*/
  declare double con;
  method p_bsa()/*2*/
    con=100;
  end;
  method h_update(in_out double h);/*3*/
    if . < h < con then h=h*con;
  end;
  method cal_bsa(double h,double w) returns double;/*4*/
    declare double result;
    result=w**0.425 * h**0.725 * 0.007184;
    return result;
  end;
endpackage;/*5*/
run;
quit;
```

パッケージを作成する場合、1 のように「`package ライブラリ名.パッケージ名/オプション;`」で開始し、5 の `endpackage` で終了する。

`work.p_bsa` としているが `work.` は省略可能である。外部ライブラリに作成する場合は `work` の部分を任意のライブラリ名にする。

2 はパッケージ名と同じメソッドだが、これはコンストラクタと言われる特別なメソッドである。コンストラクタはパッケージを呼び出した際（インスタンス生成時）に強制的に最初に行われるメソッドになる。

パッケージが大規模で、各メソッドで共通する定数が多い場合や、初期値の設定を行いたい場合などにコンストラクタをうまく使うことで、引数や記述の繰り返しを減らせ、コードの見通しをよくすることができる。

3 4 は PG4 で説明したメソッドと同じなので説明を割愛するが、3 のメソッドの式の部分はコンストラクタで設定される `con` の値を参照するようにしている。

作成済みのパッケージを呼び出して、パッケージ内のメソッドを利用している例が PG7 になる。1 の部分、`declare package ライブラリ名.パッケージ名 インスタンス名`（コンストラクタへの引数がある場合は括弧内に引数);で宣言する。

インスタンス名は、パッケージが呼び出されたステップで通用する参照名のようなもので、この例では `ib` という名前を使って `p_bsa` パッケージを参照することとしている。

1 の最後の括弧は、`p_bsa` パッケージのコンストラクタ (`p_bsa` メソッド) に引数があった場合は、そこに値を渡せる部分になる。

```
/*PG7*/
proc ds2;
  data OUT4(overwrite=yes);
  declare package work.p_bsa  ib(); /*1*/
  declare double bsa;
  method run();
    set TEST2;
    ib.h_update(HEIGHT);
    BSA=ib.cal_bsa(HEIGHT,WEIGHT);
  end;
enddata;
run;
quit;
```

なお、コンストラクタになるメソッドにもオーバーロードが可能なので、PG6の2のブロックの後に「method p\_bsa(double dcon);con=dcon;end;」といった引数があった場合のコンストラクタを追加し、PG8の呼び出し時に ib(150)などとして、引数をもってインスタンスを生成する際に値を変更することも可能である。

## システム定義パッケージ

ユーザー定義パッケージとは異なり、あらかじめDS2プロシジャで用意されているシステム定義パッケージが存在する。それぞれのシステム定義パッケージには固有のメソッドが用意されており、それらを利用することでより柔軟な処理が可能になる。

例えば、動的にSQL文を構築でき、DBMSにパススルーを行ったり、結果を受け取ることが可能なSQLSTMTパッケージ。IMLプロシジャでしかできなかった行列の処理が可能なMATRIXパッケージ。ハッシュテーブルを生成し、キールックアップ処理が可能なHASHパッケージなどがある。

中でも、HASHパッケージは旧データステップのハッシュオブジェクトに非常に似ており、さらにハッシュオブジェクトで不便であった箇所が多く改善されているため、ハッシュオブジェクトを習得していれば学習コストが低く、有用性が高い。

【データセット：TEST2 再掲】 【データセット：TEST3】

ID	SEX	HEIGHT	WEIGHT	ID	AGE	COUNTRY
1	M	167	66	1	27	JPN
2	F	155	68	2	33	JPN
3	M	177	56	4	52	USA
4	F	1.6	86	5	41	JPN
5	M	183	56			
6	F	158	46			

前述のデータセット「TEST2」に「TEST3」のAGEとCOUNTRYについてIDをキーにして取得する処理について、旧データステップのハッシュオブジェクト(PG8)とDS2のHASHパッケージ(PG9)を比較してみる

```

/*PG8 ハッシュオブジェクト*/
data OUT5;
if 0 then set TEST3; /*A*/
set TEST2;
if _N_=1 then do; /*B*/
    declare hash h1(dataset:'TEST3');
    h1.definekey('ID');
    h1.definedata('AGE','COUNTRY');
    h1.definedone();
end;
rc=h1.find();
if rc ne 0 then do; /*C*/
    call missing(AGE ,COUNTRY);
end;
run;

```

```

/*PG9 HASH パッケージ*/
proc ds2 ;
    data OUT6(overwrite=yes);
    declare package hash h1(); /*1*/
    declare double AGE; /*2*/
    declare char COUNTRY; /*3*/
    method init(); /*4*/
        h1.dataset('TEST3'); /*5*/
        h1.keys([ID]); /*6*/
        h1.data([AGE COUNTRY]); /*7*/
        h1.definedone();
    end;
    method run();
        set TEST2;
        h1.find(); /*8*/
    end;
enddata;
run;

quit;

```

PG9 の 1 が HASH パッケージの呼び出し（インスタンス生成）部分で、declare 後の hash が固定キーワードである点以外はユーザー定義パッケージの呼び出しと同様となる。

PG9 の 2,3 の箇所について、旧データステップでは変数の宣言という概念がないため、変数情報を PDV 上に用意するために PG8 の A のような書き方や length ステートメントなどが必要であったが、DS2 ではそれが変数の宣言に代わり、理解しやすくなった。

旧データステップでは、1 オブザベーションずつ処理するという SAS の特性上、ハッシュオブジェクトの定義部分が何度も繰り返されないように B の箇所のような if ステートメントでの工夫が必要であったが、DS2 では事前実行される init メソッド(4 の箇所)に定義部分を入れればよく、見通しがよくなった。

5 の箇所、旧データステップでは declare hash の際に括弧内に、「dataset:」や「ordered:(ハッシュテーブル内の昇順降順指定)」、「multidata(重複キーの許容)」などを列挙していく定義方式だったが、DS2 ではそれぞれ独立したメソッドとして定義可能となった。

6 7 の箇所についてそれぞれ key と data の指定を[]内に変数リストで指定できる。旧データステップのハッシュオブジェクトと同様に definekey メソッド、definedata メソッドも存在するが、1 メソッドごとに 1 変数のみ指定なので、複数変数がある場合いくつも繰り返す必要がある。keys メソッド、data メソッドであれば 1 変数から複数変数まで対応可能なので使いやすい。

なお、別の定義法として 5 6 7 の内容やその他オプションをすべて、1 の括弧内に決まった順番に引数指定して定義する方法も可能で、マクロ化する際には特に有用であるが本稿では割愛する。

PG8 の C の箇所は、旧データステップのハッシュオブジェクトでは find メソッドにより key マッチできなかった場合、直前の結果を retain してしまう性質があったため、このような分岐が必要であった。

しかし DS2 では 8 のように特に分岐処理をしなくても問題がなくなった。

## マルチスレッド処理

DS2 にはスレッドという仕組みが存在する。スレッドで定義した処理について、実行時に並列実行したい数を指定することによって同時に処理が実行され、処理時間を大きく短縮することができる。

スレッド定義内に set ステートメントが存在する場合（何らかのデータソースからデータを取ってくる場合）と存在しない場合によってイメージが少し異なるので簡単な例で説明したい。

```
/*PG10 set を含まないスレッドの定義*/
proc ds2;
thread th1/overwrite=yes; /*1*/
  declare double i thread;
  method run();
    do i = 1 to 15;
      thread=_threadid_; /*2*/
      put 'スレッド番号:_'_threadid_ i=;
      output;
    end;
  end;
endthread; /*3*/
run;
quit;
```

```
/*PG11 set を含まないスレッドの実行*/
proc ds2;
data OUT7(overwrite=yes);
  declare thread th1 t; /*1*/
  declare double NO;
  method run();
    set from t threads=3; /*2*/
    NO=_N_;
  end;
enddata;
run;
quit;
```

PG10 がスレッドを定義するプログラム、PG11 がスレッド数を指定して定義したスレッドを実行するプログラムになる。

PG10 について、スレッド定義の仕組みとしてはパッケージのそれに似ている。1 の thread から 3 の endthread までにスレッド化したい処理を記述する。1 の th1 の部分はスレッド名をつけるところで、任意の名前で定義できる。今回は work に作成されるが、永久ライブラリに作成することも可能である。



2 の `threadid` は、実行時に並列で処理される場合に、どのスレッドで実行されたかが入る自動変数になる（0 から連番）。

PG11 では PG10 で作成されたスレッド `th1` を使って、`t` というインスタンス名を与えて参照している部分になる。

2 の部分、`set from` はスレッドを実行して、そこからデータを取得するためのステートメントになる。`threads=` で、スレッドをいくつ実行するかを指定しており、例では 3 を指定している。

PG11 の実行で作成されるデータセット `OUT7` の内容と、ログへの出力については本稿最末の補足図に掲載している。内容をみると 45 オブザベーションであるが、これは 1 スレッドあたり 15 オブザベーション出力する処理を 3 スレッドで実行したためである。変数 `thread` には `threadid` の値を割り当てているが、そこから、`set from` によち 3 スレッドのアウトプットを統合して一つのデータソースとする際に、スレッド番号 0,2,1 の順にセットされたことがわかる。この順番はランダムになる。

またログ（巻末補足図）を見ているとそれぞれのスレッドが並行して実行されているが、ログウィンドウは一つを共有しているので、ランダムにミックスされたようなログ出力になる。

次にスレッド定義内に `set` を含む場合について説明する。PG12 は 100 変数 100 万オブザベーションのデータセット `TEST4` を作成する旧データステップのプログラムになる。PG13 はスレッド定義部分で、`term` メソッドで最後に `threadid` とオブザベーション数のカウントをログに出力する以外は単純に `TEST4` を `set` するだけの内容になっている。PG14 では定義されたスレッドについて、3 スレッド数で実行しており、その他の処理は特に行っていない。

```
/*PG12 テストデータ*/
data TEST4;
array ar{100};
do i= 1 to 10000000;
  do j= 1 to 100;
    ar{j}=rand('uniform');
  end;
  output;
  drop i j;
end;
run;
```

```
/*PG13 スレッド定義*/
proc ds2;
thread th/overwrite=yes;
  dcl double count;
  drop count;
  method run();
    set TEST4;
    count +1;
  end;
  method term();
    put 'スレッド番号' threadid 'が'
      count 'obs 読み込みました';
  end;
endthread;
run;
quit;
```

```
/*PG14 スレッド実行*/
proc ds2;
data OUT8(overwrite=yes);
  dcl thread th t;
  method run();
    set from t threads=3;
  end;
enddata;
run;
quit;
```

PG14 を実行した際のログが下図になる

```
スレッド番号 2 が 3210768 obs読み込みました
スレッド番号 0 が 3584160 obs読み込みました
スレッド番号 1 が 3205072 obs読み込みました
```

100 万オブザベーションの処理が、SAS により適当に 3 分割で並列処理が行われたことがわかる。

環境によって異なるので参考程度であるが、PG5 の処理時間は、旧データステップと同様に単純 `set` するプログラムを書いて実行した場合より 4 割程度高速化する結果を得ている。

このように単純に一度定義してから呼び出して実行するだけで、簡単に処理を高速化できるのでスレッドは非常に有用な機能といえる。

ただし、どこで分割して処理されるかは制御できず、各スレッドがそれぞれ独立して実行される。そのため、データセット全体を通して値を retain するような処理を含めると正しい結果にならないため、注意が必要である。基本的には 1 オブザベーションごとに完結している処理で構成しなければならない。

## DS2 で特徴的なコードの記述法について

旧データステップに習熟した後に DS2 を学習し始めた際に、個人的に便利に感じられた点について 2 つほど紹介したい。

### ① by ステートメントを入れただけで自動ソートが起きる

例えば、前述の TEST2 のデータセットについて、性別ごとに身長が最大のオブザベーションにフラグ変数=1 を付与したい場合以下のように記述することができる。結果のデータセットは右図になる

Code	FL	SEX	HEIGHT	ID	WEIGHT
/*PG15 by によるソート*/	.	F	1.6	4	86
proc ds2 ;	.	F	155	2	68
data OUT10(overwrite=yes);	1	F	158	6	46
declare double FL;	.	M	167	1	66
method run();	.	M	177	3	56
set TEST2;	1	M	183	5	56
by SEX HEIGHT; /*1*/					
if last.SEX then FL=1; /*2*/					
end;					
enddata;					
run;					
quit;					

旧データステップでは set ステートメント後に by ステートメントをつけて、first last の処理をする場合、事前に sort プロシジャなどでデータセットをソートするステップが必要であった。

DS2 では 1 2 の箇所のように by ステートメントが set 後に出現すると、set で指定されているデータにソートがかかるようになっている。

これによってコード上の記述が簡略化され、見易くなった。

ただし、細かい注意点として DS2 のソートアルゴリズムはソートプロシジャと異なり、SQL で order by 句をいれた場合のものと一致する。

### ② データセットの指定箇所に SQL 文を記述することができる

DS2 は SQL との親和性が高く、set ステートメントでデータセット名を指定する箇所などに SQL 文をそのまま指定することができる。

例えば前述の TEST2 データセットを用いて、性別ごとに体重と身長の平均をとり、それを使って体表面積を計算するコードを書く場合、次頁 PG16 のように記述できる。

```

/*PG16 set ステートメントに{SQL}*/
proc ds2 ;
data OUT11(overwrite=yes);
declare double BSA;
method run();
    set {select SEX,mean(HEIGHT) as HEIGHT ,mean(WEIGHT) as WEIGHT
        from TEST2
        group by SEX
    };
    BSA = WEIGHT**0.425 * HEIGHT **0.725 * 0.007184;
end;
enddata;
run;

```

また set ステートメントだけでなく、例えば HASH パッケージの dataset メソッドの指定部分にも SQL 文を記述できる。

TEST2 の SEX に、以下のデータセット TEST4 を使って算出した SEX ごとの VAL の平均値を結合したい場合、以下のように記述することができる。

#### 【TEST4】

ID	SEX	VAL
A	M	10
B	M	14
C	M	13
D	M	15
E	F	9
F	F	8
G	F	11
H	F	16

括弧{}内の SQL が先に実行され、その結果が仮想的にデータセットとみなされていることがわかる。

旧データステップにおいてもハッシュ反復子オブジェクトを利用すればあらかじめ加工した結果をハッシュテーブルに格納できたが、do while ループを変則的に利用する必要があり、難易度が高かった。しかし DS2 では SQL の知識があればかなり容易に記述することができる。

```

/*PG17 HASH パッケージのステートメントに{SQL}*/
proc ds2 ;
data OUT12(overwrite=yes);
declare package hash h1();
declare double ID V_MEAN;
drop mean;
method init();
    h1.dataset('{select SEX ,mean(VAL) as V_MEAN
                from TEST4 group by SEX}');
    h1.keys([SEX]);
    h1.data([V_MEAN]);
    h1.definedone();
end;
method run();
    set TEST2;
    h1.find();
end;
enddata;
run;
quit;

```

## まとめ

本稿では、DS2 言語でプログラミングする上で、必要となる最低限の知識を取り上げた。本稿が DS2 に興味を持つ一つのきっかけになれば幸いである。

もし興味を持たれた方は、ステートメント、オプション、関数、その他多くの点において本稿で紹介できていない DS2 独自の仕様があるので、ぜひリファレンス等で詳細を確認していただきたい。

SAS 言語は歴史が長く、バージョンアップとともに機能追加されてきた。しかし、データステップにおける根本的な文法はほとんど変わっていない。そのため再利用性の高い処理、可変的な処理、共有機能などは全て SAS マクロでまかなう習慣が完全に根付いているといえる。

SAS マクロは強力なツールではあるが、あくまでテキスト展開による処理に過ぎず、大規模かつ複雑なプログラムを組む上で必ずしも最適とは言えない部分があった。

またデータステップと、SQL プロシジャ、IML プロシジャがそれぞれ独立していたため、それらを混ぜて一つの処理を作ることが難しい。

加えて、現在主流の他の開発言語と比べて、言語的に旧世代のため、他言語から SAS に乗り換える際の学習コストが非常に高かった。

それらの問題を解決する上で DS2 言語は大きな役割を果たせると考えている。当面は全ての処理を DS2 だけで記述するという事は難しいかもしれないが、効果的な箇所を徐々に置き換え、組織として管理していく体制を整えば、開発効率の向上に結びつくはずである。一人でも多くのユーザーが DS2 を学習し、知見を共有していけるようにしていきたい。

## 参考資料

- (1) Peter Eberhardt (2016) The Ds2 Procedure: SAS Programming Methods at Work
- (2) SAS® 9.4 DS2 Language Reference Fifth Edition SAS

補足図

【データセット：OUT7】

【ログ】

NO	i	thread
1	1	0
2	2	0
3	3	0
4	4	0
5	5	0
6	6	0
7	7	0
8	8	0
9	9	0
10	10	0
11	11	0
12	12	0
13	13	0
14	14	0
15	15	0
16	1	2
17	2	2
18	3	2
19	4	2
20	5	2
21	6	2
22	7	2
23	8	2
24	9	2
25	10	2
26	11	2
27	12	2
28	13	2
29	14	2
30	15	2
31	1	1
32	2	1
33	3	1
34	4	1
35	5	1
36	6	1
37	7	1
38	8	1
39	9	1
40	10	1
41	11	1
42	12	1
43	13	1
44	14	1
45	15	1

スレッド番号: 0 i=1  
 スレッド番号: 0 i=2  
 スレッド番号: 0 i=3  
 スレッド番号: 0 i=4  
 スレッド番号: 0 i=5  
 スレッド番号: 1 i=1  
 スレッド番号: 0 i=6  
 スレッド番号: 0 i=7  
 スレッド番号: 1 i=2  
 スレッド番号: 0 i=8  
 スレッド番号: 1 i=3  
 スレッド番号: 0 i=9  
 スレッド番号: 1 i=4  
 スレッド番号: 1 i=5  
 スレッド番号: 1 i=6  
 スレッド番号: 1 i=7  
 スレッド番号: 1 i=8  
 スレッド番号: 0 i=10  
 スレッド番号: 1 i=9  
 スレッド番号: 0 i=11  
 スレッド番号: 1 i=10  
 スレッド番号: 0 i=12  
 スレッド番号: 1 i=11  
 スレッド番号: 0 i=13  
 スレッド番号: 2 i=1  
 スレッド番号: 0 i=14  
 スレッド番号: 0 i=15  
 スレッド番号: 2 i=2  
 スレッド番号: 1 i=12  
 スレッド番号: 2 i=3  
 スレッド番号: 1 i=13  
 スレッド番号: 2 i=4  
 スレッド番号: 1 i=14  
 スレッド番号: 2 i=5  
 スレッド番号: 1 i=15  
 スレッド番号: 2 i=6  
 スレッド番号: 2 i=7  
 スレッド番号: 2 i=8  
 スレッド番号: 2 i=9  
 スレッド番号: 2 i=10  
 スレッド番号: 2 i=11  
 スレッド番号: 2 i=12  
 スレッド番号: 2 i=13  
 スレッド番号: 2 i=14  
 スレッド番号: 2 i=15

# デフォルトと複数格付の同時推計による ロジットモデル構築手法の提案

木村和央\*

株式会社金融工学研究所 技術統括部

Proposals for Logit Model Building by the Simultaneous Estimation of Default and Multiple Ratings

Kazuo Kimura

Technology Management Div., Financial Technology Research Institute Inc.

## 要旨

ロジットモデル構築にあたっては、限られた時間コストの中で、人的スキルによらずに、一定精度を確保したモデルを得たいという実務上の要請がある。木村 (2015a) では、ソブリン (国) にかかわるデフォルト確率および格付該当確率モデルの構築を題材に、Weight of Evidence (WoE; 対数オッズ, Zスコア) を中心に、外れ値処理, 変数変換, 欠損値補完, 同時推計といった手法の提案がなされた。本稿では、これらの基本的手法をベースとして、格付該当確率モデルについて単一格付から複数格付に自然に拡張して同時推計するという手法を提案した。この結果を利用して、格付間の寛厳性についての比較評価がなされた。なお、SAS プログラムについては、NLP プロシジャより一般的な NLMIXED プロシジャへと移行させた。本稿で提案された手法は、一般のロジットモデル構築時にも適用可能な基本的手法の集合体であり、上記の実務上の要請を効率的に達成するための指針となれば幸いである。

キーワード: ソブリン, 外れ値処理, Yeo-Johnson 変換, 欠損値補完, 重回帰モデル, 同時推計, 寛厳性評価, NLMIXED プロシジャ, FMM プロシジャ

## 1. はじめに

企業等が一定期間後にデフォルトとなる確率, あるいは格付に該当する確率を推計するために, ロジットモデルが金融機関実務の中で広く活用されている。モデル構築にあたっては、限られた時間コストの中で、人的スキルによらずに、最良でなくても構わないが一定精度を確保したモデルを得たいという実務上の要請がある。これに対し、木村 (2015a) では、ソブリン (国) のデフォルト確率および格付該当確率モデルの構築を題材に、Weight of Evidence (WoE あるいは対数オッズ, Zスコア) を中心に、外れ値処理, 変数変換, 欠損値補完, 同時推計といった手法の提案がなされた。ソブリンはサンプルが少なく、採用される説明変数の数が少なくなることが予想され、連続型変数の離散により定義される一般の WoE を用いると、結果である

\* 本稿の内容は筆者に属し、所属組織の見解ではない。さらに、商用利用目的でないことも明記しておく。

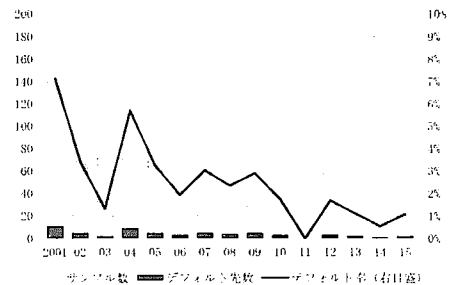
デフォルト確率も離散的になる恐れがあるため、Yeo-Johnson 変換 (Yeo and Johnson, 2000) を用いて連続型変数としての WoE に変換がなされた。また、モデル構築のベースとなる過去データは、一部の国のデータ入力が遅延することに加え、そもそも収録されないデータ項目も多数存在するため、欠損値を含むサンプルを除外してモデル構築をすることは不可能である。これについては前述の WoE を用いて重回帰モデルを繰り返し推計する単一代入法が提案された。加えて、デフォルト確率モデルと格付該当確率モデルを独立に推計したのではデフォルト確率と推計格付間の整合性がとれないため、両モデルで説明変数とウエイトを共通化し、両モデルの対数尤度加重和を最大化するという枠組みで同時推計がなされた。

本稿では、木村 (2015a) が単一の格付を想定していたところを、複数の格付にも対応できるよう自然な拡張を試みた。また、この結果を利用した格付間の寛厳性評価 (山下, 2005) として、モデルパラメータを用いた評価手法を提案した。最後に付録として SAS プログラムを添付した。なお、SAS プログラムについては、木村 (2015a) で重用された SAS/OR に含まれる NLP プロシジャを、より一般的な SAS/STAT に含まれる NLMIXED プロシジャへと移行した。

## 2. データと分析の枠組み

対象国は世界銀行 (世銀) の加盟国である 188 か国、対象期間は 2001 年から 2015 年までの 15 年間を採用した。有効なサンプル数は 2,785 先 (国・年) であった。説明変数用データとしては、世銀の World Development Indicators (WDI) と The Worldwide Governance Indicators (WGI) の年次データをベースに、WDI に収録されていない国の対外債務残高については、Quarterly External Debt Statistics (QEDS/SDDS) と CIA の The World Factbook により補完した<sup>1</sup>。なお、原データに存在する指標以外にも、いくつかの指標を計算し追加、さらにフロー型変数を利用する指標は、過去 2 年の算術ないし 3 年の幾何平均操作を実施した。

デフォルトフラグについては、主要債権国会議 (パリクラブ) へ持ち込まれた年と Cruces and Trebesch (2013) にて調査された債務リストラ (debt restructuring) 年をデフォルトと認定し作成した<sup>2</sup>。なお、デフォルト年の翌年と翌々年はデフォルトが継続している状態 (既デフォルト状態) とみなし、さらに、その翌年は非デフォルト状態に復帰したか否かの確認に用いるため、デフォルト後の合計 3 年間でデフォルト確率モデル推計時の分析対象外とした。図表 1 にデフォルト率の時系列遷移を示した。デフォルト確率モデル推計時の分析対象サンプル数は



図表 1: デフォルト率の時系列遷移

2,495 先 (全体の 89.6%)、うちデフォルトサンプル数は 60 先、平均デフォルト率は 2.405% であった。

格付については、格付投資情報センター (R&I) を含む格付機関 5 社が前年 12 月末に付与していた格付を当年の格付とした。対象期間における格付該当確率モデル推計時の分析対象サンプル数は、1 社でも格付が振られていたことを条件として、1,706 先 (全体の 64.1%) であった。

モデル構築の枠組みとしては、データの収録時期の遅れを考慮し、 $t-2$  年の WDI と  $t-3$  年の WGI により、 $t$  年のデフォルト確率および格付該当確率を推計することとした。

<sup>1</sup> 為替換算方式、基準日の相違等で、各データ間の対外債務残高は必ずしも一致しないが、補正せず利用した。

<sup>2</sup> 債務リストラ年データでは、記憶に新しい 2005 年のアルゼンチン、2012 年のギリシャがデフォルトとして認識される。

### 3. 幾つかの数学記号の事前準備

$i$ : サンプル番号,  $N$ : サンプル数,  $j$ : 変数番号,  $M$ : 変数の数,  $x_{ji}$ : 変数 (候補指標),  $u_{ji}$ : データ有無フラグ ( $x_{ji}$ が欠損値のとき 0, データが存在するとき 1),  $\delta_i$ : デフォルトフラグ (デフォルトのとき 1, 非デフォルトのとき 0)とする. 変数  $j$  について  $x_{ji}$ が欠損値でないサンプル番号の集合を  $\mathcal{N}_j = \{i | (1 \leq i \leq N) \cap (u_{ji} = 1)\}$ , サンプル  $i$  について  $x_{ji}$ が欠損値でない変数番号の集合を  $\mathcal{M}_i = \{j | (1 \leq j \leq M) \cap (u_{ji} = 1)\}$ とする. なお, サンプルにおいて, 対象国と基準年を明示的に扱う必要がある場合は,  $i = i(l, t)$ と表記し,  $l$ : 対象国,  $t$ : 基準年とする.

### 4. 外れ値処理と Yeo-Johnson 変換を用いた変数変換

WoE の定義は, サンプル全体と部分集合の (デフォルト率に関する) 対数オッズ比であるため, 一般に連続型変数の場合は離散化(カテゴリ化)が必要とされる. サンプル全体のデフォルト率を  $p_0$ , 対数オッズを  $\mu$ ,  $x_{ji}$ を離散化し,  $c$ 番目のカテゴリ ( $x_{j,c} \leq x_{ji} < x_{j,c+1}$ ) に属するサンプルのデフォルト率を  $p_{j,c}$ , 対数オッズを  $z_{j,c}$ として, このカテゴリの WoE ( $\text{WoE}_{j,c}$ ) は,

$$\text{WoE}_{j,c} = \ln\left(\frac{p_{j,c}}{1 - p_{j,c}}\right) - \ln\left(\frac{p_0}{1 - p_0}\right) = z_{j,c} - \mu,$$

$\mu$ は定数項であるので,  $\text{WoE}_{j,c}$ は各カテゴリの対数オッズ  $z_{j,c}$ と同一視しても構わないであろう. さらに, 連続型変数を離散化せず, 連続型変数のまま単変数にてロジットモデルを推計し, 得られた対数オッズ (Z スコア) を実数空間上で定義された WoE と考えてもよい. ただし, 連続型変数を説明変数とする際には, 外れ値の処理と変数変換の選択が, モデル精度向上のためには不可欠である. 前者についてはグリッドサーチ, 後者については Box-Cox 変換を負側へ拡張した Yeo-Johnson 変換 (Yeo and Johnson, 2000) を用いて対応した.

具体的な処理の流れは, 以下のとおり.

- (1) 変数  $x_{ji}$  ( $i \in \mathcal{N}_j$ ) を選択.
- (2)  $x_{ji}$  の下限値候補  $x_{j,k_1}^{\min}$  ( $k_1 = 1, \dots, K_1$ ) と上限値候補  $x_{j,k_2}^{\max}$  ( $k_2 = 1, \dots, K_2$ ) を列挙.  
実際の下限候補値は 1%, 2%, ..., 20%, 上限候補値は 80%, 81%, ..., 99% タイル値.
- (3)  $x_{j,k_1}^{\min}$  と  $x_{j,k_2}^{\max}$  を組み合わせて  $K = K_1 \times K_2$  個の候補変数  $y_{jki}$  を定義.
 
$$\left. \begin{array}{l} x_{ji} \leq x_{j,k_1}^{\min} \text{ のとき } y_{jki} = x_{j,k_1}^{\min} \\ x_{j,k_1}^{\min} < x_{ji} < x_{j,k_2}^{\max} \text{ のとき } y_{jki} = x_{ji} \\ x_{j,k_2}^{\max} \leq x_{ji} \text{ のとき } y_{jki} = x_{j,k_2}^{\max} \end{array} \right\} y_{jki} = \max(\min(x_{ji}, x_{j,k_2}^{\max}), x_{j,k_1}^{\min}).$$
- (4)  $K$  個の  $y_{jki}$  に対し, 基準点移動 ( $c_{jk}$ ) を考慮し, 累乗パラメータ ( $\lambda_{jk}$ ) を 1 つだけでもつ Yeo-Johnson 変換を施したものを説明変数とした単変数ロジットモデルを最尤法にて推計<sup>3</sup>.  
パラメータは,  $\min(x_{j,k_1}^{\min}, 0) \leq c_{jk} \leq \max(x_{j,k_2}^{\max}, 0)$ ,  $-1 \leq \lambda_{jk} \leq 1$  の範囲に制約.
- (5) 得られた  $K$  個の結果のうち, 尤度が最大となる時のパラメータ ( $\alpha_{jk}, \beta_{jk}, \lambda_{jk}, c_{jk}, x_{j,k_1}^{\min}, x_{j,k_2}^{\max}$ )

<sup>3</sup> 基準点における 2 階の微分係数の連続条件を課さなければ, 基準点の左右で累乗パラメータを異なるものとしてよいが, 実際には, ほとんどの変数において AIC 基準での精度向上 (対数尤度の 1 以上の改善) は認められなかった.



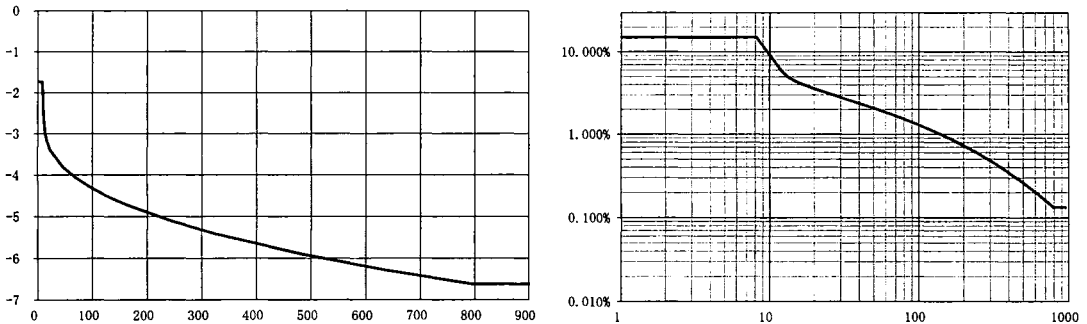
および $z_{jki}$ を採用し、変数 $x_{ji}$ からZスコア $z_{ji}$ への変換式を確定（インデックス $k$ を削除）。  
 後続の欠損値処理に利用するため、Zスコアの上下限値 $z_j^{\min}$ ,  $z_j^{\max}$ を計算。

単変数ロジットモデルの構造は以下のとおり。

$$z_{jki} = \alpha_{jk} + \beta_{jk} \cdot \left[ \frac{\{1 + \max(y_{jki} - c_{jk}, 0)\}^{(1+\lambda_{jk})} - 1}{1 + \lambda_{jk}} - \frac{\{1 - \min(y_{jki} - c_{jk}, 0)\}^{(1-\lambda_{jk})} - 1}{1 - \lambda_{jk}} \right],$$

$$p_{jki} = 1 / (1 + e^{-z_{jki}}),$$

なお、 $\lambda_{jk} = \pm 1$ のときは、 $\lambda_{jk} \rightarrow \pm 1$ の極限と解釈することで、自然対数関数となる。また、推計にあたっては、推計用デフォルトフラグ ( $\delta'_i$ ) として、非デフォルト先は一律ゼロではなく、0.030% (AAA 格), 0.035% (AA+, AA, AA-格), 0.040% (A+, A, A-格), 0.045% (BBB+, BBB, BBB-格), 0.050% (BB+格以下, 無格付) とみなし、格付機関 5 社で平均した値により微小なデフォルト率を与えた<sup>4</sup>。このことにより、デフォルトのない上位格の序列と、Zスコアが発散しないよう推計の安定化に配慮した。



図表 2：1人当たり GDP/世界平均[%]（横軸）とZスコアの関係（左図），デフォルト確率換算との関係（右図）

一例として、図表 2 に 1 人当たり GDP/世界平均の原数値[%]と Zスコアの関係（左図），デフォルト確率換算したものとの関係（右図）を示した。本変数の基準点移動パラメータは $c_j = 13.18$ 、累乗パラメータは $\lambda_j = -0.61$ であった。原数値と Zスコアの関係は、基準点の右側では 0.4 乗変換のため平方根変換に近く、左側では（区間が短いのでわかりにくいものの）1.6 乗変換となっている。また、Yeo-Johnson 変換の特性から、基準点において滑らかに接続されていることもわかる。

本作業にあたっては、非線形な変数変換パラメータのある説明変数を用いた単変数ロジットモデルの推計に対応するため NLMIXED プロシジャを利用した（詳細は付録を参照）。NLP プロシジャと同様、NLMIXED プロシジャにおけるパラメータ制約条件も、BOUNDS ステートメントにより指定すればよい。上下限値のグリッドサーチにあたっては、マクロ変数ループにて、1 回目は全探索範囲を 4%刻みのグリッドで探索し尤度基準により最適格子点を決定、2 回目は 1 回目の最適格子点の周囲を 2%刻みのグリッドで探索し最適格子点を決定、3 回目は 2 回目の最適格子点の周囲を 1%刻みのグリッドで探索し最終的な最適格子点を決定することで、総当たりでの探索を回避し総探索回数を抑制した<sup>5</sup>。

<sup>4</sup> 0.03%はバーゼル規制に定められるデフォルト確率の下限値であり、設定の参考とした。

<sup>5</sup> したがって、総当たりを実施したときの最適格子点とは異なるかもしれないが、実務的にはそれで十分である。

## 5. 重回帰モデルによる欠損値補完処理

Zスコア $z_{ji}$ が欠損値である部分 ( $i \notin \mathcal{N}_j$  または  $j \in \mathcal{M}_i$ ) につき、欠損値パターンに応じた4種類の重回帰モデルを繰り返し推計する単一代入法を用いて補完した。欠損値補完をデータ原数値でなくZスコアを利用することの利点は、原数値の欠損値補完において留意しなくてはならない各変数間での分布の相違、あるいは偏りの補正といったことから解放されることが大きい。

具体的な処理の流れは、以下のとおり。

(1)  $z_{ji}$  ( $i \in \mathcal{N}_j$ ) の初期値 (推計値) として、 $j \in \mathcal{M}_i$  のZスコア $z_{ji}$ の単純平均値などにより補完。

ただし、前章で決定されたZスコアの上下限値の範囲を超えた場合は、上下限値を採用。

(2)  $z_{ji}$  ( $i \in \mathcal{N}_j$ ) を被説明変数 (Y) に、説明変数 (X) の異なる重回帰モデル A~D を推計。

A: X は Y 以外の同一年 ( $t$ ) の Zスコア $z_{j'i}$  ( $j' \neq j$ ) 推計対象:  $i \in \mathcal{N}_j$

B: X は A の説明変数+Y の $t-1$ 年のZスコア $z_{ji^-(l,t-1)}$  推計対象:  $i^-, i \in \mathcal{N}_j$

C: X は A の説明変数+Y の $t+1$ 年のZスコア $z_{ji^+(l,t+1)}$  推計対象:  $i, i^+ \in \mathcal{N}_j$

D: X は A の説明変数+Y の $t \mp 1$ 年のZスコア $z_{ji^-(l,t-1)}, z_{ji^+(l,t+1)}$  推計対象:  $i^-, i, i^+ \in \mathcal{N}_j$

推計にあたっては、切片 ( $\alpha_j^M$ ) を除く係数パラメータは正值に制約 ( $\beta_{j'j}^M, \gamma_{-j}^M, \gamma_{+j}^M \geq 0$ )。

なお、 $z_{ji}$  ( $i \in \mathcal{N}_j$ ) が上限値 ( $z_j^{\max}$ ) または下限値 ( $z_j^{\min}$ ) をとり、前プロセスで得られた $z_{ji}$ の推計値が $z_{ji}$ と同じ上限値または下限値にある場合は、当該サンプルは推計対象外 (推計値が上下限値の外側にある場合、一般の最小二乗法では誤差がゼロと評価されないことを回避するため)。

(3)  $i \notin \mathcal{N}_j$  に欠損値パターンにしたがって、重回帰モデル A~D を適用し、Zスコアの推計値 $\hat{z}_{ji}$ を計算。

A: 時系列全データが欠損値のとき適用

B: 過去方向にのみ1時点以上データが存在するとき適用

C: 将来方向にのみ1時点以上データが存在するとき適用

D: 両方向に1時点以上データが存在するとき適用

なお、 $\hat{z}_{ji}$ の値が上限値 ( $z_j^{\max}$ ) または下限値 ( $z_j^{\min}$ ) の外側になる場合は、上限値または下限値に変更。

(4)  $z_{ji}$  と  $\hat{z}_{ji}$  を比較し、変化が一定値以下となれば終了。乖離がある場合は、 $z_{ji}$  ( $i \notin \mathcal{N}_j$ ) の値を $\hat{z}_{ji}$ 方向へ近づけて更新し、(2)の重回帰モデルの再推計。

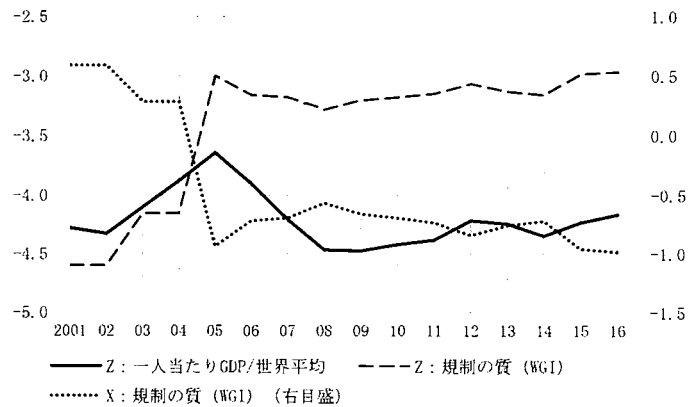
重回帰モデルの構造は以下のとおり。

$$\hat{z}_{ji}^M = \max \left( \min \left( \alpha_j^M + \sum_{j' \neq j} \beta_{j'j}^M \cdot z_{j't} + \underbrace{\gamma_{-j}^M \cdot z_{ji^-(l,t-1)}}_{M=B \text{ or } D} + \underbrace{\gamma_{+j}^M \cdot z_{ji^+(l,t+1)}}_{M=C \text{ or } D}, z_j^{\max} \right), z_j^{\min} \right).$$

ここで、 $M = A, B, C, D$ : モデルタイプである。数式では幾分わかりにくいので、コンセプトを要約する。モデル A は被説明変数 (Y) と同時点 (クロスセクション) の Zスコアを説明変数 (X) に利用したモデルである。当該国における Y の時系列データが全て欠損値の場合は、モデル A を利用して補完することになる。しかし、Y の時系列データの幾つかにデータが存在する場合は、それを利用したほうがクロスセクションのみから推計するよりも精度は高いはずである。モデル B/C/D は Y 自身の1期前のみ/後のみ/前後両方のデータを X に加えて利用したモデルである。当該国における Y を推計したい時点より前方/後方/前後両

方に1つでもデータが存在すれば、そこを起点として将来方向/過去方向/中心方向にモデルB/C/Dで当該時点まで連続的に推計がなされるため、安定的な結果を得ることが期待できる。

一例として、図表3に全期間で欠損値補完されたアルゼンチンの一人当たりGDP/世界平均Zスコアと、それと関係の強い規制の質(WGI)の原数値とZスコアを示した<sup>6</sup>。ここでの横軸はデフォルト観測年であり、実際のデータは、WDIは2年前、WGIは3年前のものを参照していることに留意しなければならないが、アルゼンチンは2005年にデフォルトしており、一人当たりGDP/世界平均Zスコアの動きと整合的である。実際の重回帰モデルAを確認したところ、係数パラメータが最大



図表3：欠損値補完されたアルゼンチンの1人当たりGDP/世界平均のZスコアと規制の質(WGI)の時系列推移(2005年にデフォルト)

であったものは、規制の質(WGI)のZスコアであり、図表より、その影響を反映したものと解釈できる。ただし、規制の質(WGI)のZスコアは、2005年以降も高止まりしていることから、別の変数の影響により、一人当たりGDP/世界平均Zスコアは2003年以前の水準まで低下したものと考えられる。

本作業にあたっては、係数パラメータを正值に制約する重回帰モデルの推計に対応するためFMMプロシージャを利用した(詳細は付録を参照)。それ以外の部分、たとえばモデルA~Dの繰り返し処理、欠損値パターンによる推計値の選択、上下限值処理、収束判定などは、マクロとデータステップで処理した。

## 6. デフォルト確率モデルと複数の格付該当確率モデルの同時推計

前章までに得られた欠損値のないZスコア $z_{jt}$ のデータをもとに、最終的なロジットモデルであるデフォルト確率モデルと格付該当確率モデルの推計を実施した。これらのモデルをそれぞれ独立に構築するという考え方もあるが、説明変数と係数パラメータが異なることになるため、デフォルト確率の高い先の格付が必ずしも低くならないという事態が発生しうる。では、デフォルト先を格付最下位として格付該当確率モデル1本で推計すればよいというアイデアもあるが、すべての先に格付が振られているわけではなく、無格付先が必ずしも格付下位先でもないために、その取扱いに苦慮することになる。

木村(2015a)では、両モデルの説明変数と重みパラメータを共通とすることでデフォルト確率と格付の整合性を維持しつつ、両モデルの対数尤度( $l_p, l_r$ )の加重平均値( $l_T$ )を最大化する同時推計により、パラメータ( $w_j, \alpha, \beta, \alpha_R^T, \beta_R$ )を決定した。この方法であれば、無格付先はデフォルト確率モデルの対数尤度には寄与するが、格付該当確率モデルの対数尤度は一律ゼロとすることで自動的に評価対象外とすることを可能とした。さて、本稿では複数の格付へと拡張する必要がある。格付機関ごとに格付該当確率モデルを設定

<sup>6</sup> WDIにおける一人当たりGDPは、購買力平価にて定義されるためか、アルゼンチンには本データが存在せず欠損値。名目GDP、GDP成長率等のデータは存在しており、GDP関連指標がすべて欠損値というわけではない。

するのが自然であるが、説明変数と重みパラメータを共通にするか否かは意見が分かれるところである。

《方法1》

全格付機関の格付該当確率モデルおよびデフォルト確率モデルで説明変数と重みパラメータは共通。このことにより、いずれの格付機関の推計格付およびデフォルト確率とも整合性がとれる。一方で、そもそも格付は、格付機関ごとの意見表明であり、それは見るべきポイント、重視するデータがそれぞれ異なっているからこそ生まれるものだとするならば、上記の共通化は現実に反する。

《方法2》

格付機関ごとに格付該当確率モデルおよびデフォルト確率モデルを構築。格付機関ごとの両モデルでは説明変数と重みパラメータは共通だが、別の格付機関とは異なってよい。このことにより、格付機関ごとに重視しているデータが特定され、その比較は興味深い。一方で、最終的には全格付機関で平均操作を行い単一の推計格付とデフォルト確率を求めることになると思われるが、その際には整合性が崩れる可能性が高い。

いずれの方法もメリットとデメリットがあるが、本稿では《方法1》にしたがい検討をすすめることとした。《方法2》は、出力結果が大量になることが想定され、これについては別の機会に譲ることとしたい。

具体的な処理の流れは、以下のとおり。

- (0) 格付該当確率モデルの分析対象先について、木村（2015a）では、既デフォルト先を除外していたが、デフォルト後の格付は急降下し、経済データの悪化も想定されるため、下位格付データを確保するべく本稿では分析対象先に加えた。なお、デフォルト確率モデルの分析対象先については変更なし。
- (1) デフォルト確率モデルと複数の格付該当確率モデルの同時推計。ただし、格付は全6格に集約。  
1格：AAA, 2格：AA+, AA, AA-, 3格：A+, A, A-, 4格：BBB+, BBB, BBB-,  
5格：BB+, BB, BB-, 6格：B+以下。
- (2) デフォルト確率モデルについては、同時推計で得られた結果で確定。
- (3) 格付該当確率モデルについては、説明変数と重みパラメータ（すなわち中間変数）を固定し、全16格の格付にて再推計。  
1格：AAA, 2格：AA+, 3格：AA, …, 14格：B+, 15格：B, 16格：B-以下。
- (4) 出力される複数の格付該当確率については、格付機関5社で単純平均して代表格付該当確率を計算。
- (5) 各サンプルの格付該当確率を上から積算して50%点が含まれる格付を、当該サンプルの推計格付と決定<sup>7</sup>。

最終的なデフォルト確率モデルと格付該当確率モデルの構造は以下のとおり。

(中間変数)

$$\hat{z}_i = \sum_j w_j (z_{ji} - \mu) = \sum_j w_j \cdot \text{WoE}_{ji},$$

(デフォルト確率モデル)

$$z_i = \alpha + \beta \cdot \hat{z}_i, \quad p_i = 1/(1 + e^{-z_i}),$$

$$l_p = \sum_i \{\delta'_i \cdot \ln(p_i) + (1 - \delta'_i) \cdot \ln(1 - p_i)\},$$

<sup>7</sup> この方式を中央値方式と呼称。これに対し、該当確率が最大の格付を推計格付とする方法もあり、最大確率方式と呼称。後者の場合、推計格付として登場しない格付が発生し得ることに注意すること。

(格付該当確率モデル)

$$z_{i,k}^r = \alpha_{R_k}^r + \beta_{R_k} \cdot \hat{z}_i, \quad p_{i,k}^r = \begin{cases} 1/(1 + e^{-z_{i,k}^r}) & r = 1 \\ 1/(1 + e^{-z_{i,k}^r}) - 1/(1 + e^{-z_{i,k}^{r-1}}) & 2 \leq r \leq r_{\max} - 1, \quad l_{R_k} = \sum_i \ln(p_{i,k}^{r_{i,k}}), \\ 1 - 1/(1 + e^{-z_{i,k}^{r_{\max}-1}}) & r = r_{\max} \end{cases}$$

(対数尤度の加重平均)

$$l_T = a_0 \cdot l_P + a_1 \cdot l_{R_1} + a_2 \cdot l_{R_2} + a_3 \cdot l_{R_3} + a_4 \cdot l_{R_4} + a_5 \cdot l_{R_5}.$$

ここで、 $w_j : z_{ji}$ にかかる重みで、制約条件 $\sum w_j = 1, w_j \geq 0$ を満たす。 $z_i, p_i$ ：サンプル $i$ のデフォルト確率モデルのZスコアと確率、 $z_{i,k}^r, p_{i,k}^r$ ：サンプル $i$ の格付機関 $k$ 、格付 $r$ にかかわる格付該当確率モデルのZスコアと確率、 $r_{\max}$ ：全格付数（6 or 16）、 $r_{i,k}$ ：サンプル $i$ に対して格付機関 $k$ が実際に付与した格付、 $a_0 = 0.7, a_1 \sim a_5 = 0.06$ ：対数尤度の加重で、木村（2015a）よりも0.1だけ格付該当確率モデルのウエイトを高めた。

本作業にあたっては、対数尤度の加重和という特殊な評価関数の最大化が必要であるため NLMIXED プロシジャを利用した（詳細は付録を参照）。NLMIXED プロシジャによるパラメータ確定後は、中間変数である $\hat{z}_i$ を固定すれば、デフォルト確率モデル、格付該当確率モデルをそれぞれ独立に推定しても同時推計と同様の結果が得られる。よって、検算を兼ねて LOGISTIC プロシジャにかけることで、豊富で有益な出力結果（順序統計量、Hosmer-Lemeshow 検定（LACKFIT オプション）など）を利用することが可能である。

## 7. 最終結果と精度確認

図表 4：中間変数を構成する説明変数の重みパラメータと説明力

### A. 中間変数を構成する説明変数の重みパラメータと説明力

パラメータ $\alpha, \beta, \alpha_{R_k}^r, \beta_{R_k}$ については結果を省略し、 $w_j$ のみ 図表 4 に示した。表中の「説明力」とは、 $p$  値だけでは判断が難しい個々の説明変数が全体の結果に与える影響度合を見積もったものである。中間変数である $\hat{z}_i$ のサンプル全体での全分散 ( $\text{Var}(\hat{z}_i)$ ) は、

$$\text{Var}(\hat{z}_i) = \text{Cov}(\hat{z}_i, \hat{z}_i)$$

$$= \text{Cov}\left(\hat{z}_i, \sum w_j \cdot \text{WoE}_{ji}\right) = \sum w_j \cdot \text{Cov}(\hat{z}_i, \text{WoE}_{ji}),$$

であるから、その構成割合として、 $w_j \cdot \text{Cov}(\hat{z}_i, \text{WoE}_{ji})/\text{Var}(\hat{z}_i)$ を共分散による説明力と定義した<sup>8</sup>。また、個々の説明変数間の相関がゼロと仮定すると、 $w_j \cdot \text{Cov}(\hat{z}_i, \text{WoE}_{ji})/\text{Var}(\hat{z}_i)$ は $w_j^2 \cdot \text{Var}(\text{WoE}_{ji})/\text{Var}(\hat{z}_i)$ となり、これを自分散による説明力と定義した。共分散と自分散の差が、他変数からの影響と捉えることができる。さて、

説明変数名称	推計値	p 値	説明力	
			共分散	自分散
債務利息支払/歳入	0.053	0.000	2.8%	0.8%
総国民貯蓄率	0.079	0.000	6.7%	1.4%
消費者物価上昇率	0.161	0.001	4.2%	0.8%
為替レート変化率の3年標準偏差	0.144	0.002	2.1%	0.6%
1人当たり GDP/世界平均	0.179	0.000	26.2%	8.8%
GDP/世界合計	0.055	0.002	6.4%	0.9%
対外債務/GDP $\div\sqrt{\text{GDP/世界合計}}$	0.080	0.000	6.9%	2.3%
政府の有効性 (WGI)	0.099	0.000	21.0%	5.8%
政治的安定と暴力の不在 (WGI)	0.049	0.053	4.1%	0.4%
規制の質 (WGI)	0.100	0.000	19.6%	5.1%

<sup>8</sup> この考え方は、与信ポートフォリオリスク（バリュエアットリスク）計量化時における非期待損失額（UL：Unexpected Loss）を個々の構成要素へ配賦する手法として知られる「標準偏差ベースのリスク寄与度」と同一である。

結果をみると、最も説明力の高い変数は、一人当たり GDP/世界平均であり、全体の 26.2%の分散が説明された。さらに、GDP/世界合計 (6.4%) を加え、GDP 関連変数が基本的なモデルの骨格を形成した。WGI からは、政府の有効性 (21.0%)、規制の質 (19.6%)、政治的安定と暴力の不在 (4.1%) が選択され、合計で 44.7%の分散が説明された。対外債務/GDP を GDP/世界合計の平方根にて補正した指標について、木村 (2015a) ではモデルに取り込むことができたものの、説明力は 2.9%にとどまったのに対し、本稿の結果では 6.9%に増加したのは、格付該当確率モデルの分析対象先数と対数尤度の加重を増加させたことが要因と考えられる。一方で、GDP 成長率は有意とならず、格付上位国が低成長であることを反映してのことと考えられる。

## B. デフォルト確率モデルにおける各種検証 (AR 値, KS 値, ダイバージェンス, HL 検定)

図表 5 に各種検証結果を示した。各種統計量と Hosmer-Lemeshow 検定 (HL 検定) の詳細については、山下・三浦 (2011)、森内・木村 (2009) を参照のこと。なお、ダイバージェンス (Div.) は、デフォルト確率

図表 5: デフォルト確率モデルにおける各種統計量 (左) と HL 検定 (右)

	統計量	AR 値	KS 値	Div.	サン プル数	デフォルト数		
						実績	推計	
	統計量	0.673	0.559	1.912	1	251	0.09	0.12
	標準誤差	0.048	—	—	2	250	0.10	0.38
ブ ー ト ス ト ラ ッ プ に よ る 統 計 量  ( $N=1,000$ )	平均値	0.672	0.568	1.934	3	250	0.11	0.90
	標準偏差	0.047	0.052	0.318	4	250	3.12	1.67
	1.0%	0.559	0.441	1.307	5	250	3.12	2.59
	2.5%	0.573	0.461	1.363	6	250	3.12	3.64
	5.0%	0.588	0.480	1.428	7	250	2.11	4.95
	10.0%	0.613	0.505	1.554	8	250	3.12	6.78
	25.0%	0.641	0.533	1.722	9	250	14.12	10.78
	50.0%	0.672	0.568	1.895	10	244	29.11	29.33
	75.0%	0.704	0.601	2.134				
	90.0%	0.733	0.637	2.362				
	95.0%	0.747	0.651	2.489				
	97.5%	0.758	0.670	2.588				
	99.0%	0.785	0.687	2.783				
				$\chi^2$ 値	自由度	$p$ 値		
				4.761	8	0.783		

の対数オッズに基づいて計算した。ここでのブートストラップは、モデルの推計結果は変更せず、ランダムサンプリングによる統計量の変化のみを分析対象とした。

AR 値は 0.673 であり、木村 (2015a) からは 0.053 低下してしまっていたが、前述の理由同様と考えられる。HL 検定の結果は、 $p$  値が 0.783 とランク別のデフォルト数の推計と実績の差は小さいことを示した。

本作業にあたって、ブートストラップについては、SURVEYSELECT プロシジャを利用した。元のサンプル数

と同数まで抽出するには「SAMPRATE=1」、本稿では利用しなかったが各回のデフォルト数と非デフォルト数を一定に保持したければ STRATA によるデフォルトフラグ別の層化抽出を行えばデフォルト率変動による影響を除外できる。

図表 6: 推計 (行) - 実績 (列) 集約格付マトリクス

## C. 推計-実績格付マトリクスと一致率, 順位相関係数

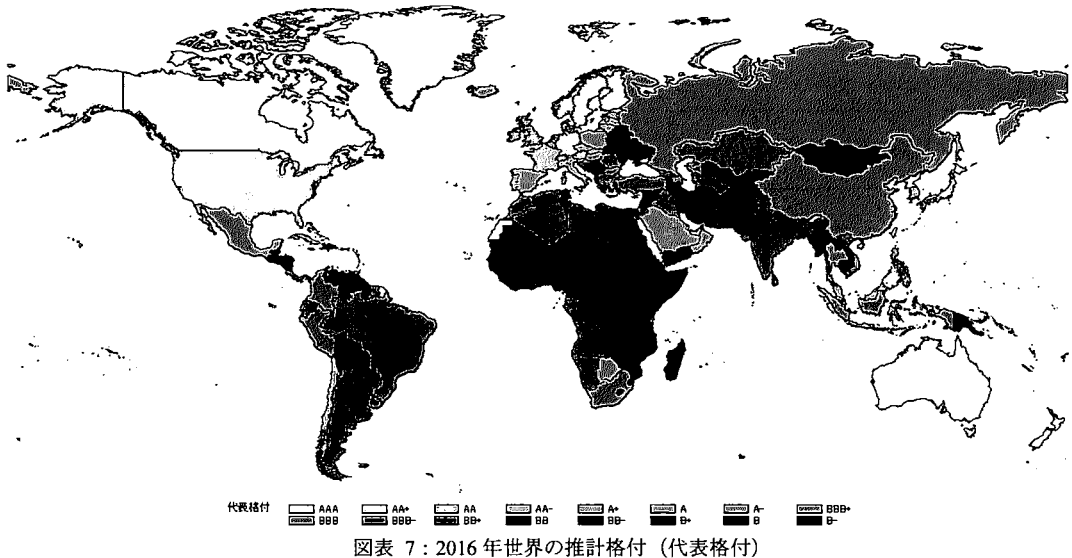
図表 6 に、全 6 格に集約した格付にて、推計と実績の比較をマトリクスにて示した。格付機関ごとに作成されたマトリクスの各セルの値を合算したものである。一致率 (対角成

	1:AAA	2:AA	3:A	4:BBB	5:BB	6:B	合計	全体
1:AAA	931	127	17	1			1,076	1,331
2:AA	113	213	148	19	3		496	706
3:A	6	202	497	122	3	2	832	1,198
4:BBB		10	187	678	167	46	1,088	2,027
5:BB			1	203	500	276	980	2,596
6:B				7	176	746	929	6,067
合計	1,050	552	850	1,030	849	1,070	5,401	13,925

分÷全体)は、 $3,565 \div 5,401 = 66.0\%$ 、順位相関係数は0.931であった。木村(2015a)と比較すると、一致率が+3.7%、順位相関係数が+0.7%、それぞれ向上したが、前述の理由同様と考えられる。なお、右端は無格付先の国も含む全体の集計結果であるが、推計格付の下位先は格付が付与されにくい(格付の需要がない)という直観と合致する結果が得られた。

#### D. 2016年推計格付(代表格付)の世界地図による確認

図表7にアウトサンプルである2016年の推計格付を世界地図に落とし込み確認した。この推計格付は、格付機関5社の格付該当確率モデルによって得られた各格付への該当確率を、確率のまま平均し、中央値によって決定した代表格付である。地図は、GMAPプロシージャにて作成した(MAPSGFK.WORLD\_ATTR)の国コードは変更がないか要確認のこと。

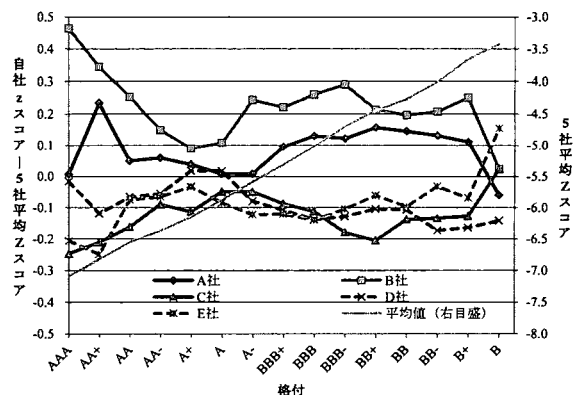


図表7: 2016年世界の推計格付(代表格付)

#### E. 格付機関間の寛厳性評価 ①格付閾値に対するZスコアの比較

本稿では、各サンプルの格付該当確率を上から積算して50%点が含まれる格付を推計格付と定めている。他方、デフォルト確率モデルと格付該当確率モデルは、共通の中間変数を用いている。この結果、各格付間の閾値に対応するデフォルト確率が求められ、この値の高低により寛厳性を評価することとした。格付 $r$ と格付 $r+1$ の間の変数の閾値は、 $z_{i,k}^r = \alpha_{R_k}^r + \beta_{R_k} \cdot \hat{z}_i = 0$ を解いて、 $\hat{z}_i = -\alpha_{R_k}^r / \beta_{R_k}$ 。これをデフォルト確率モデルの定義式に代入し、 $z_i = \alpha + \beta \cdot \hat{z}_i = \alpha - (\beta / \beta_{R_k}) \alpha_{R_k}^r$ 。右辺の各パラメータを代入すれば、左辺のデフォルト確率モデルのZスコアが計算でき、定義式からデフォルト確率に容易に変換できる。

図表8に格付閾値に対するZスコアを示した。



図表8: 格付閾値に対するZスコアの比較

ただし、単純に Z スコアを表示したのでは線が重なって比較しにくいので、閾値ごとに 5 社平均の Z スコアを計算し、それと各社との差分を表示した。高格付順であるので、閾値としての Z スコア（デフォルト確率でも増減関係は同じ）は低いほうが厳しい評価をしていると解釈される。図表によれば、A 社と B 社に比較して、明らかに C 社、D 社、E 社は厳しい評価であった。ただし、AAA 格から A 格にかけて差は縮小しつつも、A 格から B+格にかけて再度拡大するような傾向がみえた。なお、最上位格の AAA 格については、C 社と E 社が厳しく、D 社は平均に近い評価であった。ところで、5 社平均の Z スコアが直線的であるのだが、これは、非デフォルト先に対する推計用デフォルトフラグ ( $\delta_i^*$ ) が、格付別に設定された値であることを反映した人工的な結果であると考えられる。

**F. 格付機関間の寛厳性評価 ②代表格付と推計格付の比較マトリクス**

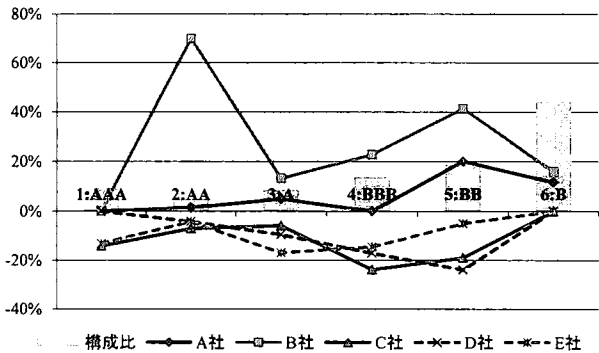
前節では、パラメータの値のみから寛厳性評価を実施したが、実際のサンプルが各格付にいかほど含まれるかも評価しておく必要がある。この場合、たとえば 2 つの格付機関ごとに比較行列を作成したのでは、5 社では 10 枚になってしまう。さらに、格付付与先のみで比較したのでは、その偏在によっては、比較が困難なケースもあろう。本稿では、格付機関 5 社の平均操作により代表格付を定義したので、これを基準格付とし、各機関の推計格付との差異にて評価した。推計格付であれば、無格付先を含み全先が評価可能である。

図表 9 に、例として A 社の代表集約格付と推計集約格付のマトリクスを示した。たとえば 2:AA 格は 2 先 (2/143=1.4%)、3:A 格は 12 先 (12/242=5.0%) しか、ずれていない。このようなずれの評価を格付機関すべてについて実施したのが図表 10 である。A 社の推計格付は 4:BBB 格以上の上位格では代表格付に近いことがわかった。B 社は前節からも予想されるように、2:AA 格以下では緩い評価であった。C 社、D 社、E 社は厳しい評価であることに変わりない。なお、低格付先のサンプル数が多いすなわち構成比が高いため、5:BB と 6:B の一致率が高い E 社が全体としても一致率が高い (94.1%、図表省略) という結果となった。

以上の比較は、推計格付、つまり推計されたモデルを前提にしたものである。図表 11 のように、実績格付とすると各格付機関の格付先が異なるため、比較が難しくなることを再度指摘しておきたい。

図表 9：代表（行）－推計（列）集約格付マトリクス（A 社）

	1:AAA	2:AA	3:A	4:BBB	5:BB	6:B	合計
1:AAA	260						260
2:AA	2	141					143
3:A		12	230				242
4:BBB				386			386
5:BB				105	418		523
6:B					143	1,088	1,231
合計	262	153	230	491	561	1,088	2,785



図表 10：代表格付に対する推計集約格付の不一致率とその方向

図表 11：代表（行）－実績（列）集約格付マトリクス（A 社）

	1:AAA	2:AA	3:A	4:BBB	5:BB	6:B	合計
1:AAA	177	32					203
2:AA	29	21	23	5			78
3:A	1	36	52	17			106
4:BBB			30	87	25	5	147
5:BB			2	40	22	10	74
6:B					8	16	24
合計	207	89	107	149	55	31	638



## 8. おわりに

本稿では、ソブリンを題材に、デフォルトと単一格付の同時推計によるロジットモデルの構築方法を提案した木村 (2015a) をベースとして、デフォルトと複数格付の同時推計へと自然に拡張がなされた。推計されたパラメータにより、格付機関間の寛厳性評価の一助となることが示された。また、木村 (2015a) では SAS/OR に含まれる NLP プロシジャを利用していたが、本稿では SAS/STAT に含まれる NLMIXED プロシジャへと移行させた。SAS/STAT は SAS ユーザーの大部分が契約するプロダクトである。よって、ユーザー自らの課題に対し、本稿の内容を参考にしながら取組み可能な態勢が整えられたと言えよう。今回は取り組めなかったが、将来課題として、格付機関ごとに説明変数が異なるモデルを構築することで、各機関の格付相違の要因が掴めれば興味深いと思われる。本稿が、ロジットモデル構築における実務上の要請を効率的に達成するための指針となり、また企業内で自らロジットモデルを構築したいと願う実務者の方々の参考になれば幸いである。

## 参考文献

- [1] 木村和央 (2015a). ロジットモデル構築における Weight of Evidence を用いた変数変換と欠損値処理方法の提案. 2015SAS ユーザー総会論文集, 211-220.
- [2] 木村和央 (2015b). ロジットモデル構築における Weight of Evidence を用いた変数変換と欠損値処理方法の提案. 第4回金融シンポジウム講演資料, 統計数理研究所リスク解析研究センター.
- [3] 細野健二, 大重斉 (2003). ソブリンの特性と評価: 審査の実例. 第2回金融シンポジウム講演資料, 統計数理研究所リスク解析研究センター.
- [4] 森内一朗, 木村和央 (2009). 格付モデルの構築と検証. 「金融工学とリスクマネジメント高度化」研究会公開資料, <http://www.ffr-plus.jp/material/pdf/100913/kinkoken.pdf> (2016年6月17日).
- [5] 山下智志 (2005). デフォルト確率推計モデルの相互比較と寛厳性の評価. FSAサーチ・レビュー 2, 59-71.
- [6] 山下智志, 三浦翔 (2011). 信用リスクモデルの予測精度—AR 値と評価指標—. 朝倉書店.
- [7] Cruces, Juan J. and Trebesch, Christoph (2013). Sovereign Defaults: The Price of Haircuts. *American Economic Journal: Macroeconomics*, 5(3), 85-117.
- [8] Yeo, In-Kwon and Johnson, Richard (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87, 954-959.

## 付録 A. NLMIXED プロシジャによる最終モデルの同時推計

2015 年の SAS ユーザー会にて、SAS/OR に含まれる NLP プロシジャの代わりに、SAS/STAT に含まれる NLMIXED プロシジャに移行することで、より多くのユーザーが利用可能なプログラムになるので好ましいという指摘をいただいた。検討を重ねた結果、NLMIXED プロシジャへの移行は可能であることがわかった。検討過程で課題となったポイントは、NLP プロシジャにて利用している線形および非線形制約条件すなわち LINCON および NLINCON ステートメントについて、当該ステートメントが存在しない NLMIXED プロシジャにてどのように回避しつつ同様の効果を得るかということであった。結論から言えば、これら制約条件を回避するべく上下限值制約条件すなわち BOUNDS ステートメントのみで表現可能となるよう推計すべきパラメータを変更し、プログラミングステートメント先頭にて本来推計したいパラメータに変換し、追加的にパラメータの検定統計量を出力する ESTIMATE ステートメントにて前述のパラメータを指定すればよい。

実際のプログラムは、説明変数の追加・変更など、試行錯誤に対応するべくマクロコーディングが多用されており説明には適さないため、ここでは、説明変数 (WoE) の数 3、格付機関数 2、格付ランク数 4 を前提とした簡易なデフォルト確率モデルと格付該当確率モデルの同時推計プログラム例を示した。NLMIXED プロシジャでは、説明変数  $i$  のウエイト ( $W_i$ ) に関する線形制約を利用できないため、 $BW_i = B \times W_i$  をパラメータとして推計するものとし、プログラミングステートメント先頭にて、 $B = \sum BW_i$ ,  $W_i = BW_i / B$  と本来推計したいパラメータを定義した。また、格付機関  $j$  の格付該当確率モデルの格付  $k$  と  $k-1$  の切片の差 ( $D_{jk}$ ) に関する非線形制約を利用できないため、 $k \geq 2$  については  $D_{jk}$  そのものをパラメータとして推計するものとし、プログラミングステートメント先頭にて、 $A_{jk} = A_j(k-1) + D_{jk}$  と本来推計したいパラメータを定義した。通常の実出力では、PARMS ステートメントに記載したパラメータ以外の検定統計量は出力されないが、後段の ESTIMATE ステートメントにより表示可能となるので、それを ODS 出力機能 (AdditionalEstimates) にてファイルへと格納した。なお、アウトサンプルデータへモデルを適用することも、しばしば実行される。この場合のパラメータ渡しは、NLP プロシジャは先頭ステートメントの INEST オプション、NLMIXED プロシジャは PARMS ステートメントにおける DATA オプションとなることに留意されたい。

### 《入力データ定義》

DF デフォルトフラグ (デフォルト先=1, 非デフォルト先=0, 既デフォルト等の分析対象外先=)  
 RK1 格付機関 1 の格付実績 (格付保有先=1~4 のいずれか, 非格付先および分析対象外先=)  
 RK2 格付機関 2 の格付実績 (格付保有先=1~4 のいずれか, 非格付先および分析対象外先=)  
 Z1 説明変数 1 の Weight of Evidence  
 Z2 説明変数 2 の Weight of Evidence  
 Z3 説明変数 3 の Weight of Evidence  
 U 分析対象フラグ

### 《出力データ定義》

A デフォルト確率モデルの切片  
 B デフォルト確率モデルの傾き  
 $W_i$  説明変数  $i$  のウエイト  
 $BW_i$  デフォルト確率モデルの傾き  $\times$  説明変数  $i$  のウエイト ( $=B \times W_i$ )  
 $A_{jk}$  格付機関  $j$  の格付該当確率モデルの格付  $k$  の切片  
 $B_j$  格付機関  $j$  の格付該当確率モデルの傾き  
 $D_{jk}$  格付機関  $j$  の格付該当確率モデルの格付  $k$  と  $k-1$  の切片の差 ( $=A_{jk} - A_j(k-1)$ )  
 L0 デフォルト確率モデルにおける対数尤度  
 L1 格付機関 1 の格付該当確率モデルにおける対数尤度  
 L2 格付機関 2 の格付該当確率モデルにおける対数尤度  
 LL 対数尤度の荷重和 ( $=0.70 \times L0 + 0.15 \times L1 + 0.15 \times L2$  を採用)  
 Z デフォルト確率モデルの Z スコア  
 P デフォルト確率  
 PKjk 格付機関  $j$  における格付  $k$  の格付該当確率

《NLP プロシジャを利用した最終モデルの同時推計》

```
/** ODS ファイル出力設定 **/  
ODS OUTPUT ParameterEstimates = ODSEST1;  
  
/** NLP プロシジャによる推計 **/  
PROC NLP DATA = DATA_IN1  
OUT=DATA_OUT1 OUTEST=DATA_EST1;  
  
/* パラメータ定義 */  
PARMS  
A=0, B=1, W1-W3=0.3333, /* デフォルト推計用 */  
A11=0, A12=0.5, A13=1, B1=-1, /* 格付機関 1 用 */  
A21=0, A22=0.5, A23=1, B2=-1; /* 格付機関 2 用 */  
  
/* 上下限値制約 */  
BOUNDS  
B >= 0, W1-W3 >= 0;  
  
/* 線形制約: 変数ウェイト合計=1 */  
LINCON W1+W2+W3=1;  
  
/* 非線形制約: 格付モデル切片は単調増加 */  
D12=A12-A11; D13=A13-A12;  
D22=A22-A21; D23=A23-A22;  
NLINCON D12-D13 >= 0.001, D22-D23 >= 0.001;  
  
/*****  
/**** 次頁の《モデル・対数尤度定義プログラム》を挿入 ****/  
/*****  
  
/* 対数尤度最大化 (LL は対数尤度荷重和) */  
MAX LL;  
  
RUN;  
  
/** アウトサンプルデータへモデルを適用 **/  
PROC NLP DATA = DATA_IN2  
TECHNIQUE = NONE MAXITER = 0 MAXFUNC = 0  
INEST = DATA_EST1  
OUT=DATA_OUT2;  
  
/* パラメータ定義 (初期値は INEST オプションで指定) */  
PARMS A, B, W1-W3, A11, A12, A13, B1, A21, A22, A23, B2;  
  
... (略) ...  
  
RUN;
```

《NLMIXED プロシジャを利用した最終モデルの同時推計》

```
/** ODS ファイル出力設定 **/  
ODS OUTPUT ParameterEstimates = ODSEST1;  
ODS OUTPUT AdditionalEstimates = ODSADD1;  
  
/** NLMIXED プロシジャによる推計 **/  
PROC NLMIXED DATA = DATA_IN1;  
  
/* パラメータ定義 */  
PARMS  
A=0, BW1-BW3=0.3333, /* デフォルト推計用 */  
A11=0, D12-D13=0.5, B1=-1, /* 格付機関 1 用 */  
A21=0, D22-D23=0.5, B2=-1; /* 格付機関 2 用 */  
  
/* 上下限値制約 */  
BOUNDS  
BW1-BW3 >= 0, D12-D13 >= 0.001, D22-D23 >= 0.001;  
  
/* 本来推計したいパラメータを定義 */  
B=BW1+BW2+BW3;  
W1=BW1/B; W2=BW2/B; W3=BW3/B;  
  
A12=A11+D12; A13=A12+D13;  
A22=A21+D22; A23=A22+D23;  
  
/*****  
/**** 次頁の《モデル・対数尤度定義プログラム》を挿入 ****/  
/*****  
  
/* 対数尤度最大化 (LL は対数尤度荷重和) */  
MODEL U=GENERAL(LL); /* U: 分析対象=1, 対象外=. */  
  
/* データ出力関係 */  
ID_ALL_; /* 下記の DATA_OUT1 に中間変数等も出力 */  
PREDICT P OUT = DATA_OUT1;  
  
/* 追加出力させたいパラメータ (ODS_ADD1 に出力) */  
ESTIMATE 'A' A; ESTIMATE 'B' B;  
ESTIMATE 'W1' W1; ESTIMATE 'W2' W2; ESTIMATE 'W3' W3;  
ESTIMATE 'A11' A11; ESTIMATE 'A12' A12;  
ESTIMATE 'A13' A13;  
ESTIMATE 'B1' B1;  
ESTIMATE 'A21' A21; ESTIMATE 'A22' A22;  
ESTIMATE 'A23' A23;  
ESTIMATE 'B2' B2;  
  
RUN;  
  
/** アウトサンプルデータへモデルを適用 **/  
PROC NLMIXED DATA = DATA_IN2  
TECHNIQUE = NONE MAXITER = 0 MAXFUNC = 0;  
  
/* パラメータ定義 (初期値は DATA オプションで指定) */  
PARMS / DATA = ODSEST1;  
  
... (略) ...  
  
PREDICT P OUT = DATA_OUT2;  
  
RUN;
```

《モデル・対数尤度定義プログラム》

```

/* 対数尤度ゼロクリア */
L0 = 0; L1 = 0; L2 = 0; LL = 0;

/* 中間変数 (Zi : 変数 i の WoE) */
WZ = W1*Z1+W2*Z2+W3*Z3;

/* デフォルト確率モデル (DF : デフォルトフラグ) */
Z = A+B*WZ; P = 1/(1+EXP(-Z));

/* 対数尤度計算 */
IF DF ^= . THEN L0 = DF*LOG(P) + (1-DF)*LOG(1-P);

/* 格付該当確率モデル (複数格付を配列にて対応) */
ARRAY AR_A[2,3] A11 A12 A13 A21 A22 A23;
ARRAY AR_B[2] B1 B2;
ARRAY AR_PK[2,4] PK11 PK12 PK13 PK14
                PK21 PK22 PK23 PK24;
ARRAY AR_L[2] L1 L2;
ARRAY AR_RK[2] RK1 RK2; /* 格付実績 */

/* 格付該当確率ゼロクリア (NLP での出力目的) */
PK11=0; PK12=0; PK13=0; PK14=0;
PK21=0; PK22=0; PK23=0; PK24=0;

/* 格付該当確率計算 */
DO J = 1 TO 2; /* 格付機関ループ開始 */
DO K = 1 TO 4; /* 格付ランクループ開始 */
/* 各格付の該当確率を計算 */
/* この計算は非常に重いので、格付ランクでループせず */
/* 格付実績のときのみ計算しておいて (K = AR_RK[J]), */
/* パラメータ確定後に再計算にて求める方法もある。 */
IF K=1 THEN DO;
  AR_PK[J,K] = 1/(1+EXP(-(AR_A[J,K]+AR_B[J]*WZ)));
END;
ELSE IF K < 4 THEN DO;
  AR_PK[J,K] = 1/(1+EXP(-(AR_A[J,K]+AR_B[J]*WZ)))
    -1/(1+EXP(-(AR_A[J,K-1]+AR_B[J]*WZ)));
END;
ELSE IF K = 4 THEN DO;
  AR_PK[J,K] = 1-1/(1+EXP(-(AR_A[J,K-1]+AR_B[J]*WZ)));
END;
END; /* 格付ランクループ終了 */

/* 対数尤度計算 */
IF AR_RK[J] ^= . THEN AR_L[J] = LOG(AR_PK[J,AR_RK[J]]);

END; /* 格付機関ループ終了 */

/* 対数尤度の加重和 (デフォルト 0.7, 格付 0.15 づつ) */
LL = 0.70*L0 + 0.15*L1 + 0.15*L2;

```

## 付録 B. FMM プロシジャによる欠損値補完処理

変数  $j$  の  $Z$  スコアを被説明変数, 変数  $j$  以外の  $Z$  スコアおよび変数  $j$  の 1 年前/後の  $Z$  スコアを説明変数とした重回帰モデルを推計し, 被説明変数が欠損値であるサンプルについて穴埋めを収束するまで繰り返した。このとき, 説明変数の  $Z$  スコアの係数を正值に制約しておくことで, 多重共線性に陥る危険性を低下させることが可能で, そのためには FMM プロシジャにおける RESTRICT ステートメントを利用した。

《FMM プロシジャによる制約条件付き重回帰モデル構築 (主要部分)》

```

%macro fmm;
%do i = 1 %to 3; /* 変数ループ開始 */
%do ptn = 0 %to 3; /* 推計パターンループ開始 */
/** ODS ファイル出力設定 **/
ODS OUTPUT ParameterEstimates = ODSEST_&ptn._Z&i;
/** FMM プロシジャによる推計 **/
PROC FMM DATA = DATA_IN3;
/* モデル定義 */
MODEL Z&i =
/* 1 期前データの利用 */
%if &ptn = 1 %then %do; Z&i._BEFORE %end;
/* 1 期後データの利用 */
%if &ptn = 2 %then %do; Z&i._AFTER %end;
/* 1 期前+1 期後データの利用 */
%if &ptn = 3 %then %do; Z&i._BEFORE Z&i._AFTER %end;
%do j = 1 %to %eval(&i - 1); Z&j %end; /* i 以前 */
%do j = %eval(&i + 1) %to 3; Z&j %end; /* i 以降 */
;
/* 対象選定① Z&i の非欠損値フラグ */
FREQ USE Z&i;
/* 対象選定② 前回の推計パターン &ptn 実行時に */
/* Z スコアが上下限値にヒットしていないフラグ */
WEIGHT WGT_&ptn._Z&i;
/* 出力 */
OUTPUT OUT = DATA_OUT3 PRED = PRED_Z&i;

/* 制約条件 (RESTRICT A B ? C → A の係数 × B ? C) */
/* 1 期前データの利用 */
%if &ptn = 1 %then %do;
  RESTRICT Z&i._BEFORE 1 >= 0.000000001;
%end;
/* 1 期後データの利用 */
%if &ptn = 2 %then %do;
  RESTRICT Z&i._AFTER 1 >= 0.000000001;
%end;
/* 1 期前+1 期後データの利用 */
%if &ptn = 3 %then %do;
  RESTRICT Z&i._BEFORE 1 >= 0.000000001;
  RESTRICT Z&i._AFTER 1 >= 0.000000001;
%end;
%do j = 1 %to %eval(&i - 1); /* i 以前 */
  RESTRICT Z&j 1 >= 0.000000001;
%end;
%do j = %eval(&i + 1) %to 3; /* i 以降 */
  RESTRICT Z&j 1 >= 0.000000001;
%end;
RUN;
%end; /* 推計パターンループ終了 */
%end; /* 変数ループ終了 */
%mend;
%fmf;

```

# ガンマ過程を用いた比例ハザードモデルによる解析

○矢田 真城<sup>1</sup>, 浜田 知久馬<sup>2</sup>

(<sup>1</sup>エイツーヘルスケア株式会社, <sup>2</sup>東京理科大学)

Proportional hazard model using gamma process

Shinjo Yada<sup>1</sup> and Chikuma Hamada<sup>2</sup>

<sup>1</sup>A2 Healthcare Corporation, <sup>2</sup>Tokyo University of Science

## 要旨

比例ハザードモデルは、生存時間解析において広く用いられている統計モデルである。生存時間分布（基準ハザード関数）に特定の分布を仮定することなく、回帰パラメータの推論を行うことが最大の利点である。比例ハザードモデルに対するベイズ流解析としては、大きく2つのアプローチが存在する。ひとつは、基準ハザード関数に確率モデルを仮定せず、回帰パラメータにのみ確率モデルをあてはめ、部分尤度に基づきパラメータの推論を行うアプローチである。もうひとつは、回帰パラメータのみならず基準ハザード関数にも確率モデルをあてはめ、パラメータの推論を行うアプローチである。前者のアプローチに関しては、PHREGプロシジャにおいてBAYESステートメントを指定することにより実行可能となった。そこで本稿では、後者のアプローチの1つである、累積基準ハザード関数にガンマ過程を想定したモデルをとりあげ、解析方法についてまとめ、SASによる使用例を紹介する。

キーワード：比例ハザードモデル，ガンマ過程，基準ハザード

## 1. はじめに

ある基準の時刻からある目的の反応（観測対象とする個体に対し一度だけ非再起的に起きる事象であるとする）が起きるまでの時間を解析対象とする解析手法の総称を、生存時間解析とよぶ（大橋・浜田,1995）。

生存時間解析において用いられる統計モデルのひとつに、Cox(1972)の比例ハザードモデルがあげられる。比例ハザードモデルには、1)基準ハザード関数が未知のままでも共変量の影響を推測することができる、2)生存時間の順位情報のみに基づいて推論を行うことができるため外れ値に対して頑健である、3)モデルが単純な場合にはログランク検定に一致する（大橋・浜田,1995）、といったいくつかの利点がある。比例ハザードモデルに対しては、基準ハザード関数に特定の分布を仮定することなく、部分尤度に基づき共変量の回帰パラメータを推定する、セミパラメトリックな方法が広く用いられている。その際、基準ハザード関数は局外パラメータとして扱われており、必要に応じて部分尤度法により求められた回帰パラメータの推定量を用

いて基準ハザード関数を推定することができる（中村,2001；大橋・浜田,1995）。

一方、比例ハザードモデルに対するベイズ流アプローチとしては、回帰パラメータのみに確率モデルを考えるアプローチと、回帰パラメータ及び基準ハザード関数に確率モデルを考えるアプローチとがある（Dey and Rao,2005）。前者では、基準ハザード関数に特定の事前分布をおくことなく、回帰パラメータのみに事前分布をおき、部分尤度を用いることで回帰パラメータの事後分布を導出する。SAS では、Ver.9.2 より PHREG プロシジャにおいて BAYES ステートメントを記述することにより、部分尤度を用いた回帰パラメータの事後分布の算出が可能となった（SAS Institute Inc., 2008；黒田,2010）。本稿では、もうひとつのアプローチである、回帰パラメータ及び基準ハザード関数に確率モデルをおき、モデルパラメータの推論を行う方法について説明する。基準ハザード関数に想定する確率過程に関して多数の方法が提案されているが、ここではそれらの先駆けとなった、累積基準ハザード関数にガンマ過程をあてはめるモデルをとりあげる。

## 2. 比例ハザードモデルに対するベイズ流解析

### 2.1. Intensity function による解析

以下、解析対象となる時間を生存時間とよび、対象とする事象をイベントとよぶことにする。

いま  $i$  番目の被験者に対して、イベントが観察されるまでの時間  $t_i$ 、観察打ち切りまでの時間  $c_i$  に対し、 $u_i = \min(t_i, c_i)$ 、 $\omega_i = I(t_i \leq c_i)$  として、 $N_i(t) = I(t_i \leq t, \omega_i = 1)$ 、 $Y_i(t) = I(t_i \geq t)$  とおく ( $i=1,2,\dots,n$ )。ここに  $I(\cdot)$  はカッコの中が成り立てば 1 を、そうでなければ 0 をとる指示関数である。 $N_i(t)$  は、 $i$  番目の被験者について時間  $t$  直前までイベントが生起しなければ 0、イベントが生起しているなら 1 をとる階段関数である。一方、 $Y_i(t)$  は  $i$  番目の被験者について時間  $t$  直前までイベントが生起していない、つまりイベントが生起するリスクにさらされていたのであれば 1 を、そうでなければ 0 をとる関数である。微小時間  $dt$  に対する  $N_i$  の増分

$$dN_i(t) = N_i(t+dt) - N_i(t)$$

は、 $i$  番目の被験者が微小時間  $dt$  の間にイベントが生起すれば 1、そうでなければ 0 の 2 つの値しかとらない。 $i$  番目の被験者における  $p$  個の共変量を  $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{ip})^T$  と表し、 $\mathbf{z}_i$  所与のもとでのハザード関数を  $\lambda_i(t | \mathbf{z}_i)$  とおくと、 $i$  番目の被験者に対し時間  $t$  直前までイベントが生起しておらず ( $N_i(t)=0$ ) リスク下にある状態 ( $Y_i(t)=1$ ) で、微小時間  $dt$  の間にイベントが生起する ( $dN_i(t)=1$ ) 確率は

$$\Pr(dN_i(t)=1 | Y_i(t)=1, N_i(t)=0, \mathbf{z}_i) = \lambda_i(t | \mathbf{z}_i) dt \quad (2.1)$$

となる。同様に、 $i$  番目の被験者が打ち切り症例であれば

$$\Pr(dN_i(t)=1 | Y_i(t)=0, N_i(t)=0, \mathbf{z}_i) = 0 \quad (2.2)$$

であり、時間  $t$  直前にイベントが生起していれば

$$\Pr(dN_i(t)=1 | Y_i(t)=0, N_i(t)=1, \mathbf{z}_i) = 0 \quad (2.3)$$

である。 $N_i(t)$ 、 $Y_i(t)$  はともに 0 か 1 の値をとるため、組み合わせとしては  $\Pr(dN_i(t)=1 | Y_i(t)=1, N_i(t)=1, \mathbf{z}_i)$  も考えられるが、当該症例において時間  $t$  直前までにイベントが生起しており、かつリスク下にある状態はありえない。時間  $t$  までに観測された利用可能なデータ  $N_i(t)$ 、 $Y_i(t)$ 、 $\mathbf{z}_i$  をまとめて  $F_{i-}$  と表記するとき、(2.1)から(2.3)をまとめて

$$\Pr(dN_i(t)=1 | F_{i-}) = Y_i(t) \lambda_i(t | \mathbf{z}_i) dt$$

と表すことができ、

$$E(dN_i(t)=1 | F_{i-}) = Y_i(t) \lambda_i(t | \mathbf{z}_i) dt \quad (2.4)$$

となる（中村,2001）。 $i$  番目の被験者に対する時間  $t$  での強度関数（intensity function）を

$$I_i(t)dt = E(dN_i(t) = 1 | F_{t-})$$

と定義するとき、(2.4)より

$$I_i(t)dt = Y_i(t)\lambda_i(t | \mathbf{z}_i)dt$$

よって

$$I_i(t) = Y_i(t)\lambda_i(t | \mathbf{z}_i)$$

となり、ハザード関数  $\lambda_i(t | \mathbf{z}_i)$  を強度関数  $I_i(t)$  に取り替えることができる。

Cox (1972) の比例ハザードモデルでは、共変量  $\mathbf{z}_i$  をもつ被験者のハザード関数  $\lambda_i(t | \mathbf{z}_i)$  に対し

$$\lambda_i(t | \mathbf{z}_i) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}_i) \quad (2.5)$$

を想定する。ここに  $\lambda_0(t)$  は基準ハザード関数であり、 $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  は共変量ベクトル  $\mathbf{z}_i$  に対する回帰パラメータベクトルである。したがって比例ハザードモデルでは強度関数  $I_i(t)$  は

$$I_i(t)dt = Y_i(t)\lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}_i)dt \quad (2.6)$$

よって

$$I_i(t) = Y_i(t)\lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}_i)$$

となる。また、共変量ベクトル  $\mathbf{z}_i$  をもつ被験者における生存関数  $S(t | \mathbf{z}_i)$  は、生存関数、ハザード関数、累積ハザード関数の関係から

$$\begin{aligned} S(t | \mathbf{z}_i) &= \exp(-\Lambda(t | \mathbf{z}_i)) = \exp\left(-\int_0^t \lambda(t | \mathbf{z}_i)dt\right) = \exp\left(-\int_0^t \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}_i)dt\right) \\ &= \exp\left(-\exp(\boldsymbol{\beta}^T \mathbf{z}_i) \int_0^t \lambda_0(t)dt\right) = \left[\exp\left(-\int_0^t \lambda_0(t)dt\right)\right]^{\exp(\boldsymbol{\beta}^T \mathbf{z}_i)} \end{aligned} \quad (2.7)$$

となる。

## 2.2. ガンマ過程を用いたモデル

$i$  番目の被験者に対する観測値  $(u_i, \omega_i)$  と  $(N_i(t), Y_i(t))$  は 1 対 1 に対応するため、 $(N_i(t), Y_i(t), \mathbf{z}_i)$  を  $i$  番目の被験者に対するデータとみなすことができる (服部, 2009)。  $n$  人のデータを  $D = \{N_i(t), Y_i(t), \mathbf{z}_i\}_{i=1}^n$  と表すとき、 $\boldsymbol{\beta}$  と  $\Lambda_0(t)$  の同時事後分布  $p(\boldsymbol{\beta}, \Lambda_0(t) | D)$  は

$$p(\boldsymbol{\beta}, \Lambda_0(t) | D) \propto p(D | \boldsymbol{\beta}, \Lambda_0(t)) p(\boldsymbol{\beta}) p(\Lambda_0(t)) \quad (2.8)$$

で与えられる。ここに  $p(\boldsymbol{\beta})$  は  $\boldsymbol{\beta}$  の事前分布、 $p(\Lambda_0(t))$  は  $\Lambda_0(t)$  に対する確率過程事前分布、 $p(D | \boldsymbol{\beta}, \Lambda_0(t))$  はデータ  $D$  が与えられたもとの尤度である。打ち切りが non-informative であるとき、被験者の生存時間が互いに独立として尤度は

$$p(D | \boldsymbol{\beta}, \Lambda_0(t)) = \prod_{i=1}^n \left[ \prod_{t>0} I_i(t)^{dN_i(t)} \right] e^{-I_i(t)dt} \quad (2.9)$$

となる。(2.9)より、尤度は平均  $I_i(t)dt$  をもつポアソン分布の核  $(I_i(t)dt)^{N_i(t)} \exp(-I_i(t)dt)$  に比例していることがわかる。本稿では、Kalbfleisch (1978) に従い、累積基準ハザード関数  $\Lambda_0(t)$  に対する事前分布としてガンマ過程 (gamma process) をとりあげる。ガンマ過程はガンマ分布に従って変化する確率過程であり、ポアソン分布の共役事前分布がガンマ分布であることを考えると、(2.9)において  $\Lambda_0(t)$  にガンマ過程をおくことは自然な発想といえる。

ガンマ過程は以下のように記述される (Ibrahim et al., 2001)。

(i)  $Z(0) = 0$

(ii)  $Z(t)$  は排反な区間において独立増分をもつ。即ち、任意の  $t_1 < t_2 < \dots < t_n$  に対し、 $Z(t_2) - Z(t_1), Z(t_3) - Z(t_2), \dots$

$Z(t_n) - Z(t_{n-1})$ は互いに独立である.

(iii)  $\alpha(0) = 0$  であり  $[0, \infty]$  で左連続な増大関数  $\alpha(t)$  と  $c > 0$  が存在するとき, 任意の  $t > s$  に対して

$$Z(t) - Z(s) \sim Ga(\alpha(t) - \alpha(s), c)$$

である. ここに  $Ga(\alpha, \beta)$  は平均  $\alpha/\beta$ , 分散  $\alpha/\beta^2$  となるガンマ分布を表す.

上記 3 つの条件を満たす確率過程  $\{Z(t) | t > 0\}$  をガンマ過程とよび, 以降, 確率過程  $\{Z(t) | t > 0\}$  がガンマ過程に従うことを,  $Z(t) \sim GP(\alpha(t), c)$  と表記する. いま,  $\Lambda_0^*(0) = 0$  であり  $[0, \infty]$  で左連続な増大関数  $\Lambda_0^*(t)$  に対し,  $c > 0$  として  $\Lambda_0(t) \sim GP(c\Lambda_0^*(t), c)$  であるとすれば, 微小時間  $dt$  での累積基準ハザード関数の増分

$$d\Lambda_0(t) = \Lambda_0(t + dt) - \Lambda_0(t)$$

については,  $d\Lambda_0^*(t) = \Lambda_0^*(t + dt) - \Lambda_0^*(t)$  として

$$d\Lambda_0(t) \sim Ga(c d\Lambda_0^*(t), c) \tag{2.10}$$

となる. (2.10) のとき  $d\Lambda_0(t)$  の平均は  $d\Lambda_0^*(t)$ , 分散は  $d\Lambda_0^*(t)/c$  となるため,  $c$  は  $d\Lambda_0^*(t)$  への確信度を表す指標であることがわかる.  $d\Lambda_0(t)$  を用いれば, 強度関数の増分  $I_i(t)dt$  は(2.6)より

$$I_i(t)dt = Y_i(t) \exp(\beta^T z_i) d\Lambda_0(t) \tag{2.11}$$

となり, 尤度(2.8)は  $dN_i(t), Y_i(t), d\Lambda_0(t)$  により表すことができる.

### 3. 文献での適用例

適用例として, Gehan (1965) の急性白血病データ (以下 Gehan のデータとよぶ) をとりあげる. 急性白血病的覚解維持のために 6-mercaptopurine (6-MP) とプラセボ群とを比較するために行われたランダム化比較試験であり, 6-MP 群 21 例, プラセボ群 21 例に対するデータが表 3.1 である. 白血病的再発をイベントとしており, 覚解状態から再発または打ち切りまでの時間が週を単位として測定されている.

表 3.1 : 急性白血病患者の再発までの時間 (週)

プラセボ群	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23
6-MP 群	6, 6, 6†, 6, 7, 9†, 10, 10†, 11†, 13, 16, 17†, 19†, 20†, 22, 23, 25†, 32†, 32, 34†, 35†

†: 打ち切り

この Gehan のデータに対し, 2.1. 及び 2.2. で示した比例ハザードモデルをあてはめ, ベイズ流アプローチによる推論を行うための SAS プログラム例を以下に示す. このプログラム例は, DATA ステップによるデータハンドリングと MCMC プロシジャによるパラメータ推定の 2 つのパートで構成されている.

プログラム 3.1 は, DATA ステップによるデータハンドリングの一例である. 大橋・浜田 (1995) に倣い, 再発または打ち切りを表す変数 WEEK, 群を表す変数 GROUP (プラセボ群であれば 0, 6-MP 群であれば 1), 打ち切りを表す変数 REMISS (打ち切りであれば 0, イベント発生であれば 1) をもつデータセット GEHAN を作成した. (2.9) に従い, MCMC プロシジャにて尤度を記載する必要があるため, SAS/STAT 9.3 (R) User's Guide, The MCMC Procedure の Example 54.14 を参考に, MCMC プロシジャで使用するためのデータセット A\_GEHAN を用意した. Gehan のデータでは, イベント発現時間は重複を除くと 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17, 22, 23, 32 の 18 時点ある. そこでこれらイベント発現時間をもたせたデータセット PARTITION を用意し, 2 つのデータセット GEHAN と PARTITION から, 42 名の被験者に対し, 各イベント発現時点でのリスクセットを変数 m\_Y に, 計数過程を変数 m\_dN に, それぞれ入力したデータセット A\_GEHAN を用



意した.

プログラム 3.1 : Gehan のデータハンドリング

```
data GEHAN ;
do drug = 0,1 ;
  do id = 1 to 21 ; input week remiss @@ ; output ; end ; end ;
cards ;
  1 1 1 1 2 1 2 1 3 1 4 1 4 1 5 1 5 1 8 1 8 1
  8 1 8 1 11 1 11 1 12 1 12 1 15 1 17 1 22 1 23 1
  6 1 6 1 6 1 6 0 7 1 9 0 10 0 10 1 11 0 13 1 16 1
  17 0 19 0 20 0 22 1 23 1 25 0 32 0 32 0 34 0 35 0
;
run ;
data PARTTITION ;
input int_1-int_18 ;
cards ;
1 2 3 4 5 6 7 8 10 11 12 13 15 16 17 22 23 35
;
run ;
data A_GEHAN ; set GEHAN ;
if _n_ eq 1 then set PARTTITION ;
array int[18] int_1-int_18 ;
array Y[18]; array dN[18];
do k = 1 to (dim(int)-1) ;
  Y[k] = (week - int[k] + 0.001 >= 0) ;
  dN[k] = Y[k] * ( int[k+1] - week - 0.001 >= 0 ) * remiss ;
end ;
output ;
drop k ;
run ;
data A_GEHAN ; set A_GEHAN ;
array Y[*] Y ;
array dN[*] dN ;
do i = 1 to (dim(y)) ;
  y_val = y[i] ;
  dn_val = dn[i] ;
  int_index = i ;
output ;
```

```

end ;
keep Y_ : dN_ : int_ : drug int_index id ;
run ;
data A_GEHAN ; set A_GEHAN ; rename y_val = m_Y dn_val = m_dN ; run ;

```

プログラム 3.2 は、データセット A\_GEHAN を用いて、MCMC プロシジャにより事後分布のモンテカルロ標本を生成するための SAS プログラム例である。WinBUGS 1.4 Examples Volume I の”Leuk: survival analysis Using Cox regression”を参考にした。Burn in を 1000、Burn in で除外された後のモンテカルロ標本におけるサンプリング間隔を 1、最終的にサンプリングされるモンテカルロ標本のサンプル数を 10000 個とした。回帰パラメータの事前分布には平均 0、分散  $10^6$  の正規分布を指定した。累積基準ハザード関数の増分  $d\Lambda_0(t)$  の事前分布にはガンマ分布(2.10)を指定した。Hobbs and Carlin (2008) に従い、 $r$  を単位時間あたりのハザード、 $dt$  を次イベント発現までの間隔として  $d\Lambda_0^*(t) = r dt$  とおき、 $r = 0.1, c = 0.001$  と設定した。(2.9)のとおり、尤度がポアソン分布の核に比例することから、MODEL ステートメントにて平均  $I_i(t)dt$  をもつポアソン分布を指定した。煩雑さを避けるため、プログラム 3.2 では  $d\Lambda_0(t)$  に対する初期パラメータの設定と事前分布の指定を SAS マクロ %dLprior により記述している。

プログラム 3.2 : 比例ハザードモデルを用いたベイズ流解析

```

%macro dLprior ;
array int[18] int_1-int_18 ;
%do i=1 %to 17 ;
parms m_dL0_&i. 1 ;
dL0star[&i.] = int[&i.+1]-int[&i.] ;
prior m_dL0_&i.~gamma(r* dL0star[&i.] * c, iscale=c) ;
%end ;
%mend dLprior ;
proc mcmc data = A_GEHAN
seed = 160721 nbi = 1000 nmc = 10000 nthin = 1 outpost = NLOUT ;
c = 0.001 ;
r = 0.1 ;
array dL0star[17] ;
array m_dL0[17] m_dL0_1-m_dL0_17 ;
parms beta 0 ;
prior beta ~normal(0.0,prec=1e-6) ;
%dLprior ;
m_ldt = m_y*exp(beta*drug)*m_dL0[int_index] ;
model m_dn~poisson(m_ldt) ;
run ;

```

表 3.2 に、回帰パラメータ  $\beta$  に対する推定結果として、事後平均、事後標準偏差、事後中央値、95%信用区間 (事後分布の 2.5%点と 97.5%点) を示した。表 2.2 には示していないが、95%最高事後密度信用区間 (highest posterior density credible interval) は[-2.2894, -0.6258]であった。データセット NLOUT に出力されたモンテカルロ標本を用いることで、あるいはプログラム 3.2 にて MCMC プロシジャのオプションとして monitor=(HR) と追加し、MCMC プロシジャの中で beginnodata; HR=exp(beta); endnodata; と指定することで、対応するハザード比 (プラセボ群に対する 6-MP 群のハザード比) の事後平均 0.2344、95%信用区間[0.089, 0.476]を得る。

表 3.2 : 比例ハザードモデルにおける回帰パラメータの推定結果

	Standard		Percentiles	
	Mean	deviation	Median	2.5 97.5
$\beta$	-1.5364	0.4187	-1.5213	-2.4223 -0.7427

比例ハザードモデルにおいて、生存関数は、累積基準ハザード関数を  $\Lambda_0(t)$  として(2.7)より

$$S(t | \mathbf{z}_i) = \exp(-\exp(\beta^T \mathbf{z}_i) \Lambda_0(t))$$

となる。累積基準ハザード関数  $\Lambda_0(t)$  に  $GP(c\Lambda_0^*(t), c)$  をあてはめたときには、ガンマ過程の定義より、時間軸  $(0, \infty)$  を、 $0 = a_0 < a_1 < a_2 < \dots < a_{j-1} < a_j = \infty$  と、 $J$  個の区間  $(0, a_1], (a_1, a_2], \dots, (a_{j-1}, a_j]$  に分割したとき、区間  $(a_{j-1}, a_j]$  での累積基準ハザード関数の増分

$$\Lambda_0(j) = \Lambda_0(a_j) - \Lambda_0(a_{j-1}) \quad (j=1, 2, \dots, J)$$

は平均  $d\Lambda_0^*(t)$ 、分散  $d\Lambda_0^*(t)/c$  をもつガンマ分布に従う。プログラム 3.3 は、上記に従い、データセット NLOUT に出力された 10000 個のモンテカルロ標本を用いて、各群の生存関数を推定する SAS プログラム例である。プログラム 3.3 により推定された生存関数の事後平均を図 3.1 に示した。

プログラム 3.3 : GP を用いた比例ハザードモデルによる生存関数の推定

```
proc sql ;
  create table SURVE as
  select * from NLOUT, PARTITION ;
quit ;
data SURVE ;set SURVE ;
array int[18] int_1-int_18 ;
array aa[17] m_dL0_1-m_dL0_17 ;
do int_index = 1 to 17 ;
  t = int[int_index] ;
  m_dL0 = aa[int_index] ;
  output ;
end ;
keep Iteration int_index t beta m_dL0 ;
run ;
data SURVE ;set SURVE ;by Iteration ;
```

```

keep iteration t S_placebo S_6MP ;
if first.iteration = 1 then m_H0 = 0 ;
m_H0+m_dL0 ;
S_6MP = exp(-m_H0)**exp(beta*1) ;
S_placebo = exp(-m_H0)**exp(beta*0) ;
run ;

```

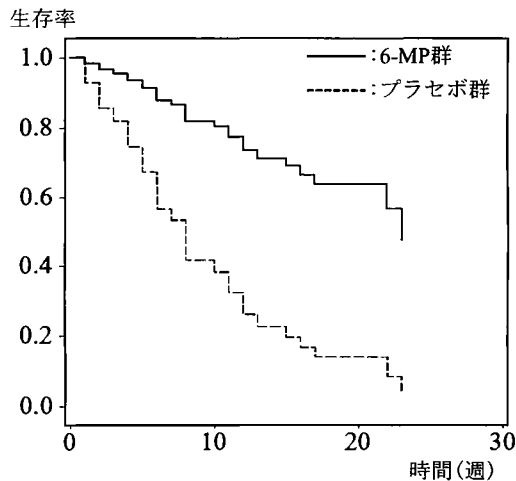


図 3.1 : Gehan のデータに対する GP-PH モデルによる生存関数

なお、基準ハザード関数に特定の事前分布をおくことなく、部分尤度に基づき推定された回帰パラメータを用いて生存関数を推定する場合には、データセット GEHAN を入力データセットとし、PHREG プロシジャにて BAYES ステートメントにより回帰パラメータの事前分布を指定し、BASELINE ステートメントにより各群に対する生存時間を出力させることができる。ただし、デフォルトの設定では共変量の平均値に対する生存時間のみが出力されるため、BASELINE ステートメントのオプション COVARIATE= で予め用意した各群を表す共変量の値をもたせたデータセットを指定する必要がある。プログラム 3.4 にその一例を示した。BAYES ステートメントのオプション COEFFPRIOR オプションを用い、回帰パラメータの事前分布に平均 0、分散  $10^6$  の正規分布を指定している。プログラム 3.2 と同様、Burn in を 1000、Burn in で除外された後のモンテカルロ標本におけるサンプリング間隔を 1、最終的にサンプリングされるモンテカルロ標本のサンプル数を 10000 個とした。

プログラム 3.4 : 部分尤度に基づく比例ハザードモデルによる生存関数の推定

```

data Prior ;
input _TYPE_ $ drug ;
cards ;
Mean 0
var 1e+6

```

```

;
run ;
data Cov_vals ; do drug = 0,1 ; output ; end ; run ;
proc phreg data = GEHAN ;
model week*remiss(0) = drug ;
bayes seed = 160721 coeffprior = normal(input = Prior)
      nbi = 1000 nmc = 10000 thin = 1 outpost = Post ;
baseline out = Estsurv survival = Surv
covariates = Cov_vals /nomean ;
run ;

```

#### 4. おわりに

ベイズ流解析では、パラメトリックな推論がよく用いられている。パラメトリックな推論では、あてはめる統計モデルにおける分布関数を有限次元のパラメータにより表現する。有限個のパラメータに対して事前分布を設定し、ベイズの定理を用いて得られる事後分布からパラメータの推論を行う。分布関数が有限個のパラメータで定式化できないとき、ノンパラメトリックな推論となる。ノンパラメトリックな推論では、通常、確率過程を用いて分布関数を表現する。得られたデータを確率過程からのサンプルパスと捉え、あてはめた確率過程からの経路によって決定される確率分布に従うと考える。Gehan のデータに対する適用例では、時間軸をイベントが発現した時点で分割し、各区間での累積基準ハザード関数の増分がガンマ分布に従うものとしてパラメータを推定した。このように、ノンパラメトリックとは、パラメータをおかないことではなく、無限個のパラメータをもつという意味である。想定したモデルでは無限個のパラメータであっても、得られたデータは“有限”であるため、得られたデータに対して解析するには有限個のパラメータでよく、有限個のパラメータにおとしこむことができれば、事後分布の計算が可能となる。

Kalbfleisch (1978) は、累積基準ハザード関数にガンマ過程やディリクレ過程をあてはめることの問題点として、確率 1 で累積基準ハザード関数が離散的な値をとること、隣接する区間で累積基準ハザード関数の増分を独立とみなすことをあげている。Dykstra and Laud (1981) は、基準ハザード関数の増分に拡張ガンマ過程をあてはめる方法を提案した。このことにより、上記に示した、ガンマ過程やディリクレ過程を累積基準ハザード関数にあてはめたときの問題点を回避できる。基準ハザード関数に対するモデリングとしては、本稿で取り上げたガンマ過程を用いる以外にも多数存在し、それらについて Ibrahim et al.(2001)によって紹介されている。こうした方法の多くは、1)基準ハザード関数に厳格なパラメトリック仮定をおかない、2)基準ハザード関数に対する推測がなされる、3)適当な事前分布を選ぶことにより推定された基準ハザード関数はもっともらしい滑らかな曲線となる、といった利点を有している (Dey and Rao,2005)。

本稿では、ベイズ流アプローチとして、累積基準ハザード関数に対しガンマ過程をあてはめた比例ハザードモデルをとりあげ、SAS を用いてこのモデルをあてはめ解析を行うための具体的な方法を紹介した。比例ハザードモデルに対するベイズ流ノンパラメトリックな方法について検討する際に、本稿がその一助になれば幸いである。

## 参考文献

- [1] Cox,D.R.(1972).Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society, Series B.* **34**:187–220.
- [2] Dey,D.K. and Rao,C.R. (2005). *Bayesian Thinking, Modeling and Computation*. Elsevier, Amsterdam. 繁榊算男・岸野洋久・大森裕浩監訳(2011). ベイズ統計分析ハンドブック. 朝倉書店.
- [3] Dykstra,R.L. and Laud,P.W. (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics.* **9**, 356-367.
- [4] Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrary single-censored samples. *Biometrika.* **52**: 203–223.
- [5] Hobbs, B.P. and Carlin, B.P. (2008). Practical Bayesian design and analysis for drug and device clinical trials. *Journal Biopharmaceutical Statistics.* **18**:54-80.
- [6] Ibrahim, J.G., Chen,M.H., Sinha,D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York.
- [7] Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society: Series B.* **40**:214–221.
- [8] SAS Institute Inc. (2008). *SAS/STAT(R) 9.2 User's Guide*, Cary, NC, USA: SAS Institute Inc.
- [9] SAS Institute Inc. (2011). *SAS/STAT(R) 9.3 User's Guide*, Cary, NC, USA: SAS Institute Inc.
- [10]Sinha, D. Ibrahim, J. G. and Chen, M. H. (2003). A Bayesian justification of Cox's partial likelihood. *Biometrika.* **90**(3): 629–641
- [11]WinBUGS – The BUGS Project. WinBUGS 1.4, Examples Volume I, Leuk: survival analysis Using Cox regression. Available at [http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS\\_Vol1.pdf](http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS_Vol1.pdf).
- [12]大橋靖雄・浜田知久馬 (1995). 生存時間解析 –SASによる生物統計. 東京大学出版.
- [13]黒田普吾 (2010). Cox 回帰における Bayes 推定と PHREG プロシジャ.日本 SAS ユーザー総会 論文集,23-44.
- [14]服部聡 (2009). 生存時間解析におけるセミパラメトリック推測とその周辺. 統計数理. **57**(1):119-138.
- [15]中村剛 (2001). Cox 比例ハザードモデル. 朝倉書店.

## 連絡先

E-mail: yada-s@a2healthcare.com

# 落語協会の有力一門の勢力分布をD3.jsで「分かりやすく」ビジュアライズする －JSONプロシジャを用いて－

○坂部 裕美子

(公益財団法人 統計情報研究開発センター)

## (1) 表示させる内容について

「D3.js」とはデータを印象的に可視化できるJavaScriptライブラリの一つで、無料でコードを使用、改造、修正できる。<http://d3js.org/>で多数の利用事例が紹介されているが、この中の「Circle Packing」(<http://bl.ocks.org/mbostock/4063530>)という表示プログラムは、グループ内の階層構造およびグループ全体の大きさを明瞭に表示できる上に、図を見ただけでは「最上位はどれか」という明確な順位が判別できない(落語家はこの辺りのことに非常に敏感である)という利点がある。そこで、Webデザインについては全くの初心者ながら、プログラムの指示内容を判読し、自身で集計した「落語協会の寄席定席興行への登場回数の一門別表示」への書き換えを試みた。

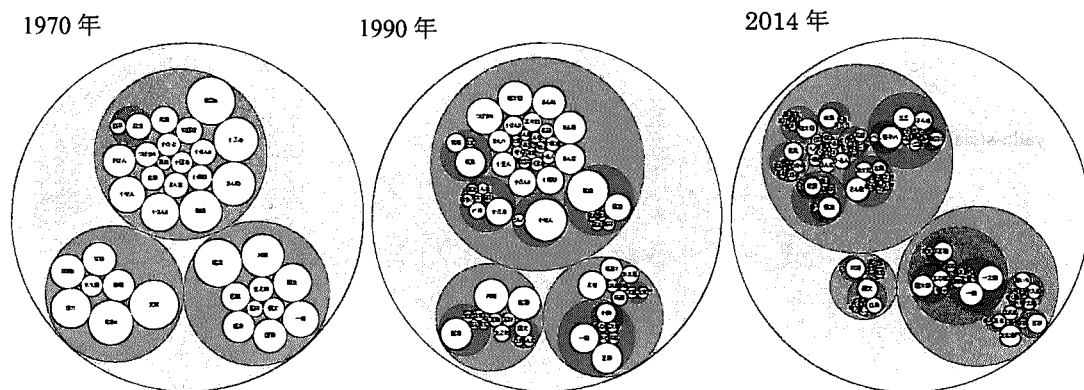
このプログラムではJSON形式ファイルを使用しているため、SAS9.4から使用可能になったJSONプロシジャを用いて、集計結果DSを書き出した。

## (2) 表示プログラム

すべての落語家について描画すると紛雑になるので、登場回数の動向が特徴的な「柳家小さん一門」「三遊亭圓生一門」「林家彦六一門」の3つのみを表示させることとする。

JSONプロシジャ内では、「write ~;」という書き出し方に関する指示と、DSのexportの指示しか使えないため、事前に「書き出しやすいDS」を作成しておく必要がある。今回は、円の階層構造を示したDSと、師匠別に所属メンバーと各々の回数をまとめたDSを作成することとし、これらを適宜マクロで呼び出しながらファイルを書き出した。

作成したJSONファイルを読み込ませて図示したところ、総計・一門計でグラフ表示した場合よりも情報量の多い図を描画できた。



※元データの集計内容については、「SASを用いた寄席定席興行の現状分析」(SASユーザー会発表論文集 2005年、p.273～)を参照頂きたい。

SASシステム／その他関連分野



## SASによる迷路の自動生成

○松沢 享

(株式会社インターサイエンティフィックリサーチ)

### 【はじめに】

SASのデータハンドリング機能は日々進化しており、これまでSASでは面倒とされていた処理が容易に行えるようになった。

本発表では、その進化の一例として、従来のデータステップと新機能を用いて迷路を自動生成するSASプログラムを作成したので紹介する。

### 【方法】

本発表では迷路生成の入門として比較的単純なアルゴリズムかつプログラミングが容易な棒倒し法と穴掘り法<sup>[1]</sup>の2つの手法を用いる。

これらの手法は配列とハッシュオブジェクトを用いて実現し、迷路の描画にはSAS9.4から機能追加されたReport Writing Interfaceを用いる。

### 【結果】

棒倒し法による結果を図1に、穴掘り法による結果を図2に示す。

図2の穴掘り法は線を細くする等の処理を加えて、より迷路らしい見た目になっている。

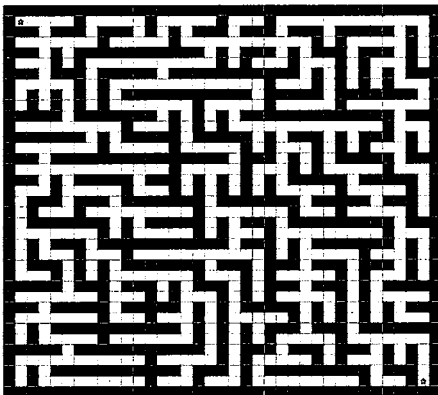


図1 棒倒し法による迷路の生成結果

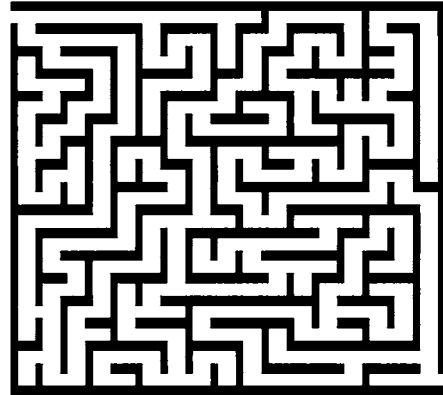


図2 穴掘り法による迷路の生成結果

### 【総括】

ハッシュオブジェクトやReport Writing Interface等の新たな機能を用いることで、迷路の生成を行うプログラムを容易に作成する事が出来た。

本発表によりSASのデータステップの進化と魅力を感じて頂ければ幸いです。

### 【参考文献】

[1] 柳井 政和(2014). プログラマのためのコードパズル ~JavaScriptで挑むコードゴルフとアルゴリズム. 技術評論社

生物統計・医薬関連

## Server版SASを基盤にした「生物統計解析システム」の紹介

○柄澤 欣之、角元 慶二、佐々木 和典、菱垣 晴次

(大塚製薬株式会社 徳島研究所)

### 概略

「生物統計解析システム」は、個人の統計スキルに依存せずに統計解析を行えるシステムである。各個人 PC のマウス操作より、エクセルファイルの実験データに対し、統計解析手法を選択し実行するものとなっている。

### 要旨

近年、臨床試験でのデザイン上の問題や、統計方法の誤用、データの改竄といった問題が起り、改めて統計家の意識の向上を考えさせられた。この問題は非臨床試験に対しても同様であり、1999年には、「非臨床試験に関する統計学的提言」が日本製薬工業協会・医薬品評価委員会から発表されている。非臨床試験については、統計家の関与が増えつつあるものの、臨床試験に比べ統計家の人数が少ないのが現状である。非臨床試験ではデータの種類が雑多であり、市販の統計ソフトウェアを使う場合において、ブラックボックス的に統計処理が行われていることが問題である。特に、委託試験においては、適切に統計処理が行われているかを確認する必要も増えている。我々は、1992年より医薬品の研究所における SAS を用いた社内システムの活用事例を紹介するとともに、ブラックボックス的な統計処理の使用を避けるための使用上の検討も行ってきた。従来のシステムでは各部署に統計解析用の PC を用意する必要があり、同時使用の制限や、SAS の毎年の更新作業の手間などが生じるという問題があった。

今回、我々は Server 版 SAS を導入し、「生物統計解析システム」を構築した。「生物統計解析システム」では、各研究員のネット環境のある個人 PC のマウス操作により、エクセルファイルの実験データに対し、統計解析内容を選択して統計検定を行える。そのため、個人の統計スキルに依存せずに統計解析を行えるシステムとして社内 350 人が使用している。今回、非臨床試験、特に薬効薬理試験等に関して、研究の質の向上をはかるための道具として、Server 版 SAS を基盤にした「生物統計解析システム」を紹介する。本抄録では、「生物統計解析システム」のシステムおよび統計解析内容の紹介を行い、次に使用方法を説明することで解析ツールとしての利便性がどの程度であるかを議論したい。

「生物統計解析システム」について

統計家が開発した最新統計解析処理プログラムをリアルタイムで各研究員が利用し統計検定できるシステムとなっている。「生物統計解析システム」では、サーバに開発した統計解析プログラムを導入することで、社内ネットを利用して各研究員が利用できるよになっている(図 1)。

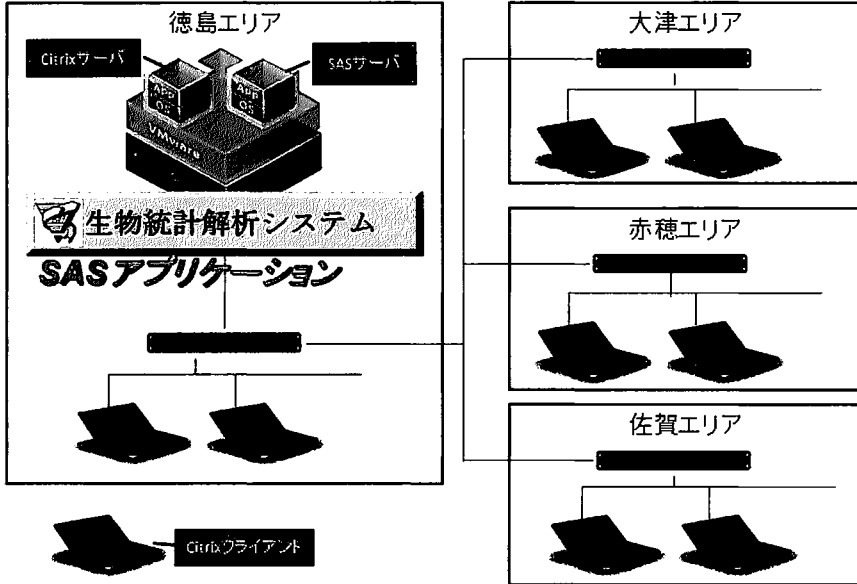


図 1 社内 LAN 環境について

各研究員が例えば SAS などを利用して統計解析を行うことは困難であるため、エクセルファイルの実験データを用意すれば統計解析内容を選択して統計検定を行えるように、VBA, COM, DDE, SAS を利用して、エクセルファイルから SAS データセットを作成し、結果をエクセルファイルで出力されるようなシステムを構築した(図 2)。このことにより統計学やプログラミングのスキルが高くない研究員においても統計検定が各自で行えるようになった。

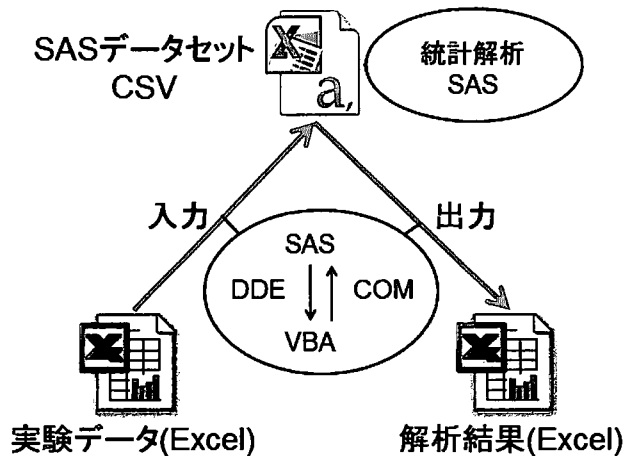


図 2 解析ワークフロー

生物統計解析システムでは、非臨床試験の生物統計に関する検定は一通り網羅されており(表1)、更に必要に応じて随時、新しい統計解析処理の追加対応を行うことができる。このことにより、各研究員の様々な要望に瞬時に対応することができるシステムとなっている。

表1 実装統計手法(抜粋)

用語	手法	分散対数分布	正規分布	正規分布以外	多変数	非正規分布	その他
群間比較	検定(仮定あり/なし)	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
	Wilcoxon 順位和検定	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
	Dunnnett or Turkey or Helm or Bonferroni 検定	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
	多変数 Dunnnett or Turkey	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
回帰分析	混合モデル	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
	Williams 検定	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
	線形計画法(線形検定)	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
	PKC-メータ	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
累積データ	Combination Index	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
	Log-Rank 検定	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
因子分析	Man-Whitney 検定	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
	Anderson-Darling 検定	多変数	正規分布	正規分布以外	多変数	正規分布	多変数
その他	検定(仮定あり/なし)	多変数	正規分布	正規分布以外	多変数	正規分布	多変数

統計解析処理の追加対応事例の具体例の一つとして、Log-Rank 検定における Dunnnett 型の多重性調整アルゴリズムに関して説明する(図3)。Dunnnett 型の多重性調整を行うには、検定を行うデータにおいて「Factor-Analytic Structure」と呼ばれる特定の形式になる必要がある。検定を行うデータが「Factor-Analytic Structure」である場合は Dunnnett 型の多重性調整を行い、そうでない場合は Dunnnett-Hsu 近似を行う。ここまでは、SAS を用いた Log-Rank 検定で標準的に搭載されている機能である。しかし、データセットによっては Dunnnett-Hsu 近似にも収束した結果を得ることができない場合がある。そこで、我々は Dunnnett-Hsu 近似でも収束した結果を得られない場合においては、Simulation による近似を行って得られた結果を出力するようにしている。このように、生物統計解析システムでは実際のデータ解析を行う上で統計解析処理機能に随時追加対応することができる。

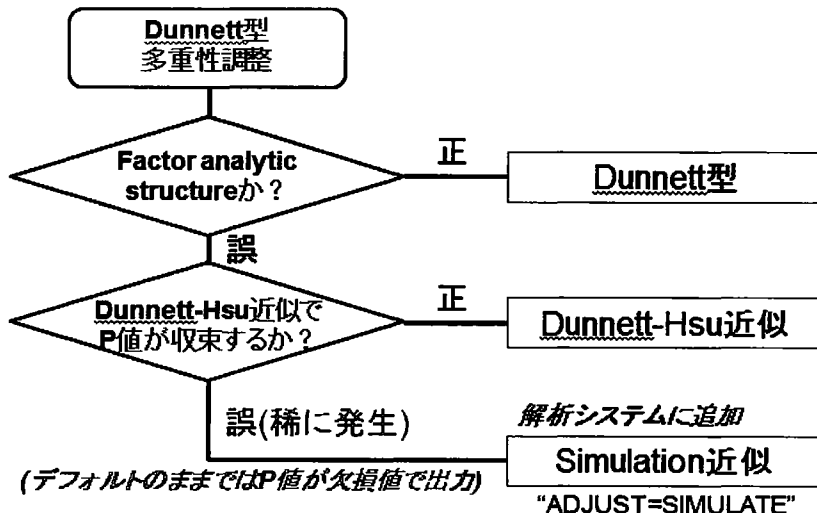
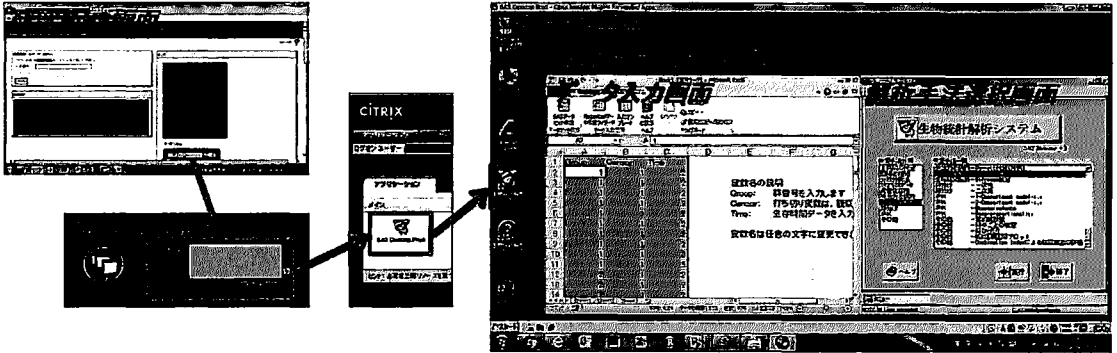


図3 統計解析処理の追加対応事例(Log-Rank 検定における Dunnnett 型の多重性調整)

## 「生物統計解析システム」の使用方法

### A. Server に接続し、生物統計解析システムを起動する。



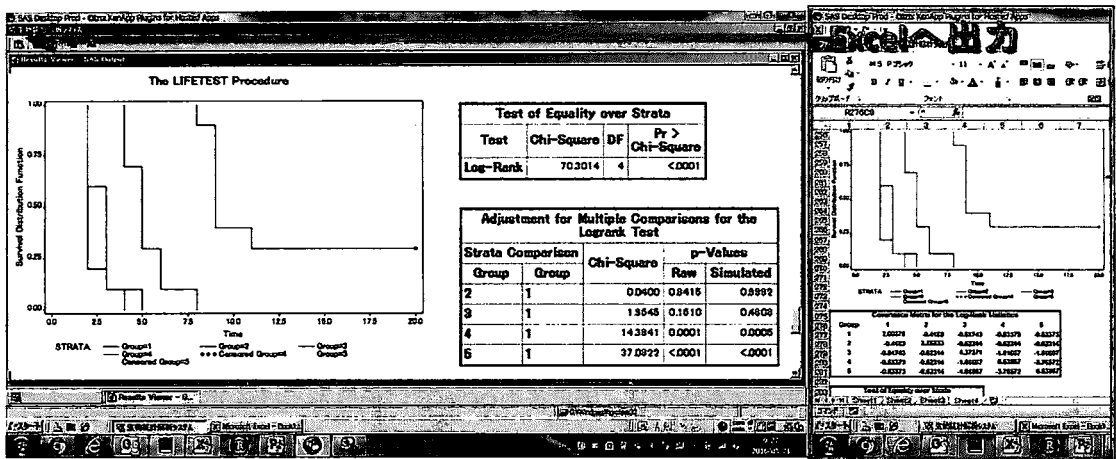
### B. データの入力方法

1. 「入力テンプレート」をクリック
2. 検定手法を選択し実行。テンプレートが作成される。
3. データを入力
4. 範囲を指定
5. 「SASデータセット作成」をクリック
6. SASデータセット作成

### C. 解析手法の選択と、変数の設定



#### D. 解析結果(抜粋)



#### 参考資料

- [1]角元慶二, 佐々木和典, 小原直樹(1997). 研究所データ解析における SAS システム活用事例, 日本 SAS ユーザー会論文集
- [2]角元慶二(1997). 各種の実験デザインにおける PROC GLM, PROC MIXED の利用, 日本 SAS ユーザー会論文集
- [3]川崎美保, 佐々木和典, 小原直樹, 角元慶二(1997). SAS/AF ソフトウェアを用いた前臨床解析システムの紹介, 日本 SAS ユーザー会論文集
- [4]佐々木和典, 角元慶二, 小原直樹, 岡美保(2001). SAS バージョン 8e による非臨床統計解析システムの開発, 日本 SAS ユーザー会論文集
- [5]Kakumoto K., Yamanaka S., Hamada C., Yoshimura I.(2004)A statistical analysis of an effective method to conduct in silico screening for active compounds. Chem-Bio Informatics Journal, 4, pp.121—132
- [6]共著: 菅波秀規, 角元慶二, 長谷川要, 他 (2003). SAS における統計解析バリデーションと解析計画書, 報告書作成 技術情報協会
- [7]SAS 社ホームページ, テクニカルサポート. Dunnett 法による多重比較

## SASで作成した統計解析帳票の柔軟なQC体制構築に向けた取り組みの事例紹介

○仲川 慎太郎, 吉本 拓矢, 吉田 寛輝, 久力 洋, 河田 祐一

(中外製薬株式会社 臨床開発企画部)

弊社の統計解析帳票（以下、帳票）作成業務では、データ固定後（以下、固定後）の帳票QC作業が帳票納品までのクリティカルパスであり、QC期間をできるだけ短縮することが望ましい。当該期間を短縮するためには、固定後迅速に帳票を作成しQCを完了させることが重要となる。一方、データ固定前は比較的作業時間に余裕があるため、従来固定後に実施していたQC作業の一部を固定前に前倒しすることによる帳票提供までの期間の短縮可能性を考案した。具体的には、重要度と複雑さの観点から各帳票のリスクを特定し、固定後のQC内容に緩急をつけるリスクベースアプローチの有用性を検討した。また、統計解析担当者以外でも帳票をQCできれば、社内リソースを最大限活用した柔軟なQC体制を構築可能であると考え、そのためにどのようなトレーニングが必要になるか等も併せて検討した。

以前、QC作業の一部前倒しによる固定後QC時間の削減効果やQCの質が従来法から保持されているかを検討するために、弊社の新入社員を対象として単純な比較試験（従来のQC法 vs. 前倒しQC法）を試みたが、QC方法の差異だけでなく固体間差（ベースラインのQC能力（訓練度）、集中力の持続性等の差）やQC環境等の差異といった要因もQC時間に影響した可能性を否定できなかった。そこで今回、「QC作業の前倒しにより帳票提供期間をどの程度削減可能か」という課題に対してより真に近い推定値が得られるよう、上述したバイアス要因をデザイン段階で調整し、比較可能性を可能な限り保証できるような試験を再実施した。具体的にはwash-out期間を十分に設けた2期のクロスオーバー比較試験を実施した。また、ベースライン時点でのQC能力やデータ入力をミスする確率等を個人間でできるだけ揃えるためにトレーニングを一定期間設けたり、集中力がQC時間に影響したりしないような工夫を施した。

本発表では、SASで作成した帳票のQC時間短縮効果検討にあたり、新入社員にどのような説明資料を用意してトレーニングを実施したのか等を説明し、本実験結果も簡単に報告する。さらに、実験終了後に実施した新入社員のアンケート結果を分析し、どのようなトレーニングを事前に実施すればより柔軟なQC体制（具体的には統計解析担当者以外によるQC体制）を構築可能なのかといった点を考察して今後の展望を述べる。



# 区間打ち切りを考慮したノンパラメトリックな生存関数推定法の推定性能を向上させる診断スケジュールの検討

○野村怜史、寒水孝司、浜田知久馬

(東京理科大学)

## 背景

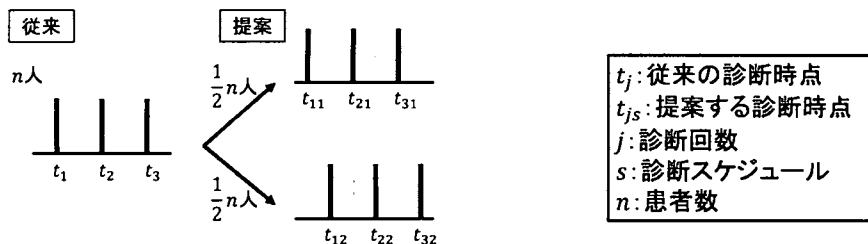
がんの臨床試験において、無増悪生存期間という評価項目が用いられることがある。無増悪生存期間とは試験への登録時点から死亡または増悪が認められた時点までの期間をあらわす。無増悪生存期間を評価項目とする場合、増悪についてはCT検査等の画像診断によって判定するため正確な発生時点が特定できず、ある区間で起きたことしかわからないデータ（区間打ち切りデータ）となる。区間打ち切りデータにおいて増悪確認時点を増悪発生時点として評価してしまうと、生存時間が過大に評価されてしまう。しかし、従来の臨床試験では、増悪確認時点を増悪発生時点として評価を行っており、生存時間の過大評価が起きているのが現状である。

## 研究目的

本研究は無増悪生存期間を評価項目とした際の生存関数の推定性能を向上させる診断スケジュールを提案することを目的としている。評価指標としては、生存関数の代表値として用いられるMST(Median Survival Time)に関し、MST Bias、MSE(Mean Squared Error)、MST95%信頼区間被覆確率を用いた。生存関数の推定にはSASのPROCEDUREのLIFETEST、ICLIFETEST[1,2]を用いた。ICLIFETESTは区間打ち切りを考慮して生存関数を求めることができる。

## 提案方法

提案方法は、各患者を複数の診断スケジュールにそれぞれ割り当てることである。従来、診断スケジュールは全ての患者が同じスケジュールである。提案方法は患者全体で見た際の診断時点を増加させることが狙いである。



## 結果・まとめ

シミュレーション実験により評価したところ、提案法はICLIFETESTを用いた際に、従来法と比較して、MSTの推定値のBiasやMSEを減少させ、MSTの95%信頼区間の被覆確率においても、従来の診断スケジュールに比較して提案スケジュールが安定した性能を示した。よって無増悪生存期間を評価項目とする際には、提案するスケジュールを用いICLIFETESTによる区間打ち切りを考慮した生存関数推定をすることを推奨する。

## 参考文献

- [1] Peto, R. Experimental Survival Curves for Interval-censored Data. Applied Statistics. 1973; **22**: 86-91.
- [2] Sun, J. The Statistical Analysis of Interval-censored Failure Time Data. Springer. 2006.

## PROC SGPLOTによるSwimmer Plot –見やすさを求めて–

吉田 秀幸<sup>2</sup>、韓 士榮<sup>1</sup>、若菜 明<sup>1</sup>

○福田 裕章<sup>1</sup> (発表のみ)

(<sup>1</sup>MSD株式会社、<sup>2</sup>株式会社タクミインフォメーションテクノロジー)

swimmer plot は、患者ごとのレスポンスやそれに至るまでの時間等を図示する手法であり、その最大の特徴は、これらの情報を一目で把握することが可能という点である。これまで、SAS による swimmer plot 作成の方法はさまざまに報告されており、SAS version 9.3 の SGPLOT プロシジャや TEMPLATE プロシジャ及び SGRENDER プロシジャを利用した報告等もなされているが、未だ明瞭性や情報量の点で十分とは言えず、臨床家や論文執筆者の要求に応え切れていないことが現状である。特に、近年「がん免疫療法」ががんの画期的な治療法として注目を集めているが、安定した効果が長期間持続することが期待されるため、従来のがん化学療法と異なり、腫瘍縮小や増悪の遅延によって評価されるだけでなく、患者ごとの治療継続の是非が重要な情報である。したがって、がん治療に対するレスポンスを患者ごとに一目で把握することのできる swimmer plot は当該治療法に対する臨床試験の解析において、非常に有力なツールとなると考えられるが、従来の swimmer plot は、患者の投与継続情報が含まれていない、あるいは含まれていたとしても明瞭性に欠けるものであった。一方、SGPLOT プロシジャの annotation 機能を利用することで、複雑な色調配分や線種、シンボル等の詳細な規定が可能となるため、当該機能を用いることでより情報量の多い作図が可能である。そこで、annotation 機能を利用した SGPLOT プロシジャによる、患者の治療継続状況を含む、より明瞭な swimmer plot 作成プログラムの開発を試みた。

本プログラムは、規定の様式のデータセットに対し、患者のレスポンスを示すシンボル、そのシンボルの説明や投与群の情報を示したレジェンドという大きく 2 つの成分を含む annotation データセットを利用し、SGPLOT プロシジャにより swimmer plot の作成を行うものである。scatter を用いてシンボルを表示せず annotation を利用することにより、unicode を用いた特殊記号を使えるなどシンボル表示の柔軟性を高めた。また、レジェンド部分についても annotation を利用することで、SGPLOT プロシジャで自動的に付与されるレジェンド以上に詳細な説明を行うことを可能とした。その結果、臨床家や論文執筆者が求める理想的なものに限りなく近い swimmer plot を作成できるプログラムを開発することができた。さらに、当該プログラムの簡単な説明書を作成し、インプットする規定のデータ形式や annotation データに対する仕様を示すことによって、プログラムに精通していない作業従事者でも容易に取り扱うことを可能とした。しかしながら、本 swimmer plot は、がん領域における Phase 1 又は Phase 2 試験を想定したものであり、症例数が数百例に及ぶ国際共同比較試験のような規模を想定していない。ゆえに、そのような規模の大きい試験に対し、全症例の情報を 1 つの plot としてどのように表示するかについての対応が今後の課題と考えている。

【企画セッション】 欠測のあるデータの解析のためのSASプログラムの紹介  
～データ発生・DIAマクロとプロシジャの進展～

(1) セッションの概要

○土居正明<sup>1,2</sup>

(<sup>1</sup>日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4  
欠測のあるデータの解析検討チーム, <sup>2</sup>東レ株式会社)

欠測のあるデータの解析に対する背景と、これまでにSASユーザー総会で行った製薬協 欠測チームの企画セッションの内容を概説した上で、本セッションの課題と概要を説明する。

【企画セッション】 欠測のあるデータの解析のためのSASプログラムの紹介  
～データ発生・DIAマクロとプロシジャの進展～  
(2) シミュレーションデータ発生方法

○横山 雄一<sup>1,2</sup>, 大浦 智紀<sup>1,3</sup>, 土居 正明<sup>1,4</sup>

(<sup>1</sup>日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4

欠測のあるデータの解析検討チーム, <sup>2</sup>持田製薬株式会社,

<sup>3</sup>日本イーライリリー株式会社, <sup>4</sup>東レ株式会社)

応答変数が経時的に測定される臨床試験において、データの欠測はよく起こる。欠測のあるデータに対しては、様々な解析手法があり、試験や解析の目的に合わせた使い分けが必要となる。そのため、臨床試験の計画段階において、試験及び薬剤の特徴を捉えたシミュレーションデータを発生させ、目的に合わせた検討を行うことは有用である。本発表では、文献データから、欠測のある連続量経時データを発生させる方法について、SASプログラムとともに紹介する。

【企画セッション】 欠測のあるデータの解析のためのSASプログラムの紹介  
～データ発生・DIAマクロとプロシジャの進展～  
(3) MNARの仮定の下でのSelection Modelに基づく解析

○黒田 晋吾<sup>1,2</sup>

(<sup>1</sup>日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4  
欠測のあるデータの解析検討チーム、<sup>2</sup>武田薬品工業株式会社)

DIAマクロ「%Selection\_Model2」を利用することによって、Diggle and Kenward (1994)により提案されたMNARの仮定の下でのSelection Modelに基づく解析が実行可能である。本発表では、「%Selection\_Model2」で使用されている解析手法を解説し、二変量正規分布に従う欠測のあるデータを例としてSAS/IML®を利用した実装方法を示すとともに、マクロの利用方法及びマクロを利用する際の留意点も紹介する。また、SASのMCMCプロシジャを利用してBayes流の方法により同様の解析を実行する方法についてもあわせて紹介する。

【企画セッション】 欠測のあるデータの解析のためのSASプログラムの紹介  
～データ発生・DIAマクロとプロシジャの進展～

(4) Pattern Mixture ModelとMultiple Imputationに基づく解析1 (Delta adjustmentとTipping Point Analysis)

○大浦 智紀<sup>1,2</sup>

(<sup>1</sup>日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4  
欠測のあるデータの解析検討チーム、<sup>2</sup>日本イーライリリー株式会社)

Delta adjustmentおよびTipping Point Analysisを行うためのDIAマクロである%delta\_pmmおよび%delta\_and\_tip (missingdata.org.ukで公開されているPMM Delta Tipping Point and CBI macros に含まれる) の背景となる考え方およびマクロのパラメータの設定について解説を行う。またシミュレーションデータへの適用結果も示す。

【企画セッション】 欠測のあるデータの解析のためのSASプログラムの紹介  
～データ発生・DIAマクロとプロシジャの進展～

(5) Pattern Mixture ModelとMultiple Imputationに基づく解析2 (Reference-based imputation)

○藤原 正和<sup>1,2</sup>

(<sup>1</sup>日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4  
欠測のあるデータの解析検討チーム、<sup>2</sup>塩野義製薬株式会社)

DIAマクロであるFive macros (%part1A, %part1B, %part2A, %part2B, %part3) を利用することによって実装可能なReference-based imputationである, CIR (Copy Increment from Reference), J2R (Jump to Reference), 及びCR (Copy Reference) について詳細を説明する. また, 補完後に実施されるDelta methodsについても説明する. その後, Five macros内の引数の意味合いとシミュレーションデータへの適用例を紹介する.

【企画セッション】 欠測のあるデータの解析のためのSASプログラムの紹介  
～データ発生・DIAマクロとプロシジャの進展～

(6) MIプロシジャで実行可能なPattern Mixture ModelとMultiple Imputationに基づく解析

○渡邊 大丞<sup>1,2</sup>

(<sup>1</sup>日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4  
欠測のあるデータの解析検討チーム、<sup>2</sup>サノフィ株式会社)

欠測のあるデータの解析を行う際、「応答変数の分布が異なるであろう部分集団」（以後「パターン」と呼ぶ）ごとに分布を想定し、その複数のパターンが混在した状況を扱うモデルに、Pattern-Mixture Model (PMM)がある。PMMとして、確率モデルに含まれている推定不可能なパラメータに対して、何らかの制約条件を加えて、Multiple Imputation (MI) の解析を利用して行う方法があり、Missing Not At Random (MNAR) に基づく解析として利用されている。またSAS Ver9.4からProc MIでMNARステートメントが使用可能となり、①CCMV、NCMV等の制約条件を用いた解析、②controlled imputationや③delta-adjustmentに基づく解析が実行可能となった。本発表は応答変数が単調な欠測である場合のMNARステートメント使用時の留意点を示し、シミュレーションデータを使用した解析結果を提示する。



【企画セッション】 欠測のあるデータの解析のためのSASプログラムの紹介  
～データ発生・DIAマクロとプロシジャの進展～

(7)セッションのまとめと質疑応答

○大江基貴<sup>1,2</sup>, 土居正明<sup>1,3</sup>

(<sup>1</sup>日本製薬工業協会 医薬品評価委員会 データサイエンス部会 タスクフォース4  
欠測のあるデータの解析検討チーム, <sup>2</sup>株式会社大塚製薬工場, <sup>3</sup>東レ株式会社)

セッションのまとめと質疑応答を行う。

# SAS での Network Meta-Analysis の実施例

## ～ 頻度論に基づくアプローチ ～

○舟尾 暢男、黒田 晋吾

武田薬品工業株式会社 日本開発センター 生物統計室

Conducting Network Meta-Analysis Using SAS

Nobuo Funao, Shingo Kuroda

Takeda Pharmaceutical Company, Ltd.

### 要旨

SAS による Network Meta-Analysis の実施方法として、均質性 (homogeneity) や一貫性 (consistency) の評価、各薬剤の Ranking の算出方法を含めて紹介する。

キーワード：Network Meta-Analysis、均質性(homogeneity)、一貫性(consistency)、Ranking、頻度論、ベイズ

### 1. はじめに

Meta-Analysis を実施することで、複数の試験を統合した上で薬剤の効果を比較することが出来るが、通常の Meta-Analysis で比較することが出来る薬剤は2種類のみである。近年、複数の試験を統合した上で3種類以上の薬剤間比較を行うことが出来る手法として Network Meta-Analysis が提案されており、この方法では、直接比較の結果と間接比較の結果を統合したうえで薬剤間比較を行い、また、直接比較を実施していない薬剤の組についても間接的な薬剤間比較 (間接比較) の結果を得られる点が特徴的である。Network Meta-Analysis はベイズの枠組みで解析が実施されることが多いが、最近では頻度論の枠組みでの解析手法も Krahn (2013) 等により提案されている。

本稿では、頻度論の枠組みでの解析手法による Network Meta-Analysis の実施方法として、SAS を利用した薬剤間差の推定量の算出方法、均質性 (homogeneity) や一貫性 (consistency) の評価方法、及び各薬剤の Ranking の算出方法について紹介する。参考までに、Bayesian Network Meta-Analysis の結果との比較も行う。

### 2. Network Meta-Analysis とは

通常の Meta-Analysis では、複数の試験を統合した上で2種類の薬剤効果を比較するが、Network Meta-Analysis では、複数の試験を統合した上で3種類以上の薬剤効果を比較することが出来る。Network Meta-Analysis を行う手順は通常の Meta-Analysis と同様であり、例えば下記の手順で解析を行うこととなる。本稿では、下記③の I. ～ II. を SAS で行う方法を紹介する。

- ① 研究の背景と目的、試験の選択基準、解析手法や感度分析の方法を事前規定する  
(統合する試験は通常、無作為化試験とする)
- ② 統合する試験を選定し、バイアスが入るリスクの評価(割付手法、盲検性の維持、評価項目の欠測の取り扱い、公表バイアス等の評価)を行う(通常、独立した2名以上の評価者を設定)
- ③ Network Meta-Analysis を実施する
  - I. 試験結果を統合し、薬剤間の比較を行う
  - II. 均質性 (homogeneity) や一貫性 (consistency) の評価を行う
  - III. 感度分析、部分集団解析、メタ回帰等を行う
  - IV. 必要に応じて追加解析や個々の試験の再評価を行う
- ④ 結果をまとめ、報告する

この Network Meta-Analysis を適用することで、薬剤間の「直接比較」に加え、薬剤間の直接的な比較結果がデータに含まれていない場合でも、薬剤間の「間接比較」が出来るのが特徴である。また、薬剤間の「直接比較」と「間接比較」の結果を用いることで、エビデンスの質が強化される<sup>1</sup>。さらに、統合する試験は、薬剤数が3以上のもの (multi-arm study) があってもよく、右の様な「パス図」を用いて Network の状態を表すことが出来る。直線は薬剤間の「直接比較」があることを示し、直線の太さで「直接比較がある試験数」や「薬剤間差の標準誤差の逆数」等を表すことが出来る。

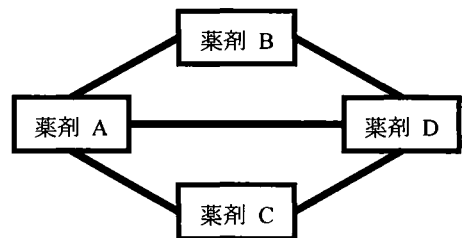


図 2.1 パス図の例

- 直接比較：薬剤 A vs. 薬剤 B、薬剤 A vs. 薬剤 C 等
- 間接比較：薬剤 B vs. 薬剤 C
- 直接比較 + 間接比較：薬剤 A vs. 薬剤 D 等

具体的な手法の説明の前に、下記の内容を定義する。Network Meta-Analysis の結果の解釈においては、一貫性 (consistency) の成立の有無が特に注目される点であることに留意する。

- 均質性 (homogeneity)
  - 同一の比較を行った試験 (≒同一のデザイン) 間での結果が類似していること
  - 反意語は異質性 (heterogeneity)
- 類似性 (transitivity, similarity とも呼ぶ)
  - 間接比較により、薬剤間差が適切に推定可能であること
    - Network Meta-Analysis を行う際は常に気にすべき問題
    - 間接比較の妥当性、つまり Network Meta-Analysis の結果の妥当性に直結する
  - 効果修飾因子 (effect modifier) の分布が試験間で類似している
  - 間接比較の結果の妥当性についてはデータからの確認が難しいことがあるため、あらかじめ試験選定時に留意する必要がある

<sup>1</sup> この点が Network Meta-Analysis を適用するメリットであるものの、このメリットを享受するためには類似性に関する強い仮定が必要であることに留意する。

- 一貫性 (consistency)
  - 直接比較の結果と間接比較の結果が一致していること
  - 一貫性 (consistency) は「直接比較」と「間接比較」の結果がある (ループがある) 時にのみ確認可能
  - 反意語は不一致性 (inconsistency)

### 3. データセット「Senn2013」

Senn et al. (2013) で用いられている糖尿病治療薬に関する臨床試験データ及び本データに関するパス図を紹介する (詳細は Senn et al. (2013) を参照)。試験数は 26 で全て無作為化試験、うち multi-arm study (3 群での比較試験) は 1 試験となっている。解析対象となるデータは Appendix 2 に示す。

- 薬剤は以下の 10 種類
  - 1 : **acarbose**、2 : **benfluorex**、3 : **metformin**、4 : **migliitol**、5 : **pioglitazone**、6 : **rosiglitazone**、
  - 7 : **sitagliptin**、8 : **sulfonylurea alone**、9 : **vildagliptin**、10 : **placebo** (以下、先頭 4 文字にて略記)
- 変数は以下の 5 変数
  - **STUDLAB** : 試験ラベル
  - **TREAT1** : 薬剤 1
  - **TREAT2** : 薬剤 2
  - **TE** : HbA1c 変化量 (%) の薬剤間差 (TREAT1 - TREAT2)
  - **SETE** : TE の標準誤差

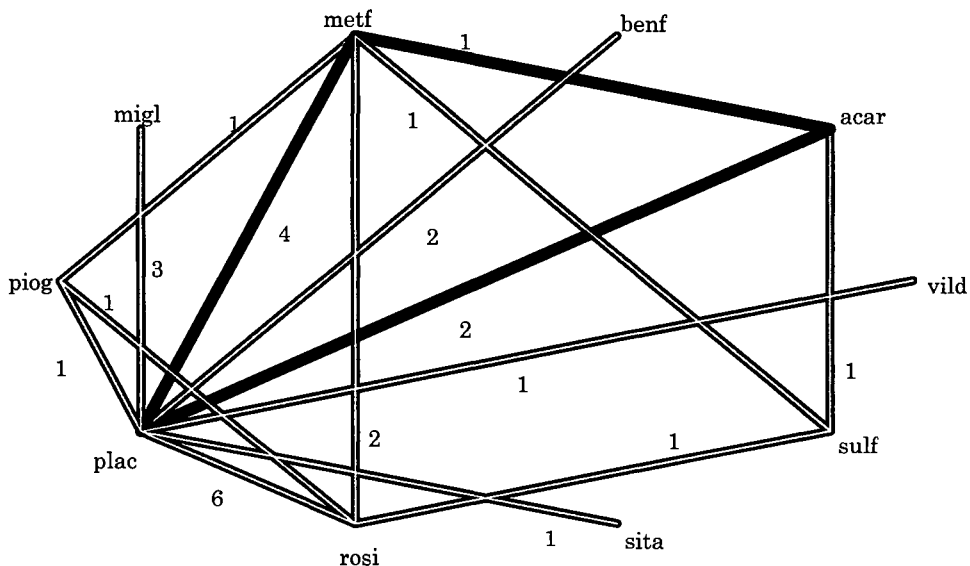


図 3.1 「Senn2013」のパス図 (太線 : multi-arm 試験を含むパス)

表 3.1 は、データセット「Senn2013」のうち、計算例を紹介するために 7 レコード (5 試験) のみ抽出したデータである。なお、表 3.1 の ID=5 の行は、分散共分散行列の計算にのみ使用する。また、抽出したデータセットに関するパス図は右図のようになり、「metf と plac の比較試験」は 3 試験、multi-arm の試験が 1 試験となっている。

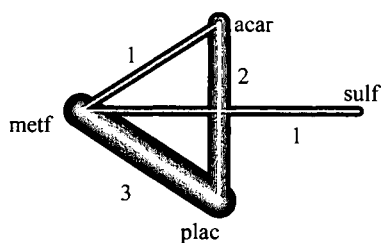


図 3.2 「Senn2013」(7 レコード抽出) のパス図

表 3.1 「Senn2013」のうち 7 レコードを抽出したデータ

ID	TE	SETE	TREAT1	TREAT2	STUDLAB	試験 s	デザイン d
1	-0.37	0.1184	metf	sulf	Alex1998	1	1
2	-0.80	0.1432	acar	plac	Costa1997	2	2
3	-1.90	0.1414	metf	plac	DeFronzo1995	3	3
4	-0.40	0.4356	metf	plac	Gonzalez-Ortiz2004	4	3
5	0.20	0.3579	acar	metf	Willms1999	5	4
6	-1.20	0.3758	metf	plac	Willms1999	5	4
7	-1.00	0.4669	acar	plac	Willms1999	5	4

## 4. 頻度論に基づく Network Meta-Analysis

### 4.1. 固定効果モデル

まず、以下の一般化線形モデルを考える。

$$Y = X\theta^{net} + \epsilon \quad (E[\epsilon] = 0, \text{Cov}[\epsilon] = V)$$

「Senn2013 (7 レコード抽出)」に適用する場合は以下となる。このとき、multi-arm study (Willms1999) に関する  $V_5$  については、対比較間の共分散を適切に考慮する必要がある点に留意する。

- $\theta^{net} = (\theta_{\text{metf-plac}}, \theta_{\text{acar-plac}}, \theta_{\text{sulf-plac}})'$  (plac を参照薬剤とする)
- $V = \text{diag}(V_1, \dots, V_4, V_5)$
- 試験  $s = 1, \dots, 5$ 、デザイン  $d = 1, \dots, 4$ 、 $S_d$ : デザイン  $d$  の試験の集合

$$Y = \begin{pmatrix} -0.37 \\ -0.80 \\ -1.90 \\ -0.40 \\ -1.20 \\ -1.00 \end{pmatrix} \begin{matrix} Y_1 \\ \vdots \\ Y_4 \\ Y_5 \end{matrix}, \quad X = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{matrix} \text{metf} & \text{acar} & \text{sulf} \end{matrix}, \quad V = \begin{pmatrix} 0.014 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.021 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.020 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.191 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.141 & 0.116 \\ 0 & 0 & 0 & 0 & 0.116 & 0.218 \end{pmatrix} \begin{matrix} V_1 \\ \dots \\ V_4 \\ V_5 \end{matrix}$$

Network Estimate  $\hat{\theta}^{net}$  を算出する手順は 1-step での推定と 2-step での推定の 2 つがある。

- **1-step での推定**

- Network Estimate:  $\hat{\theta}^{net} = (X'V^{-1}X)^{-1}X'V^{-1}Y = (-1.64, -0.88, -1.27)'$

• 2-step での推定

- $\hat{\theta}_d^{dir} = (\sum_{s \in S_d} V_s^{-1})^{-1} \sum_{s \in S_d} V_s^{-1} Y_s$ ,  $V_d^{dir} = \text{Cov}(\hat{\theta}_d^{dir}) = (\sum_{s \in S_d} V_s^{-1})^{-1}$
- Direct Estimate:  $\hat{\theta}^{dir} = (\hat{\theta}_1^{dir}, \dots, \hat{\theta}_3^{dir}, \hat{\theta}_4^{dir})' = X_a \theta^{net} + \epsilon_a$   
 $(E[\epsilon_a] = \mathbf{0}, \text{Cov}[\epsilon_a] = V_a, V_a = \text{diag}(V_1^{dir}, \dots, V_3^{dir}, V_4^{dir}))$
- Network Estimate:  $\hat{\theta}^{net} = (X_a' V_a^{-1} X_a)^{-1} X_a' V_a^{-1} \hat{\theta}^{dir} = (-1.64, -0.88, -1.27)'$

$$\hat{\theta}^{dir} = \begin{pmatrix} -0.37 \\ -0.80 \\ -1.76 \\ -1.20 \\ -1.00 \end{pmatrix} \begin{matrix} \hat{\theta}_1^{dir} \\ \vdots \\ \hat{\theta}_3^{dir} \\ \hat{\theta}_4^{dir} \end{matrix}, X_a = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{matrix} X_1 \\ \vdots \\ X_3 \\ -X_4 \end{matrix}, V_a = \begin{pmatrix} 0.014 & 0 & 0 & 0 & 0 \\ 0 & 0.021 & 0 & 0 & 0 \\ 0 & 0 & 0.018 & 0 & 0 \\ 0 & 0 & 0 & 0.141 & 0.116 \\ 0 & 0 & 0 & 0.116 & 0.218 \end{pmatrix}$$

いずれの手順でも、HbA1c 変化量 (%) に関する「metf と plac の薬剤間差」「acar と plac の薬剤間差」「sulf と plac の薬剤間差」はそれぞれ-1.64%、-0.88%、-1.27%と推定される。

4.2. 変量効果モデル

試験間の均質性 (homogeneity) を仮定しない変量効果モデルにて解析を行う場合は、以下の一般化線形モデルを考えればよい。すなわち、固定効果モデルに、変量効果  $b$  を追加すればよい<sup>2</sup>。

$$Y_s = X_s \theta^{net} + b_s + \epsilon_s \quad \left( E[\epsilon_s] = \mathbf{0}, \text{Cov}[\epsilon_s] = V_s, E[b_s] = \mathbf{0}, \text{Cov}[b_s] = \tau^2 \begin{pmatrix} 1 & \frac{1}{2} & \dots & \frac{1}{2} \\ \frac{1}{2} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1}{2} \\ \frac{1}{2} & \dots & \frac{1}{2} & 1 \end{pmatrix} \right)$$

「Senn2013 (7レコード抽出)」に適用する場合は以下となる。

- 試験  $s = 1, \dots, 5$ 、デザイン  $d = 1, \dots, 4$ ,  $S_d$ : デザイン  $d$  の試験の集合
- 例えば、 $\tau$  は DerSimonian and Laird の方法にて算出 ( $\tau = 0.583$ )
- $\text{Cov}(Y) = V_\tau$

$$V_\tau = \begin{pmatrix} 0.014 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.021 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.020 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.191 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.141 & 0.116 \\ 0 & 0 & 0 & 0 & 0.116 & 0.218 \end{pmatrix} + \tau^2 \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 1 \end{pmatrix} V_s$$

以降は  $V$  を  $V_\tau$  に置き換えて、固定効果モデルと同様の計算を行えばよい。

<sup>2</sup> multi-arm の試験の変量効果の分散共分散行列の非対角成分は  $\tau^2/2$  とする。そうすることで ID=5 (acar vs. metf) の変量効果の分散も  $\tau^2$  となる。

### 4.3. 均質性と一貫性の評価

Network Estimate  $\hat{\theta}^{net}$  が得られた後は、下記指標により均質性 (homogeneity) と一貫性 (consistency) が評価出来る。以下、 $D$  はデザインの数とする。

- $Q^{net} = (\mathbf{Y} - \mathbf{X}\hat{\theta}^{net})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\theta}^{net})$
- $Q^{net} = Q^{het} + Q^{inc}$ 
  - $Q^{het}$ : デザイン内の異質性 (heterogeneity) を評価するための指標 (within-designs Q statistic)  
 $\Rightarrow Q^{het} = \sum_{d=1}^D Q_d^{het}$ ,  $Q_d^{het} = \sum_{s \in S_d} (\mathbf{Y}_s - \hat{\theta}_d^{dir})' \mathbf{V}_s^{-1} (\mathbf{Y}_s - \hat{\theta}_d^{dir})$
  - $Q^{inc}$ : デザイン間の異質性 = 直接比較と間接比較の不一致性 (inconsistency) を評価するための指標 (between-designs Q statistic)  
 $\Rightarrow Q^{inc} = (\hat{\theta}^{dir} - \mathbf{X}_a \hat{\theta}^{net})' \mathbf{V}_a^{-1} (\hat{\theta}^{dir} - \mathbf{X}_a \hat{\theta}^{net})$   
 $\Rightarrow Q^{inc} = \sum_{d=1}^D Q_d^{inc}$ ,  $Q_d^{inc} = (\hat{\theta}_d^{dir} - \mathbf{X}_d \hat{\theta}^{net})' \mathbf{V}_d^{-1} (\hat{\theta}_d^{dir} - \mathbf{X}_d \hat{\theta}^{net})$

### 4.4. P スコアによる各薬剤の Ranking

Bayesian Network Meta-Analysis では Surface Under the Cumulative Ranking (SUCRA, Salanti G, Ades AE, and Ioannidis JPA (2011)) 等により薬剤間の順位付け (Ranking) を行うが、頻度論に基づく手法においても  $\hat{\theta}^{net}$  と  $\text{Cov}(\hat{\theta}^{net})$  から「P スコア」という指標を算出して各薬剤の Ranking を計算することが出来る。なお、この P スコアは SUCRA に対応するものとなる (Rucker G and Schwarzer G (2015))。例えば、metf に関する P スコアは以下の様に算出すればよい。

1. metf 以外の全ての薬剤との薬剤間差に関する片側 p 値 ( $p_{\text{metf},1}, \dots, p_{\text{metf},T-1}$ ) を算出する ( $T$ : 薬剤数)
  2. P スコア:  $P_{\text{metf}} = \frac{1}{T-1} \sum_i (1 - p_{\text{metf},i})$  を計算する
- 他の薬剤についても P スコアを算出し、この大小関係により Ranking を行う (P スコアが大きい方が良い)。

### 4.5. SAS を用いた解析例

以上の解析を SAS で実行する例として、固定効果モデルに関する iml プロシジャでの解析例を紹介する。

プログラム 4.1 「Senn2013」(7 レコード抽出) に対する解析

```
proc iml ;
  y = {-0.37, -0.8, -1.9, -0.4, -1.2, -1} ;
  X = {1 0 -1, 0 1 0, 1 0 0, 1 0 0, 1 0 0, 0 1 0} ;
  V = {0.01401856 0 0 0 0 0,
       0 0.02050624 0 0 0 0,
       0 0 0.01999396 0 0 0,
       0 0 0 0.19053225 0 0,
       0 0 0 0 0.14122564 0.11556442,
       0 0 0 0 0.11556442 0.21799561} ;

  theta_net = inv(t(X) * inv(V) * X) * t(X) * inv(V) * y ;
  v_theta_net = inv(t(X) * inv(V) * X) ;
  result1 = theta_net || (theta_net-probit(0.975)#sqrt(vecdiag(v_theta_net)))
             || (theta_net+probit(0.975)#sqrt(vecdiag(v_theta_net))) ;

  theta_dir = j(5, 1, 0) ; Va = j(5, 5, 0) ; Xa = X[1 2 3 5 6,] ;
  theta_dir[1,1] = sum(inv(V[1,1]) * y[1]) # V[1,1] ; Va[1,1] = V[1,1] ;
  theta_dir[2,1] = sum(inv(v[2,2]) * y[2]) # V[2,2] ; Va[2,2] = V[2,2] ;
  theta_dir[3,1] = sum(inv(V[{3 4}], {3,4}) * y[{3 4}]) / sum(vecdiag(inv(v[{3 4}], {3 4}))) ;
  Va[3,3] = 1/sum(vecdiag(inv(V[{3 4}], {3 4}))) ;
  theta_dir[4 5,1] = y[5 6] ; Va[4 5,4 5] = V[5 6,5 6] ;
```

```

theta_net_a = inv(t(Xa) * inv(Va) * Xa) * t(Xa) * inv(Va) * theta_dir ;
v_theta_net_a = inv(t(Xa) * inv(Va) * Xa) ;
result2      = theta_net_a || (theta_net_a+probit(0.975)#sqrt(vecdiag(v_theta_net_a)))
              || (theta_net_a-probit(0.975)#sqrt(vecdiag(v_theta_net_a))) ;

Q_net = t(y - X * theta_net) * inv(V) * (y-X * theta_net) ;
Q_inc = t(theta_dir - Xa * theta_net_a) * inv(Va) * (theta_dir - Xa * theta_net_a) ;
Q_het = Q_net - Q_inc ;
CC = -1#i(3) ; CC[,1] = 1 ; E = CC * theta_net ; z = j(1,3,0) ;
do i=1 to 3 ;
  c = CC[i,] ;
  z[i] = sqrt(t(c * theta_net) * inv(c * v_theta_net * t(c)) * (c * theta_net)) ;
  if (E[i] < 0) then z[i] = -1*z[i] ;
end ;
p_metf = 1-(probnorm(z)) ; p_score_metf = p_metf[:] ;
title "Treatment estimate (Fixed effect model)" ;
print result1 result2, Q_net Q_het Q_inc p_score_metf ;
quit ;

```

実行結果は以下の通り。行列 result1 は 1-step で計算した  $\theta^{net}$  の推定値（1 列目）とその両側 95%信頼区間（2~3 列目）、行列 result2 は 2-step で計算した  $\theta^{net}$  に関する推定結果となっている。また、変数 Q\_net、Q\_het 及び Q\_inc はそれぞれ  $Q^{net}$ 、 $Q^{het}$  及び  $Q^{inc}$ 、変数 p\_score\_metf は metf の P スコアとなっている。

出力 4.1 「Senn2013」（7 レコード抽出）に対する解析結果  
Treatment estimate (Fixed effect model)

result1			result2		
-1.64293	-1.882384	-1.403476	-1.64293	-1.882384	-1.403476
-0.880131	-1.141532	-0.61873	-0.880131	-1.141532	-0.61873
-1.27293	-1.606381	-0.939478	-1.27293	-1.606381	-0.939478
Q_net	Q_het	Q_inc	p_score_metf		
15.000466	10.687505	4.3129608	0.9997023		

#### 4.6. Detaching a single design を用いた不一致性の原因調査とヒートマップ

均質性 (homogeneity) と一致性 (consistency) を評価した際、 $Q^{inc}$  が大きく、不一致性 (inconsistency) が示唆された場合、不一致性 (inconsistency) が生じている原因を調べる必要がある。ここでは Krahn (2013) の「Detaching a single design」の方法を用いて、各デザイン  $d$  の影響を計算するために  $Q$  を再計算した後、「デザイン  $d$  を除いた (detach した) ときのデザイン  $d'$  の  $Q_{d'}^{inc}$  の変化」を計算する。

1.  $\theta_{(d)}^{net} = (\theta^{net'}, \theta_d^{dir-ind'})'$  とする
  - $\theta_d^{dir-ind}$  : デザイン  $d$  に関する「デザイン  $d$  の薬剤数-1」個の新たなパラメータ (デザイン  $d$  における直接比較と間接比較の結果の差を意味する)
2. モデル  $\hat{\theta}^{dir} = (X_a, \mathbf{1}_{\{d\}})\theta_{(d)}^{net} + \epsilon$  ( $E[\epsilon] = \mathbf{0}$ ,  $Cov[\epsilon] = V_a$ ) を考える
  - $\mathbf{1}_{\{d\}}$  : デザイン  $d$  の当該対比較が 1、それ以外は 0 となっているベクトル又は行列
3. デザイン  $d$  を除いたときのモデルの不一致性 (inconsistency) は以下の統計量で表現出来る

$$Q_{(d)}^{inc} = R'_{(d)} V_a^{-1} R_{(d)} \quad (R_{(d)} = \hat{\theta}^{dir} - (X_a, \mathbf{1}_{\{d\}})\hat{\theta}_{(d)}^{net})$$



4. 「デザイン  $d$  を除いたときのデザイン  $d'$  の  $Q_{d',d}^{inc}$  の変化」を以下で定義する

- $Q_{d',d}^{diff} = Q_{d'}^{inc} - Q_{d'(d)}^{inc}$
- $Q_{d'}^{inc} = (\hat{\theta}_{d'}^{dir} - X_{d'} \hat{\theta}^{net})' V_{d'}^{-1} (\hat{\theta}_{d'}^{dir} - X_{d'} \hat{\theta}^{net})$
- $Q_{d'(d)}^{inc} = R_{d'(d)}' V_{d'(d)}^{-1} R_{d'(d)}$  ( $R_{d'(d)} = \hat{\theta}_{d'}^{dir} - (X_a, \mathbf{1}_{(d)})_{d'} \hat{\theta}^{net}$ )  
 -  $(X_a, \mathbf{1}_{(d)})_{d'} : (X_a, \mathbf{1}_{(d)})$  のうちデザイン  $d'$  の部分を取り出した行列
- ちなみに、 $Q^{inc} = \sum_{d'=1}^D Q_{d'}^{inc}$  となる

$Q^{inc}$  の値が大きい場合、すなわち不一致性 (inconsistency) が示唆された場合、 $Q_{d',d}^{diff}$  を利用し、例えば下図のヒートマップを用いて、どのデザインが不一致性 (inconsistency) の原因なのかを調査することが出来る。本稿はグレースケールであるため分かりにくいですが、実際のヒートマップは「寒色系」と「暖色系」による図示となる。

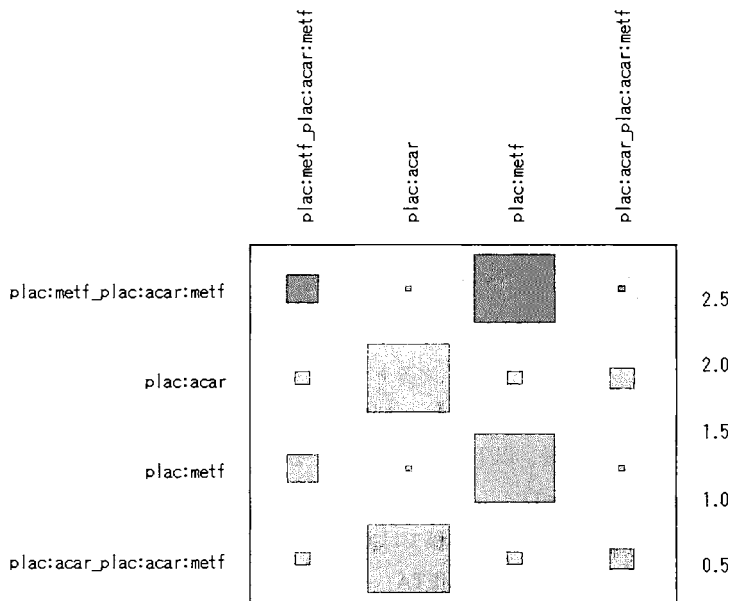


図 4.1 「Senn2013」 (7レコード抽出) に関するヒートマップ

### ■の大きさ

ハット行列  $H = X_a(X_a'V_a^{-1}X_a)^{-1}X_a'V_a^{-1}$  にて計算され、各デザインの direct estimate  $\hat{\theta}_d^{dir}$  の network estimate への寄与の度合いを表す。すなわち、network estimate が「直接比較」と「間接比較」をどのように組み合わせて推定されているかを表現する。

$$X_a \hat{\theta}^{net} = H \hat{\theta}^{dir}$$

### 対角成分のセルの色

- 各デザインの  $Q_{d',d}^{inc}$  を表す ( $Q_{d'(d)}^{inc} = 0$  であるため暖色系しか発生しない)

### 非対角成分のセルの色

- 暖色系:  $Q_{d',d}^{diff} > 0$ ; 列のデザイン  $d$  を除いたときの行のデザイン  $d'$  に関する  $Q_{d',d}^{inc}$  が減る (色の濃さ=度合い、inconsistent evidence であることを示す)
- 寒色系:  $Q_{d',d}^{diff} < 0$ ; 列のデザイン  $d$  を除いたときの行のデザイン  $d'$  に関する  $Q_{d',d}^{inc}$  が増す (色の濃さ=度合い、supportive evidence であることを示す)

図 4.1 では、multi-arm 試験における「metf vs. plac」の結果に関して  $Q_d^{inc}$  が大きく（暖色で色が濃い）、不一致性（inconsistency）が示唆される。次に、■の大きさを見ると、「metf vs. plac」の network estimate に対する direct estimate の寄与が大きいのは、「metf vs. plac」が含まれる 3 試験だが、特に 2-arm 試験の寄与が大きい。この 2 つの情報を組み合わせると、「metf vs. plac」が含まれる 3 試験間の何らかの差異によって不一致性（inconsistency）が生じた示唆されることから、これらの試験に関してさらなる調査を行う必要がある。ヒートマップの見方の詳細は Krahn (2013) を参照されたい。

## 5. SAS での Network Meta-Analysis の実施

さて、Network Meta-Analysis を実行する SAS プログラムの構築を試みる。まず、固定効果モデルと変量効果モデルによる  $\theta^{net}$  の推定値とその分散、 $Q^{net}$ 、 $Q^{het}$  及び  $Q^{inc}$ 、各薬剤の ranking 等を計算する iml プロシジャによるプログラムを作成した（SAS プレミアムラウンジの本ユーザー総会サイトにて公開予定）。しかし、この段階でプログラムが約 500 行となり、 $Q^{het}$  及び  $Q^{inc}$  の分解を行おうとすると行数が 2~3 倍に膨れる恐れがあり、ここでプログラム作成を打ち切った。ところで、フリーで使うことが出来る統計ソフト「R」には「netmeta」パッケージがあり、本パッケージの関数を用いることで頻度論による Network Meta-Analysis を実行することが出来る。本項では SAS から R を呼び出して Network Meta-Analysis を実行する方法を紹介する。

### 若干の準備

- ① 以下のサイトより R をダウンロードし、インストールする（Ver. 3.2.5 を想定）  
<http://ftp.yz.yamagata-u.ac.jp/pub/cran/bin/windows/base/>
- ② R を「管理者権限」で起動した後、以下の命令を実行する  

```
install.packages("netmeta", dep=T)
```
- ③ C ドライブに作業フォルダを準備する（本稿では C ドライブの直下に「temp」フォルダを作成）
- ④ Appendix 2 で紹介するマクロ「%MYNETMETA」を定義し、実行する。

プログラム 5.1 マクロ「%MYNETMETA」の引数

---

%macro MYNETMETA(dataset	= ,
sm	= MD,
level	= 0.95,
reference	= ,
seq	= ,
small	= good,
path	= C:/temp) ;

---

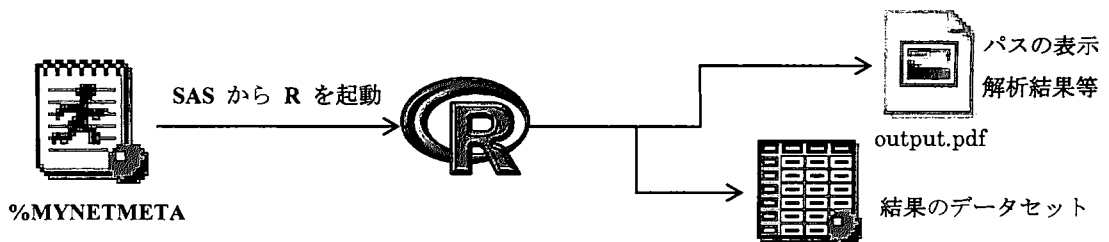


表 5.1 マクロ「%MYNETMETA」の引数

引数名	機能
dataset	SAS データセット名を指定 (変数名は TE、SETE、TREAT1、TREAT2、STUDLAB で固定)
sm	TE の指標を指定 (RD : Risk Difference、RR : Risk Ratio、OR : Odds Ratio、MD : Mean Difference、SMD : Standardized Mean Difference、IRR : Incidence Rate Ratio、IRD : Incidence Rate Difference 等)
level	信頼係数を指定
reference	参照薬剤を指定する場合に指定 (例 : reference=plac)
seq	薬剤の順番を指定
small	指標について、小さい方が改善 (good) 又は悪化 (bad) のいずれかを指定
path	作業ディレクトリを指定 (出力結果の output.pdf や中間ファイルの格納場所)

データセット「Senn2013\_char (Appendix 2 のプログラム A.2 参照)」に対して、マクロ「%MYNETMETA」を用いて Network Meta-Analysis を実行する場合は以下の様にすればよい。

プログラム 5.2 「Senn2013\_char」に対する Network Meta-Analysis の実行例①

```
%MYNETMETA(dataset = Senn2013_char,
             sm      = MD,
             level  = 0.95) ;
```

また、引数 reference に参照薬剤を指定し (以下では plac を指定)、引数 seq に薬剤の順番を指定する場合は以下の様にすればよい<sup>3</sup>。

プログラム 5.3 「Senn2013\_char」に対する Network Meta-Analysis の実行例②

```
%MYNETMETA(dataset = Senn2013_char,
             sm      = MD,
             level  = 0.95,
             reference = plac,
             seq    = %str('plac', 'acar', 'benf', 'metf', 'mig1', 'piog', 'rosi', 'sita',
                          'sulf', 'vild')) ;
```

実行結果として、作業フォルダ (例えば「C:\temp」) に解析結果となる output.pdf が出力される。まず、pdf ファイルの 1~2 頁にて固定効果モデルと変量効果モデルによる薬剤間差の推定結果 (Network Estimate:  $\hat{\theta}^{net}$ ) に関する Forest Plot が示される。なお、両モデルの推定結果のうち点推定値は同様であるが、信頼区間は変量効果モデルの方が広がっている。

<sup>3</sup> 引数 seq に順番を指定する際、%str() 中で指定する内容を途中で改行しないこと (エラーとなる)。

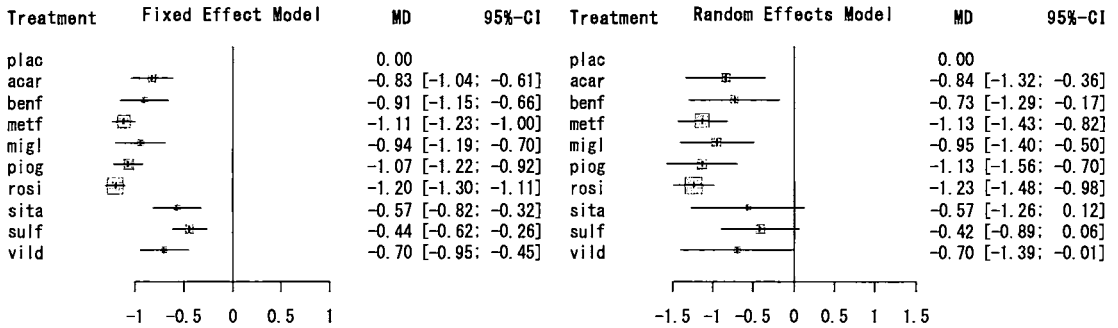


図 5.1 固定効果モデルの解析結果

図 5.2 変量効果モデルの解析結果

次に、pdf ファイルの 3~4 頁にて固定効果モデルと変量効果モデルに関するパス図が示される。パスの線の太さは、各モデルにて算出した「薬剤間差の推定値の標準誤差の逆数」を表している。

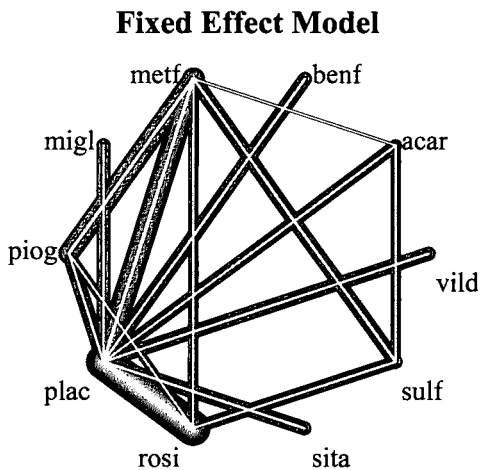


図 5.3 固定効果モデルに関するパス図

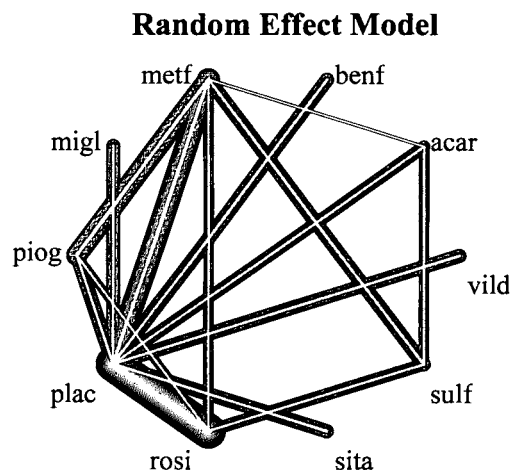


図 5.4 変量効果モデルに関するパス図

最後に、pdf ファイルの 5~6 頁にて、固定効果モデルと変量効果モデルにおいて、不一致性 (inconsistency) を評価するためのヒートマップが示される。固定効果モデルではいくつかのデザインにおいて不一致性 (inconsistency) が示唆されているが、変量効果モデルではその度合いが薄れていることが伺える。

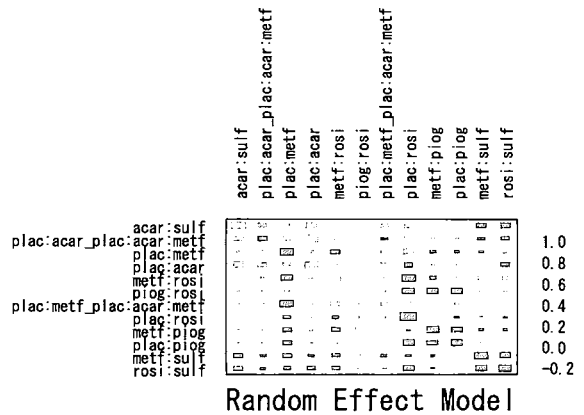
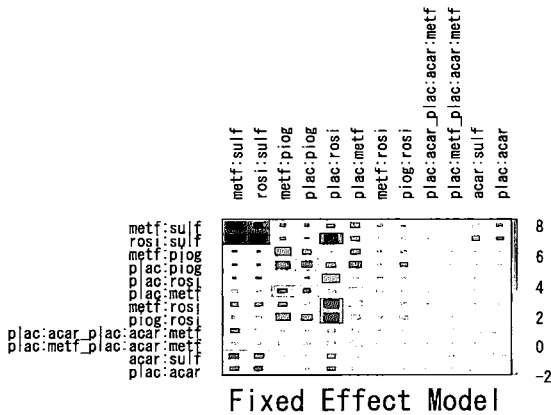


図 5.5 固定効果モデルに関するヒートマップ

図 5.6 変量効果モデルに関するヒートマップ

他にも、SAS 上に以下のデータセットが生成される。

- TE\_FIXED、TE\_RANDOM : 薬剤間差の推定値 (前者:固定効果モデル、後者:変量効果モデル、以下同様)
- TE\_FIXED\_LCL、TE\_RANDOM\_LCL : 薬剤間差の推定値の信頼下限
- TE\_FIXED\_UCL、TE\_RANDOM\_UCL : 薬剤間差の推定値の信頼上限
- Q1\_STATISTICS、Q2\_STATISTICS : Q 統計量とその分解結果 ( $Q^{het}$ : homogeneity /  $Q^{inc}$ : consistency)
- Q1\_WITHINDESIGNS、Q2\_WITHINDESIGNS :  $Q^{het}$  の分解結果 (Design-specific decomposition of within-designs Q statistic)
- Q1\_BETWEENDESIGNS、Q2\_BETWEENDESIGNS :  $Q^{inc}_{(d)}$  の結果 (Between-designs Q statistic after detaching of single designs)
- RANK\_FIXED、RANK\_RANDOM : Network 内における各薬剤の P スコア

Q statistics to assess homogeneity / consistency			
	Q	df	pval
Whole network	96.985553005	18	7.871481E-13
Within designs	74.455281736	11	1.729943E-11
Between designs	22.53027127	7	0.0020570142

Design-specific decomposition of within-designs Q statistic			
design	Q	df	pval
acar:plac	0	0	NA
acar:sulf	0	0	NA
benf:plac	4.3813771308	1	0.036939630425004
metf:piog	0	0	NA
metf:plac	42.164314083	2	6.98450297598185E-10
metf:rosi	0.1869507963	1	0.665466653458945
metf:sulf	0	0	NA
nigl:plac	6.4492777799	2	0.0399770140380112
piog:plac	0	0	NA
piog:rosi	0	0	NA
plac:rosi	21.273361946	5	0.000719161596152484
plac:sita	0	0	NA
plac:wild	0	0	NA
rosi:sulf	0	0	NA
acar:metf:plac	0	0	NA

Between-designs Q statistic after detaching of single designs			
design	Q	df	pval
acar:plac	22.444462972	6	0.0010055771
acar:sulf	22.517140236	6	0.000975428
metf:piog	17.12869346	6	0.008219225
metf:plac	22.069517381	6	0.0011762461
metf:rosi	22.52029252	6	0.0009741406
metf:sulf	7.5134869505	6	0.275955102
piog:plac	17.246842419	6	0.0084176157
piog:rosi	22.483554439	6	0.009892487
plac:rosi	16.293710202	6	0.0122738101
rosi:sulf	6.7716569389	6	0.3424815166
acar:metf:plac	22.384427651	5	0.0004424241

出力 5.1 固定効果モデルでの Q 統計量と分解結果

Q statistics to assess homogeneity / consistency (Random)			
	Q	df	pval
Whole network	16.367703819	18	0.566901947
Within designs	13.533882217	11	0.2598663221
Between designs	2.8338216013	7	0.8999377058

Design-specific decomposition of within-designs Q statistic (Random)			
design	Q	df	pval
acar:plac	0	0	NA
acar:sulf	0	0	NA
benf:plac	1.7075771528	1	0.191300024523713
metf:piog	0	0	NA
metf:plac	7.159046125	2	0.0278889963923192
metf:rosi	0.0608177632	1	0.805208070055138
metf:sulf	0	0	NA
nigl:plac	2.0293285373	2	0.362524120110668
piog:plac	0	0	NA
piog:rosi	0	0	NA
plac:rosi	2.5771126393	5	0.764839597544153
plac:sita	0	0	NA
plac:wild	0	0	NA
rosi:sulf	0	0	NA
acar:metf:plac	0	0	NA

Between-designs Q statistic after detaching of single designs (Random)			
design	Q	df	pval
acar:plac	2.8084423245	6	0.8324767107
acar:sulf	2.8235343867	6	0.8306396172
metf:piog	2.4237916465	6	0.8763503442
metf:plac	2.7410481737	6	0.840575178
metf:rosi	2.8099156107	6	0.8322983412
metf:sulf	1.1162771459	6	0.9808180416
piog:plac	2.4522678877	6	0.8737709355
piog:rosi	2.8336147362	6	0.8234215162
plac:rosi	2.2848077505	6	0.8917310851
rosi:sulf	0.8260294783	6	0.9913575108
acar:metf:plac	2.7958829014	5	0.7406250738

出力 5.2 変量効果モデルでの Q 統計量と分解結果

	×		×
acar	0.4791835536	acar	0.5202644159
benf	0.5726641851	benf	0.4358170235
metf	0.8512771649	metf	0.7817765316
migl	0.6200044122	migl	0.6136555282
piog	0.7686325656	piog	0.7745697307
plac	6.4907432E-7	plac	0.0138950538
rosi	0.9789032445	rosi	0.8933681813
sita	0.2386146419	sita	0.3330992064
sulf	0.1394916587	sulf	0.2103076597
vild	0.3512279244	vild	0.4232466695

出力 5.3 固定効果モデルでの各薬剤の P スコア

出力 5.4 変量効果モデルでの各薬剤の P スコア

ところで、手持ちのデータが「Senn2013\_char」の様な薬剤間差に関する形式 (contrast-based format) ではなく、各薬剤の推定値に関する形式 (arm-based format) である場合は、事前にデータの変換を行う必要がある。本稿では、各薬剤の推定値に関する形式 (arm-based format) のデータを薬剤間差に関する形式 (contrast-based format) に変換するマクロ「%MYCONVERT」の定義を Appendix 2 にて紹介する。頁数の制限のため、引数等の詳細な説明はユーザー会当日の発表にて紹介する<sup>4</sup>。

## 6. Bayesian Network Meta-Analysis の結果との比較

本総会では既に、SAS を用いた Bayesian Network Meta-Analysis の方法が提案されている (福井他 (2014))。よって本稿では、Bayesian Network Meta-Analysis の解析結果と頻度論に基づく Network Meta-Analysis の解析結果を比較するにとどめる。

### 6.1. 固定効果モデル

まず、固定効果モデルに関する Bayesian Network Meta-Analysis を実施する。

- パラメータと事前分布

- $\theta_i \sim N(0, 10000)$  ( $i = 1, \dots, 9$ )
  - ◇  $\theta_1$ : acar と plac の薬剤間差、 $\dots$ 、 $\theta_9$ : vild と plac 薬剤間差
  - ◇  $\theta_{10}$ : 便宜上 0 とおき、plac の効果を表す (推定は行わない)

- モデル

- 2-arm study:  $y_{ij} \sim N(\theta_i - \theta_j, \sigma_{ij}^2)$  ( $i, j = 1, \dots, 10$ ;  $\sigma_{ij}^2$ :  $y_{ij}$  の分散)
- 3-arm study:  $\begin{pmatrix} y_{ij} \\ y_{ik} \end{pmatrix} \sim N \left( \begin{pmatrix} \theta_i - \theta_j \\ \theta_i - \theta_k \end{pmatrix}, \begin{pmatrix} \sigma_{ij}^2 & (\sigma_{ij}^2 + \sigma_{ik}^2 - \sigma_{jk}^2)/2 \\ (\sigma_{ij}^2 + \sigma_{ik}^2 - \sigma_{jk}^2)/2 & \sigma_{ik}^2 \end{pmatrix} \right)$

解析対象となるデータ「Senn」は Appendix 2 にて、解析を行う SAS プログラムは以下に示す。

プログラム 6.1 Bayesian Network Meta-Analysis (固定効果モデル)

```
ods graphics on ;
proc mcmc data=Senn nbj=5000 nmc=500000 thin=10 seed=777 missing=ac diagnostics=all plots=all
monitor=(theta) ;
```

<sup>4</sup> データ「Parkinson」の出典: Dias S, Sutton AJ, Ades AE and Welton NJ (2013)。

```

array te[2] te1 te2 ;
array theta[9] ;
array s[2,2] ;
array mu[2] mu1 mu2 ;
parms theta: 0 ;
prior theta: ~ normal(0, var=10000) ;

if studlab=3 then do ;
  do i=1 to 9 ;
    if treat1_1=i then mu1_1=theta[i] ;
    if treat2_1=i then mu2_1=theta[i] ;
    if treat1_2=i then mu1_2=theta[i] ;
    if treat2_2=i then mu2_2=theta[i] ;
  end ;
  if treat1_1=10 then mu1_1=0 ;
  if treat2_1=10 then mu2_1=0 ;
  if treat1_2=10 then mu1_2=0 ;
  if treat2_2=10 then mu2_2=0 ;

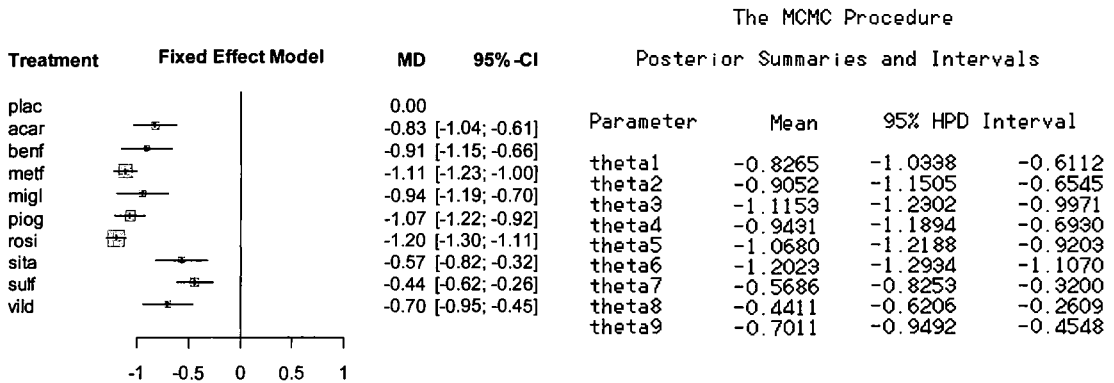
  mu[1]=mu1_1-mu2_1 ; mu[2]=mu1_2-mu2_2 ;
  s[1,1]=v1 ; s[2,2]=v2 ; s[1,2]=(v1+v2-v3)/2 ; s[2,1]=s[1,2] ;
  ll=lpdfmvn(te,mu,s) ;
end ;
else do ;
  do i=1 to 9 ;
    if treat1_1=i then mu1_1=theta[i] ;
    if treat2_1=i then mu2_1=theta[i] ;
  end ;
  if treat1_1=10 then mu1_1=0 ;
  if treat2_1=10 then mu2_1=0 ;

  mu[1]=mu1_1-mu2_1 ;
  ll=lpdfnorm(te[1],mu[1],sqrt(v1)) ;
end ;

model general(ll) ;
run ;
ods graphics off ;

```

結果は以下となり、頻度論での解析結果と Bayesian Network Meta-Analysis の結果は同様である。



出力 6.1 固定効果モデルの解析結果 (頻度論)

出力 6.2 固定効果モデルの解析結果 (ベイズ)

## 6.2. 変量効果モデル

次に、変量効果モデルに関する Bayesian Network Meta-Analysis を実施する。

### • パラメータと事前分布

- $\theta_i \sim N(0, 10000)$  ( $i = 1, \dots, 9$ )
  - ◇  $\theta_1$ : acar と plac の薬剤間差、 $\dots$ 、 $\theta_9$ : vild と plac の薬剤間差
  - ◇  $\theta_{10}$ : 便宜上 0 とおき、plac の効果を表す (推定は行わない)
- $V_h \sim \text{igamma}(1, 0.000001)$

### • モデル

- 2-arm study:  $y_{ij} \sim N(\theta_i - \theta_j, \sigma_{ij}^2 + V_h)$  ( $i, j = 1, \dots, 10$ ;  $\sigma_{ij}^2$ :  $y_{ij}$  の分散)
- 3-arm study:  $\begin{pmatrix} y_{ij} \\ y_{ik} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_i - \theta_j \\ \theta_i - \theta_k \end{pmatrix}, \Sigma\right)$

$$\Sigma = \begin{pmatrix} \sigma_{ij}^2 & (\sigma_{ij}^2 + \sigma_{ik}^2 - \sigma_{jk}^2)/2 \\ (\sigma_{ij}^2 + \sigma_{ik}^2 - \sigma_{jk}^2)/2 & \sigma_{ik}^2 \end{pmatrix} + V_h \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$$

解析対象となるデータ「Senn」は Appendix 2 にて、解析を行う SAS プログラムは以下に示す。

プログラム 6.2 Bayesian Network Meta-Analysis (変量効果モデル、下線部は固定効果モデルとの相違部分)

---

```
ods graphics on ;
proc mcmc data=Senn nbi=10000 nmc=5000000 thin=50 seed=777 missing=ac diagnostics=all plots=all
  stats(percent=(2.5 97.5))=all monitor=(theta var_h sd) ;
  array te[2] te1 te2 ;
  array theta[9] ;
  array s[2,2] ;
  array g[2,2] ;
  array mu[2] mu1 mu2 ;
  array delta[2] delta1 delta2 ;
  parms theta: 0 ;
  parms var_h 1;

  prior theta: ~ normal(0, var=10000) ;
  prior var_h ~ igamma(1, scale=0.000001) ;

  if studlab=3 then do ;
    do i=1 to 9 ;
      if treat1_1=i then mu1_1=theta[i] ;
      if treat2_1=i then mu2_1=theta[i] ;
      if treat1_2=i then mu1_2=theta[i] ;
      if treat2_2=i then mu2_2=theta[i] ;
    end ;
    if treat1_1=10 then mu1_1=0 ;
    if treat2_1=10 then mu2_1=0 ;
    if treat1_2=10 then mu1_2=0 ;
    if treat2_2=10 then mu2_2=0 ;

    mu[1]=mu1_1-mu2_1 ; mu[2]=mu1_2-mu2_2 ;
    s[1,1]=v1 ; s[2,2]=v2 ; s[1,2]=(v1+v2-v3)/2 ; s[2,1]=s[1,2] ;
    g[1,1]=var_h ; g[2,2]=g[1,1] ; g[1,2]=var_h/2 ; g[2,1]=g[1,2] ;
    random delta ~ mvn(mu, g) subject=_obs_ ;
    ll=lpdfmvn(te, delta, s) ;
  end ;
```

---



```

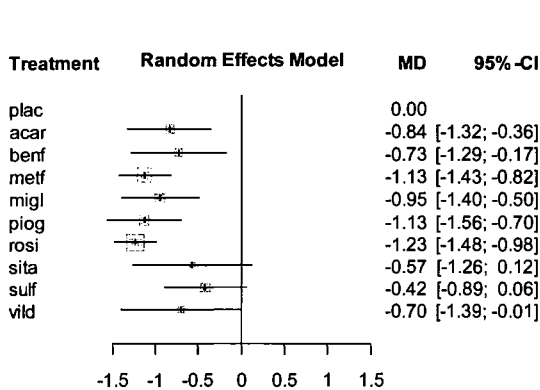
else do ;
  do i=1 to 9 ;
    if treat1_1=i then mu1_1=theta[i] ;
    if treat2_1=i then mu2_1=theta[i] ;
  end ;
  if treat1_1=10 then mu1_1=0 ;
  if treat2_1=10 then mu2_1=0 ;

  mu[1]=mu1_1-mu2_1 ;
  vt=sqrt(v1) ;
  random delta3 ~ normal(mu[1],v=var_h) subject=obs_ ;
  ll=lpdfnorm(te[1],delta3,vt) ;
end ;

model general(ll) ;
sd=sqrt(var_h) ;
run ;
ods graphics off ;

```

結果は以下となり、固定効果モデルと同様、頻度論での解析結果と Bayesian Network Meta-Analysis の結果は同様である。



出力 6.3 変量効果モデルの解析結果 (頻度論)

The MCMC Procedure

Posterior Summaries

Parameter	Mean	Percentiles	
		2.5	97.5
theta1	-0.8417	-1.3152	-0.3694
theta2	-0.7436	-1.2851	-0.1872
theta3	-1.1300	-1.4189	-0.8312
theta4	-0.9522	-1.3996	-0.4996
theta5	-1.1303	-1.5504	-0.7066
theta6	-1.2314	-1.4746	-0.9863
theta7	-0.5717	-1.2613	0.1051
theta8	-0.4169	-0.8756	0.0501
theta9	-0.6997	-1.3754	-0.0104
var h	0.0997	0.0383	0.2187
sd	0.3081	0.1958	0.4677

出力 6.4 変量効果モデルの解析結果 (ベイズ)

## 7. まとめ

本稿では、頻度論の枠組みでの解析手法による Network Meta-Analysis の実施方法として、SAS を利用した薬剤間差の推定量の算出方法、均質性 (homogeneity) や一貫性 (consistency) の評価方法、及び各薬剤の Ranking の算出方法を紹介した。本稿では平均値の差に関するデータに対して解析を行ったが、本稿で紹介した SAS マクロを用いることで、リスク差やリスク比、発現率の差や比に関するデータについても解析を行うことが出来る。参考までに、Bayesian Network Meta-Analysis の結果との比較を行い、薬剤間差の推定結果が同様であることも示した。ちなみに、Bayesian Network Meta-Analysis は WinBUGS にて実施されることが多いが、WinBUGS のプログラムの性質上、汎用的なプログラムを作成することが難しい。R では Bayesian Network Meta-Analysis を行う汎用関数が含まれるパッケージ「gemtc」があり、これを用いることで大抵のデータ形式に対して簡単に Bayesian Network Meta-Analysis を行うことが出来る。本稿で紹介した方法によりパッケージ

「gemtc」と SAS を簡単に組み合わせることが可能で、SAS でも Bayesian Network Meta-Analysis を簡単に実施することが出来ることを申し添える。

連絡先

[nobuo.funao@takeda.com](mailto:nobuo.funao@takeda.com)

参考文献

- **Caldwell DM (2014)**  
An overview of conducting systematic reviews with network meta-analysis. *Systematic Reviews*, 3:109.
- **Dias S, Sutton AJ, Ades AE and Welton NJ (2013)**  
Evidence synthesis for decision making 2: A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making*, 33, 607-617.
- **Higgins JPT, Jackson D, Barrett JK, Lu G, et. al. (2012)**  
Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research Synthesis Methods*, 3(2), 98-110.
- **Hutton B, Salanti G, Caldwell DM, Chaimani A, et. al. (2015)**  
The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of Internal Medicine*, 162(11):777-784.
- **Krahn U, Binder H and Konig J (2013)**  
A graphical tool for locating inconsistency in network meta analyses. *BMC Medical Research Methodology*, 13:35.
- **Mills EJ, Thorlund K and Ioannidis JP (2013)**  
Demystifying trial networks and network meta-analysis. *BMJ*. May14; 346:f2914.
- **DerSimonian R and Laird N (1986)**  
Meta-analysis in clinical trials. *Controlled Clinical Trials* 7:177-188.
- **Rucker G, Schwarzer G, Krahn U and Konig J (2014)**  
Netmeta: network meta-analysis with R. R package (version 0.8-0).
- **Rucker G and Schwarzer G (2015)**  
Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Medical Research Methodology*, 15:58.
- **Salanti G, Ades AE, and Ioannidis JPA (2011)**  
Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* 64, 163-171.
- **Senn S, Gavini F, Magrez D, and Scheen A (2013)**  
Issues in performing a network meta-analysis. *Statistical Methods in Medical Research*, 22 (2), 169-189.
- **福井 伸行, 乙黒 俊也, 磯崎 充宏 (2014)**  
ネットワークメタアナリシスによる 無作為化比較試験の統合. SAS ユーザー総会

以上

## Appendix 1. Q 統計量の自由度

均質性 (homogeneity) や一致性 (consistency) を評価する際に算出した  $Q^{net}$  や  $Q^{het}$ 、 $Q^{inc}$  等は全て  $\chi^2$  分布に従う。以下に各 Q 統計量に従う  $\chi^2$  分布の自由度を示す。

表 A.1 各 Q 統計量の自由度

統計量	自由度
$Q^{net}$	$\sum_{d=1}^D \sum_{s \in S_d} (N_s - 1) - (T - 1)$
$Q^{het}$	$\sum_{d=1}^D (Q_d^{het} \text{ の自由度})$
$Q_d^{het}$	$\sum_{s \in S_d} (N_s - 1) - (N_d - 1)$
$Q^{inc}$	$\sum_{d=1}^D (N_d - 1) - (T - 1)$
$Q_{(d)}^{inc}$	$(Q^{inc} \text{ の自由度}) - (N_d - 1)$

$T$  : ネットワーク内の薬剤数、 $N_s$  : 試験  $s$  の薬剤数、 $N_d$  : デザイン  $d$  の薬剤数

## Appendix 2. 本文に掲載していないが引用したプログラム

Rucker et. al. (2014) による R パッケージ「netmeta」は、License は GPL-2+ (プログラムの調査及び改変、公衆への頒布が可) に従っており、本稿中のマクロ「%MYNETMETA」及び「%MYCONVERT」は本パッケージ中の関数を参考に作成した。よって、本稿中のマクロの License も GPL-2+ とする。

プログラム A.1 マクロ「%MYNETMETA」

```
%macro MYNETMETA(dataset=, sm= MD, level= 0.95, reference=, seq=, small= good, path= C:/temp) ;
options noxwait xsync ;
%let Rexepath='C:\Program Files\RXR-3.2.5\bin\i386\RX.exe' ; * R.exe のパスを指定!! ;
%let datapath=&path./data.csv ;
%let pdfpath =&path./output.pdf ;
%let textpath=&path./output.txt ;
%let prgpath =&path./program.R ;

options validvarname=uppercase ;
proc export data=&dataset. outfile="&datapath." dbms=csv replace ; run ;
options validvarname=v7 ;

filename xxx "&prgpath." ;
data _null_ ;
file xxx ;
put "library(netmeta) ; setwd("&path.") ; options(width=1000)" ;
put "d <- read.csv('data.csv', head=T)" ;
put "x <- netmeta(d$TE, d$SETE, d$TREAT1, d$TREAT2, d$STUDLAB, sm="&sm.", " ;
put "level=&level., level.comb=&level., comb.fixed=TRUE, comb.random=TRUE, " ;
put "reference.group="&reference.", seq=c(&seq.), warn=FALSE)" ;
put "d1 <- decomp.design(x)" ;
put "d2 <- netmeta:::decomp.tau(x, tau=x$tau)" ;
put "r2 <- netranks(x, small="&small.")" ;
put "pdf('output.pdf', paper='a4')" ;
put "forest(x, pooled='fixed', reference.group="&reference.")" ;
put "forest(x, pooled='random', reference.group="&reference.")" ;
put "netgraph(x, thickness='se.fixed') ; title('Fixed Effect Model')" ;
put "netgraph(x, thickness='se.random') ; title('Random Effect Model')" ;
put "netheat(x, random=F) ; mtext('Fixed Effect Model', 1, 1, cex=1.8)" ;
put "netheat(x, random=T, tau=x$tau) ; mtext('Random Effect Model', 1, 1, cex=1.8)" ;
put "dev.off() ; x$TE.random <- x$TE.fixed ; x$pval.random <- x$pval.fixed" ;
put "r1 <- netranks(x, small="&small.")" ;
put "write.csv(x$TE.fixed, 'TEfixed.csv', quote=F)" ;
```

---

```

put "write.csv(x$lower.fixed, 'TEfixedLCL.csv', quote=F)";
put "write.csv(x$upper.fixed, 'TEfixedUCL.csv', quote=F)";
put "write.csv(x$TE.random, 'TERandom.csv', quote=F)";
put "write.csv(x$lower.random, 'TERandomLCL.csv', quote=F)";
put "write.csv(x$upper.random, 'TERandomUCL.csv', quote=F)";
put "write.csv(d1$Q.decomp, 'Q1statistics.csv', quote=F)";
put "write.csv(d1$Q.het.design, 'Q1withindesigns.csv', quote=F)";
put "write.csv(d1$Q.inc.detach, 'Q1betweendesigns.csv', quote=F)";
put "write.csv(d2$Q.decomp, 'Q2statistics.csv', quote=F)";
put "write.csv(d2$Q.het.design, 'Q2withindesigns.csv', quote=F)";
put "write.csv(d2$Q.inc.detach, 'Q2betweendesigns.csv', quote=F)";
put "write.csv(r1$Pscore, 'rankfixed.csv', quote=F)";
put "write.csv(r2$Pscore, 'rankrandom.csv', quote=F)";
run ;
x "&Rexepath. --no-restore --no-save < ""&prgpath."" " ;
x ""&pdfpath."" ;

proc import out=TE_FIXED datafile="&path./TEfixed.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=TE_FIXED_LCL datafile="&path./%TEfixedLCL.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=TE_FIXED_UCL datafile="&path./%TEfixedUCL.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=TE_RANDOM datafile="&path./TERandom.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=TE_RANDOM_LCL datafile="&path./TERandomLCL.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=TE_RANDOM_UCL datafile="&path./TERandomUCL.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=Q1_STATISTICS datafile="&path./Q1statistics.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=Q1_WITHINDESIGNS datafile="&path./Q1withindesigns.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=Q1_BETWEENDESIGNS datafile="&path./Q1betweendesigns.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=Q2_STATISTICS datafile="&path./Q2statistics.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=Q2_WITHINDESIGNS datafile="&path./Q2withindesigns.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=Q2_BETWEENDESIGNS datafile="&path./Q2betweendesigns.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=RANK_FIXED datafile="&path./rankfixed.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;
proc import out=RANK_RANDOM datafile="&path./rankrandom.csv"
  dbms=csv replace ; getnames=yes; datarow=2 ; run ;

title "Treatment estimate (Fixed effect model)";
proc print data=TE_FIXED noobs ; run ;
title "Lower confidence limit (Fixed effect model)";
proc print data=TE_FIXED_LCL noobs ; run ;
title "Upper confidence limit (Fixed effect model)";
proc print data=TE_FIXED_UCL noobs ; run ;
title "Treatment estimate (Random effect model)";
proc print data=TE_RANDOM noobs ; run ;
title "Lower confidence limit (Random effect model)";
proc print data=TE_RANDOM_LCL noobs ; run ;
title "Upper confidence limit (Random effect model)";
proc print data=TE_RANDOM_UCL noobs ; run ;
title "Q statistics to assess homogeneity / consistency";
proc print data=Q1_STATISTICS noobs ; run ;
title "Design-specific decomposition of within-designs Q statistic";
proc print data=Q1_WITHINDESIGNS noobs ; run ;
title "Between-designs Q statistic after detaching of single designs";
proc print data=Q1_BETWEENDESIGNS noobs ; run ;
title "Q statistics to assess homogeneity / consistency (Random)";
proc print data=Q2_STATISTICS noobs ; run ;

```

---

```

title "Design-specific decomposition of within-designs Q statistic (Random)";
proc print data=Q2_WITHINDESIGNS noobs ; run ;

title "Between-designs Q statistic after detaching of single designs (Random)";
proc print data=Q2_BETWEENDESIGNS noobs ; run ;
title "Rank treatments in network (Fixed effect model)";
proc print data=RANK_FIXED noobs ; run ;
title "Rank treatments in network (Random effect model)";
proc print data=RANK_RANDOM noobs ; run ; title "";
%mend ;

```

---

プログラム A.2 データセット「Senn2013\_char」

---

```

data Senn2013_char ;
  length STUDLAB $100. ;
  input TE SETE TREAT1 $ TREAT2 $ STUDLAB $ ;
  cards ;
    -1.90 0.1414 metf plac DeFronzo1995
    -0.82 0.0992 metf plac Lewin2007
    -0.20 0.3579 metf acar Willms1999
    -1.34 0.1435 rosi plac Davidson2007
    -1.10 0.1141 rosi plac Wolffenbuttel1999
    -1.30 0.1268 piog plac Kipnes2001
    -0.77 0.1078 rosi plac Kerenyi2004
    0.16 0.0849 piog metf Hanefeld2004
    0.10 0.1831 piog rosi Derosa2004
    -1.30 0.1014 rosi plac Baksi2004
    -1.09 0.2263 rosi plac Rosenstock2008
    -1.50 0.1624 rosi plac Zhu2003
    -0.14 0.2239 rosi metf Yang2003
    -1.20 0.1436 rosi sulf Vongthavaravat2002
    -0.40 0.1549 acar sulf Oyama2008
    -0.80 0.1432 acar plac Costa1997
    -0.57 0.1291 sita plac Hermansen2007
    -0.70 0.1273 vild plac Garber2008
    -0.37 0.1184 metf sulf Alex1998
    -0.74 0.1839 migl plac Johnston1994
    -1.41 0.2235 migl plac Johnston1998a
    0.00 0.2339 rosi metf Kim2007
    -0.68 0.2828 migl plac Johnston1998b
    -0.40 0.4356 metf plac Gonzalez-Ortiz2004
    -0.23 0.3467 benf plac Stucci1996
    -1.01 0.1366 benf plac Moulin2006
    -1.20 0.3758 metf plac Willms1999
    -1.00 0.4669 acar plac Willms1999
  ;
run ;

```

---

プログラム A.3 マクロ「%MYCONVERT」

---

```

%macro MYCONVERT(dataset = ,
  studlab = ,
  treat = ,
  event = ,
  n = ,
  mean = ,
  sd = ,
  TE = ,
  seTE = ,
  time = ,
  sm = ,
  path = C:/temp) ;
options noxwait xsync ;
%let Rexepath='C:\Program Files\RR-3.2.5\bin\RR.exe' ; * R.exe のパスを指定!! ;
%let datapath=&path./data.csv ;
%let prgpath =&path./program.R ;

```

---

```

proc export data=&dataset. outfile="&datapath." dbms=csv replace ; run ;

filename xxx "&prgpath." ;
data _null_ ;
file xxx ;
put "library(netmeta); setwd(""&path."")" ;
put "d <- read.csv('data.csv', head=T)" ;
put "p1 <- pairwise(data=d, studlab=&studlab.," ;
%if &event. ne %then %do ; put "event=list(&event.)," ; %end ;
%if &n. ne %then %do ; put "n=list(&n.)," ; %end ;
%if &mean. ne %then %do ; put "mean=list(&mean.)," ; %end ;
%if &sd. ne %then %do ; put "sd=list(&sd.)," ; %end ;
%if &TE. ne %then %do ; put "TE=list(&TE.)," ; %end ;
%if &seTE. ne %then %do ; put "seTE=list(&seTE.)," ; %end ;
%if &time. ne %then %do ; put "time=list(&time.)," ; %end ;
%if &sm. ne %then %do ; put "sm="""&sm."" ; %end ;
put "treat=list(&treat.)" ;
put "p1$SETE <- p1$seTE; p1$TREAT1 <- p1$treat1" ;
put "p1$TREAT2 <- p1$treat2; p1$STUDLAB <- p1$studlab" ;
put "p2 <- p1[,c('TE','SETE','TREAT1','TREAT2','STUDLAB')]" ;
put "write.csv(p2, 'result.csv', quote=F, row.names=F)" ;
run ;

x "&Rexepath. --no-restore --no-save < ""&prgpath."" " ;

proc import out=&dataset._CONTRAST datafile="&path./result.csv"
dbms=csv replace ; getnames=yes; datarow=2 ; run ;

title ""&dataset._CONTRAST (Contrast-based)"" Converted from ""&dataset"" ;
proc print data=&dataset._CONTRAST ; run ; title "" ; title2 "" ;
%mend ;

data Parkinson ;
length STUDY $100. ;
input STUDY $ Group1 $ TE1 seTE1 Group2 $ TE2 seTE2 Group3 $ TE3 seTE3 ;
cards ;
1 1 -1.22 0.50 3 -1.53 0.44 . . .
2 1 -0.70 0.28 2 -2.40 0.26 . . .
3 1 -0.30 0.50 2 -2.60 0.51 4 -1.20 0.48
4 3 -0.24 0.27 4 -0.59 0.35 . . .
5 3 -0.73 0.34 4 -0.18 0.44 . . .
6 4 -2.20 0.20 5 -2.50 0.19 . . .
7 4 -1.80 0.20 5 -2.10 0.25 . . .
;
run ;

%MYCONVERT(dataset = Parkinson,
studlab = STUDY,
treat = %str(Group1, Group2, Group3),
TE = %str(TE1, TE2, TE3),
seTE = %str(seTE1, seTE2, seTE3)) ;

```

---

プログラム A.4 データセット「Senn2013」

```

proc format ;
  value treatf 1 = "acar"
                2 = "bent"
                3 = "metf"
                4 = "migl"
                5 = "piog"
                6 = "rosi"
                7 = "sita"
                8 = "sulf"
                9 = "vild"
                10 = "plac" ;
  value studlab 1 = "DeFronzo1995"
                2 = "Lewin2007"
                3 = "Wilms1999"
                4 = "Davidson2007"
                5 = "Wolffenbuttel1999"
                6 = "Kipnes2001"
                7 = "Kerenyi2004"
                8 = "Hanefeld2004"
                9 = "Derosa2004"
                10 = "Baksi2004"
                11 = "Rosenstock2008"
                12 = "Zhu2003"
                13 = "Yang2003"
                14 = "Vongthavaravat2002"
                15 = "Oyama2008"
                16 = "Costa1997"
                17 = "Hermansen2007"
                18 = "Garber2008"
                19 = "Alex1998"
                20 = "Johnston1994"
                21 = "Johnston1998a"
                22 = "Kim2007"
                23 = "Johnston1998b"
                24 = "Gonzalez-Ortiz2004"
                25 = "Stucci1996"
                26 = "Moulin2006" ;
run ;

data Senn2013 ;
  input TE seTE treat1 treat2 studlab ;
  format treat1 treatf. treat2 treatf.
         studlab studlabf. ;
  cards ;
    -1.90 0.1414 3 10 1
    -0.82 0.0992 3 10 2
    -0.20 0.3579 3 1 3
    -1.34 0.1435 6 10 4
    -1.10 0.1141 6 10 5
    -1.30 0.1268 5 10 6
    -0.77 0.1078 6 10 7
     0.16 0.0849 5 3 8
     0.10 0.1831 5 6 9
    -1.30 0.1014 6 10 10
    -1.09 0.2263 6 10 11
    -1.50 0.1624 6 10 12
    -0.14 0.2239 6 3 13
    -1.20 0.1436 6 8 14
    -0.40 0.1549 1 8 15
    -0.80 0.1432 1 10 16
    -0.57 0.1291 7 10 17
    -0.70 0.1273 9 10 18
    -0.37 0.1184 3 8 19
    -0.74 0.1839 4 10 20
    -1.41 0.2235 4 10 21
     0.00 0.2339 6 3 22
    -0.68 0.2828 4 10 23
    -0.40 0.4356 3 10 24
    -0.23 0.3467 2 10 25
    -1.01 0.1366 2 10 26
    -1.20 0.3758 3 10 3
    -1.00 0.4669 1 10 3
  ;
run ;

```

プログラム A.5 データセット「Senn2013」 → 「Senn」の作成 (mcmc プロシジャ適用のための前処理)

```

proc sort data=Senn2013 ;
  by studlab ; run ;
proc transpose data=Senn2013 out=TE(rename=(col1=TE1 col2=TE2 col3=TE3) drop=_name_) ;
  by studlab ; var TE ; run ;
proc transpose data=Senn2013 out=seTE(rename=(col1=seTE1 col2=seTE2 col3=seTE3) drop=_name_) ;
  by studlab ; var seTE ; run ;
proc transpose data=Senn2013 out=treat1(rename=(col1=treat1_1 col2=treat1_2 col3=treat1_3) drop=_name_) ;
  by studlab ; var treat1 ; run ;
proc transpose data=Senn2013 out=treat2(rename=(col1=treat2_1 col2=treat2_2 col3=treat2_3) drop=_name_) ;
  by studlab ; var treat2 ; run ;

data Senn ;
  merge TE seTE treat1 treat2 ;
  by studlab ;

  v1 = seTE1**2 ;
  v2 = seTE2**2 ;
  v3 = seTE3**2 ;
run ;

```

## ナニワデータサイエンス研究会 ～第1回ダイジェスト版～

○高浪 洋平<sup>12</sup>、北西 由武<sup>13</sup>、藤原 正和<sup>13</sup>、坂井 絵理<sup>13</sup>、黒田 晋吾<sup>12</sup>、舟尾 暢男<sup>12</sup>

(<sup>1</sup>ナニワデータサイエンス研究会、<sup>2</sup>武田薬品工業株式会社、<sup>3</sup>塩野義製薬株式会社)

2016年3月に「ナニワデータサイエンス研究会」なる関西発の研究会が発足し、第1回目の研究会が大阪にて開催された。本セッションではそのダイジェスト版をお送りする。なお、最新情報に応じて、内容を一部アップデートする。

### 第一部：CDISCセッション

「もしも、日米で申請電子データ提出するとしたら…」高浪 洋平（武田薬品工業）

「もしも、とある開発品目Xを申請電子データ提出するとしたら…」坂井 絵理（塩野義製薬）

### 第二部：統計解析セッション

「もしも、この機に統計担当者がNASレポートを振り返ったら…（前編）」藤原 正和（塩野義製薬）

「もしも、この機に統計担当者がNASレポートを振り返ったら…（後編）」黒田 晋吾（武田薬品工業）

### 第三部：グラフ頂上決戦

「もしも、SASのsgplotとRのggplot2を比較したら…」舟尾 暢男（武田薬品工業）

「もしも、Visual AnalyticsでOpenDataを眺めてみたら…」北西 由武（塩野義製薬）



# Sales Analyticsの可能性

－医薬品営業を例として－

武藤 猛

MarkeTech Consulting 代表

On the Possibility of Sales Analytics with Applications to Pharmaceutical Sales

Takeshi Muto

President, MarkeTech Consulting

## 要旨

Sales Analytics (SA) という用語から、一般に用いられている SFA (営業支援システム) のための分析ツールという枠を越えて「営業分析学」という広い意味で再定義することによりその可能性を引き出す方法論を提案する。本来、営業にはデータを用いて科学的な意思決定を行う「サイエンス」的な側面と、営業担当者のスキルに依存した「アート」的側面とが存在する。SA が実践的かつ科学的であるためには、データ分析のプロセスの中に、「アート」的側面をいかに取り込むかがポイントとなる。本論文では「売上げアップの公式」をベースにした営業分析学を説明する。このモデルは、成果(売上高、生産性、顧客満足度など)を生み出す要因を、①ターゲティング、②ディテール回数、および③ディテールの質、の3つに集約するものである。3つの要因のうちターゲティングとディテール回数とは容易に数量化でき、「サイエンス」化が比較的容易である。一方ディテールの質には「アート」的な要素が多く、その数量的な扱いには慎重なステップが必要である。本論文では、医薬品営業に対する営業分析学的アプローチの事例を3つ紹介する。医薬品営業を取り巻く環境は、従来の医療施設や医師を独立に扱うアプローチから、地域医療ネットワークを前提にしたアプローチへと大きく変貌しつつある。本論文で提案した営業分析学は、今後重要となる営業の「チーム力」など、従来のデータ分析では無視されがちな「アート」的側面の数量的アプローチなども取り込むことで、顧客(医師など医療従事者)のニーズと同時に営業現場のニーズにも応えうるのではないかと考えている。

キーワード：営業分析学、Sales Analytics、SFA、医薬品営業、事例による検証

## 1. Sales Analytics：IT ツールから営業分析学へ

### 1.1 本論文の目的

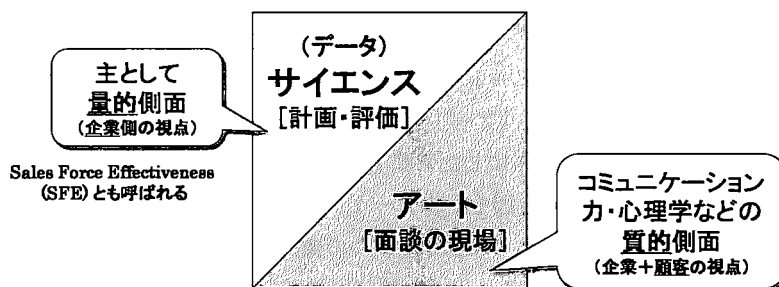
営業・マーケティングは、企業経営の重要分野の一つである。このうちの営業については属人的な要素が多く、マーケティングに較べると学術的な研究が少ない。企業における営業分野のIT活用に関しては、SFA (Sales Force Automation=営業支援システム)の導入が盛んである。最近はそのSFAのデータ分析ツールとしてSales Analytics (SA)が、ITベンダーにより喧伝されている。ところがAnalyticsという言葉は本来「分

析学」という意味を持っている。本論文では、SA から SFA のための分析ツールという枠を越えて「営業分析学」という広い意味で再定義することによりその可能性を引き出す方法論を提案する。医療用医薬品営業を例として、営業分析学の具体的な方法論を説明し、また3つの事例を紹介する。これらの事例から、SA の可能性を一層引き出すための課題を述べる。なお、本論文が対象とする営業は、営業担当者が個別に顧客と面談する「対面営業」である。対面営業の例は、本論文が対象とする医療用医薬品営業（営業担当者は MR=Medical Representative、つまり医薬情報担当者と呼ばれる）の他に、生産財営業や高額消費財（乗用車、住宅、金融商品など）営業などが該当する。

## 1.2 営業における「サイエンス」対「アート」

営業にはデータを用いて科学的な意思決定を行う「サイエンス」的な側面と、営業担当者のスキルに依存した「アート」的側面とが存在する。サイエンス的側面は、営業活動を量的に計画・評価するものであり、従来から SFE (Sales Force Effectiveness) と呼ばれてきた。一方、アートの側面は、面談の現場におけるコミュニケーション力やスキル、心理学などの質的側面を指す。サイエンス的側面は専ら企業側の視点であるのに対して、アートの側面は企業に加えて顧客の視点を含むことに留意が必要である。これら両側面の関係を図表1に示す。

図表1 営業における「サイエンス」対「アート」



「営業は科学であるか?」という設問に対して、古くから「営業は科学であるはずがない」という意見と、それに対して「営業にも科学的アプローチが成り立つ」という意見との対立がある。これらの意見の相違はアートまたはサイエンスのどの側面を重視するかの違いによるものである。著者は、両者の意見をともに尊重し、第三の道を切り拓くことが重要だと考えている。

## 1.3 営業の科学に関する先行研究

前述のように、マーケティングに比較すると営業に関する学術的研究は圧倒的に少ない。この理由は、営業に関する詳細情報には企業秘密の要素が多く、外部からのアクセスが困難なことが一因と考えられる。日本においては営業の全体モデルに関する研究がいくつか発表されている<sup>(1)~(5)</sup>。数は少ないが事例研究も行われている<sup>(6)</sup>。営業活動の数量的研究となると非常に限られている<sup>(7)~(10)</sup>。ビジネス誌の特集号では何回か営業が取り上げられている<sup>(11)~(15)</sup>。

欧米でも、営業に関する研究書は少ない。ここでは、研究と実践（コンサルティング）を両立させている Zoltners-Sinha による一連の著書を挙げておく<sup>(16)~(19)</sup>。特に「The Power of Sales Analytics」と題した近著<sup>(19)</sup>は、「Sales Analytics」を題した現在のところほとんど唯一の本であり、その狙いは本論文の趣旨に近い。

## 1.4 SFA の分析ツールとしての SA の功罪

SFA は多くの IT ベンダーが販売しているが、その基本的な機構は、営業担当者が顧客情報、営業プロセス情報（訪問顧客名、日時、面会者、面談内容など）、および商談情報（顧客要件、商談規模、商品名、競合、

提案書・見積提出期限など)を毎日データベース(クラウド上の場合も多い)に記録し、それを集計・分析可能とするものである。これらの集計・分析結果は、営業担当者が自己管理に活用したり、営業マネージャーにより管理業務や営業担当者へのアドバイスに活用されたりしている。

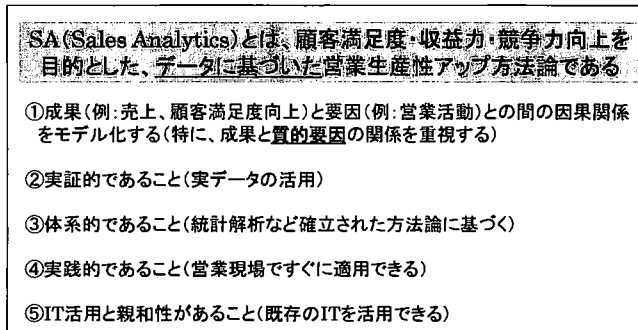
ある SFA ベンダーから公表されている導入事例 33 件を著者が分析したところ、認識された導入効果(複数カウント)のうち 86%は効率(Efficiency)向上効果、14%は効果(Effectiveness)向上効果であった。前者の代表例は「データ入力作業が効率化された」や「社内の情報共有」など、後者の例は「顧客満足度向上」と「売上高・成約率向上」である。

このように、SFA およびその分析ツールである SA に関して、一般的な導入効果としては営業活動の効率化につながる事が挙げられる。一方限界としては、収集対象データは企業の視点に立った数量化が容易なデータがほとんどであり、顧客の反応に関する質的データは限られていることや、「成果モデル」が組み込まれていないので、成果をもたらす要因分析が困難なことなどが挙げられる。

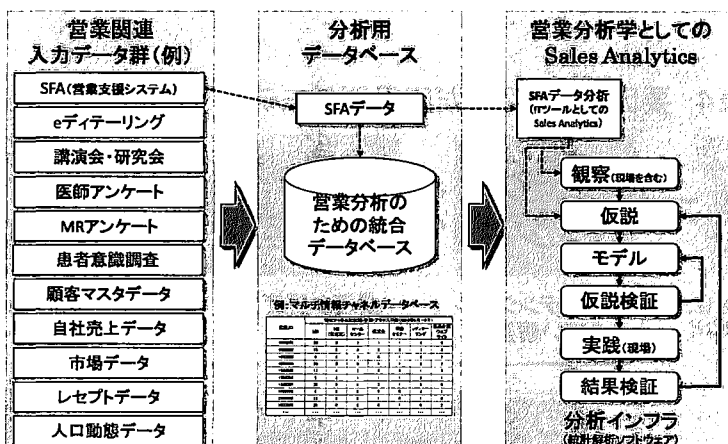
### 1.5 営業への科学的アプローチとしての SA へ

以上述べたように、SFA の分析ツールとしての SA にはメリットも多いが、限界もある。そこで SA を IT ツールという枠に留めず、「サイエンス(データに基づく科学)」と「アート(技能・スキル)」の共働を目指す科学的分析方法論(Analytics の原義)として捉えることを提案したい。このようなアプローチを本論文では「営業分析学」と名付ける。営業分析学としての SA の定義を図表 2 に示す。また、実装する場合のイメージを図表 3 に示す。

図表2 営業分析学としての Sales Analytics の定義(提案)



図表3 営業分析学としてのSales Analytics のイメージ



図表 3 に示したように、営業分析学の方法論は次のように要約できる：①観察（データや営業現場）と課題認識、②課題解決のための仮説創出、③数理的モデル構築、④仮説検証（および数理的モデル改良）、⑤現場における実践、⑥結果検証とフィードバック。入力データは SFA だけではなく、アンケートなど質的データを含む多様なデータが対象となる。SFA データは、上記方法論の①および②において、現状を的確に把握するために重要な役割を果たす。なお対面営業の場合は、顧客に紐づけられない「ビッグデータ」は、現在のところ活用方法が限定される。分析インフラとしては、SAS や JMP など、データ加工や多様なモデル作成とその検証に適した汎用統計解析ソフトウェアが適している。

## 2. 医薬品営業における定式化

### 2.1 医薬品営業は SA の研究と実践に適している

医療用医薬品は国民医療の重要な一環を担っている。また各製薬企業の MR は、自社医薬品に関する情報（効能や副作用）を医師を始めとする医療従事者に提供するとともに、自社医薬品の普及のための宣伝活動を行っている。医薬品営業は、次のような理由により、SA の研究と実践に適している：①安全な薬物療法の推進および営業生産性向上が一貫した経営課題であること（質と量の両立）、②MR はこの二つの課題に深く関わっていること、③製薬業界はグローバル化が進み「データに基づく経営」が普及していること、④研究開発分野で統計解析が普及していてデータ分析リテラシーが高いこと、⑤社内・社外で入手可能な営業関連データが豊富であること、⑥営業生産性向上が業界全体の重要な課題となっていること。

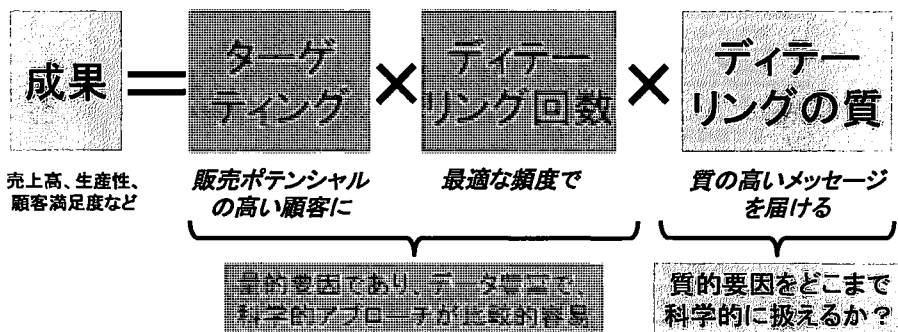
そこで本論文では、以下、医薬品営業を対象にした、営業への科学的アプローチについて論じる。

### 2.2 医薬品営業における成果達成のための因果関係モデル

図表 2 に示したように、SA において最も重要なのは、成果（売上高、営業生産性や顧客満足度など）と要因（営業活動）との間の因果関係をモデル化することである。特に質的要因もモデルに取り込むことが重要である。営業の因果関係モデルについては、Zoltners-Sinha が「営業生産性ドライバーモデル」と呼ばれる、営業活動における要因を網羅したモデルを提案している<sup>(16)~(18)</sup>。また、ポーター・ヘンリーは、システム工学的アプローチで営業における因果関係モデルを提案している<sup>(12)</sup>。これらはいずれも示唆に富む優れたモデルであるが、大規模かつ複雑で、営業現場で使い易いとは言えない。

そこで著者は従来から、要因を絞り込んだ因果関係モデルを「売上アップの公式」として提案している<sup>(20)</sup>。これは、成果に対して即効性が大きい 3 要因にフォーカスして因果関係をモデル化したものである。これを図表 4 に示す。

図表 4 要因を絞り込んだ因果関係モデル: 売上アップの公式

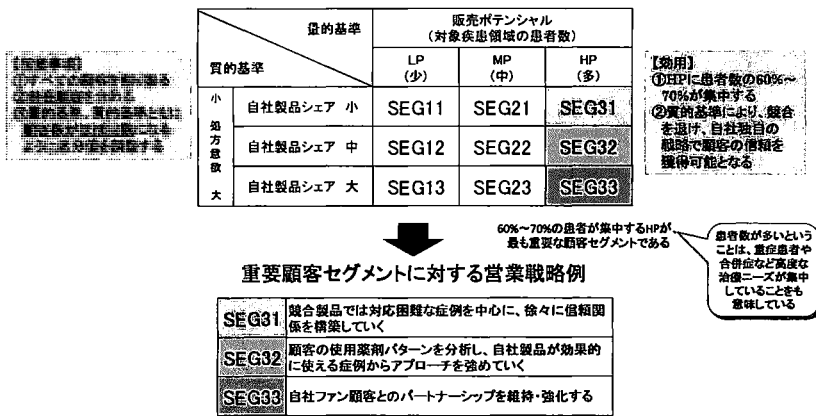


図表4は、現在の製品構成を前提としている。このモデルは、成果（売上高、営業生産性、顧客満足度など）を生み出す要因を、①ターゲティング、②ディテール回数、および③ディテールの質、の3つに集約するものである。このモデルは「販売ポテンシャルの高い顧客に、最適な頻度で、質の高いメッセージを届ける」ことにより成果を最大化できることを主張している。3つの要因のうちターゲティングとディテール回数とは容易に数量化でき、「サイエンス」的の取扱いが比較的容易である。一方ディテールの質には「アート」的な要素が多く、その数量的な扱いには慎重なステップが必要である。

図表4の「公式」は、「数式」ではなく、「売上を決定する要因は3つに集約される」という意味である。この公式は、MR活動に関して、多くの薬効領域で検証済みである（検証例は後述）。図表4のモデルは、営業分析学の課題である、営業の量的側面と質的側面の両方を簡潔にモデルに取り込んでいる。公式の対象は、対面営業だけでなく、eディテールやエリアマーケティングにも適用できる汎用性を備えている。

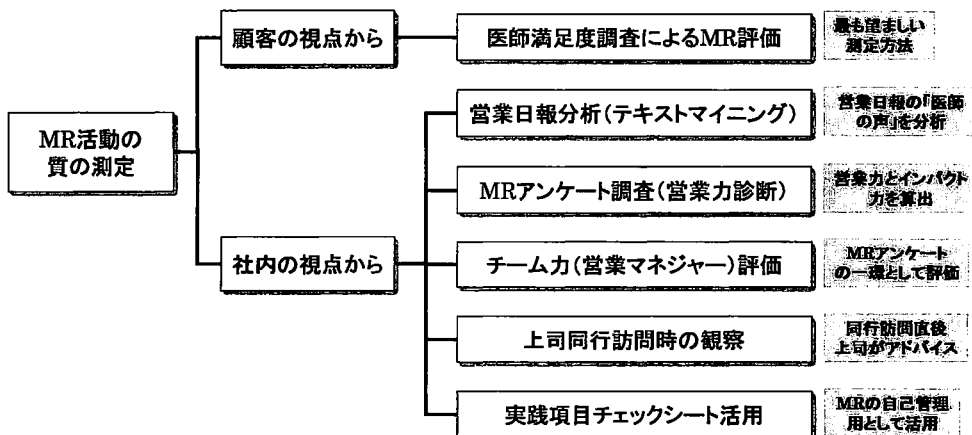
図表4の3つの要因のうち量的要因であるターゲティングとディテール回数をまとめて扱うために、著者は図表5に示す「顧客バリューマトリックス（CVM）」を提唱している。

図表5 顧客バリューマトリックス(CVM)



一方、図表4の3つの要因のうち質的要因であるディテールの質については、図表6に示す測定方法がある。

図表6 「MR活動の質」の測定方法



「MR 活動の質」について最も重要なことは、一般的な「質」の定義を求めるのではなく、自社に最も相応しい「MR 活動の質」の定義を確立することである。このためには、MR アンケート調査が有用である。

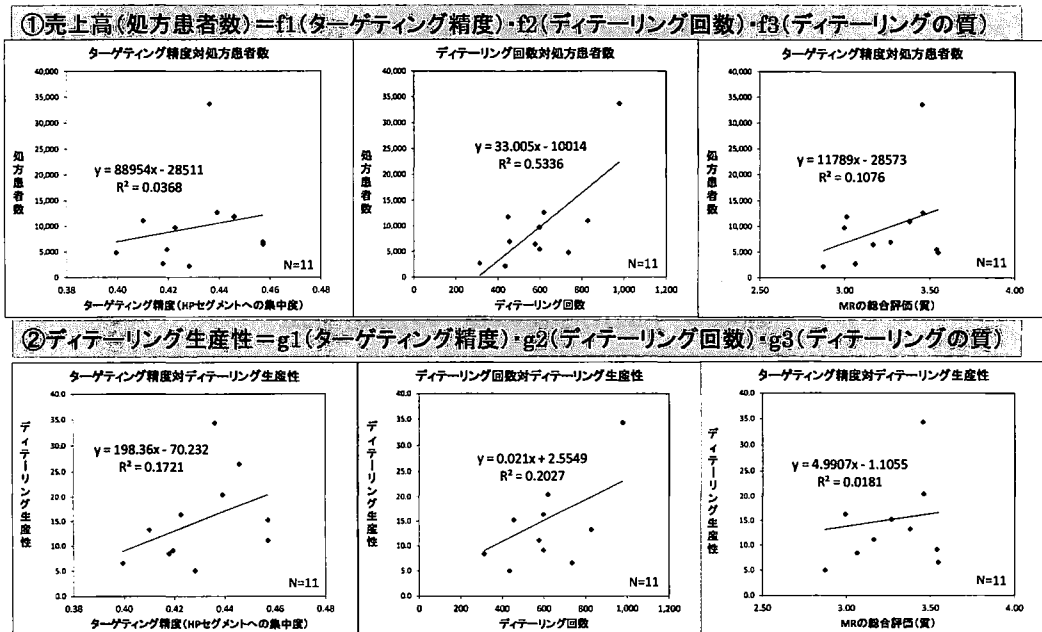
### 3. 医薬品営業における検証事例

#### 3.1 売上アップの公式の検証

第 1 の検証事例として、あるスペシャリティ領域（特定診療科の医師が処方する医薬品）に関する医師アンケートを用いて、図表 4 の売上アップの公式の検証を行った事例を紹介する<sup>(20)</sup>。これは 2011 年 10 月実施したもので、N=900、製薬企業数は 11 である。アンケート結果から、図表 5 に示した CVM を作成した。売上アップの公式の第一の要因であるターゲティング精度として、CVM の高ポテンシャル（HP）3 セグメントへの集中度を用いた。第二の要因であるディテールリング回数は、製薬企業毎の回数をそのまま用いた。第三の要因であるディテールリングの質は、医師から見た MR の総合的満足度を用いた。図表 4 の目的変数である「成果」については、処方患者数およびディテールリング生産性（処方患者数÷ディテールリング回数）の 2 種類を採用した。

検証結果を図表 7 に示す。目的変数が処方患者数およびディテールリング生産性いずれの場合も、三つの要因はいずれも目的変数と有意な正の相関がある。したがって、売上アップの公式が検証された。ただし、ディテールリングの質に関しては、他の二つの要因に比べて相関係数が小さい。これは、「質」の数量化に課題があることを示唆している。

図表 7 「売上アップの公式」の検証結果



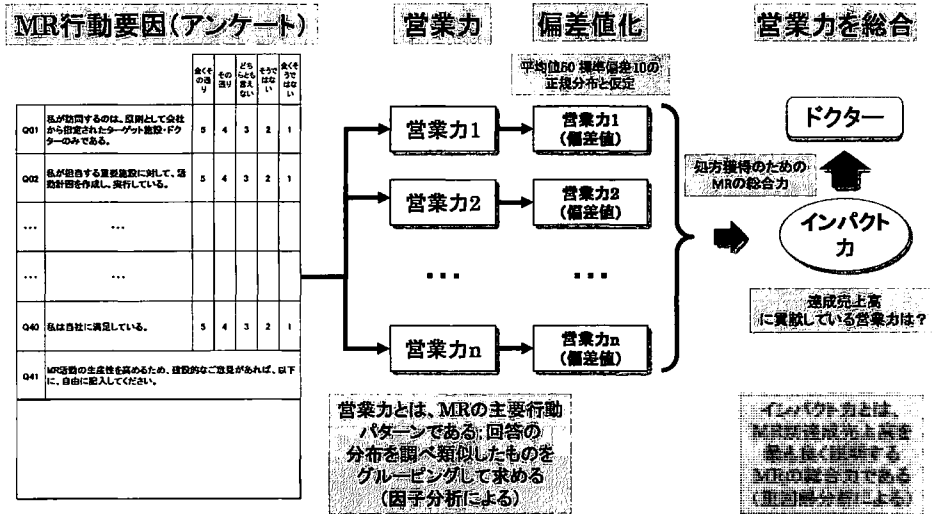
#### 3.2 MR アンケートによる「質」の把握

MR アンケートを用いて「MR 活動の質」を把握し、営業活動の改善に活用した事例を紹介する<sup>(21)</sup>。全体の流れは、準備（ヒアリング、アンケート原案作成、アンケート最終案作成）、アンケート実施、分析（基本集計、問題構造の把握、営業力の分析）のように進めた。このうち営業力の分析の詳細を図表 8 に示す。

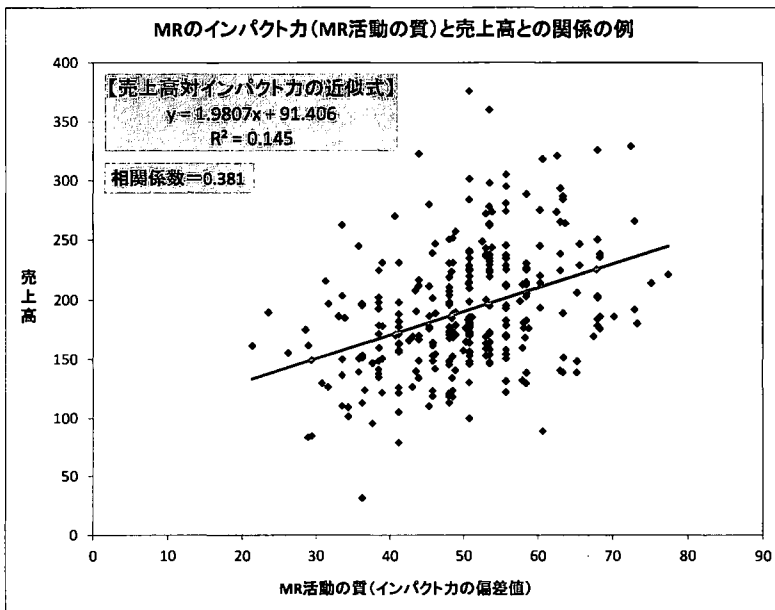
図表 8 に示すように、因子分析により MR 活動の基本要素である「営業力」を抽出し、それを成果（ここでは売上高）とリンクさせることで「インパクト力」として指標化する。インパクト力とは、MR 別達成売上高を最も良く説明する MR の総合力であり、目的変数として売上高、説明変数として営業力を用いた重回帰分析により求めたものを偏差値化した指標である。この事例では、「医師の信頼獲得」、「営業所マネジメント」など 7 つの要因を偏差値化した。一例として、MR のインパクト力と売上高との関係を図表 9 に示す。

アンケート分析の結果は、MR 教育の改善や MR の個別指導に活かす他、ハイパフォーマーの行動要因を「MR 活動チェックシート」として要約し MR 自身の自己管理に活用することができる。

図表 8 MRアンケートによる「質」の数量化



図表 9 インパクト力と売上高との関係



### 3.3 営業日報のテキストマイニングによるMRの「質」の把握

SFAには日報としてテキストデータで医師との面談内容を報告する機能がある。営業日報のテキストデータは、MR活動に対する医師など医療従事者の反応を質的に把握するための有力な手段である。ここでは営業日報のテキストマイニングによるMRの「質」の把握の事例を紹介する<sup>(22)</sup>。

テキストマイニングの対象は、ある製薬企業の半年分の営業日報のテキストデータの中で、医師と納入高の対応可能な日報データ（開業医）20,000件である。これをテキストマイニングツール（VextMiner使用）でキーワードおよびクラスター分析した。この結果、営業活動はその内容により14個のクラスターに分類された。ここではそのクラスターを更に6つの営業ステップ（顧客ニーズ把握、製品紹介、ディテリング、クロージング、フォローアップ、および人間関係構築・維持）にグループ分けし、各営業ステップの件数を説明変数に、納入高を目的変数にして重回帰分析を行った。結果を図表10に示す。

## 図表10 営業日報のテキストマイニング結果

■ある製品に関する営業プロセス別データの重回帰分析結果（営業所別データ半年分）

・目的変数＝納入高、説明変数＝テキストマイニングの結果抽出された営業ステップ（GR1～GR6）

#### ①分析の要約

R2乗	0.998
自由度調整R2乗	0.998
誤差の標準偏差(RMSE)	1208256
Yの平均	5777851
オブザベーション(または重みの合計)	94

#### ②パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob> t )
切片	-2921	127303	-0.02	0.982
GR1_顧客ニーズ把握	-13611	12173	-1.12	0.267
GR2_製品紹介	21217	12842	1.65	0.102
GR3_ディテリング	57896	23357	2.48	0.015
GR4_クロージング	73941	33536	2.20	0.030
GR5_フォローアップ	25710	5243	4.90	<.0001
GR6_人間関係構築・維持	9812	1578	6.22	<.0001

図表10によれば、納入高には幅広い営業ステップ（ディテリング、クロージング、フォローアップ、および人間関係構築・維持）が有意な影響を及ぼしている。このことは従来から「MR活動の質」を代表すると考えられてきたディテリングの質はもちろんであるが、人間関係構築・維持も含めて「MR活動の質」を幅広く捉えることの重要性を示している。この結果は、医師がMRを評価する際には、製品や疾患に関する知識だけでなく、人間性も同程度に重視するという調査結果<sup>(23)</sup>とも整合している。

## 4. まとめ：Sales Analyticsの可能性

### 4.1 Sales Analytics (SA) 活用のための今後の課題

残念なことではあるが、最近医薬品営業における不祥事が何件か報道された。売上至上主義の営業活動は、顧客に迷惑を掛け、かつ不正の温床ともなる。顧客・企業・営業担当者が「三方良し」となる営業活動のためには、数量データ（売上、シェア等）だけを目標とするのではなく、質的データにも着目することが重要である。「あるべき営業活動の質」は、各社独自に決める必要がある。このようにして定めた独自の「質」の追及により、売上至上主義からの脱却が可能になると同時に、顧客からの評価の向上により成果にも繋がるこ



とが期待される。質的データの取り扱いには、実地の営業活動の観察や営業担当者のインタビューに基づくモデル化とその数量化など、一般論に囚われない創造的な方法論と分析が必要である。SA を営業分析学として幅広く捉えることは、このような方法論の確立に有効と考えられる。

今後の課題として、営業における「アート」を「サイエンス」の視点から把握するには、「営業活動の質」を多角的に研究する（心理学、コミュニケーション学を含め）ことが必要である。地域医療ネットワークの時代を迎えて今後は、営業担当者だけでなく、営業マネジャーの役割がキーとなる。特に、「チーム力」が営業業績に及ぼす影響の研究が重要となる。そのためには営業アンケートとその多変量解析や、営業日報のテキストマイニングが分析手法として考えられる。

一方、「サイエンス」で得られた知見を「アート」の分野で活かすためには、従来からの数量データの分析に加えて、科学的根拠に基づいて営業教育を再構築する必要がある。単に製品知識だけでなく、顧客（医療チーム）との信頼関係構築のためのスキル向上を含めるべきである。営業現場で取り入れやすい分析結果の例としては、分析で得られた結果を、営業所単位で少数の実践項目に絞ったり、「営業活動チェックリスト」に落とし、活用することが考えられる。

#### 4.2 まとめ：Sales Analytics を成功させるために

Sales Analytics を成功させるための要因としては、次の5点が考えられる。(1)「営業を科学化する」ことへの経営者の強い持続的な意志、(2)分析スタッフ養成とデータ収集・分析用 IT インフラの整備、(3)自社の「営業－成果モデル」構築、(4)市場・顧客・営業活動に関する多種多様なデータの蒐集とデータベース化、(5)効果的な営業教育・コーチングと投資効果の測定。

一方、留意事項として、次の4点を挙げたい。(1)営業における「アート」の側面の重要性を忘れない（「サイエンス」の側面を強調しすぎると、顧客も営業現場も疲弊する）、(2)自社の優秀な営業担当者をよく観察する（自社の営業活動における「センターピン」の把握とその活用）、(3)営業現場が考慮すべき指標の数は最小限にする（3つ以内を目標とする）、(4)IT は万能ではなく、その機能が営業現場から乖離すると投資は無駄になる（営業管理だけを主要目的とする IT 投資は現場の恣意的なデータ入力を誘発し、ほとんどの場合投資は失敗する）。

## 参考文献

- (1) 石井淳蔵・嶋口充輝：営業の本質－伝統と革新の相克、有斐閣（1995）
- (2) 石井淳蔵：営業が変わる－顧客関係のマネジメント、岩波書店（2004）
- (3) 田村正紀：機動営業カーブ時代時代の市場戦略、日本経済新聞社（1999）
- (4) 高嶋克義：営業改革のビジョナー失敗例から導く成功のカギ、光文社（2005）
- (5) 松尾睦・早川勝夫：営業のプロセス改革と人材育成－日本ベーリンガーインゲルハイムの事例、神戸大学大学院経営学研究科 Discussion Paper Series（2010）
- (6) 松尾睦・早川勝夫・高嶋克義：改善志向の営業プロセス管理－日本ベーリンガーインゲルハイムの事例、季刊マーケティングジャーナル 119号（2011）
- (7) 日経産業消費研究所：営業の革新－混迷の時代に求められる「強さ」の研究、日本経済新聞社（1998）
- (8) 松尾睦：営業組織における文化的特性と業績、流通研究 1巻1号（1998）
- (9) 清宮政宏：営業管理様式と営業成果の因果関係に関する研究、流通研究、Vol.7 No. 2（2004）
- (10) 細井謙一・竹村正明：営業生産性尺度の開発、広島経済大学研究双書 第27冊、広島経済大学地域経済

研究所 (2006)

- (11) 一橋大学ビジネスレビュー：営業を科学する、東洋経済新報社 (2006年6月号)
- (12) ダイヤモンド・ハーバード・ビジネス編集部：営業力のプロフェッショナル、ダイヤモンド社 (2004年1月号)
- (13) ダイヤモンド・ハーバード・ビジネス編集部：一流の営業力、ダイヤモンド社 (2005年1月号)
- (14) ダイヤモンド・ハーバード・ビジネス編集部：最強の営業力、ダイヤモンド社 (2006年10月号)
- (15) ダイヤモンド・ハーバード・ビジネス編集部：強い営業、ダイヤモンド社 (2012年12月号)
- (16) Andris A. Zoltners, Prabhakant Sinha, et al.: The Complete Guide to Accelerating Sales Force Performance, AMACOM (2001)
- (17) Andris A. Zoltners, Prabhakant Sinha, et al.: Sales Force Design for Strategic Advantage, Palgrave (2004)
- (18) Andris A. Zoltners, Prabhakant Sinha, et al.: Building a Winning Sales Force, AMACOM (2009)
- (19) Andris A. Zoltners, Prabhakant Sinha, et al.: The Power of Sales Analytics, ZS Associates, Inc. (2014)
- (20) 武藤 猛: MRの生産性に関する考察、SAS ユーザ総会 (2012)
- (21) 武藤 猛: MR アンケートによる製薬企業の営業力診断、SAS ユーザ総会 (2013)
- (22) 武藤 猛: MRの生産性アップと最適配置戦略、アンドテック社 (2012)
- (23) 月刊誌「Monthly ミクス」2011年4月号～2014年4月号の記事「薬剤部長から評価される『印象派MR』の特徴」(計33回)における各薬剤部長の発言を、著者がテキストマイニングして集計した結果(複数カウント)による。

## JMP Clinicalを用いたセントラルモニタリングの経験 ～RBMを推進するために～

○福升 悠一<sup>1</sup>、近藤 秀宣<sup>1</sup>、樋口 天裕<sup>1</sup>、小幡 智憲<sup>1</sup>、畑山 知慶<sup>2</sup>、林 行和<sup>3</sup>

(<sup>1</sup>エイツーヘルスケア株式会社 開発戦略本部 セントラルモニタリング部、

<sup>2</sup>エイツーヘルスケア株式会社 開発戦略本部 生物統計部、

<sup>3</sup>エイツーヘルスケア株式会社 開発戦略本部)

【背景】RBMとは、限られたリソースで臨床試験の品質を効果的に効率よく担保するモニタリング手法である。これまでのモニタリングでは、CRA 個人の力量に依存しており、出口管理的なSDVを中心として定期的な施設訪問が行われてきた。RBMにおけるセントラルモニタリングでは、施設横断的にモニタリングを行うことでリスクを早期に発見することが可能である。リスクを事前に把握することで、ターゲットを絞った適切な指示をモニターに出すことが出来るが、そのためには、データの集約的なレビューが必要である。我々はデータの集約的なレビューを実施するツールとしてJMP Clinicalの使用が有益であると考えている。

【目的】セントラルモニタリングにおけるJMP Clinicalの利用方法を紹介し、有用性を議論する。

【方法】JMP Clinicalに標準搭載されているサンプル試験データを用いて、JMP Clinicalによるセントラルモニタリングでの実践方法を紹介する。その有用性を考察するサンプル試験は、腎排泄型の薬剤の臨床試験である。そのため、重度な腎機能障害患者を除外するエントリー基準があると仮定する。この場合、被験者の安全性を確保するために定められている除外基準に抵触しそうな被験者を集中的に登録することは試験の運用上好ましくないと考えられる。そこで、本発表で例示するセントラルモニタリングでは、重度な腎機能障害患者を集中的に登録している施設を検出することを目的として、スクリーニング時点の臨床検査値に基づいて偏った登録傾向のある施設の検出を試みることにした。

【結果】施設別および測定項目（クレアチニン、BUN）別にスクリーニング期の測定値を箱ひげ図で示した。これにより施設間で腎機能に関するそれぞれの測定値の分布をビジュアルに確認することが出来た。さらに、クレアチニンとBUNの値を同時に見て、施設別にどのような差異があるかを検討するために、クレアチニンおよびBUNの散布図を作成し、施設IDで色分けした。JMP Clinicalのフィルタリング機能を用いることで、特定の施設のみをハイライトで表示させることが出来る。この機能を用いることで、全被験者の中での特定の施設の患者の値の分布をより容易に検討することが出来る。ここまでの検討で、ある施設は、他の施設と比較するとクレアチニンおよびBUNが高い傾向にある被験者を多く登録する傾向があったことが確認された。実際のRBMでは、ここまでの検討によって、施設スタッフの薬剤特性・プロトコル理解度の確認指示、重度な腎機能障害に対する判断基準の確認指示、スクリーニング期の測定値がチェックされるプロセスの確認指示等のモニタリングアクションを検討することが出来たと考えられる。

【考察】JMP Clinicalを利用すると、プログラムを行わずに多種多様なKRIを容易に可視化することが可能となる。また、関心のある対象について、容易にその対象の関連情報を調べることが可能であることから、スピーディに適切なモニタリング指示が発行できるようになる。

【結論】JMP Clinicalを実際に利用した結果、JMP Clinicalの導入はセントラルモニタリングを効率的に運用するのに有益と考えられた。

## SAS XML Mapperを用いたDefine.xmlのメタ情報の有効活用及び業務効率化の取り組み

○林 三男、山本 繁

(大日本住友製薬株式会社 開発本部 データサイエンス部 生物統計グループ)

本邦では、2016年10月より医薬品の承認申請時にCDISC標準に準拠した電子データ提出が開始される。解析業務では、データセット (SDTM、ADaM)、データガイド (SDRG, ADRG)、定義ファイル (define.xml)を規制当局の要求通りに作成することが求められ、データセット (SDTM、ADaM) 及び define.xmlに関してバリデーションルールが公表されている。現在、複数のベンダーから define.xmlを作成するツールやバリデーションルールに基づいたチェックを行うツールが提供されている。これらは多機能・高機能なツールである一方で、「バリデーションルール以外のad-hocチェックができない」、「define.xmlのメタ情報から解析用データセット (またはSASコード) が生成されない」など、発展途上の側面もあると考える。

本発表では、SAS社のXML Mapperを用いたdefine.xmlのメタ情報をSASデータセットとして取り込むことで、define.xmlのメタ情報とデータセットの一貫性を取るためのad-hocチェックや、define.xmlのメタ情報を活用して解析用データセットの作成をセミオートメーション化するための可能性を探り、define.xmlのメタ情報の有効活用及び業務効率化の実現に向けた取り組みを紹介する。

申請時電子データのLengthにまつわる実務上の問題点と対処法  
Business Issues for Variable Length and a Solution  
on Electronic Submission of Study Data

○片山 雅仁、平野 勝也、鈴木 ゆら  
(塩野義製薬株式会社 解析センター)

2016年10月から開始されるPMDAへの申請時電子データ提出に向けて、申請時電子データのバリデーションルールが2015年11月にPMDAより公開された。その中に、SDTMデータセットの変数のLengthについて、規定した変数のLengthが実際にSDTMへ格納された値のLengthより大きいことを指摘し、データセットの容量を抑えるためのルールがある。PMDAおよびFDAで利用されているバリデーションツールのPinnacle21でバリデートすると、このルールに関してPMDA版ではWarningメッセージが、FDA版ではErrorメッセージが出力される。このルール自体はFDAのStudy Data Technical Conformance Guideやバリデーションルールにも掲載されて既に周知されている。しかしながら、WarningあるいはErrorメッセージへの対応が実務者にとっては非常に悩ましい問題になり得るにもかかわらず、これらの問題への対処法が十分に議論されているとはいえない。

上記Length問題への一つの対処法として、実際のRawデータの値のLengthに応じて、SDTMデータセットの作成仕様書で変数のLengthを試験途中で定義しなおすことが考えられる。しかし、データベース固定前に一旦Lengthを定義した後、Rawデータがそれ以上のLengthになった場合、定義したLengthを超えたデータを欠落させる懸念が生じる。加えて、ADaMの作成時にRawデータの状況に応じて異なるLengthのSDTMを扱うことになり、ADaM作成のタスクにも影響を及ぼすことになる。他方、データベース固定後にRawデータを確認してLengthを再定義することも考えられるが、厳しいスケジュールの中、スピードと正確性が要求される場面でプログラム修正を伴う仕様変更は、スケジュール遅延や解析結果の信頼性を低下させるリスクが高い。さらに、仕様書に定義された変数の基本情報をDefine.xmlに反映させることが多いと思うが、仕様書内の調整済みのLengthをDefine.xmlに反映させる必要があり、成果物の整合性を確保しようとする手間がかかる。また、併合解析でLengthが異なる複数のデータセットを結合する際にデータを欠落させる懸念、SASのログにWARNINGのメッセージが出力されることなど、CSR作成時だけでなく将来の併合時までを考慮に入れて対処するとなれば、SDTMおよびADaM担当者の負担が相当大きくなることが予想される。

本発表では、Lengthの制約に関する実務上の問題を解決する方法として、①通常通りに、SDTMおよびADaMの仕様書に定義したLengthで、sas7bdat形式のデータセットを作成し、ADaMおよびTLF作成でそれを用いる、②申請資料として提出するXPTファイルおよびDefine.xmlについては、sas7bdat形式のデータセットと中身が同一のLength調整済みのXPTファイルとLength調整済みのDefine.xmlを用意する方法を提案する。この手順を採用することにより、(1)仕様書とsas7bdat形式のデータセットの整合性、XPTファイルとDefine.xmlの整合性を確保しつつも、(2)前者と後者を切り離して考えることで作成・バリデーション手順はシンプルになり、(3)Length調整前のsas7bdat形式のデータセットを利用することで後工程への影響はなくなり、(4)Length調整後のXPTファイルとDefine.xmlでバリデーションすることでLength関連の大量のWarning/Errorが検出されなくなる。

sas7bdat形式のデータセットからLength調整済みのXPTファイルを作成する部分については、簡単なSASプログラムを紹介する。

## 欠測を含む順序カテゴリカル経時データの解析-MIプロシジャの有用性-

○藤原 正和<sup>1</sup>、駒寄 弘<sup>2</sup>

(<sup>1</sup>塩野義製薬株式会社、<sup>2</sup>マルホ株式会社)

順序カテゴリカル経時データに欠測値が含まれる場合、MIプロシジャによる補完が可能である。本講演では欠測を含む経時的な順序カテゴリカルデータの発生方法に加えて、MIプロシジャで実装可能な補完方法 (Sequential regression, MCMC, FCS) の理論を紹介して、シミュレーションデータを用いた性能評価を行う。

なお、経時的なカテゴリカルデータの発生方法はLee (1997), 及びIbrahim (2011) によって提案された方法を紹介する。

### 文献

- Lee, A. J. (1997). Some simple methods for generating correlated categorical variates. *Computational statistics & data analysis*, 26(2), 133-148.
- Donneau, A. F., Mauer, M., Molenberghs, G., and Albert, A. (2015). A simulation study comparing multiple imputation methods for incomplete longitudinal ordinal data. *Communications in Statistics-Simulation and Computation*, 44(5), 1311-1338.
- Ibrahim, N. A., and Suliadi, S. (2011). Generating correlated discrete ordinal data using R and SAS IML. *Computer methods and programs in biomedicine*, 104(3), e122-e132.
- 日本製薬工業協会 医薬品評価委員会 データサイエンス部会 欠測のあるデータの解析チーム (2016). 欠測のある連続量経時データに対する統計手法について (Ver1.0), <http://www.jpma.or.jp/information/evaluation/allotment/statistics.html>
- 藤原正和, 高橋 文博 (2015). 2015年度 計量生物セミナー, 臨床試験における estimand と感度分析, 欠測のあるデータに対する解析手法の基礎 ~ (2) 主解析の検討~.

ADaM/TLF経験者がSDTMを作成できるまでのサクセスストーリー  
A success story for achievement of SDTM program and specification  
by programming skill of ADaM/TLF

○小林 千鶴<sup>1</sup>, 神谷 亜香里<sup>1</sup>, 坂本 佳代子<sup>2</sup>, 勘場 大<sup>1</sup>, 江川 広明<sup>1</sup>, 角谷 伸一<sup>1</sup>  
(<sup>1</sup>塩野義製薬株式会社 解析センター, <sup>2</sup>株式会社エス・シー・エー)

PMDAへの承認申請用CDISC標準電子データ提出を見据えて、弊社では全ての臨床試験のプログラミング業務をRaw→SDTM→ADaM→TLFの流れで作成することを目指し、2013年にSDTM作成体制の再構築を行った。さらにはSASプログラミングでのSDTM作成を可能にするため、2014年度は、SDTMマッピング仕様書（以下、SDTM仕様書）を改良し、それを効果的に運用することでより効率的にSDTMを作成することができるようになった。これにより、SASプログラミングの知識を有してはいるが、SDTMの作成が未経験であるプログラミング担当者がSDTMの作成に取り組むことが容易になった。

SDTM作成がより身近になったことから ADaM/TLFの作成しか経験のなかった著者は、SDTM仕様書の作成ができるようになることを最終の目標として、まずはSDTMの内容の理解を深めることを優先し、SDTM仕様書を利用してSDTMプログラミングに挑戦した。この仕様書は、「SDTMおよびRawデータを知らないプログラマでもSDTMを作成できる」をコンセプトに作成されていることもあり、SDTMプログラムの作成は予想よりも容易にできた。SDTMプログラムはADaMプログラムよりも複雑な変換処理がないこともその一因と考えられた。

SDTMの内容を理解できたことから、次にSDTM仕様書作成にも取り組んだ。SDTMプログラム作成時には、プログラミングを容易にするSDTM仕様書に大きく助けられたが、今度はその仕様書を作成する立場となった。しかし、SDTMプログラム作成時とは異なり、非常に大きな困難に直面した。とくに、ADaMデータ作成時に利用しないドメイン（例えば、Trial Designシリーズのドメインなど）についてその仕様書の作成は非常に難しいものであった。SDTM IGやガイダンス等を参照しそれに従いながら作成することが求められるが、その経験が浅い著者は、SDTMの知識および経験をもつメンバーと議論し、取扱いを理解して仕様書の作成に努めた。一方、SDTMは、Rawデータの状況も考慮して作成することも必要であるため、DM担当者との連携は不可欠である。SDTMの作成にあたっては関係者と連携を密にとることが重要であると認識した。

ところで、著者は今回のSDTM作成の経験を通して、ADaM/TLF経験者の観点から、解析時に利用しやすいSDTMを意識してSDTM仕様書を作成した。SDTMの情報は、どのように帳票に出力されるか、そのためにどのようなADaMが作成されるかを確認した。DM担当者には、どのような想定で取得されているデータなのかを確認した。SDTM, ADaM, TLF を効率よく作成するには、各IGやガイダンスなどを理解することが重要であるが、関係者と議論し解決していくことも重要であることを認識した。

最後に、弊社の取り組みとして、SDTMプログラミングおよび仕様書作成を効率よく実施できることを意識して、疑問点とその解決方法をまとめた「SDTM仕様書作成の虎の巻」を作成し、知識経験を共有できるようにしている。また、SDTM, ADaM, TLFを効率よく作成するため、SASマクロ化の検討も行っている。

本発表では、SDTM作成の経験はないが、ADaM/TLF プログラミングスキルのある著者が、SDTMおよびその仕様書作成ができるまでの道のりを紹介する。

## 欠測を含む順序カテゴリカル経時データの解析-GEEプロシジャの有用性-

○駒寄 弘<sup>1</sup>、藤原 正和<sup>2</sup>

(<sup>1</sup>マルホ株式会社、<sup>2</sup>塩野義製薬株式会社)

順序カテゴリカル経時データの解析手法として、これまでは GENMOD プロシジャ、CATMOD プロシジャ、GLIMMIX プロシジャが主に使われてきた。近年、SAS STAT14.1 より、GEE プロシジャに dist=multinomial オプションが追加され、GEE プロシジャによる実行も可能となった。本プロシジャでは Liang and Zeger (1986)及び Prentice(1988)の GEE 法に加え、相関パラメータの推定の際、alternating logistic regression(ALR) 法を使用している点が特徴である (Heagerty and Zeger,1996)。

本発表ではGEEプロシジャ及びCATMODプロシジャでの順序カテゴリカル経時データの解析仕様を紹介すると共に、シミュレーションデータを用いて性能評価を行う。

### 文献

- Heagerty, P. J., & Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association*, 91(435), 1024-1036.
- Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13-22.
- Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 1033-1048.



## Multi-Regional Clinical Trialの生存時間解析における地域ごとのイベント発現数の予測

○淀 康秀<sup>1</sup>、浜田 知久馬<sup>2</sup>

(<sup>1</sup>大日本住友製薬株式会社、<sup>2</sup>東京理科大学)

ICH-E17ガイドラインで取り扱われるMulti-Regional Clinical Trial (MRCT) は数年前まで国際共同治験あるいは多地域共同治験として日本でも多くの検討がされてきた。MRCTでは多くの国(地域)が治験に参加するが、被験者の登録状況は必ずしも均一的にはならない。仮にアメリカ、カナダ、イギリス、日本の4ヵ国で試験を計画しても、全ての国で被験者の登録が一斉にスタートするとは考えにくい。更に登録速度の違いや、特定地域で目標とする被験者数が設定されているケースもあり得る(例えば日本人だけで100名は登録させるといった非競合的な被験者登録など)。

本発表では、有効性評価項目がTime-to Event (Overall Survivalなど)である抗がん剤開発を目的としたMRCTを想定し、特定のイベント発現数が観察された時点で中間解析を実施することも考える。この中間解析は開発プロジェクトの中でも大きな関心事であり、その実施時期はある程度の確度をもって予想しておくことが求められる。そこで地域ごとの登録状況の違いを考慮した経時的なイベント発現数を予測するシミュレーションプログラムをSASで作成した。地域ごとの登録速度、登録開始時期をシミュレーション条件として設定し、競合的/非競合的な被験者登録も地域単位で設定可能とした。シミュレーション結果は地域ごとのイベント発現数をグラフ及び表形式で出力させた。

SASユーザー総会ではシミュレーション結果として地域間の登録状況を考慮した場合と考慮しなかった場合の中間解析時期の予測結果の違い、中間解析時の地域ごとのイベント発現数、シミュレーションプログラムの今後の拡張等について発表する。

# 用量設定試験における MCP-Mod の 3 つの実行例 ～ 多重比較と用量反応関係のモデリング ～

○加藤 雅章、清水 康平、佐々木 英麿、舟尾 暢男  
武田薬品工業株式会社 日本開発センター 生物統計室

Three Examples of Conducting MCP-Mod in Dose-Finding Studies

Masafumi Katou, Kouhei Shimizu, Emma Sasaki, Nobuo Funao  
Takeda Pharmaceutical Company, Ltd.

## 要旨

「多重比較法」と「用量－反応モデルを用いた解析」を組み合わせた手法「MCP-Mod」を SAS で実行する 3 つの方法 (EAST 6.3 PROCs for SAS、SAS & R、SAS のみ) を紹介する。

キーワード: 用量設定試験、多重比較法、用量－反応モデルを用いた解析、最適な対比係数、EAST 6.3 PROCs for SAS、R

## 1. はじめに

「MCP-Mod (Multiple Comparison Procedure - Modelling)」は、主に第 2 相試験における有効性評価に適用され、「多重比較法」と「用量－反応モデルを用いた解析」を組み合わせた手法である。本手法を用いることで、多重性を調整した上で目的とする適応に対する薬剤の用量－反応関係を明らかにすることが出来、第 3 相 (検証的) 試験で用いる適切な用量 (Target Dose) を選択することも出来る。また、内部補完により、第 2 相試験で設定していない用量に関する推定も実施することが可能となる。本稿では、MCP-Mod を SAS で実行するための 3 つの方法 (EAST 6.3 PROCs for SAS の `mcpmod` プロシジャを用いる方法、SAS と R を組み合わせる方法、SAS のみを用いる方法) を紹介する。

## 2. MCP-Mod とは

ICH-E8 ガイドライン「臨床試験の一般指針」では、第 2 相試験に関する以下の記述がある (一部改変)。

第 2 相の重要な目的は、第 3 相で行われる試験の用法・用量を決定することである。第 2 相の初期的試験では、用量反応の初期的推測のために、用量の漸増デザインがしばしば用いられる。それに続く試験では、並行用量反応デザイン (第 3 相で実施されることもある) を用いて目的とする適応に対する用量－反応関係が確認されることとなる。

有効性評価に論点を絞ると、上記ガイドラインにおける第 2 相試験の目的は以下となる。

- 薬剤の有効性を示すとともに、有効な用量を明らかにする
- 目的とする適応に対する薬剤の用量－反応関係を明らかにする
- 第 3 相 (検証的) 試験で用いる適切な用量を選択する

最近の第 2 相試験では、共分散分析 (ANCOVA) や MMRM (Mixed Models for Repeated Measures) と多重比較法を組み合わせる手法がよく用いられるが、これらの解析では通常、投与量を質的変数として解析が行われるため、試験で設定した用量に関する推測しか行えないこととなる。また、用量間の情報を用いる目的で、

用量-反応モデル（例えば  $E_{\max}$  モデル）を用いて解析を行う場合、通常、事前に 1 つのモデルを選択することとなり、推測の妥当性は「適切なモデルを選択するかどうか」に依存することとなる。

さて、MCP-Mod は Bretz et. al. (2005) により提案された手法で、「多重比較法」と「用量-反応モデルを用いた解析」を組み合わせたものとなっている。本手法を用いることで、以下が達成出来る。

- 多重性を調整した上で、目的とする適応に対する薬剤の用量-反応関係を明らかにする
- 第 3 相試験で用いる適切な用量 (Target Dose) を選択することが出来、内部補完により、試験で設定していない用量に関する推定も実施可能

まず、応答変数（連続量）を  $y_{ij}$ 、用量を  $d_i$ 、用量  $d_i$  の例数を  $n_i$ 、用量の数を  $k$  とする。

- $i = 1, \dots, k$  ( $d_1$  をプラセボとする)
- $j = 1, \dots, n_i$  (被験者)
- $n = \sum_{i=1}^k n_i$  (試験全体の例数)

次に、以下の一元配置分散分析モデルを考える。

$$y_{ij} = \mu_{d_i} + \varepsilon_{ij}$$

ある用量-反応モデル  $f(\cdot)$  について、用量  $d_i$  の平均応答を  $\mu_{d_i} = f(d_i, \theta)$  とし、 $\varepsilon_{ij} \sim N(0, \sigma^2)$  を誤差項とする。ここで、多くのパラメトリックな用量-反応モデル  $f(d, \theta)$  は以下の様に表現出来る。なお、用量-各モデルに対する最適な対比係数を算出する際は、 $f(\cdot)$  の代わりに基準化された用量-反応モデル  $f^0(\cdot)$  を用いても良いことに留意する。本稿では表 2.1 の用量-反応モデルを考える。

$$f(d, \theta) = \theta_0 + \theta_1 f^0(d, \theta^*)$$

- $\theta_0$ : 位置パラメータ
- $\theta_1$ : 尺度パラメータ
- $f^0(d, \theta^*)$ : 基準化された用量-反応モデル ( $\theta^*$  が決まればモデル曲線の形が決まる)

表 2.1 用量-反応モデル

モデル	$f(d, \theta)$	$f^0(d, \theta^*)$
$E_{\max}$	$E_0 + E_{\max} d / (ED_{50} + d)$	$d / (ED_{50} + d)$
Linear log-dose	$E_0 + \delta \log(d + c)$	$\log(d + c)$
Linear	$E_0 + \delta d$	$d$
Exponential	East: $E_0 + E_1 \exp\left(\frac{d}{\delta}\right)$ , East 以外: $E_0 + E_1 \left\{ \exp\left(\frac{d}{\delta}\right) - 1 \right\}$	East: $\exp\left(\frac{d}{\delta}\right)$ , East 以外: $\exp\left(\frac{d}{\delta}\right) - 1$
Quadratic	$E_0 + \beta_1 d + \beta_2 d^2$	$d + \frac{\beta_2}{ \beta_1 } d^2$
Logistic	$E_0 + E_{\max} / \left\{ 1 + \exp\left(\frac{ED_{50} - d}{\delta}\right) \right\}$	$1 / \left\{ 1 + \exp\left(\frac{ED_{50} - d}{\delta}\right) \right\}$
BetaMod	$E_0 + E_{\max} B(\delta_1, \delta_2) \left(\frac{d}{D}\right)^{\delta_1} \left(1 - \frac{d}{D}\right)^{\delta_2}$	$B(\delta_1, \delta_2) \left(\frac{d}{D}\right)^{\delta_1} \left(1 - \frac{d}{D}\right)^{\delta_2}$
Sigmoid $E_{\max}$	$E_0 + E_{\max} d^h / (ED_{50}^h + d^h)$	$d^h / (ED_{50}^h + d^h)$

$$c = 0.01 \times \text{maxdose}, D = 1.2 \times \text{maxdose}, B(\delta_1, \delta_2) = (\delta_1 + \delta_2)^{\delta_1 + \delta_2} / (\delta_1^{\delta_1} \delta_2^{\delta_2})$$

### 3. MCP-Mod を行う手順

さて、MCP-Mod にて解析を行う手順は以下の通りである。

- ① 候補となる用量-反応モデルをいくつか設定する
- ② 最適な対比係数 (optimum contrast coefficients) を計算する
- ③ 有意な用量-反応関係があるかどうかを検定する
- ④ ③の結果が有意であれば、最適なモデルを選択する
- ⑤ 用量-反応関係の推定や Target Dose の推定等を行う

#### ① 候補となる用量-反応モデルをいくつか設定する

まず、候補となる用量-反応モデルの数を  $M$  とし、 $f_m(d, \theta_m)$ ,  $m = 1, \dots, M$  を考える。ここで、実際に解析を行う際は  $f^0(d, \theta_m^*)$  を用い、パラメータ  $\theta_m^*$  にて形状が決まる点に留意する。後で、各モデルについて、次項で定義する対比係数  $c_{mi}$  を用いた以下の統計量にて検定を行うこととなる。

$$t_m = \frac{\sum_{i=1}^k c_{mi} \bar{y}_i}{s \sqrt{\sum_{i=1}^k c_{mi}^2 / n_i}}, \quad m = 1, \dots, M$$

$$s^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 / (n - k), \quad \bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}, \quad \sum_{i=1}^k c_{mi} = 0$$

#### ② 最適な対比係数 (optimum contrast coefficients) を計算する

各モデルに対して検出力が最大となるような対比係数  $c_{mi}$  を計算する。まず、 $m$  番目のモデルに関する平均応答を以下とする。

$$(\mu_{m1}^0, \dots, \mu_{mk}^0) = (f_m^0(d_1, \theta_m^*), \dots, f_m^0(d_k, \theta_m^*))$$

$m$  番目のモデルに関する最適な対比係数  $c_{mi}$  は以下に比例することとなる。また、下記で対比係数を計算した後、 $\sum_{i=1}^k c_{mi}^2 = 1$  になるよう調整すると後々都合が良い。

$$n_i(\mu_{mi}^0 - \bar{\mu}), \quad i = 1, \dots, k; \quad \bar{\mu} = n^{-1} \sum_{i=1}^k \mu_{mi}^0 n_i$$

#### ③ 有意な用量-反応関係があるかどうかを検定する

Max-t 統計量  $t_{max} = \max\{t_1, \dots, t_M\}$  に基づいて有意性を判定する。

- 帰無仮説  $H_0: \mu_{d_1} = \dots = \mu_{d_k}$  の下で、検定統計量  $(t_1, \dots, t_M)$  は自由度  $n - k$ 、以下の相関行列をもつ多変量  $t$  分布に従う。

$$\rho_{ij} = \frac{\sum_{l=1}^k c_{il} c_{jl} / n_l}{\sqrt{\sum_{l=1}^k c_{il}^2 / n_l \sum_{l=1}^k c_{jl}^2 / n_l}}$$

- $t_{max} > u_{1-\alpha}$  (多変量  $t$  分布から求めた  $100(1-\alpha)\%$  棄却点) ならば、有意であると判定する。

#### ④ ③の結果が有意であれば、最適なモデルを選択する

「AIC が最も小さいモデル」又は「 $t$  統計量が最も大きいモデル」のいずれかの基準を用いて最適なモデル

を選択する<sup>1</sup>。

⑤ 用量-反応関係の推定や Target Dose の推定等を行う

最適なモデルについてパラメータ推定を行い、モデルを同定する。また、推定後のモデルについて、以下の基準により Target Dose の推定を行う。

- $\Delta$  を「プラセボに対して臨症的に意味のある差」とすると、TD (Target Dose) は以下のとおり。
  - ◇  $TD_{\Delta} = \min\{d \in (d_1, d_k): f(d) > f(d_1) + \Delta\}$
- なお、本稿では TD (Target Dose) に注目することとするが、 $p$  を「最大効果に対して何%以上の効果を期待するかを示す割合」として、ED (Effective Dose) を推定することも可能。
  - ▷  $ED_p = \min\{d \in (d_1, d_k): f(d) > f(d_1) + p(\max_{d \in (d_1, d_k)} f(d) - f(d_1))\}$

4. 準備：データセット「biom」の紹介と候補モデル

Bretz et. al. (2005) にて用いられたサンプルデータ「biom (biometrics dose response data)」を紹介する。変数は「subjectID (被験者番号)」「dose (用量)」「resp (応答変数、連続量)」、用量は 0.00 (プラセボ)、0.05、0.20、0.60、1.00 の 5 用量、レコード数は各用量 20、計 100 レコードとなっている。

表 4.1 データセット「biom」の抜粋

subjectID	dose	resp
1	0.00	0.354621644
2	0.00	0.136528014
3	0.00	-0.017221273
:	:	:
100	1.00	0.948335871

データセット「biom」の要約統計量を表 4.2 に、平均値のグラフを図 4.1 にて紹介する。

表 4.2 データセット「biom」の要約統計量

用量	例数	平均	標準偏差	最小値	最大値
0.00	20	0.34	0.52	-0.45	1.56
0.05	20	0.46	0.49	-0.36	1.53
0.20	20	0.81	0.74	-0.52	2.45
0.60	20	0.93	0.76	0.06	2.92
1.00	20	0.95	0.95	-1.11	2.56

次項以降では、本データセットに対して MCP-Mod を適用し、薬剤の有効性を示し、薬剤の用量-反応関係を明らかにするとともに、第 3 相試験で用いる適切な用量 (Target Dose) を選択することを考える。候補となる用量-反応モデルは図 4.2 に示すとおり以下の 6 つ ( $E_{\max}$  ( $ED_{50} = 0.2$ )、Linear log-dose、Linear、Exponential

( $\delta = 1.13$ )、Quadratic ( $\frac{\beta_2}{\beta_1} = -0.73$ )、Logistic ( $ED_{50} = 0.5, \delta = 0.13$ ) とする。

<sup>1</sup> AIC の値に基づいて、有意となったモデルに関する重み付け平均をとり、結果を統合するという方法もある。

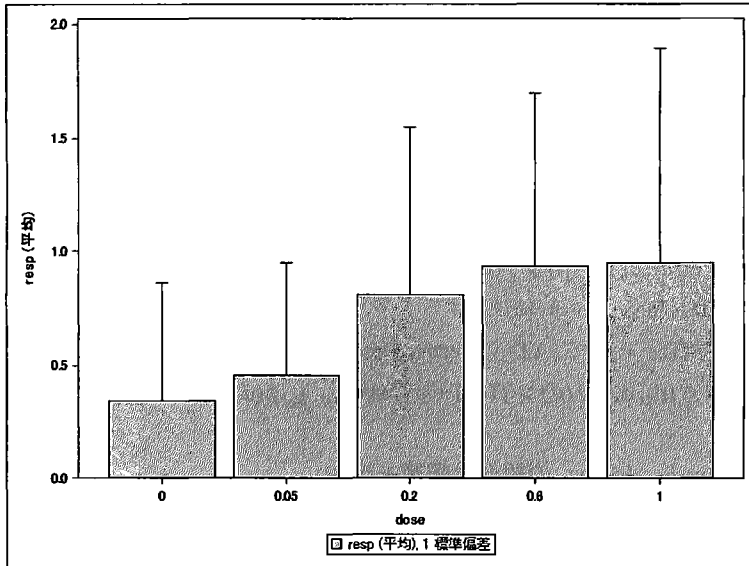


図 4.1 データセット「biom」の平均値のグラフ

### 5. 実行例① : EAST 6.3 PROCs (32-bit) for SAS による解析

まず、Cytel 社 ([www.cytel.com/](http://www.cytel.com/)) が開発したソフトウェア「East」にて提供されている mcpmod プロシジャを用いる方法を紹介します。East では、SAS ユーザーが臨床試験の計画と解析を行うためのプロシジャを「EAST PROCs」という形で提供しており、現在「eastmonitor」と「mcpmod」の2つのプロシジャがある。本プロシジャは East がインストールされていない SAS では実行できない点に留意されたい。

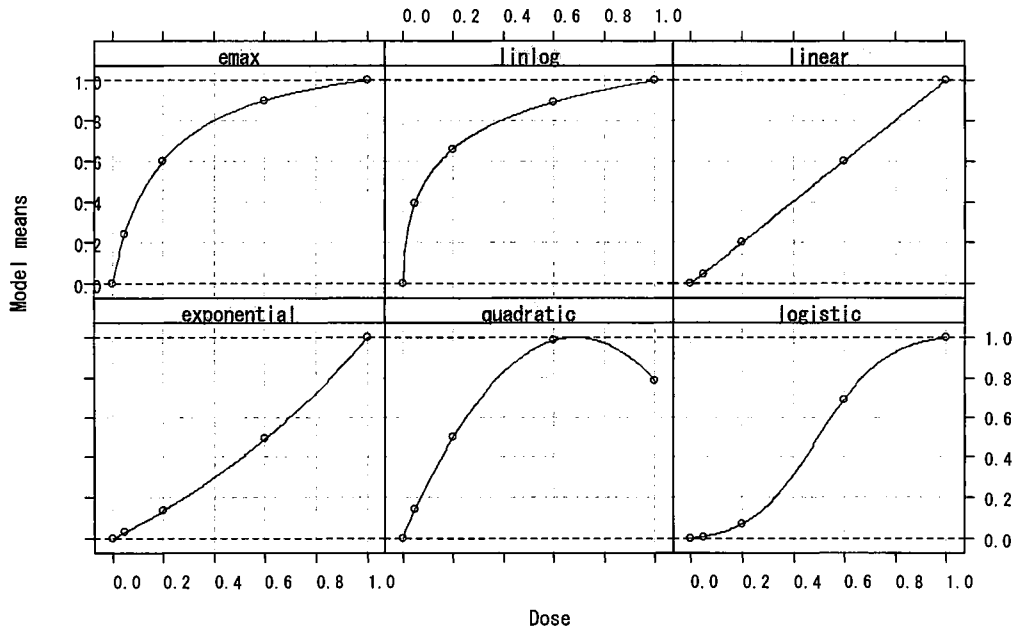


図 4.2 候補となる用量-反応モデル

まず、mcpmod プロシジャのコードは以下の通り。

#### プログラム 5.1 mcpmod プロシジャの実行

```
proc mcpmod data=biom out=_overall compmethod=pval alpha=0.025 direction=increasing ;
  response resp ;
  dose dose ;
  modeldoses 0 0.05 0.2 0.6 1 ;
  emax: model emaxs(0.2) ;
  linlog: model lin_ldoses ;
  linear: model linears ;
  expn: model expns(1.13) ;
  quad: model quads(-0.73) ;
  logit: model logits(0.5,0.13) ;
  dosesel td (delta=0.5) ;
  selmodel aic ;
  dossumry out=_dossumry ;
  optcont out=_optcont ;
  cormat out=_cormat ;
run ;
```

次に、mcpmod プロシジャで出てくるステートメントとオプションを概説する。

#### PROC MCPMOD ステートメント

##### 【指定が必須】

- DATA：使用するデータセットを指定
- ALPHA：棄却限界値又は調整済み p 値の算出に使用する有意水準を指定

##### 【指定が任意】

- COMPMETHOD
  - モデルの選択に使う手法を指定（以下の2通りから選択可能）
    - PVAL：調整済み p 値が有意なモデルが選択される（デフォルトは"PVAL"）
    - CRITV：指定された棄却限界値に基づいてモデルが選択される（棄却限界値を自動的に計算させるか、自ら指定することが可能）
- DIRECTION
  - 棄却限界域を指定（"INCREASING"又は"DECREASING"の方向を選択、デフォルトは"INCREASING"）

#### RESPONSE ステートメント

##### 【指定が必須】

- 応答変数を指定

#### DOSE ステートメント

##### 【指定が必須】

- 用量の変数を指定

#### MODELDOSES ステートメント

##### 【指定が必須】

- 目的用量（Target Dose）を推定するにあたって候補となる用量のリスト
- プラセボは0で表示、その他の用量は昇順に番号を振られていなければならない

## MODEL SPECIFICATION ステートメント

### 【指定が必須】

- モデルを指定（フルモデル又は標準化されたモデル、最小1個、最大20個のモデルが指定可能）
  - 形：「&名前：MODEL &モデル (&パラメータ)」
    - 例：emax: model emaxs(0.2);
  - &名前は好きなラベルをつけてよい
  - &モデルと&パラメータは下表の「SAS コード」を参照

表 5.1 モデルと SAS コードの対応表

モデル	$f^0(d, \theta^*)$	SAS コード
E <sub>max</sub>	$d/(ED_{50} + d)$	EmaxS (ED50)
Linear log-dose	$\log(d + c)$	Lin_LdoseS
Linear	$d$	LinearS
Exponential	$\exp\left(\frac{d}{\delta}\right)$	ExpnS (delta)
Quadratic	$d + \frac{\beta_2}{ \beta_1 } d^2$	QuadS (beta2 /  beta1 )
Logistic	$1/\left\{1 + \exp\left(\frac{ED_{50} - d}{\delta}\right)\right\}$	LogitS (ED50, delta)
BetaMod	$B(\delta_1, \delta_2) \left(\frac{d}{D}\right)^{\delta_1} \left(1 - \frac{d}{D}\right)^{\delta_2}$	Beta_ModS (delta1, delta2)
Sigmoid E <sub>max</sub>	$d^h/(ED_{50}^h + d^h)$	SigEmaxS (ED50, h)

$$c = 0.01 \times \text{maxdose}, D = 1.2 \times \text{maxdose}, B(\delta_1, \delta_2) = (\delta_1 + \delta_2)^{\delta_1 + \delta_2} / (\delta_1^{\delta_1} \delta_2^{\delta_2})$$

## DOSESEL ステートメント

### 【指定が必須】

- Target Dose 等の推定に用いる基準を指定
  - DOSESEL=TD (DELTA=X Y Z)
    - プラセボと比較して Δ の目的効果が得られる用量
  - DOSESEL=ED (P=X.XX, Y.YY, Z.ZZ)
    - プラセボと比較したときの最大効果のうち (100×p) % の効果が得られる用量

## SELMODEL ステートメント

### 【指定が任意】

- モデル選択の方法を指定（指定しない場合は全通りの結果が出力される）
  - AIC : AIC が最小のモデルが選択される
  - MAXT : t 統計量が最大のモデルが選択される
  - AVGAIC : 有意となったモデルの重み付け平均値を計算し使用する

さて、mcpmod プロシジャを実行すると、様々な解析結果が表示される。まず、MCP-Mod を行う事前情報「データセットに関する要約統計量」「最適な対比係数」「対比係数間の相関行列」が表示されるが、ここでは「最適な対比係数」のみを紹介する。



実行結果 5.1 最適な対比係数

OPTIMAL CONTRAST COEFFICIENTS						
Dose	EMAX	LINLOG	LINEAR	EXPN	QUAD	LOGIT
0.0000	-0.6431	-0.7278	-0.4367	-0.3939	-0.5785	-0.3839
0.0500	-0.3615	-0.2470	-0.3776	-0.3562	-0.4098	-0.3732
0.2000	0.0610	0.0891	-0.2006	-0.2325	0.0195	-0.3053
0.6000	0.4131	0.3752	0.2714	0.1902	0.6021	0.3642
1.0000	0.5305	0.5105	0.7435	0.7923	0.3668	0.6982

次に、MCP-Mod の実行結果のうち、MCP（多重比較）部分の結果を紹介する。本結果は「有意な用量-反応関係があるかどうかの検定結果」となっており、モデルごとに多変量 t 分布から求めた片側 p 値及び調整済み片側 p 値が出力されている。ここでは、調整済み p 値がすべて 0.025 より小さいため、どのモデルも「有意な用量反応関係が検出された」とし、次のステップ（Mod 部分）へ進むことが可能となっている。

実行結果 5.2 多重比較結果

MULTIPLE CONTRAST TEST				
	t-Stat	Raw p-value	Adj. p-value	
EMAX	3.4641	0.0004	0.0008	
LINLOG	3.4106	0.0005	0.0011	
LINEAR	2.9715	0.0019	0.0041	
EXPN	2.7780	0.0033	0.0073	
QUAD	3.3877	0.0005	0.0012	
LOGIT	2.8404	0.0028	0.0059	

今回の解析では、モデル選択の方法として AIC（AIC が最小のモデルを選択）を指定しているため、最適なモデルとして AIC が最も小さい「Linear log-dose」モデルが選択されている。

実行結果 5.3 モデル選択結果

MODEL SELECTION	
Model family	AIC
Emax	219.1785
LinLog	217.6543
Linear	220.5389
Exponential	223.1707
Quadratic	219.7595
Logistic	220.8691

Selected Model: **LinLog**

mcpmod プロシジャでは、モデル選択の結果によらず全てのモデルに対してパラメータ推定が行われるが、ここでは簡単のため最適なモデルとして選択された「Linear log-dose」モデルの推定結果のみ紹介する。結果は「 $f(d, \theta) = 0.9749 + 0.1458 \log(d + 0.01)$ 」と推定されている。

実行結果 5.4 「Linear log-dose」モデルに対するパラメータ推定結果

DOSE RESPONSE MODEL FITTING				
Fitted Model 2:				
Model family: <b>LinLog</b>				
Parameter Estimates:				
	Estimate	Std. Error		
E0	0.9749	0.1065		
Delta	0.1458	0.0422		
Residuals:				
Min	1Q	Median	3Q	Max
-2.0216	-0.4058	0.1433	0.4302	2.0894
Residual standard error: 0.7041				
Degrees of freedom: 98				

最後に、Target Dose の推定結果を紹介する。本結果では、モデル選択の結果によらず全てのモデルにおいて Target Dose の推定値が計算され、指定した  $\Delta$  ごとに推定値のリストを出力する。最適なモデルとして選択された「Linear log-dose」モデルについては、Target Dose が 0.2988 と推定されている。

実行結果 5.5 Target Dose の推定結果

DOSE ESTIMATION		
-----		
Table 1: For Delta =	0.5000	
Model family	Estimated TD (cont)	Estimated TD (disc)
EMax	0.2886	0.6000
LinLog	0.2988	0.6000
Linear	0.8951	1.0000
Exponential	0.9402	1.0000
Quadratic	0.3871	0.6000
Logistic	0.2255	0.6000

## 6. 実行例②：SAS と R を組み合わせた解析

前項で紹介した mcpmmod プロシジャは、East がインストールされていない SAS では実行できない。ところで、フリーで使うことが出来る統計ソフト「R」には「DoseFinding」パッケージがあり、本パッケージの関数を用いることで MCP-Mod を実行することが出来る。本項では SAS から R を呼び出して MCP-Mod を実行し、4~5 項と同じ解析を行う方法を紹介する。

### 若干の準備

I. 以下のサイトより R をダウンロードし、インストールする (Ver. 3.2.5 を想定)

<http://ftp.yz.yamagata-u.ac.jp/pub/cran/bin/windows/base/>

II. R を「管理者権限」で起動した後、以下の命令を実行する

```
install.packages("DoseFinding", dep=T)
```

III. C ドライブに作業フォルダを準備する (本稿では C ドライブの直下に「temp」フォルダを作成)

IV. Appendix で紹介するマクロ「%MYMCPMOD」を定義する。

プログラム 6.1 マクロ「%MYMCPMOD」の引数

```

%macro MYMCPMOD (doses      = ,
                  models     = ,
                  maxEff     = 1,
                  doseval    = ,
                  respval    = ,
                  dataset    = ,
                  S          = NULL,
                  type       = normal,
                  addCovars  = ~1,
                  placAdj    = FALSE,
                  selModel   = AIC,
                  alpha      = 0.025,
                  df         = NULL,
                  critV      = TRUE,
                  doseType   = TD,
                  Delta      = 0.5,
                  p          = 0.8,
                  pVal       = TRUE,
                  alternative = one, ...);
    
```

表 6.1 マクロ「%MYMCPMOD」の引数

引数名	機能
doses	用量をカンマ区切りで指定
models	「モデル名 = パラメータ値」の組み合わせをカンマ区切りで指定 (モデル名及びパラメータ値に関しては、表 6.2 を参照) パラメータが 2 個ある場合は c(パラメータ 1, パラメータ 2)の形式で指定
maxEff	プラセボに対する最大効果 1 ならば応答変数が正の反応、-1 ならば応答変数が負の反応を示すことを仮定
doseval	用量を表す変数名を指定
respval	応答変数を表す変数名を指定
dataset	SAS データセット名を指定
addCovars	共変量を追加する場合は「~ 変数名」の形で指定
placAdj	プラセボ調整後のデータを使用する場合は TRUE を指定
selModel	モデル選択時の指標 (AIC、maxT、aveAIC) を指定
alpha	有意水準を指定
critV	棄却限界値を指定
doseType	最適な用量を選択する際の指標 (TD、ED) を指定
Delta	「プラセボと比較して $\Delta$ の目的効果が得られる用量」を決める際の $\Delta$
p	「プラセボと比較した時の最大効果のうち $(100 \times p)$ %の効果が得られる用量」を決める際の p
pVal	p 値を計算する場合は TRUE、しない場合は FALSE を指定
alternative	片側検定 (one) 又は両側検定 (two) を指定

表 6.2 引数 models に指定するパラメータ

モデル名	$f^0(d, \theta^*)$	パラメータ値
emax	$d / (ED_{50} + d)$	ED <sub>50</sub>
linlog	$\log(d + c)$	NULL
linear	$d$	NULL

モデル名	$f^0(d, \theta^*)$	パラメータ値
exponential	$\exp\left(\frac{d}{\delta}\right) - 1$	$\delta$
quadratic	$d + \frac{\beta_2}{ \beta_1 } d^2$	$\delta (= \beta_2 /  \beta_1 )$
logistic	$1 / \left\{ 1 + \exp\left(\frac{ED_{50} - d}{\delta}\right) \right\}$	$c(ED_{50}, \delta)$
betaMod	$B(\delta_1, \delta_2) \left(\frac{d}{D}\right)^{\delta_1} \left(1 - \frac{d}{D}\right)^{\delta_2}$	$c(\delta_1, \delta_2)$
sigEmax	$d^h / (ED_{50}^h + d^h)$	$c(ED_{50}, h)$

$$c = 0.01 \times \text{maxdose}, D = 1.2 \times \text{maxdose}, B(\delta_1, \delta_2) = (\delta_1 + \delta_2)^{\delta_1 + \delta_2} / (\delta_1^{\delta_1} \delta_2^{\delta_2})$$

### マクロ%MYMCPMODの実行

データセット「biom」に対してマクロ「%MYMCPMOD」を実行し、MCP-Modによる解析を行う<sup>2</sup>。

#### プログラム 6.2 マクロ「%MYMCPMOD」の実行

```
%MYMCPMOD(
doses      = %str(0, 0.05, 0.2, 0.6, 1),
models     = %str(emax=0.2, linlog=NULL, linear=NULL, exponential=1.13, quadratic=-0.73,
                 logistic=c(0.5, 0.13)),
doseval    = dose,
respval    = resp,
dataset    = biom,
selModel   = AIC,
alpha      = 0.025,
doseType   = TD,
Delta      = 0.5,
pVal       = TRUE,
alternative = one
);
```

結果は前項の mcpmod プロシジャのものと同様であることと頁数の制限のため出力の詳細は行わないが、まず作業フォルダ（例えば「C:\temp」）に解析結果となる「output.txt」が出力される。

#### 実行結果 6.1 output.txt の中身（抜粋）

```
MCPMod

*****
MCP part
*****
Multiple Contrast Test
```

<sup>2</sup> 引数 seq に順番を指定する際、%str() 中で指定する内容を途中で改行しないこと（エラーとなる）。

Contrasts:

	emax	linlog	linear	exponential	quadratic	logistic
0	-0.643	-0.728	-0.437	-0.394	-0.578	-0.384
0.05	-0.361	-0.247	-0.378	-0.356	-0.410	-0.373
0.2	0.061	0.089	-0.201	-0.232	0.019	-0.305
0.6	0.413	0.375	0.271	0.190	0.602	0.364
1	0.530	0.510	0.743	0.792	0.367	0.698

Contrast Correlation:

	emax	linlog	linear	exponential	quadratic	logistic
emax	1.000	0.989	0.912	0.867	0.965	0.884
linlog	0.989	1.000	0.875	0.830	0.937	0.837
linear	0.912	0.875	1.000	0.994	0.840	0.988
exponential	0.867	0.830	0.994	1.000	0.774	0.978
quadratic	0.965	0.937	0.840	0.774	1.000	0.844
logistic	0.884	0.837	0.988	0.978	0.844	1.000

Multiple Contrast Test:

	t-Stat	adj-p
emax	3.464	< 0.001
linlog	3.411	< 0.001
quadratic	3.388	0.00131
linear	2.972	0.00382
logistic	2.840	0.00612
exponential	2.778	0.00709

Critical value: 2.276 (alpha = 0.025, one-sided)

(以下略)

また、作業フォルダ (例えば「C:\temp」) に「output.pdf」が出力され、MCP-Mod の候補としたモデルと、これらのモデルに対してパラメータ推定を行い、モデルの当てはめを行った結果が表示される。

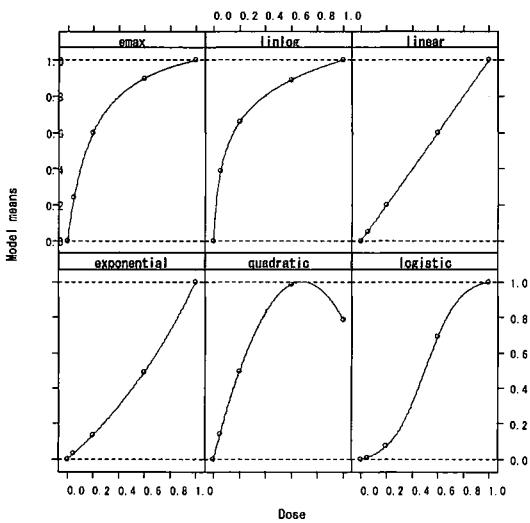


図 6.1 候補となるモデル

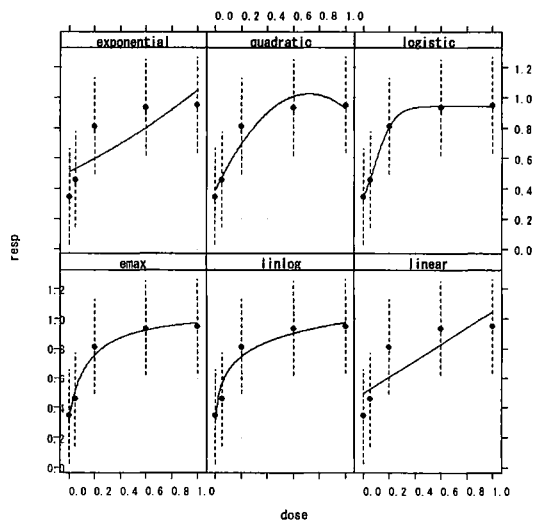


図 6.2 当てはめ結果

ところで、各モデルのパラメータ（例えば  $E_{\max}$  モデルのパラメータ  $ED_{50}=0.2$ ）の決め方については言及しなかったが、実際の試験デザインを検討する際、値を特定するのは結構大変である。R では「0.2mg ならば最大効果の 30%の効果が期待出来る」等の事前情報から、モデルのパラメータを推定する関数があるので、SAS でもこの機能を使えるようにするため、下記のマクロ「%MYGUESS」を定義した。書式例は以下の通り。

- %MYGUESS(d=0.3, p=0.8, model=emax, Maxd=1) ;
  - 用量 0.3 で 80% の効果が得られると仮定したときの  $E_{\max}$  モデル
- %MYGUESS(d=%str(0.2, 0.5), p=%str(0.6, 0.95), model=logistic) ;
  - 用量 0.2 で 60%、用量 0.5 で 95% の効果が得られると仮定したときの Logistic モデル
- %MYGUESS(d=0.5, p=0.5, model=betaMod, dMax=0.8, scal=1.2, Maxd=1) ;
  - 用量 0.5 で 50% の効果が得られると仮定したときの BetaMod モデル
- %MYGUESS(d=%str(0.2, 0.5), p=%str(0.6, 0.95), model=sigEmax) ;
  - 用量 0.2 で 60%、用量 0.5 で 95% の効果が得られると仮定したときの Sigmoid  $E_{\max}$  モデル
- %MYGUESS(d=0.7, p=1, model=quadratic) ;
  - 用量 0.7 で 100% の効果が得られると仮定したときの Quadratic モデル
- %MYGUESS(d=0.8, p=0.5, model=exponential, Maxd=1) ;
  - 用量 0.8 で 50% の効果が得られると仮定したときの Exponential モデル

さらに、各モデルの検出力を求めるマクロ「%MYPPOWER」と例数設計を行うマクロ「%MYSAMPLE」の定義も合わせて Appendix にて紹介する。頁数の制限のため、引数の説明及び実行例はユーザー一會当日の発表にて紹介する。

## 7. 実行例③：SAS による解析

Menon et. al. (2015) では、普通の SAS にて MCP-Mod を行う方法が紹介されており、最適な対比係数を求めるマクロ「%optcont」が紹介されている。これを用いて glimmix プロシジャ等で MCP-Mod が実行出来るのだが、本マクロでは使用出来るモデルが前述の 2 つの方法と比較して若干少なくなることに、実際の解析で用いた分散共分散行列を使った対比の計算が出来ない。そこで本項では、最適な対比係数を求める新たなマクロ「%MYCONTRAST」を定義した上で、Menon et. al. (2015) で紹介されているものと同様の方法にて、4～5 項と同じ解析を実施する。MCP-Mod を実行することにする。まず、マクロ「%MYCONTRAST」の引数は以下の通り<sup>3</sup>。

プログラム 7.1 マクロ「%MYCONTRAST」の引数

```
%macro MYCONTRAST (models=%str('emax' 'linear' 'linlog' 'exponential' 'quadratic' 'logistic'
                               'betaMod' 'sigEmax'),
  params=%str(0.3 . . . . . 0.7 . . -0.85 . . 0.3 0.1,
              0.3 1.3, 0.5 2.0),
  doses =%str(0.00 0.05 0.20 0.60 1.00),
  w      =%str(1),
  S      =0) ;
```

<sup>3</sup> %str() 中で指定する値について、途中で改行しないこと（エラーとなる）。

表 7.1 マクロ「%MYMCPMOD」の引数

引数名	機能
model	モデルをスペース区切りで設定（詳細は次表）
params	各モデルのパラメータをカンマ区切りで指定（詳細は次表）
doses	用量をスペース区切りで指定
w	各用量の例数比（重み）を不等例数にする場合は、スペース区切りで指定（指定は任意；例：w=%str(2 1 1 1 2)）
S	実際の解析で用いた分散共分散行列を用いる場合は、データセットを予め準備した上で、データセット名を指定（指定は任意）

表 7.2 引数 models と引数 params に指定するパラメータ

models に指定するパラメータ	$f^0(d, \theta^*)$	params に指定するパラメータ
'emax'	$d / (ED_{50} + d)$	ED50 .
'linlog'	$\log(d + c)$	. .
'linear'	$d$	. .
'exponential'	$\exp\left(\frac{d}{\delta}\right) - 1$	delta .
'quadratic'	$d + \frac{\beta_2}{ \beta_1 } d^2$	beta2 /  beta1  .
'logistic'	$1 / \left\{ 1 + \exp\left(\frac{ED_{50} - d}{\delta}\right) \right\}$	ED50 delta
'betaMod'	$B(\delta_1, \delta_2) \left(\frac{d}{D}\right)^{\delta_1} \left(1 - \frac{d}{D}\right)^{\delta_2}$	delta1 delta2
'sigEmax'	$d^h / (ED_{50}^h + d^h)$	ED50 h

$c = 0.01 \times \text{maxdose}$ ,  $D = 1.2 \times \text{maxdose}$ ,  $B(\delta_1, \delta_2) = (\delta_1 + \delta_2)^{\delta_1 + \delta_2} / (\delta_1^{\delta_1} \delta_2^{\delta_2})$ 、「.」は欠測を表す

まず、マクロ「%MYCONTRAST」を下記の通り実行すると、文字列マクロ変数「&ctr」に最適な対比係数が格納される。

プログラム 7.2 マクロ「%MYCONTRAST」の実行

```
%MYCONTRAST (models=%str('emax' 'linlog' 'linear' 'exponential' 'quadratic' 'logistic'),
  params=%str(0.2 . . . . . 1.13 . . -0.73 . . 0.5 0.13),
  doses =%str(0.00 0.05 0.20 0.60 1.00)) ;
```

次に、glimmix プロシジャにて MCP-Mod のうち MCP（多重比較）部分を実施することが出来る。

プログラム 7.3 glimmix プロシジャによる MCP-Mod の実施

```
data _null_ ;
call execute ("
proc glimmix data=biom ;
class dose ;
```

```

model resp = dose / noint solution covb ;
estimate &ctr. / adjust=simulate(nsamp=1000000 seed=1) uppertailed ;
run ;
");
run ;

```

上記の実行結果を、マクロ「%MYCONTRAST」の実行結果（最適な対比係数）と共に紹介する。

実行結果 7.1 マクロ「%MYCONTRAST」と glimmix プロシジャの実行結果

SAS システム							
c							
	emax	linlog	linear	exponential	quadratic	logistic	
0	-0.643115	-0.727771	-0.436656	-0.393882	-0.578494	-0.383908	
0.05	-0.361459	-0.247029	-0.377648	-0.356166	-0.409831	-0.373166	
0.2	0.0610255	0.089096	-0.200626	-0.232466	0.0194833	-0.305336	
0.6	0.4130955	0.3752056	0.2714349	0.1901794	0.6020559	0.3642147	
1	0.5304521	0.5104985	0.7434955	0.7923346	0.3667862	0.6981966	

GLIMMIX プロシジャ							
推定値							
多重度の調整 : Simulated							
ラベル	推定値	標準誤差	自由度	t 値	Pr > t	調整済 P	
1 emax	0.5518	0.1593	95	3.46	0.0004	0.0009	
2 linlog	0.5433	0.1593	95	3.41	0.0005	0.0011	
3 linear	0.4733	0.1593	95	2.97	0.0019	0.0041	
4 exponential	0.4425	0.1593	95	2.78	0.0033	0.0071	
5 quadratic	0.5396	0.1593	95	3.39	0.0005	0.0012	
6 logistic	0.4524	0.1593	95	2.84	0.0028	0.0060	

この後、Mod（各モデルのパラメータや Target Dose の推定等）部分の解析を nlmixed プロシジャ等で実施することが出来、例えば最適なモデルとして選択された「Linear log-dose」モデルのパラメータ推定は以下のコードで実行出来るが、これ以降の詳細な解説は Menon et. al. (2015) に譲る。

プログラム 7.4 nlmixed プロシジャによる「Linear log-dose」モデルのパラメータ推定

```

title "Linear log-dose model" ;
proc nlmixed data=biom ;
parms E0=0.01 Delta=0.01 SD=1.0 ;
Mu = E0 + Delta*log(dose+0.01) ;
model resp ~ normal(Mu, SD**2);
run ;

```

## 8. まとめ

本稿では、MCP-Mod を実行するために「EAST 6.3 PROCs (32-bit) for SAS を用いる方法」「SAS と R を組み合わせる方法」「SAS のみで実行する方法」の 3 つの方法を紹介した。1 番目の方法はデータセットのみを指定



するだけで様々な解析結果を出力することが出来る。2番目の方法は、外部ソフト（R）を呼び出す必要がある煩雑さはあるが、グラフ出力が自動で得られ、かつ事前に設定する用量-反応モデルのパラメータ推定や例数設計を行うことが出来る。3番目の方法は、普通のSASのみで解析を行える簡便さがあり、かつ本稿で提案したマクロを適用することで、1番目や2番目の方法と同じ数の用量-反応モデルを使用することが出来る。場面に応じてこれらの方法を使い分けるのが良い。

ちなみに、MCP-Modを適用する際は、下記の点に留意する必要があることを最後に申し添える。

- 前項までの流れで解析を行った場合、対照群（例：プラセボ群）との対比較の結果は得られない。特に、「MCP-Modにて有意となる」検出力を計算する際、「対比較にて有意となる」ことを考慮せずに計算することとなる。
- 本手法を適用する際、経験則として実薬群は4~7群、用量幅は10倍以上が望ましい。
- 設定する用量-反応モデルの数は3~7個が望ましい。

連絡先

[masafumi.katou@takeda.com](mailto:masafumi.katou@takeda.com)

参考文献

- **F. Bretz, J. C. Pinheiro, and M. Branson. (2005),**  
"Combining multiple comparisons and modeling techniques in dose-response studies", *Biometrics*, 61, 738-748.
- **F. Bretz, T. Hothorn, and P. Westfall (2010),**  
"Multiple Comparisons Using R", Chapman and Hall/CRC Press.
- **Pinheiro, J. C., Bornkamp, B., and Bretz, F. (2006),**  
Design and analysis of dose finding studies combining multiple comparisons and modeling procedures, *Journal of Biopharmaceutical Statistics*, 16, 639-656.
- **EMA (2014),**  
"Qualification Opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty".
- **Sandeep M. Menon and Richard C. Zink (2015),**  
"Modern Approaches to Clinical Trials Using SAS: Classical, Adaptive, and Bayesian Methods", SAS Institute.
- **ICH-E8 ガイドライン**  
臨床試験の一般指針
- **EAST 6.3 PROCs for SAS Users**  
MCPMod 1.1 ユーザーマニュアル
- **Bjoern Bornkamp, Jose Pinheiro, Frank Bretz (2015),**  
R: DoseFinding パッケージ・ユーザーマニュアル

以上

## Appendix 本文に掲載していないが引用したプログラム

### プログラム A.1 データセット「biom」

data biom ;	33	0.2	1.233325825	68	1	0.62533415
input subjectID dose resp ;	34	0.2	1.867825704	69	1	1.330258137
cards ;	35	0.2	1.057835057	70	1	0.149454689
1 0.05 1.353800979	36	0.2	0.34534894	71	1	0.887139996
2 0.05 0.154980239	37	0.2	0.404851371	72	1	0.563064088
3 0.05 -0.070838037	38	0.2	0.48420331	73	1	0.730863126
4 0.05 0.583740909	39	0.2	0.571199878	74	1	-0.771473861
5 0.05 0.962598527	40	0.2	1.010696185	75	1	0.494778358
6 0.05 0.381388876	41	0.6	1.172221004	76	1	0.303110251
7 0.05 -0.004695991	42	0.6	0.667558102	77	1	0.425231331
8 0.05 0.519304935	43	0.6	1.777681617	78	1	2.557773784
9 0.05 -0.36332213	44	0.6	0.309735795	79	1	1.12889658
10 0.05 0.338903336	45	0.6	0.05646191	80	1	0.948335871
11 0.05 0.145620717	46	0.6	0.900065687	81	0	0.354621644
12 0.05 0.442803406	47	0.6	0.738302949	82	0	0.136528014
13 0.05 -0.021949664	48	0.6	0.227879049	83	0	-0.017221273
14 0.05 0.37259824	49	0.6	1.385106066	84	0	0.367483563
15 0.05 1.528976072	50	0.6	0.912390582	85	0	0.395127618
16 0.05 0.149471298	51	0.6	0.299300126	86	0	0.357965911
17 0.05 0.374063362	52	0.6	0.278917304	87	0	0.312936459
18 0.05 1.0116782	53	0.6	0.270043807	88	0	-0.044915902
19 0.05 0.281306267	54	0.6	0.569625173	89	0	0.623091387
20 0.05 0.994656155	55	0.6	1.604328006	90	0	0.03951218
21 0.2 1.198081819	56	0.6	1.576944531	91	0	-0.30840357
22 0.2 1.56768394	57	0.6	2.924508962	92	0	-0.452964305
23 0.2 -0.155607279	58	0.6	0.171532376	93	0	-0.201709236
24 0.2 0.20658777	59	0.6	2.160523738	94	0	1.560869091
25 0.2 1.853465242	60	0.6	0.685611744	95	0	0.561286293
26 0.2 0.995682873	61	1	2.248445841	96	0	0.336241137
27 0.2 2.452234157	62	1	2.173688799	97	0	-0.159375068
28 0.2 -0.52039887	63	1	1.254509265	98	0	0.992604728
29 0.2 0.048856237	64	1	1.864626603	99	0	0.806987008
30 0.2 0.633128014	65	1	-1.113082058	100	0	1.237442253
31 0.2 0.528413238	66	1	1.198415091			;
32 0.2 0.422902023	67	1	1.97485829			run ;

### プログラム A.2 データセット「biom」の要約統計量と平均値のグラフ

```
ods output summary=out ;
proc means data=biom nonobs n mean stddev min max ;
  class dose ;
  var resp ;
run ;
ods output close ;

proc sgplot data=biom ;
  vbar dose / response=resp limitstat=stddev limits=upper stat=mean ;
run ;
```

### プログラム A.3 SAS と R を組み合わせて MCP-Mod を行うマクロ「%MYMCPMOD」

Bornkamp et. al. (2015) による R パッケージ「DoseFinding」は、License は GPL-3 (プログラムの調査及び改変、公衆への頒布が可) に従っており、本稿中のマクロは本パッケージ中の関数を参考に作成した。よって、本稿中のマクロの License も GPL-3 とする。以下のプログラム A.4~A.6 も同様とする。

```

%macro MYMCPMOD(doses      = ,
                 models     = ,
                 maxEff     = 1,
                 doseval    = ,
                 respval    = ,
                 dataset    = ,
                 S          = NULL,
                 type       = normal,
                 addCovars  = ~1,
                 placAdj    = FALSE,
                 selModel   = AIC,
                 alpha      = 0.025,
                 df         = NULL,
                 critV      = TRUE,
                 doseType   = TD,
                 Delta      = 0.5,
                 p          = 0.8,
                 pVal       = TRUE,
                 alternative = one.sided,
                 datapath   = 'C:/temp/data.csv',
                 pdfpath    = "C:/temp/output.pdf",
                 outtextpath = "C:/temp/output.txt",
                 programpath = "C:/temp/program.R") ;
options noxwait xsync ;

*--- R の実行ファイルのパス ;
%let Rexepath='C:\Program Files\R\R-3.2.5\bin\i386\R.exe' ;

proc export data=&dataset. outfile=&datapath. dbms=csv replace ;
run ;

filename aaa &programpath. ;
data _null_ ;
  file aaa ;
  put "library(DoseFinding)" ;
  put "d      <- read.csv(&datapath., head=T)" ;
  put "models <- Mods(&models., doses=c(&doses.), maxEff=&maxEff.)" ;
  put "set.seed(777)" ;
  put "results <- MCPMod(&doseval., &respval., d, models, S=&S., " ;
  put "  type=""&type."" , addCovars=&addCovars., " ;
  put "  placAdj=&placAdj., selModel=""&selModel."" , " ;
  put "  alpha=&alpha., df=&df., critV=&critV., " ;
  put "  doseType=""&doseType."" , Delta=&Delta., p=&p., pVal=&pVal., " ;
  put "  alternative=""&alternative."" , na.action=na.omit)" ;
  put "sink(file=""&outtextpath."" )" ;
  put "summary(results)" ;
  put "sink()" ;
  put "pdf(""&pdfpath.", paper='a4')" ;
  put "plot(models, maxEff=&maxEff.)" ;
  put "plot(results, plotData='meansCI')" ;
  put "dev.off()" ;
run ;

x "&Rexepath. --no-restore --no-save < ""&programpath."" " ;
x ""&pdfpath."" ;
x "notepad ""&outtextpath."" " ;
%mend ;

```

#### プログラム A.4 各モデルのパラメータ設定を行うマクロ「%MYGUESS」

---

```
%macro MYGUESS(d          = ,
                p          = ,
                model      = ,
                less       = TRUE,
                local      = FALSE,
                dMax       = NULL,
                Maxd       = NULL,
                scal       = NULL,
                outtextpath = "C:/temp/output.txt",
                programpath = "C:/temp/program.R") ;
options noxwait xsync ;

*--- R の実行ファイルのパス ;
%let Rexepath='C:\Program Files\R\R-3.2.5\bin\i386\R.exe' ;

filename aaa &programpath. ;
data _null_ ;
  file aaa ;
  put "library(DoseFinding)" ;
  put "sink(file="&outtextpath.")" ;
  put "guesst(d=c(&d.), p=c(&p.), model=""&model.""", less=&less., local=&local., dMax=&dMax.,
Maxd=&Maxd., scal=&scal.)" ;
  put "sink()" ;
run ;

x "&Rexepath. --no-restore --no-save < "&programpath. " " ;
x "notepad "&outtextpath."" ;
%mend ;
```

---

#### プログラム A.5 SAS と R を組み合わせて検出力の計算や例数設計を行うマクロ

---

```
*--- 検出力の計算を行うマクロ ;
%macro MYPOWER(doses      = ,
               models      = ,
               maxEff      = 1,
               alpha       = 0.025,
               n           = ,
               sigma       = NULL,
               alternative  = one,
               placAdj     = FALSE,
               critV       = TRUE,
               outtextpath = "C:/temp/output.txt",
               programpath = "C:/temp/program.R") ;
options noxwait xsync ;

*--- R の実行ファイルのパス ;
%let Rexepath='C:\Program Files\R\R-3.2.5\bin\i386\R.exe' ;

filename aaa &programpath. ;
data _null_ ;
  file aaa ;
  put "library(DoseFinding)" ;
  put "models <- Mods(&models., doses=c(&doses.), maxEff=&maxEff.)" ;
  put "set.seed(777)" ;
  put "contMat <- optContr(models, w=c(&n.))" ;
  put "sink(file="&outtextpath.")" ;
```

---

```

put "(x <- powMCT(contMat, alpha=&alpha., altModels=models, n=c(&n.), sigma=&sigma., " ;
put "placAdj=&placAdj., alternative=""&alternative.""", df=NULL, critV=&critV.))" ;
put "cat('¥n'); y <- mean(x); names(y) <- 'Mean Power'; y" ;
put "sink()" ;
run ;

```

```

x "&Rexepath. --no-restore --no-save < "&programpath." " ;
x "notepad "&outtextpath."" ;
%mend ;

```

\*--- 例数設計を行うマクロ ;

```

%macro MYSAMPLE(doses      = ,
                models     = ,
                maxEff     = 1,
                upperN     = ,
                lowerN     = ,
                power      = 0.8,
                sigma      = 1,
                alRatio    = ,
                alpha      = 0.025,
                alternative = one,
                Ntype      = total,
                sumFct     = mean,
                outtextpath = "C:/temp/output.txt",
                programpath = "C:/temp/program.R") ;

```

```
options noxwait xsync ;
```

\*--- R の実行ファイルのパス ;

```
%let Rexepath='C:¥Program Files¥R¥R-3.2.5¥bin¥i386¥R.exe' ;
```

```
filename aaa &programpath. ;
```

```

data _null_ ;
file aaa ;
put "library(DoseFinding)" ;
put "models <- Mods(&models., doses=c(&doses.), maxEff=&maxEff.)" ;
put "set.seed(777)" ;
put "contMat <- optContr(models, w=c(&alRatio.))" ;
put "sink(file=""&outtextpath.")" ;
put "sampSizeMCT(upperN=&upperN., lowerN=&lowerN., contMat=contMat, altModels=models," ;
put "power=&power., sigma=&sigma., alpha=&alpha., alternative=""&alternative.""", ;
put "tol=0.001, alRatio=c(&alRatio.), Ntype=""&Ntype.""", verbose=FALSE, sumFct=&sumFct.)" ;
put "sink()" ;

```

```
run ;
```

```
x "&Rexepath. --no-restore --no-save < "&programpath." " ;
```

```
x "notepad "&outtextpath."" ;
```

```
%mend ;
```

## プログラム A.6 最適な対比係数を求めるマクロ「%MYCONTRAST」

```

%macro MYCONTRAST(models=%str('emax' 'linear' 'linlog' 'exponential' 'quadratic' 'logistic' 'betaMod' 'sigEmax'),
                  params=%str(0.3 . . . . . 0.7 . . . -0.85 . . 0.3 0.1. 0.3 1.3. 0.5 2.0),
                  doses =%str(0.00 0.05 0.20 0.60 1.00),
                  w      =%str(1),
                  S      =0) ;
proc iml ;
flag = 0 ;
%if &S. ^= 0 %then %do ;

```

---

```

flag = 1 ;
use work. &S. ;
read all into S ;
close work. &S. ;
%end ;
models = {&models.} ;
params = {&params.} ;
w      = {&w.} ;
doses  = {&doses.} ;
scale  = 1.2 * max(doses) ;
off    = 0.01* max(doses) ;
nModels = ncol(models) ;
nDoses = ncol(doses) ;
mu     = repeat(0, nDoses, nModels) ;
c      = repeat(0, nDoses, nModels) ;
do j=1 to nModels ;
  do i=1 to nDoses ;
    if (models[j]='emax')      then mu[i, j] = doses[i]/(params[j, 1]+doses[i]) ;
    if (models[j]='linear')   then mu[i, j] = doses[i] ;
    if (models[j]='linlog')   then mu[i, j] = log(doses[i]+off) ;
    if (models[j]='exponential') then mu[i, j] = exp(doses[i]/params[j, 1])-1 ;
    if (models[j]='quadratic') then mu[i, j] = doses[i]+params[j, 1]*(doses[i]**2) ;
    if (models[j]='logistic') then mu[i, j] = 1/(1+exp((params[j, 1]-doses[i])/params[j, 2])) ;
    if (models[j]='sigEmax')  then
      mu[i, j] = doses[i]**params[j, 2]/((params[j, 1]**params[j, 2])+(doses[i]**params[j, 2])) ;
    if (models[j]='betaMod')  then do ;
      maxDens = (params[j, 1]+params[j, 2])**params[j, 1]+params[j, 2]/
        ((params[j, 1]**params[j, 1])*(params[j, 2]**params[j, 2])) ;
      tmpdose = doses[i]/scale ;
      mu[i, j] = maxDens*(tmpdose**params[j, 1])*(1-tmpdose)**params[j, 2] ;
    end ;
  end ;
end ;
if (flag=1) then Sinv = inv(S) ;
else if (ncol(w)=1) then Sinv = i(ncol(doses)) ;
else Sinv = i(ncol(w))*w ;
do j=1 to nModels ;
  aux = Sinv[, j] ;
  mn = sum(mu[, j] # aux)/sum(aux) ;
  val = Sinv * (mu[, j] - repeat(mn, nDoses, 1)) ;
  val2 = val - repeat(sum(val), nDoses, 1) ;
  c[, j] = val2/sqrt(sum(val**2)) ;
end ;
qu1 = repeat(" ", nModels, 1) ;
qu2 = repeat(" dose", nModels, 1) ;
tx1 = catx(" ", qu1, catx(" ", t(1:nModels), t(models))) ;
tx2 = catx(" ", tx1, qu2) ;
tx3 = catx(" ", tx2, rowcat(char(t(c))+" ")) ;
ctr = tx3[1, 1] ;
do i=2 to nModels ; ctr = ctr + ", " + tx3[i, 1] ; end ;
doses = char(doses) ;
mattrib c rowname=doses colname=models ;
print c ;
options noquotelenmax ;
%global ctr ;
call symputx("ctr", ctr) ;
quit ;
%mend ;

```

---

# 指数分布に従うデータへの層別ログランク検定の検出力の検討

○小松 邦岳 (株式会社アスクレップ)

(臨床評価研究会 基礎解析分科会)

Impact of power of the stratified Log-Rank test for survival data based on exponential distribution.

Kunitake Komatsu (ASKLEP Inc.)

The association of clinical evaluation

## 要旨

層別生存時間解析とは、生存時間解析で比較する群の中に生存関数に違いのある因子が認められた場合、その因子で層別し個別に生存時間解析を実施、それぞれの結果を併合する手法である。層別を行わなかった場合に起こる検出力の低下を、層別を行うことにより防ぐことができるが、状況によっては層別を行うことにより逆に検出力が低下してしまう。

この演題は、さまざまな指数分布に従った乱数を用いたシミュレーションによって、層別を行った場合と行わなかった場合の検出力の違いを示すことを目的とする。

キーワード：層別生存時間解析、層別ログランク検定、検出力

## 1 はじめに

ログランク検定は、症例が不均一ではないことを仮定しており、症例が不均一な場合には検出力が低下する<sup>[1]</sup>。症例が不均一とは予後に影響のある因子、すなわち生存関数に影響を与える因子が症例ごとに異なって存在していることである。ある因子が生存関数に影響を与えているとき、その因子で層別を行ったログランク検定により、層間の生存関数の違いを認めたいうでの解析結果を得ることができ、層別を行わなかった場合と比べ、検出力の低下を抑止できる<sup>[1]</sup>。ただし、層間には治療効果が一樣であること、つまり治療効果と層別因子間に交互作用がないことが前提となる<sup>[2]</sup>。ではどのような場合にどの程度検出力は変化するのだろうか。

指数分布に従った乱数を用いて作成した生存時間データのパラメータをさまざまに変化させ、層別を行わないログランク検定、層別を行ったログランク検定のシミュレーションを実施し、検出力の違いを検討した。

なお、以降本稿中では層別を行わないログランク検定を「単純ログランク検定」、層別を行うログランク検定を「層別ログランク検定」と呼称する。また、シミュレーションデータは全て2群、2層とした。

## 2 層別ログランク検定について

層別ログランク検定の実施方法を以下に紹介する。なお、ログランク検定そのものについてはここでは触れない。

### 2.1 層別ログランク検定の計算方法

手順としては、それぞれの層ごとにスコア統計量  $u$  およびその分散成分  $V$  を求め、それを元にして併合した統計量 ( $\chi^2$  値) を算出する。併合した統計量は自由度 [層の数-1] の  $\chi^2$  分布に従うことを利用し統計的仮説検定を行う。

層ごとのスコア統計量  $u$  およびその分散成分  $V$  は、層で分割したデータからそれぞれに算出すればよい。2 群の場合の各層における算出方法は以下のとおりである<sup>[1]</sup>。

スコア統計量：

$$u_j = \sum_i (d_{ij} - e_{ij})$$

スコア統計量の分散成分：

$$V_j = \sum_i \frac{(n_i - n_{ij})n_{ij}d_i(n_i - d_i)}{n_i^2(n_i - 1)}$$

$$\left\{ \begin{array}{l} e_{ij} : \text{イベント数の期待値} \\ d_{ij} : \text{時点 } i \text{ におけるイベント数} \\ n_{ij} : \text{時点 } i \text{ における } j \text{ 群のリスク集合の大きさ} \\ n_i : \text{時点 } i \text{ における群を併合したリスク集合の大きさ} \end{array} \right.$$

S 個の層ごとに算出したスコア統計量  $u$  および  $V$  から、併合した統計量 ( $\chi^2_t$ ) は以下のように求める<sup>[2]</sup>。

$$\chi^2_t = \left( \sum_{k=1}^S u_k \right)^2 / \left( \sum_{k=1}^S V_k \right) \quad \left\{ \begin{array}{l} \chi^2_t : \text{併合した統計量} \\ u_k : \text{層ごとに算出したスコア統計量 } u \\ V_k : \text{層ごとに算出したスコア統計量の分散成分 } V \end{array} \right.$$

この統計量 ( $\chi^2_t$ ) を自由度 [層の数-1] の  $\chi^2$  乗分布の任意の%点と比較することにより、有意か否かを得ることができる。

### 2.2 SAS program

lifetest プロシジャによる層別ログランク検定実施のプログラム例を以下の通り提示する<sup>[4][5]</sup>。

SAS program2.2.1 : lifetest プロシジャによる層別ログランク検定例

```
proc lifetest data=入力データセット名 ;
    time 時間変数 * 打ち切り変数(打ち切り値) ;
    strata 層別変数 / group=群別変数 ;    run;
```

## 3 シミュレーション方法

### 3.1 概要・手順

指数分布に従った生存時間データの乱数データセットを 4 つ生成し、縦結合をした 1 つのデータセットについて単純ログランク検定および層別ログランク検定を実施する。これを繰り返し、検出力の算出および検討を行う。

4 つの乱数データセットは群間の差、層別の差を考慮しハザードなどのパラメータを変化させる。結合した乱数データに対して単純ログランク検定と層別ログランク検定を実施し、それぞれの検定結果で有意差の有



無を確認する。これを、乱数データセットを変えながら複数回繰り返したときに有意となった割合が検出力となる。検討のため、検出力の比（単純ログランク検定の検出力÷層別ログランク検定の検出力）も合わせて算出する。特に断りが無い限りシミュレーションの回数は1万回とした。

なお、データおよびシミュレーションは特定の疾患や状況を仮定していないことを断っておく。

### 3.2 指数分布に従う乱数列の生成方法

指数分布に従う生存時間データを、以下の方法で生成した<sup>[1]</sup>。

指数分布に従う生存関数は $S(t) = \exp[-\lambda t]$ と表され、これをイベント発現までの期間  $t$  について解くと以下の数式が得られる。

$$t = -\frac{\log(S(t))}{\lambda} \quad (\text{式 3.2.1})$$

$$\left\{ \begin{array}{l} t : \text{生存時間} \\ S(t) : \text{生存関数} \\ \lambda : \text{ハザード} \end{array} \right.$$

式 3.2.1 に対して、 $\lambda$ に任意の値、 $S(t)$ に 0~1 の値をとる一様乱数を与えることにより、指数分布モデルに従ったシミュレーションデータとなるイベント発現時期 $t$ を得た。

### 3.3 指数分布に従う乱数列の生成方法（SAS プログラム）

乱数列発生の SAS プログラム例を以下に示す。このプログラムでは、任意の時点(マクロパラメータ:CNS)以降は全てその時点で打ち切りとしている。

SAS program3.3 : 指数分布に従う生存時間シミュレーションデータ作成 SAS プログラム例

```

%macro DATAGEN(LAMBDA=,N=,CNS=);
do I=1 to &N;
  ST = rand("uniform");
  TIME = ((-1)*(log(ST)))/(&LAMBDA);
  CENSOR=0;
  if(TIME>&CNS)then do;
    CENSOR=1;
    TIME = &CNS;
  end; output; end;
%mend DATAGEN;

```

・マクロパラメータ

LAMBDA =ハザード ( $\lambda$ ) を設定。

N=生成するレコード数を設定。

CNS=打ち切り時点を設定。

### 3.4 層別生存時間解析のシミュレーションデータセットの生成方法

指数分布に従う乱数を 4 種類生成し、それぞれを群・層とした1つのデータセットとして縦結合する。基本となるデータセットを 3.5 の通り定め、そこから様々な乱数発生のパラメータを変化させ、それぞれの場合での検出力をシミュレーションにより導出、検討する。

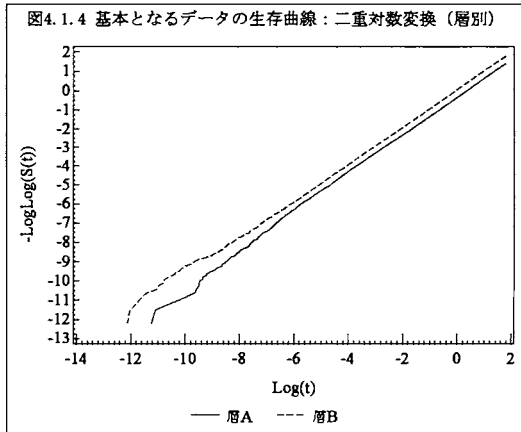
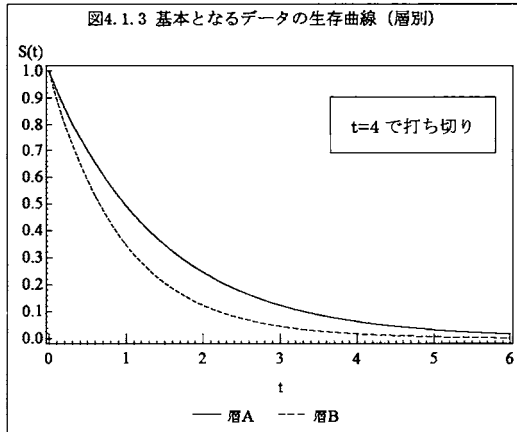
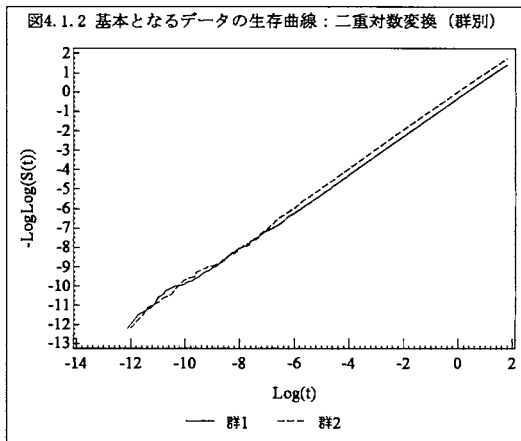
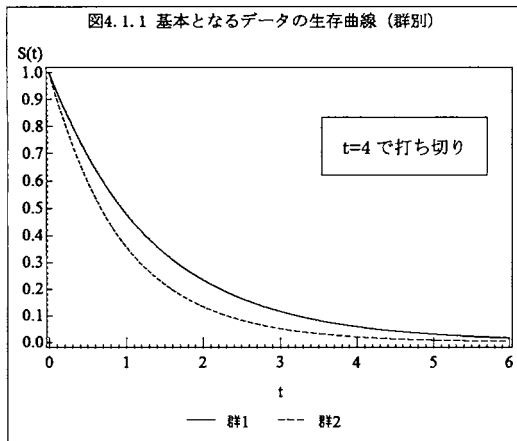
### 3.5 基本となる乱数データセット

基本となる乱数データセットは以下のように定めた。

	層 A	層 B
群 1	例数 : $N_{1A} = 75$ ハザード : $\lambda_{1A} = 0.6$	例数 : $N_{1B} = 75$ ハザード : $\lambda_{1B} = 0.9$
群 2	例数 : $N_{2A} = 75$ ハザード : $\lambda_{2A} = 0.84$	例数 : $N_{2B} = 75$ ハザード : $\lambda_{2B} = 1.26$
観察期間 : $t=4$ で打ち切り。		

このデータセットは以下のような特徴を持たせている。

- 層 A、B による層別ログランク検定の検出力が約 0.8 となる。(10 万回のシミュレーションによる検出力計算を 10 回繰り返した結果、検出力の平均±標準偏差は  $0.8092 \pm 0.0015$ 。)
- 群、層において症例数が同数である。
- 各群に共通で層 A、層 B 間のハザードの比が 1.5 倍。
- 各層に共通で群 1、群 2 間のハザードの比が 1.4 倍。
- このデータセットと同じ分布に従う生存曲線について、1 層あたり 10 万症例としてシミュレーションした結果のグラフを、理論曲線に換えて示す (打ち切りはしない)。群ごと、層ごとに示した。また、二重対数変換したプロットも合わせて示した。



## 4 シミュレーション結果

### 4.0 シミュレーションを行った検討項目

シミュレーションによって、以下の内容を検討した。

- 4.1. 症例数の変化に伴う検出力の検討
- 4.2. 観察期間の変化に伴う検出力の検討
- 4.3. 層間のハザードの変化に伴う検出力の検討
- 4.4.1 層間の例数がアンバランスな時の検出力の検討(層内ではバランス)
- 4.4.2 層間の例数がアンバランスな時の検出力の検討(層内でアンバランス)
- 4.5. 層間の一様性が仮定できないときの検出力の検討

### 4.1 症例数の変化に伴う検出力の検討

1層あたりの症例数を変化させ、検出力の推移を確認した。その際、各群・層の症例数は同数とした。

#### 【データの説明】

	層 A	層 B
群 1	例数 : $N_{1A}$ ハザード : $\lambda_{1A}=0.6$	例数 : $N_{1B}$ ハザード : $\lambda_{1B}=0.9$
群 2	例数 : $N_{2A}$ ハザード : $\lambda_{2A}=0.84$	例数 : $N_{2B}$ ハザード : $\lambda_{2B}=1.26$
$t=4$ で打ち切り。 $N_{1A}=N_{1B}=N_{2A}=N_{2B}=5, 10, 25, 50, 75, 100, 125, 150, 200$		

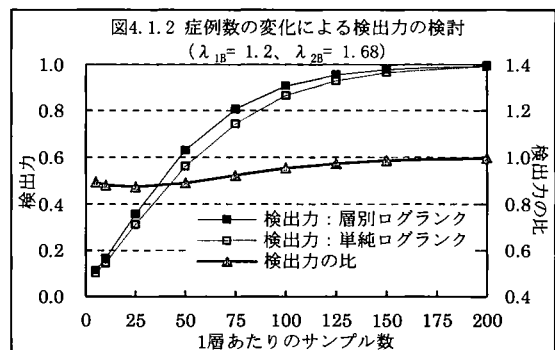
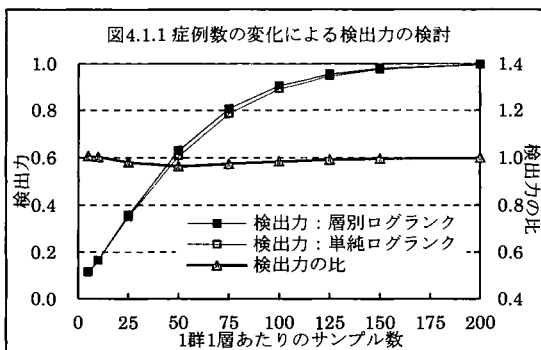
#### 【シミュレーション結果と検討】

結果を図 4.1.1 に示した。全体の傾向として、例数が増えるほど検出力が向上した。

症例数が 25~125 の間で層別ログランク検定よりも単純ログランク検定の検出力が低かった。それ以外の症例数では、検出力が過大 (1 に近い) または検出力が過小 (0 に近い) の場合となり、検出力の違いは確認できなかった。

このデータでは双方の検出力が拮抗していたため、さらに  $\lambda_{1B}=1.2$ 、 $\lambda_{2B}=1.68$  とした場合の検出力についても示す (図 4.1.2)。

こちらについても、同様の傾向が見られたが、層別を行った場合と行わない場合の検出力の違いについてはより顕著となった。1層あたりの症例数が 25 の場合には層別を行った場合の検出力が 0.35 なのに対し、層別を行わない検出力は 0.31 であり、検出力の比は 0.87 倍と低下している。層別を行った場合の検出力が約 0.8 となる 1層あたりの症例数が 75 の場合でも、層別を行わない検出力は 0.74 となり、0.92 倍となった。



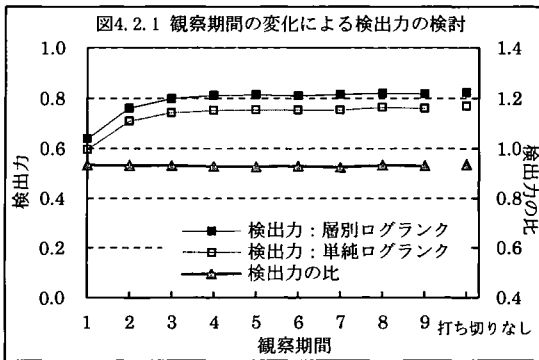
#### 4.2 観察期間の変化に伴う検出力の検討

観察期間を変化させ、検出力の推移を確認した。

##### 【データの説明】

	層 A	層 B
群 1	例数 : $N_{1A} = 75$ ハザード : $\lambda_{1A} = 0.6$	例数 : $N_{1B} = 75$ ハザード : $\lambda_{1B} = 0.9$
群 2	例数 : $N_{2A} = 75$ ハザード : $\lambda_{2A} = 0.84$	例数 : $N_{2B} = 75$ ハザード : $\lambda_{2B} = 1.26$
観察期間 : $t=1,2,3,4,5,6,7,8,9,10$ で打ち切り。および打ち切りなし。		

##### 【シミュレーション結果と検討】



結果を図 4.2.1 に示した。全体の傾向として、観察期間が非常に短い場合において検出力が低下する様子が見られる。観察期間=4 以降において、検出力は一定した。

層別を行った場合と行わない場合の検出力の比較については、層別を行わないことによる検出力の低下が見られたが、検出力の比は観察期間の設定を問わず約 0.93 倍で一定であった。

#### 4.3 層間のハザードの変化に伴う検出力の検討

層 B のハザードを変化させ、検出力の推移を確認した。

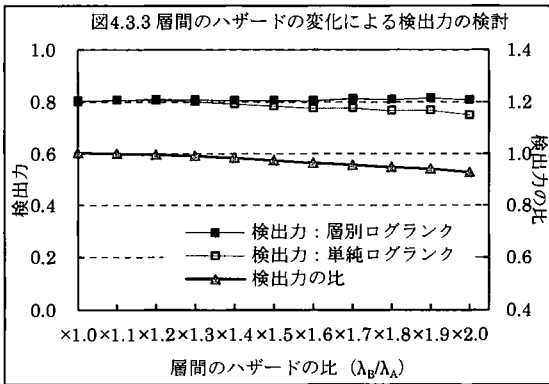
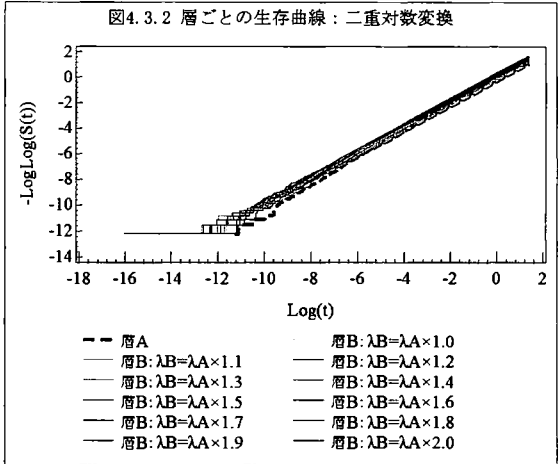
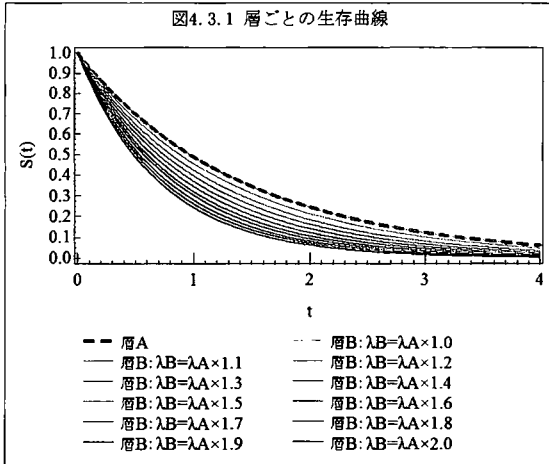
##### 【データの説明】

	層 A	層 B
群 1	例数 : $N_{1A} = 75$ ハザード : $\lambda_{1A} = 0.6$	例数 : $N_{1B} = 75$ ハザード : $\lambda_{1B}$ 下記
群 2	例数 : $N_{2A} = 75$ ハザード : $\lambda_{2A} = 0.84$	例数 : $N_{2B} = 75$ ハザード : $\lambda_{2B}$ 下記
観察期間 : $t=5$ で打ち切り。 $\lambda_{1B}, \lambda_{2B} = \lambda_{1A}, \lambda_{2A}$ の 1 倍, 1.1 倍, 1.2 倍, ..., 1.9 倍, 2.0 倍		

層 B のハザードを変化した時の層ごとの生存曲線について、1 層あたり 10 万症例としてシミュレーションした結果のグラフを、理論曲線に換えて図 4.3.1 および図 4.3.2 に示す。

##### 【シミュレーション結果と検討】

結果を図 4.3.3 に示した。層間のハザードの比が増加するにつれて、単純ログランク検定の検出力は低下した。一方、層別ログランク検定では約 0.8 を維持した。このことより、層間でのハザードの比が大きくなればなるほど、層別ログランク検定を実施する意義が明確となることが示唆された。



しかしながら、層間のハザードの比が1.3倍までの検出力は拮抗しており、層間のハザードの違いがわずかであれば層別を行うことの意義は低いことも示唆された。

また、 $\times 1.0$ 、 $\times 1.1$ における検出力の比はそれぞれ1.00337、1.00062となり、ごくわずかではあるが層別を行う検定の検出力が低下した。

#### 4.4.1 層間の例数がアンバランスな時の検出力の検討(層内ではバランス)

各群の層A+層Bの症例数を150から変化させず、層A層Bの症例数の割合を変化し、検出力の推移を確認した。ここでは、群1群2それぞれにおける層A層Bの症例数の割合は等しく保つ。

##### 【データの説明】

	層A	層B
群1	例数： $N_{1A}$ = 下記の通り増加 ハザード： $\lambda_{1A} = 0.6$	例数： $N_{1B}$ = 下記の通り減少 ハザード： $\lambda_{1B} = 0.9$
群2	例数： $N_{2A}$ = 下記の通り増加 ハザード： $\lambda_{2A} = 0.8$	例数： $N_{2B}$ = 下記の通り減少 ハザード： $\lambda_{2B} = 1.26$
観察期間： $t=4$ で打ち切り。 $(N_{1A}:N_{1B}, N_{2A}:N_{2B}) : (5:145, 5:145), (10:140, 10:140), (25:125, 25:125), (50:100, 50:100),$ $(75:75, 75:75), (100:50, 100:50), (125:25, 125:25), (140:10, 140:10), (145:5, 145:5)$		

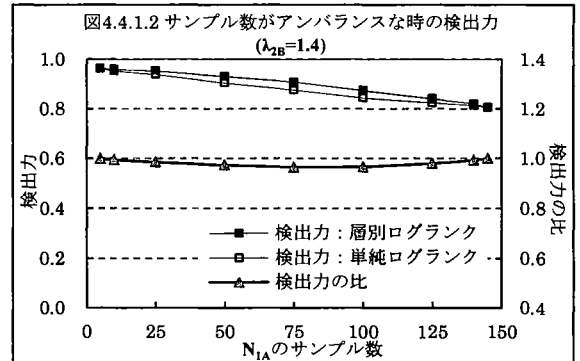
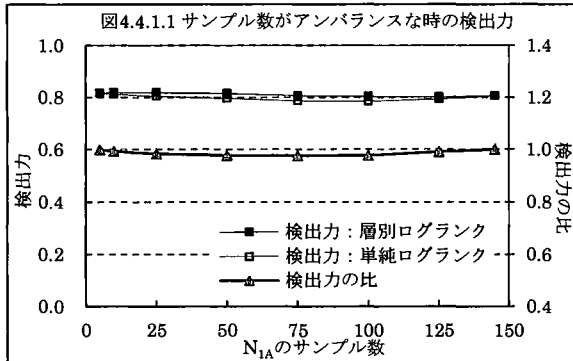
##### 【シミュレーション結果と検討】

結果を図4.4.1.1に示す。検出力はほぼ一定を保ったが、これは $\lambda_A : \lambda_{2A} = \lambda_B : \lambda_{2B}$ と設定しているためと考えられる。例えば $\lambda_{1A} : \lambda_{2A} < \lambda_B : \lambda_{2B}$ と設定していれば層Bの例数が多いほど検出力は向上するはずである。本件についても検討を行うため、 $\lambda_{2B} = 1.4$ とし、その他は先ほどの検証から変更しないシミュレーションを行

った。

シミュレーション結果を図 4.4.1.2 に示す。 $\lambda_{2B}$  を増加したことにより全体的に検出力は向上している。

$\lambda_{2B}=1.4$  とした場合、層 B の症例数が多い (=  $N_{1A}$  が少ない) ほど検出力が高く、層 B の症例数が少ないほど検出力が低くなる傾向となった。単純ログランク検定の検出力と層別ログランク検定の検出力の違いについては、全体的に層別ログランク検定の検出力が高い状態を維持したが、層の例数が均等に近いほど検出力の差は広がった。



#### 4.4.2 層間の例数がアンバランスな時の検出力の検討(層内でアンバランス)

各群の層 A+層 B の症例数を 150 から変化させず、層 A 層 B の症例数の割合を変化し、検出力の推移を確認した。ここでは、群 1 は層 A を徐々に増加、群 2 は層 B を徐々に増加させていく。なお、検出力の違いが大きく出たため、ここでは検出力の比の算出は行わなかった。

##### 【データの説明】

	層 A	層 B
群 1	例数: $N_{1A}$ = 下記の通り増加 ハザード: $\lambda_{1A}=0.6$	例数: $N_{1B}$ = 下記の通り減少 ハザード: $\lambda_{1B}=0.9$
群 2	例数: $N_{2A}$ = 下記の通り減少 ハザード: $\lambda_{2A}=0.84$	例数: $N_{2B}$ = 下記の通り増加 ハザード: $\lambda_{2B}=1.26$
観察期間: $t=4$ で打ち切り。 $(N_{1A}:N_{1B}, N_{2A}:N_{2B}) : (5:145, 145:5), (10:140, 140:10), (25:125, 125:25), (50:100, 100:50), (75:75, 75:75), (100:50, 50:100), (125:25, 25:125), (140:10, 10:140), (145:5, 5:145)$		

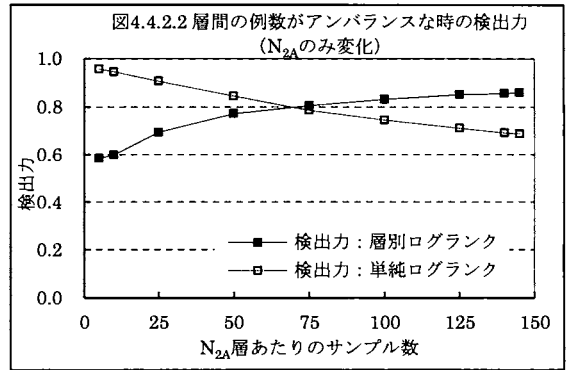
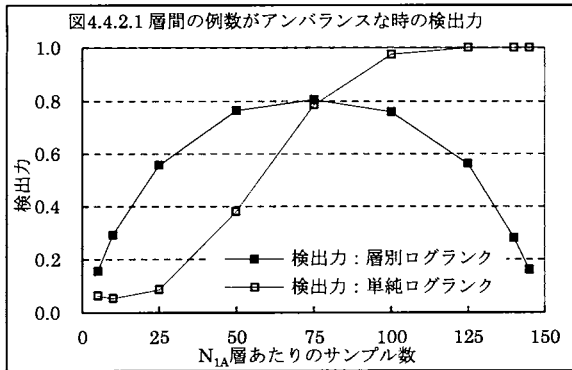
##### 【シミュレーション結果と検討】

結果を図 4.4.2.1 に示す。単純ログランク検定では、 $N_{1A}$  および  $N_{2B}$  が増加するほど検出力が向上した。これは  $\lambda_{1A}:\lambda_{2B} > \lambda_{2A}:\lambda_{1B}$  のため、 $\lambda_{1A}$  と  $\lambda_{2B}$  の症例数が増えるほど群間でのハザードの比が大きくなり、検出力が向上したためと思われる。対して層別ログランク検定では、層 A、層 B ごとに統計量を算出するため、例数がアンバランスな状態での統計量算出となってしまう、検出力低下を招いた<sup>[1]</sup>。

症例数のアンバランスが 1 層のみ発生する場合でも検討した。 $N_{1A}=N_{1B}=N_{2B}=75$  で固定し、 $N_{2A}$  のみ  $\{5, 10, 25, 50, 75, 125, 140, 145\}$  と変化させた。シミュレーション結果を図 4.4.2.2 に示す。 $\lambda_{2A} < \lambda_{2B}$  のため、単純ログランク検定では  $N_{2A}$  が少ないほどハザードに群間差が生じることとなり検出力が高い。逆に  $N_{2A}$  が多いほ

どハザードに群間差は小さくなるため検出力は減少した。

一方、層別ログランク検定では、層 A のアンバランスおよび例数過小のため、 $N_{2A}$  が少数の時に検出力が低下した。その後、 $N_{2A}$  の増加に合わせて検出力は向上した。



#### 4.5 層間の一様性が仮定できないときの検出力の検討

生存時間データを層別するにあたって、層間で治療効果が同様（同じ）であることが求められる<sup>[2]</sup>。ここでは層間の治療効果の一様性が仮定できないデータでの検出力について検討する。なお、治療効果が一律でないとは、層別因子と治療との間に交互作用があるということであり、この場合は交互作用について検討する必要がある。

層間の一様性を検定するプログラム以下の通り提示する<sup>[6]</sup>。本稿では、phreg プロシジャを用い、層間の交互作用について検定を行った。

SAS program4.5.1 : phreg プロシジャによる層間の一様性を検定プログラム例

```
proc phreg data=入力データセット名 ;
  class 群別変数 層別変数 /ref=first param=ref;
  model 時間変数*打ち切り変数(打ち切り値) = 群別変数
        群別変数*層別変数 /rl type3(score);
  strata 層別変数 ; run ;
```

#### 【データの説明】

乱数データをそれぞれ 10 万通り作成、それぞれに層間の一様性の検定を実施し、有意 ( $p < 0.05$ ) となったデータのみを対象として単純ログランク検定および層別ログランク検定を行い、検出力を算出・検討した。なお、層間の一様性の検定で優位になったということは、一様性があるとは言えないということである。

4.1 で示したデータと同様の乱数を使用して検討を行った。また、4.1 と同様、 $\lambda_{1B} = 1.2$ 、 $\lambda_{2B} = 1.68$  とした結果も示す。

10 万通りのデータ中、層間の一様性の検定で有意となったデータセット数を表 4.5 に示した。

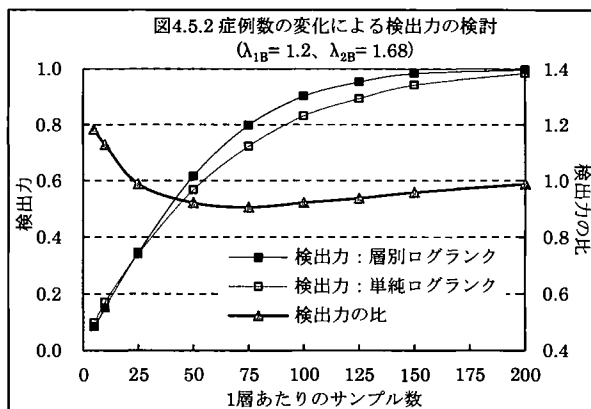
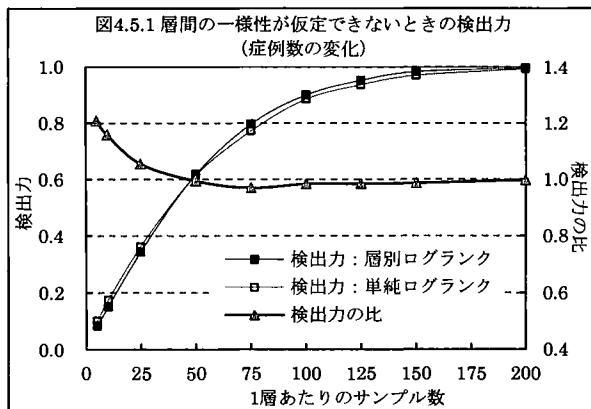
#### 【シミュレーション結果と検討】

結果を図 4.5.1 および図 4.5.2 に示す。全体的な推移は図 4.1.1 および図 4.1.2 と同様である。しかし、図 4.5.1 では 1 層あたりの例数が 50 症例未満の場合において、単純ログランク検定の検出力が層別ログランク検定の検出力を上回り、逆に層の例数が 50 症例以上の場合においては層別ログランク検定の検出力が単純ログラン

ク検定を上回っていた。図 4.5.2 では 25 症例以上にて層別ログランク検定の検出力が単純ログランク検定の検出力を上回っており、図 4.5.1 の場合と比べてより少ない症例数が閾値となった。このことより、層間の一様性の検定で有意となった場合も、層間のハザードの差が大きいほど、また症例数が多いほど層別ログランク検定の検出力が単純ログランクを上回ることが示唆された。なお、検出力が上回っていたとしても、交互作用に関する検討を放棄してよい理由にはならないと考えている。交互作用が認められた場合は、検出力とは全く別件として、交互作用に関する検討を行うべきと考える。

＜表 4.5 一様性の検定で有意となったデータセット数＞

層あたりの例数	有意になったデータセット数		層あたりの例数	有意になったデータセット数	
	$\lambda_{1B}=0.9, \lambda_{2B}=1.26$	$\lambda_{1B}=1.2, \lambda_{2B}=1.68$		$\lambda_{1B}=0.9, \lambda_{2B}=1.26$	$\lambda_{1B}=1.2, \lambda_{2B}=1.68$
5	7,943	7,941	100	5,079	5,089
10	6,605	6,602	125	5,203	5,208
25	5,726	5,725	150	5,082	5,089
50	5,300	5,309	200	4,961	4,988
75	5,133	5,136			



## 5 まとめ

本稿ではシミュレーションにより、以下の内容が示唆された。

- 多くの場合、層間のハザードの差が大きいとき、層別ログランク検定を行うことにより検出力の低下を抑制する。
- 症例数のアンバランスが生じた場合や、層間の一様性が仮定できない場合において、層別ログランク検定を行うとかわって検出力が低下する場合がある。
- 層別ログランク検定を行うにあたっては、症例数が均一であることや、特に例数や層間のハザードの比が小さいときにおいては層間の一様性に留意する必要があることが示された。

最後に、本稿で示したシミュレーションデータが今後のシミュレーション実験や症例数設計の際に役立てば幸いである。



## 参考文献

- [1] 赤澤宏平・柳川 堯 著. サバイバルデータの解析ー生存時間とイベントヒストリデーター. 近代科学社
- [2] 大橋靖雄・浜田知久馬 著. 生存時間解析 SAS による生物統計. 東京大学出版会
- [3] 監修/浜田知久馬,執筆/臨床評価研究会 (ACE) 基礎解析分科会. 実用 SAS 生物統計ハンドブック[SAS8.2 および SAS9.1 対応]. サイエンティスト社
- [4] 張方紅. SAS による生存時間解析の実務. SAS ユーザー総会 論文集 2012 167-184
- [5] SAS Institute Inc. SAS/STAT(R) 9.2 User's Guide, Second Edition, The LIFETEST Procedure.  
[http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#lifetest\\_toc.htm](http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#lifetest_toc.htm)
- [6] 浜田知久馬・中西豊支・松岡伸篤. SASV9 の TPHREG を用いたメタアナリシス. SAS ユーザー総会 論文集 2004 165-191

## PPK解析時の潜在的なデータ不備が推定値に与える影響の検討

○川崎勝己、碓新里、新城博子

(株式会社ベル・メディカルソリューションズ)

### 【背景・目的】

2015年12月に母集団薬物動態・薬力学解析ガイドライン(案)が発出され、母集団薬物動態(PPK)解析は、今後の承認申請時又は添付文書への情報提供における活用について大きな期待が寄せられている。それに伴いPPK解析に用いる入力データの品質の高さが求められている。

PPK解析用データセットは、複雑なデータハンドリングを伴うことが多く、SASのプログラミングを用いてデータ加工をすることが標準的と思われる。解析対象として、10試験以上の複数試験のデータベースを統合し作成することも少なくない。対象となる試験の中には過去に実施されたレガシーな試験も含まれている場合もあり、さらに海外試験となる場合、データベース構造が標準化されていない、データベース定義書が十分に揃っていない等といったデータ加工に様々なハードルが存在する場面にぶつかることもある。複雑なデータベース構造から必要な情報を取得するため、イレギュラーな処理を施しながらPPK解析用データセットを作成することも多く、データ上からは検出しにくい不備が存在する可能性を否定できないことがある。また、服薬アドヒアランスや食事の時間が明らかとならない、データベースの設計上の不備が存在することもある。本検討では、そのような不備を「潜在的なデータ不備」とし、意図的にデータ不備を内包したデータセットをシミュレーションによって作成し、母集団推定値にどのような影響を与えるのか検討した。

### 【方法】

シミュレーションの設定として、薬剤情報は1日1回、投与量30 mgの経口投与製剤であり、PPK解析のモデルは1次吸収過程のある1-コンパートメントモデルとした。薬物動態パラメータとして全身クリアランスが10 L/h、分布容積が200 L、吸収速度定数が1.5 /hとなるシミュレーション濃度データを発生させた。シミュレーションは、1症例につき5点の採血ポイントの血中濃度を40症例分1000セット発生させた。また、潜在的なデータ不備は以下の3ケースを想定した。初回投与時の採血データを定常状態時の採血とした場合(Case 1)、定常状態時の採血データを初回投与時の採血とした場合(Case 2)、Case 1とCase 2を同じ割合で混在させた場合(Case 3)の3つのケースで、各ケースともに不備の割合を変えた(0%、10%、20%、30%、40%とした5パターン)データを作成し、それぞれのパターンで1000回ずつPPK解析を実施することにより、薬物動態パラメータの推定値及びその個体間変動に与える影響を検討した。

### 【結果】

Case 1及びCase 2では、潜在的なデータ不備の割合を高めることにより、各薬物動態パラメータの推定値に一定の傾向を伴ったバイアスが見られたが、Case 3では推定値に大きな偏りは見られなかった。また、全てのケースでパラメータの個体間変動に傾向を持ったバイアスが見られた。

以上のことから、PPK解析用データセットに潜在的なデータ不備がある場合、PPK解析で得られた推定値には、何らかのバイアスが含まれる可能性があることをシミュレーションにより示した。今回は、単純な構造モデルを用いたシミュレーションであったが、潜在的なデータ不備を抽出する効果的な診断方法の検討を今後の課題としたい。

# ポアソン分布に従うと仮定された総ケース数が 固定された条件の下でのrate ratioの条件付き信頼区間

伊藤要二

(アストラゼネカ株式会社バイオメトリックス部)

Confidence interval of rate ratio conditioning on  
the total number of cases which follows a Poisson distribution

Yohji Itoh

Biometrics Department, AstraZeneca K.K.

## 要旨

2つの処理の比較試験においてケースの数が Poisson 分布に従うと仮定した場合の、両処理の総ケース数が固定された条件の下での rate ratio の exact な条件付きの信頼区間を紹介し、それを SAS で計算する方法について示す。また、それに対応する統計的検定やサンプルサイズ計算の方法についても解説する。これらの方法は SAS ではサポートされていないが、SAS の関数を用いて容易に計算できることを示す。さらに、この exact な条件付き信頼区間を修正することにより、それとは異なる rate ratio の信頼区間を提案し、被覆確率について比較評価を行い、他の方法に比べてその方法が優れていることを示す。

キーワード： 例数設計、Clopper-Pearson 信頼区間、Wilson スコア信頼区間、被覆確率

## 1. 序文

臨床試験やコホート調査においては、観察された人・年(person-years)に対して、注目するイベントの発現したケースの数の比を event rate とし、その比である rate ratio を用いて群間比較を行うことがしばしばある。その際に、event rate が時点によらず一定であるという前提が成り立つ場合には、ケースの数は Poisson 分布に従うという仮定を置き、その仮定の下に群間比較のための統計的推測がなされる。そのような場合の rate ratio の信頼区間の計算方法としてよく用いられる方法の1つとして、総ケース数が与えられたという条件の下での rate ratio の exact な条件付き信頼区間がある。

本論文の目的は、この rate ratio の exact な条件付き信頼区間について紹介し、それを SAS を用いてどのように計算するかを示すことである。さらに、それに対応する統計的検定、またそのための

サンプルサイズの計算方法についても紹介する。また、最後に、この exact な条件付き信頼区間を修正し、より優れていると考えられる方法を提案し、他の方法との比較評価を行う。

## 2. Rate ratio の exact な条件付き信頼区間

表 1 に示したような 2 群比較の臨床試験を想定し、人・年 ( $y_1$  及び  $y_2$ ) が観察され、対象となるイベントが発現した被験者、すなわち「ケース」の例数 ( $C_1$  及び  $C_2$ ) が得られるものとする。また、各群における年あたりイベントの発現率の真値が  $\lambda_1$  及び  $\lambda_2$  であると仮定する。

表 1. 2 群比較臨床試験における想定

群	観察された 人・年	ケース数	年当たりの 発現率の真値	期待 ケース数
群 1 (試験群)	$y_1$	$C_1$	$\lambda_1$	$\lambda_1 y_1$
群 2 (対照群)	$y_2$	$C_2$	$\lambda_2$	$\lambda_2 y_2$
計	$y_{\text{total}} = y_1 + y_2$	$C_{\text{total}} = C_1 + C_2$	-	$\lambda_1 y_1 + \lambda_2 y_2$

各群のケース数は発現率  $\lambda_1$  及び  $\lambda_2$  の Poisson 分布に従うものと仮定すると、各群のケース数が  $c_1$  及び  $c_2$  である確率は

$$P(C_1 = c_1 | \lambda_1) = \frac{(\lambda_1 y_1)^{c_1} e^{-\lambda_1 y_1}}{c_1!}, \quad P(C_2 = c_2 | \lambda_2) = \frac{(\lambda_2 y_2)^{c_2} e^{-\lambda_2 y_2}}{c_2!} \quad (1)$$

と表される。ここで両群のケース数の総和  $C_{\text{total}} = C_1 + C_2$  が与えられたものとみなし、 $C_{\text{total}} = c_{\text{total}}$  の条件下での  $C_1$  の条件付き分布を導くと、

$$P(C_1 = c_1 | C_{\text{total}} = c_{\text{total}}, \lambda_1, \lambda_2) = \binom{c_{\text{total}}}{c_1} \left( \frac{\lambda_1 y_1}{\lambda_1 y_1 + \lambda_2 y_2} \right)^{c_1} \left( 1 - \frac{\lambda_1 y_1}{\lambda_1 y_1 + \lambda_2 y_2} \right)^{c_{\text{total}} - c_1} \quad (2)$$

が得られる(Przyborowski & Wilenski, 1940)。式(2)はまさに成功確率  $\lambda_1 y_1 / (\lambda_1 y_1 + \lambda_2 y_2)$ 、試行回数  $c_{\text{total}}$  の二項分布となっている。このように、2 つの Poisson 分布の総ケース数が固定された場合に各群のケース数の条件付き分布が二項分布になるという点がこの方法論の重要な部分であるが、ここでの冗長な数式の展開を避けるために、その誘導については付録に記した。

式(2)の二項分布における成功確率を

$$p_1 = \frac{\lambda_1 y_1}{\lambda_1 y_1 + \lambda_2 y_2}$$

と表すこととする。この  $p_1$  について実際のデータから推定される値を  $\hat{p}_1$  と表すこととすると、 $\hat{p}_1$  は

$$\hat{p}_1 = \frac{c_1}{c_1 + c_2}$$

として得られる。これより次式が導かれる。

$$\frac{c_1}{c_2} = \frac{\hat{p}_1}{1 - \hat{p}_1} \quad (3)$$

ところで、Rate ratio の推定値は、各群の event rate、 $c_1/y_1$  及び  $c_2/y_2$  の比として、次式によって得られる。

$$\hat{r} = \frac{c_1/y_1}{c_2/y_2} = \frac{y_2}{y_1} \cdot \frac{c_1}{c_2}$$

この式の最後の  $c_1/c_2$  の部分に式(3)を代入して、rate ratio 推定値は

$$\hat{r} = \frac{y_2}{y_1} \cdot \frac{\hat{p}_1}{1 - \hat{p}_1} \quad (4)$$

と変形することができる。

ここで、確率  $p_1$  のなんらかの信頼区間が得られれば、その信頼限界を式(4)の  $\hat{p}_1$  に代入することにより、rate ratio の信頼区間を計算することができる。例えば、ここで Clopper-Pearson 信頼区間(二項分布に基づく exact な信頼区間)を用いることにより、rate ratio の exact な信頼区間を得ることができる。F 分布を用いて  $p_1$  の Clopper-Pearson の信頼限界は

$$\hat{p}_{1,CP,lower} = \frac{c_1}{c_1 + (c_2 + 1)F_{1-\alpha/2}(2c_2 + 2, 2c_1)}, \quad \hat{p}_{1,CP,upper} = \frac{(c_1 + 1)F_{1-\alpha/2}(2c_1 + 2, 2c_2)}{c_2 + (c_1 + 1)F_{1-\alpha/2}(2c_1 + 2, 2c_2)}$$

で与えられる。ここで、 $F_{1-\alpha/2}(m, n)$  は自由度  $m, n$  の F 分布の  $(1-\alpha/2) \times 100\%$  点を表す。この信頼限界を式(4)の  $\hat{p}_1$  に代入して、下記のような rate ratio の exact な信頼区間を得る：

$$\hat{r}_{lower} = \frac{y_2}{y_1} \cdot \frac{\hat{p}_{1,CP,lower}}{1 - \hat{p}_{1,CP,lower}}, \quad \hat{r}_{upper} = \frac{y_2}{y_1} \cdot \frac{\hat{p}_{1,CP,upper}}{1 - \hat{p}_{1,CP,upper}}$$

(Breslow & Day, 1987, Section 3.4, p.95)

### 3. Rate ratio の exact な条件付き信頼区間のための数値例及び SAS プログラム

下の表 2 の数値例を用いて rate ratio の exact な条件付き信頼区間を求める。この数値例では、観察された人・年は群 1 が群 2 の約 2 倍となっており、割付け比が 2 : 1 であることを想定している。

表 2. 数値例

群	観察された人・年	観察されたケース数
群 1	$y_1 = 720.2$	$c_1 = 8$
群 2	$y_2 = 362.5$	$c_2 = 11$

この数値例を用いて rate ratio の exact な条件付き信頼区間を計算するための SAS プログラムをプログラム 1 に示す。ここで変数 `cpower` 及び `cpupper` は Clopper-Pearson 信頼区間の信頼限界を表す。なお、Clopper-Pearson 信頼区間は FREQ プロシジャでも計算可能であるが、ここではプログラムを簡潔にするために、データステップにおいて FINV 関数を用いて計算している。

プログラム 1. Rate ratio の exact な信頼区間の計算のための SAS プログラム及びその出力

```

data exmpl;
  input y1 c1 y2 c2;
  datalines;
720.2 8 362.5 11
;
run;

data ex1;
  set exmpl;
  p1=c1/(c1+c2);
  cplower=c1/(c1+(c2+1)*finv(0.975, 2*c2+2, 2*c1));
  cpupper=(c1+1)*finv(0.975, 2*c1+2, 2*c2)/(c2+(c1+1)*finv(0.975, 2*c1+2, 2*c2));
  rr =y2/y1*p1/(1-p1);
  lower=y2/y1*cplower/(1-cplower);
  upper=y2/y1*cpupper/(1-cpupper);
  keep rr lower upper;
run;

proc print noobs;
run;

```

出力

rr	lower	upper
0.36606	0.12782	0.99916

この数値例における rate ratio の両側 95%信頼区間の上限は 0.99916 であり、1 をぎりぎり下回っており、群 2 に比べて群 1 が 5%水準でかろうじて有意に優れていることを示している。

#### 4. 総ケース数が与えられた条件の下での exact な検定

次に、この exact な条件付き信頼区間に対応する統計的検定を考える。

帰無仮説  $H_0: \lambda_1 = \lambda_2$  の下での群 1 のケース数の条件付き分布は、式(2)に  $\lambda_1 = \lambda_2$  を代入することによって得られ、次のような次の二項分布となる。

$$P(C_1 = c_1 | C_{\text{total}} = c_{\text{total}}, H_0: \lambda_1 = \lambda_2) = \binom{c_{\text{total}}}{c_1} \left( \frac{y_1}{y_1 + y_2} \right)^{c_1} \left( 1 - \frac{y_1}{y_1 + y_2} \right)^{c_{\text{total}} - c_1} \quad (5)$$

これに基づいて帰無仮説  $H_0: \lambda_1 = \lambda_2$  に対する二項検定 (binomial test) を構築することができる。そのためには、帰無仮説の下で、実際に観測された  $c_1$  ないしそれよりも極端な値を得る確率を合計すればよく、この場合は、式(5)の二項分布の  $c_1$  までの累積分布を用いて片側 p 値を計算することができる (Chan & Bohidar, 1998)。すなわち、

$$\text{片側 p 値} = \sum_{k=0}^{c_1} \binom{c_{\text{total}}}{k} \left( \frac{y_1}{y_1 + y_2} \right)^k \left( 1 - \frac{y_1}{y_1 + y_2} \right)^{c_{\text{total}} - k}$$

前出の数値例を用いて二項検定を行うための SAS プログラムをプログラム 2 に示す。といっても、これはプログラムと言うほどのものではなく、本質的な部分は二項分布の累積分布関数である「CDF('BINOMIAL', c1, y1/(y1+y2), c1+c2)」だけであり、これにより片側 p 値 one\_sided\_p を計算している。このように、たったの 1 行で p 値を計算できてしまうのは興味深い。

プログラム 2. 二項検定のための SAS プログラム及びその出力

```
data ex2;
  set exmpl;
  one_sided_p = CDF('BINOMIAL', c1, y1/(y1+y2), c1+c2);
  keep one_sided_p;
run;

proc print noobs;
run;
```

出力

```
one_sided_p
0.024894
```

この数値例における片側 p 値は 0.024894 であり、片側 2.5%水準でぎりぎり有意差があることを示している。このことは、上で求めた 95%信頼区間の上限 0.99916 がぎりぎり 1 を下回っていたことと対応している。

## 5. 検出力及びサンプルサイズ計算

これまでと同様、ケース数は Poisson 分布に従うと仮定し、2 群の総ケース数が与えられた条件の下での統計的検定を考える。その際の条件設定を表 3 にまとめている。

表 3. サンプルサイズ計算のための臨床試験の想定

群	割付け 被験者数	人・年	発現率	帰無仮説 $H_0: \lambda_1 = \lambda_2 = \lambda$ の下で		対立仮説 $H_1: \lambda_1 = r\lambda_2$ の下で	
				期待ケース数	割合*	期待ケース数	割合*
群 1	$n_1 = \phi n_2$	$y_1 = \phi n_2 t$	$\lambda_1$	$\lambda y_1 = \lambda \phi n_2 t$	$\phi / (1 + \phi)$	$\lambda_1 y_1 = r \lambda_2 \phi n_2 t$	$\phi r / (1 + \phi r)$
群 2	$n_2$	$y_2 = n_2 t$	$\lambda_2$	$\lambda y_2 = \lambda n_2 t$	$1 / (1 + \phi)$	$\lambda_2 y_2 = \lambda_2 n_2 t$	$1 / (1 + \phi r)$
計	$n_{\text{total}} = (1 + \phi) n_2$	$(1 + \phi) n_2 t$		$(1 + \phi) \lambda n_2 t$	1	$(1 + \phi r) \lambda_2 n_2 t$	1

$\phi$ : 割付け比,  $r$ : rate ratio,  $t$ : 追跡期間 (全ての被験者で等しいと仮定)

\*: 総ケース数に対して各群のケース数が占める割合を示す。

群 1 の群 2 に対する割付け比を  $\phi$ 、全ての被験者の追跡期間が  $t$  であるとした場合、表 3 に示すように、帰無仮説  $H_0: \lambda_1 = \lambda_2$  の下では、総ケース数に対して群 1 のケース数が占める割合は  $\phi / (1 + \phi)$

となるものと期待され、よって、総ケース数が与えられた条件の下では、群 1 のケース数  $C_1$  は確率  $\phi/(1+\phi)$  の二項分布に従うと想定される。一方、対立仮説  $H_1:\lambda_1=r\lambda_2$  ( $r$  は rate ratio を表す) の下では、総ケース数に対して群 1 のケース数が占める割合は  $\phi r/(1+\phi r)$  となるものと期待され、よって、 $C_1$  は確率  $\phi r/(1+\phi r)$  の二項分布に従うと想定される。これらをまとめると次のようになる。

$$\begin{aligned} \text{帰無仮説 } H_0:\lambda_1=\lambda_2 \text{ の下: } P(C_1=c_1 | C_{\text{total}}=c_{\text{total}}, H_0:\lambda_1=\lambda_2) &= \binom{c_{\text{total}}}{c_1} \left(\frac{\phi}{1+\phi}\right)^{c_1} \left(1-\frac{\phi}{1+\phi}\right)^{c_{\text{total}}-c_1} \\ \text{対立仮説 } H_1:\lambda_1=r\lambda_2 \text{ の下: } P(C_1=c_1 | C_{\text{total}}=c_{\text{total}}, H_1:\lambda_1=r\lambda_2) &= \binom{c_{\text{total}}}{c_1} \left(\frac{\phi r}{1+\phi r}\right)^{c_1} \left(1-\frac{\phi r}{1+\phi r}\right)^{c_{\text{total}}-c_1} \end{aligned}$$

これら 2 つの式の右辺の違いは rate ratio の  $r$  が入っているか、いないかだけである。これらに基づき、下記に示すようにサンプルサイズの計算を行うことができる (Chan & Bohidar, 1998)。

まず、片側検定を考え、第 1 種の過誤の確率が  $\alpha/2$  以下となるようにするために、帰無仮説  $H_0:\lambda_1=\lambda_2$  の下で群 1 のケース数が  $c_1$  以下である確率が  $\alpha/2$  以下となるような  $c_1$  の最大値を求め、それを critical value ( $c_v$ ) とする。それは  $c_{\text{total}}$  に依存するため、 $c_v(c_{\text{total}})$  と記すこととすると、

$$c_v(c_{\text{total}}) = \max \left\{ c_1 \mid \sum_{k=0}^{c_1} \binom{c_{\text{total}}}{k} \left(\frac{\phi}{1+\phi}\right)^k \left(1-\frac{\phi}{1+\phi}\right)^{c_{\text{total}}-k} \leq \alpha/2, c_1 \in \{0, 1, 2, \dots\} \right\}$$

検出力は、対立仮説  $H_1:\lambda_1=r\lambda_2$  の下での二項分布の  $c_v$  までの累積分布

$$\text{Power}(c_{\text{total}}) = \sum_{k=0}^{c_v(c_{\text{total}})} \binom{c_{\text{total}}}{k} \left(\frac{\phi r}{1+\phi r}\right)^k \left(1-\frac{\phi r}{1+\phi r}\right)^{c_{\text{total}}-k}$$

によって計算できる。必要な総ケース数は、上式の検出力が  $1-\beta$  以上になるような  $c_{\text{total}}$  の中から最小値を選べばよい。すなわち、

$$c_{\text{total}}^* = \min \left\{ c_{\text{total}} \mid \text{Power}(c_{\text{total}}) \geq 1-\beta, c_{\text{total}} \in \{1, 2, 3, \dots\} \right\}$$

この必要な総ケース数  $c_{\text{total}}^*$  を得るために必要な割付け被験者数  $n_{\text{total}}^*$  は、表 3 の最下行にある  $(1+\phi r)\lambda_2 n_2 t$  と  $c_{\text{total}}^*$  とが等しいと置くことから導かれる次式により求めることができる。

$$n_{\text{total}}^* = \frac{(1+\phi)c_{\text{total}}^*}{(1+\phi r)\lambda_2 t}$$

このサンプルサイズ計算のための SAS プログラムをプログラム 3 に示す。ここでは、割付け比 (allocratio) を 2、rate ratio (rr) を 2、対照群における 1 年当たりのイベント発現率 (eventrate2) を 0.06、追跡期間 (t) を一律に 1 年、片側有意水準を 2.5%、検出力 90% と設定している。8 行目の DO ループでは、検出力 (power) が 90% を超えるまで (その条件判定に UNTIL 句を使用) 総ケース数 (ctotal) を 1 ずつ増加させている。9 行目の DO ループでは (ループの中味は空であるが)、帰無仮説の下での第 1 種の過誤の確率が  $\alpha/2=0.025$  を超えるまで群 1 のケース数 (c1) を 1 ずつ増加させてゆき、次に、11 行目で 1 を引いて critical value (cv) としている。12 行目では、対立仮説の下での二項分布の cv までの累積分布関数を計算し、検出力としている。14 行目では、計算された必要な総ケース数 (ctotal) から、必要な被験者数 (ntotal) を計算している。



プログラム 3. サンプルサイズ計算のための SAS プログラム及びその出力

```

1  data ex3;
2  allocratio=2;
3  rr=0.3;
4  eventrate2=0.06;
5  t=1;
6  p1_0=allocratio/(1+allocratio);
7  p1_1=allocratio*rr/(1+allocratio*rr);
8  do ctot=1 to 999 until (power>0.9);
9      do c1=0 to ctot until (CDF('BINOMIAL',c1,p1_0,ctot)>0.025);
10     end;
11     cv=c1-1;
12     power=CDF('BINOMIAL',cv,p1_1,ctot);
13 end;
14 ntotal=ceil((1+allocratio)*ctot/(1+allocratio*rr)/eventrate2/t);
15 keep cv ntotal ctot power;
16 run;
17
18 proc print noobs;
19 run;

```

出力

ctotal	power	cv	ntotal
33	0.92929	16	1032

## 6. Wilson スコア信頼区間を用いた rate ratio の信頼区間の提案

割合の信頼区間として Clopper-Pearson 信頼区間は保守的であり、それに比べ、平均的な被覆確率 (coverage probability) がより名義的信頼水準に近いという意味で、Wilson スコア信頼区間の方が優れていることが Newcombe (1998) 及び Agresti & Coull (1998) によって示されている。このことから、割合  $p_1$  の信頼区間として、Clopper-Pearson 信頼区間ではなく、代わりに Wilson スコア信頼区間を用いることにより、より優れた rate ratio の信頼区間が得られるものと考えられる。

Wilson スコア信頼区間の下限及び上限は

$$\hat{p}_{1,Wilson,lower}, \hat{p}_{1,Wilson,upper} = (2c_{total}\hat{p}_1 + z_{1-\alpha/2}^2 \pm z_{1-\alpha/2} \sqrt{z_{1-\alpha/2}^2 + 4c_{total}\hat{p}_1(1-\hat{p}_1)}) / 2(c_{total} + z_{1-\alpha/2}^2)$$

で与えられる。ここで  $z_{1-\alpha/2}$  は標準正規分布の  $(1-\alpha/2) \times 100\%$  点を表す。これら信頼限界を式(4)の  $\hat{p}_1$  に代入し、次のような新たな rate ratio の信頼限界を得る。

$$\hat{r}_{Wilson,lower} = \frac{y_2}{y_1} \cdot \frac{\hat{p}_{1,Wilson,lower}}{1-\hat{p}_{1,Wilson,lower}}, \quad \hat{r}_{Wilson,upper} = \frac{y_2}{y_1} \cdot \frac{\hat{p}_{1,Wilson,upper}}{1-\hat{p}_{1,Wilson,upper}}$$

この信頼区間を計算するための SAS プログラムをプログラム 4 に示す。これはプログラム 1 の Clopper-Pearson 信頼区間の部分を Wilson スコア信頼区間で置き換えただけのものである。

プログラム 4. Wilson スコア信頼区間を用いた Rate ratio の信頼区間のための SAS プログラム

```

data ex4;
  set exmpl;
  p1=c1/(c1+c2);
  cttotal=c1+c2;
  z=probit(0.975);
  wilsonlower=(2*cttotal*p1+z**2-z*sqrt(z**2+4*cttotal*p1*(1-p1)))/2/(cttotal+z**2);
  wilsonupper=(2*cttotal*p1+z**2+z*sqrt(z**2+4*cttotal*p1*(1-p1)))/2/(cttotal+z**2);
  rr =y2/y1*p1/(1-p1);
  lower=y2/y1*wilsonlower/(1-wilsonlower);
  upper=y2/y1*wilsonupper/(1-wilsonupper);
  keep rr lower upper;
run;

proc print noobs;
run;

```

出力

rr	lower	upper
0.36606	0.15155	0.88418

予想通り、Wilson スコア信頼区間を用いた場合のほうが、より保守的と考えられる Clopper-Pearson 信頼区間を用いた場合よりも rate ratio の信頼区間の幅が狭くなっている。

この新たに提案した信頼区間の被覆確率を exact な信頼区間と比較評価するために、モンテカルロ・シミュレーションを行った。その際の参考とするために、漸近的な unconditional な信頼区間とも比較をおこなった。その信頼区間は

$$\exp \left[ \log \left( \frac{c_1 / y_1}{c_2 / y_2} \right) \pm z_{1-\alpha/2} \sqrt{\frac{1}{c_1} + \frac{1}{c_2}} \right]$$

で与えられる（例えば、Rothman & Greenland, 2008, p.295 参照）。

問題を簡単にするため、2 群間の人・年の比  $y_1 / y_2$  を 1 に固定した。そして各群の人・年を 100 ~ 1000 まで 10 刻みで変化させた。2 群の総ケース数の期待値が把握できるようにするために、2 群の平均的な年当たりの発現率を固定することとし、ここではそれを 0.05 に設定した。その上で、種々の rate ratio に対して 2 群の発現率の平均が 0.05 になるように各群の発現率を決定した。

シミュレーションの 1 試行においては、各群独立にケース数を Poisson 乱数を用いて発生させ (SAS 関数 RANPOI を使用)、3 種類の 95% 信頼区間を計算し、それらが設定した真の rate ratio を含むかどうかを判定した。それを 10000 回反復し、各信頼区間が真の rate ratio を含む確率、すなわち被覆確率を計算した。ただし、期待ケース数が少ない場合には、生成されたケース数  $c_1$  ないし  $c_2$  がゼロになることがあるが、その場合は被覆確率の計算から除外した。

被覆確率は rate ratio の値に左右されるため、1 つの rate ratio の値についての被覆確率を評価するのではなく、種々の rate ratio の値についての被覆確率の平均を取ることにした。このように複数のパラメタ値についての平均被覆確率を用いる方法は Newcombe (1998) や Agresti & Coul (1998) が用い

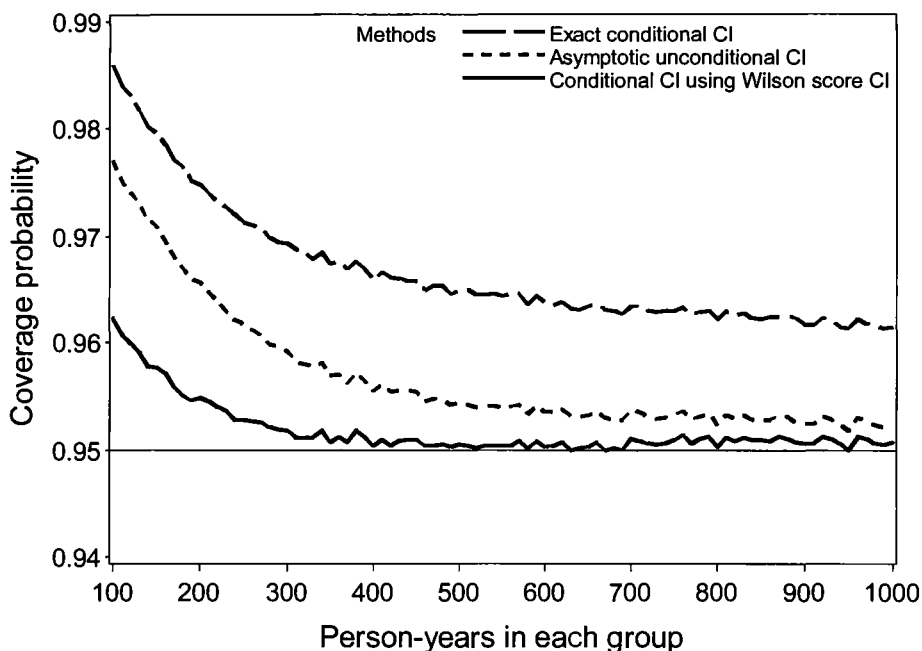


図 1. Rate ratio について 3 種類の信頼区間の平均被覆確率  
 2 群の年当たりの平均発現率を 0.05 に固定  
 0.25~4 の範囲の種々の rate ratio の被覆確率を平均

た方法と類似したものである。ここで用いた rate ratio の範囲としては、現実の研究でも起こりそうな 0.25~4 の範囲を想定し、この範囲を対数スケール上で 40 等分し、評価すべき rate ratio の値を決定した。このようにして種々の人・年に対する平均被覆確率を評価した結果を図 1 に示す。

この図から、どのような人・年の場合についても、Exact な信頼区間は他の 2 つに比べて保守的であることがわかる。一方、Wilson スコア信頼区間を用いた信頼区間の被覆確率は名義的信頼水準である 95% を下回ることなく、常に他の 2 つよりも名義的信頼水準に近い値となっている。このことは、被覆確率の観点からは、3 つの信頼区間の中では、この方法が最も優れたものであることを示している。一方、漸近的な unconditional な信頼区間は他の 2 つの中間的な被覆確率を示している。

このように Wilson スコア信頼区間を用いた rate ratio の信頼区間は望ましい性質を有するが、1 つ問題がある。それは、この方法に対応した統計的検定法やサンプルサイズ計算の方法が存在しないことである。そのため、実際の臨床試験や調査でこの方法を用いる際には、これらの点について何らかの工夫をする必要があるかもしれない。

## まとめ

本論文では、2 つの群のケース数が Poisson 分布に従うと仮定した場合の、2 群の総ケース数が固定された条件の下での exact な条件付きの rate ratio の信頼区間の計算方法について述べた。また、それに対応する統計的検定やサンプルサイズ計算方法についても述べた。さらに、exact な条件付きの rate ratio の信頼区間で用いていた Clopper-Pearson 信頼区間を Wilson スコア信頼区間で置き換えることにより、被覆確率の観点から、より優れていると考えられる信頼区間を得た。これらの計算

方法はいずれも SAS ではサポートされているものではない。しかし、それらは SAS のデータステップを用いて容易に計算できることを示した。

## 引用文献

- 1) Agresti, A., Coull, B. A., Approximate is better than exact for interval estimation of binomial proportions. *The American Statistician*, 52: 119-126, 1998.
- 2) Breslow, N. E., Day, N. E., *Statistical methods in cancer research. Volume II. The design and analysis of cohort studies.* International Agency for Research on Cancer, 1987.
- 3) Chan, I. S. F., Bohidar, N. R., Exact power and sample size for vaccine efficacy studies. *Commun. Statist.-Theory Meth.*, 27: 1305-1322, 1998.
- 4) Newcombe, R. G., Two-sided confidence intervals for the single proportion – comparison of seven methods, *Statist. Med.*, 17: 857-872, 1998.
- 5) Przyborowski, J., Wilenski, H., Homogeneity of Results in Testing Samples from Poisson Series: With an Application to Testing Clover Seed for Dodder, *Biometrika* 31: 313-323, 1940
- 6) Rothman, K. J., Greenland, S., Lash, T. L., *Modern epidemiology.* 3rd ed., Lippincott Williams & Wilkins, 2008.

## 付録. 2つの独立な Poisson 分布する変数の和を固定した場合の条件付き分布

2つの独立に Poisson 分布する変数  $C_1$  及び  $C_2$  があり、その確率関数が式(1)で与えられているものとする。よく知られているように、2つの変数の和  $C_{\text{total}} = C_1 + C_2$  もやはり Poisson 分布に従い、それが  $c_{\text{total}}$  となる確率は次式で与えられる。

$$P(C_{\text{total}} = c_{\text{total}} | \lambda_1, \lambda_2) = \frac{(\lambda_1 y_1 + \lambda_2 y_2)^{c_{\text{total}}} e^{-(\lambda_1 y_1 + \lambda_2 y_2)}}{c_{\text{total}}!}$$

ここで  $C_1$  及び  $C_{\text{total}} = C_1 + C_2$  の同時分布が

$$P(C_1 = c_1, C_{\text{total}} = c_{\text{total}} | \lambda_1, \lambda_2) = P(C_1 = c_1 | \lambda_1) \cdot P(C_2 = c_{\text{total}} - c_1 | \lambda_2) = \frac{(\lambda_1 y_1)^{c_1} e^{-\lambda_1 y_1}}{c_1!} \cdot \frac{(\lambda_2 y_2)^{c_{\text{total}} - c_1} e^{-\lambda_2 y_2}}{(c_{\text{total}} - c_1)!}$$

となることを用いて、 $C_{\text{total}} = c_{\text{total}}$  の条件の下での  $C_1$  の条件付き分布は次式のようにになる。

$$\begin{aligned} P(C_1 = c_1 | C_{\text{total}} = c_{\text{total}}, \lambda_1, \lambda_2) &= \frac{P(C_1 = c_1, C_{\text{total}} = c_{\text{total}} | \lambda_1, \lambda_2)}{P(C_{\text{total}} = c_{\text{total}} | \lambda_1, \lambda_2)} = \frac{\frac{(\lambda_1 y_1)^{c_1} e^{-\lambda_1 y_1}}{c_1!} \cdot \frac{(\lambda_2 y_2)^{c_{\text{total}} - c_1} e^{-\lambda_2 y_2}}{(c_{\text{total}} - c_1)!}}{\frac{(\lambda_1 y_1 + \lambda_2 y_2)^{c_{\text{total}}} e^{-(\lambda_1 y_1 + \lambda_2 y_2)}}{c_{\text{total}}!}} \\ &= \frac{c_{\text{total}}!}{c_1! (c_{\text{total}} - c_1)!} \cdot \frac{(\lambda_1 y_1)^{c_1} (\lambda_2 y_2)^{c_{\text{total}} - c_1}}{(\lambda_1 y_1 + \lambda_2 y_2)^{c_{\text{total}}}} \\ &= \binom{c_{\text{total}}}{c_1} \left( \frac{\lambda_1 y_1}{\lambda_1 y_1 + \lambda_2 y_2} \right)^{c_1} \left( 1 - \frac{\lambda_1 y_1}{\lambda_1 y_1 + \lambda_2 y_2} \right)^{c_{\text{total}} - c_1} \end{aligned}$$

このように成功確率  $\lambda_1 y_1 / (\lambda_1 y_1 + \lambda_2 y_2)$  の二項分布が導かれる (Przyborowski & Wilenski, 1940)。

## JMPスクリプト言語（JSL）を用いた前臨床における割付け手法のアドイン開発

○原茂恵美子、佐藤拓哉、篠津和夫、佐藤耕一  
(株式会社タクミインフォメーションテクノロジー)

探索的なデータ分析ができるソフトウェアJMPでは、プログラムを書くことなく、マウス操作によりさまざまな統計解析を行うことができる。しかし、前臨床試験で必要な個体の割付け（群分け）に関する手法は、JMPには搭載されていない。そのため、JMPスクリプト言語（以下、JSL）を用いて、JMPの既存手法と同様に簡単なマウス操作により割付け結果を得ることが出来るアドイン「多変数によるブロック化割付け」を開発し、アドインとして実装したので、その内容を報告する。

ブロック化割付けとは、実験の目的となる測定項目に影響を与える事前情報項目（調整項目）の分布が各群均質になるよう、積極的にバランスをとって個体を割付ける方法である。本手法では、複数の調整項目の平均値と分散が、各群で均質になるよう割付けを行っている。群分けアルゴリズムは「谷本学校毒性質問箱第3号」（株）サイエンティスト社(2000)の「実験動物の群分けアルゴリズムとプログラム」芳賀敏郎(p. 2)を引用した。

本手法の入力データは、EXCELあるいはJMPのデータテーブルで用意する。本手法をJMPのメニューバーから起動し、入力データを読み込み、割付けのパラメータを設定し、解析実行ボタンをクリックすることにより、解析が行われ、データテーブルに割付け結果が出力される。これらは全て、マウス操作あるいはキーボード入力で行われる。

本手法の開発は、解析部分とインターフェース部分にわけて行った。解析部分はJSLによるプログラミングで作成し、インターフェース部分はアプリケーションビルダーを使用した。アプリケーションビルダーは、簡単なマウス操作によりインターフェース機能を作成することができ、JSLが自動生成される。また、自動生成されるJSLに、エラーチェックロジックをプログラミングで追加することにより、入力データあるいはパラメータの設定に不整合がある場合や、解析が実行されないなどのエラー処理機能を実装することができた。解析部分とインターフェース部分を別々に作成した後、解析JSLとインターフェースJSLを結合し、アドインビルダーを用いてJMPアドインを作成する。作成したJMPアドインにより、JMPの既存手法のように、メニューバーから手法を選択し、実行することができるようになる。また、既存手法では、手法名にカーソルを合わせるとその手法に関する説明が表示されるが、今回開発したアドインビルダーにおいてもヒントが表示されるように設定した。

今回の開発により、JSL、アプリケーションビルダーおよびアドインビルダーを用いることで、JMPに搭載されていない手法を実装でき、ユーザーはスクリプトやプログラミングを意識することなく、既存の手法と同様のマウス操作でJMPに搭載されていない「多変数によるブロック化割付け」を解析することが可能となった。

今回、前臨床試験の割付けに関するアドイン開発の事例を紹介したが、Williams型多重比較など、一般的に前臨床試験において頻繁に使用されるがJMPに搭載されていない手法は、他にも存在するため、同様に、アドイン開発することにより利用範囲が広がることになる。また、他の分野においても、必要な手法がJMPに搭載されていない場合には、今回紹介したように、JMPアドインを使用することにより、JMPを効果的、有効的に利用されることが示唆された。

# Kaplan-Meierプロット・Forestプロット作成の応用：

## グラフ出力範囲内・範囲外への数値出力

○魚住 龍史<sup>1</sup> 吉田 早織<sup>2</sup> 平井 隆幸<sup>2</sup> 浜田 知久馬<sup>3</sup>

<sup>1</sup>京都大学大学院医学研究科 医学統計生物情報学

<sup>2</sup>日本化薬株式会社 医薬開発本部 解析チーム

<sup>3</sup>東京理科大学 工学部 情報工学科

Advancement in both Kaplan-Meier and forest plots: quantitative results output inside or outside the graph area

Ryuji Uozumi<sup>1</sup>, Saori Yoshida<sup>2</sup>, Takayuki Hirai<sup>2</sup>, and Chikuma Hamada<sup>3</sup>

<sup>1</sup>*Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine*

<sup>2</sup>*Biostatistics Team, Pharmaceutical Development Division, Nippon Kayaku Co.,Ltd*

<sup>3</sup>*Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science*

## 要旨

Kaplan-Meier プロットおよび Forest プロットは、いずれも医薬品開発でよく用いられるグラフである。本稿では、ODS GRAPHICS による機能を用いて、様々な修飾を加えた Kaplan-Meier プロットおよび Forest プロットを作成する方法を報告する。Kaplan-Meier プロットにリスク集合の大きさを修飾させて作成する場合、LIFETEST プロシジャのデフォルトではグラフ出力範囲内、すなわち横軸の上段にリスク集合の大きさが出力される。リスク集合の大きさをグラフ出力範囲外、すなわち横軸の下段に出力させる方法として、V9.4 の SGPLOT プロシジャから追加された XAXISTABLE ステートメントを用いる方法を取りあげる。次に、Forest プロットの作成を考えると、グラフの横に実際の点推定値と信頼区間も示されることが多い。V9.4 の SGPLOT プロシジャから追加された複数のステートメントを併用することで、実際の学術論文に用いられるようなクオリティの高い Forest プロットを作成する方法を報告する。

キーワード: 生存時間解析 Kaplan-Meier プロット Forest プロット サブグループ解析 SGPLOT 修飾を加えたプロット リスク集合 XAXISTABLE YAXISTABLE X2AXIS Y2AXIS

## 1 はじめに

データのグラフ化は、データ解析の結果を可視化するために有用な手段である。SAS では、V9.2 から ODS (Output Delivery System) GRAPHICS による機能が正規版として追加され、グラフィックベースのきれいなプロットを出力できるようになった。例えば、ODS GRAPHICS ON の状態で LIFETEST プロシジャを実行させれば、Kaplan-Meier プロットが自動的に出力される。オプションを指定すれば、様々な修飾を加

えることが可能である。また、V9.2 から追加された SG (Statistical Graphics) プロシジャである、SGPLOT プロシジャや SGPANEL プロシジャを用いれば、より見栄えの良いプロットを作成することができる。

本稿では、医薬品開発で用いられるグラフを SAS で作成することに焦点を当てる。これまで SAS によるグラフ作成に関して多くの報告が行われている。西本 (2013) は GGPLOT プロシジャと SGPLOT プロシジャを比較した上で、SGPLOT プロシジャによるグラフ作成について解説した。豊泉ら (2014) によって、SGPLOT プロシジャによるデータの可視化の有用性について述べられている。さらに、魚住・浜田 (2012) は SG Annotation の機能を用いて、医薬品開発でよく用いられるグラフを作成する有用性を示した。また、より実務的な応用として、SG プロシジャ、GTL (Graph Template Language) および ODS PDF を用いた簡易解析帳票の作成方法事例や、SAS と HTML アプリケーションによる CDISC ADaM (Clinical Data Interchange Standards Consortium Analysis Data Model) 形式の解析用データセットを用いた解析帳票・グラフ簡易作成ツールの開発事例についても報告されている (高浪, 2011, 2012)。加えて、V9.4 から利用できるようになった ODS PowerPoint を用いて、SAS で作成したグラフを PowerPoint にアウトプットさせる方法も報告されている (吉田ら, 2015)。

医薬品開発で用いられるグラフのうち、Kaplan-Meier プロットは生存時間データの視覚化の手段として 20 年以上前から用いられており (大橋・浜田, 1995)、V9.2 リリース以降、SAS による Kaplan-Meier プロットの作成に関する多くの報告が行われている。長島・佐藤 (2010) は、LIFETEST プロシジャおよび GGPLOT プロシジャを用いて、Kaplan-Meier プロットに付加情報を加えるマクロを開発している。魚住・浜田 (2011) は、SGPLOT プロシジャおよび SGRENDER プロシジャを用いて、Kaplan-Meier プロットを作成する手順を示している。LIFETEST プロシジャではラスタ形式のグラフ出力であり、ベクター形式の出力ができない一方、SGPLOT プロシジャではベクター形式の出力も可能である点からも有用といえる (平井ら, 2015)。

長島・佐藤 (2010) および魚住・浜田 (2011) のいずれの報告においても、Kaplan-Meier プロットのリスク集合の大きさ (number of subjects at risk) を出力させる方法について言及されている。リスク集合の大きさは、LIFETEST プロシジャのオプションを指定すれば、ODS GRAPHICS による出力で自動的に付加されるが、より工夫を凝らした出力をするための方法が報告されている。例えば、学術論文において、リスク集合の大きさは Kaplan-Meier プロットの横軸の下段に示されることが多いが、LIFETEST プロシジャのオプションで ATRISK を指定して出力すると、Kaplan-Meier プロットの横軸の上部に示される。よって、リスク集合の大きさを Kaplan-Meier プロットの下段に示す方法を考える意義があるといえる。Kaplan-Meier プロットに限らず、その他にも、例えば Forest プロットの作成においても工夫した出力を考える意義がある。Forest プロットはメタアナリシスやサブグループ解析の報告によく用いられるグラフであり、プロットに加えて、グラフの横に実際の推定値と信頼区間を示すことが多い。以上のようなことから、医薬品開発に携わる統計家および SAS プログラマにとって、グラフ出力範囲外にリスク集合の大きさや信頼区間などの情報を出力させる方法を考える必要が生じる。

本稿では、医薬品開発でよく用いられる Kaplan-Meier プロットおよび Forest プロットを作成するにあたり、グラフ出力範囲外に数値を出力させる方法として、LIFETEST プロシジャおよび SGPLOT プロシジャを用いた方法をレビューした上で、新たにできるようになった方法を報告する。

なお、本稿において示すプロットは、モノクロ印刷されても識別しやすいよう、ODS の出力としてジャーナルスタイル "STYLE=JOURNAL" を指定している。例えば、群別のグラフを示す場合、デフォルトでは群の違いが色で分けて出力されてしまう。ジャーナルスタイルの場合、モノクロ印刷で群の違いを識別できるよう、実線と破線で分けて出力される。

## 2 Kaplan-Meier プロットの作成

本節では、リスク集合の大きさの情報を修飾した Kaplan-Meier プロットを作成することを考える。

Kaplan-Meier プロットを作成する対象データとして、肺癌のデータ（データセット名：VALung）を用いる。データセット VALung (Veterans Administration Lung cancer trial) は、Kalbfleisch and Prentice (2002) で使用されたデータを一部抽出したものであり ( $n = 137$ )、SAS/STAT PHREG プロシジャのマニュアル、および大橋ら (2016) 第 3 章においても用いられている。この研究の目的は、男性の進行性肺癌患者を対象としたランダム化比較試験であり、治療法として、試験治療と標準治療を比較するために行われた。評価項目は死亡までの時間 (日) (変数名: Time) で、共変量の 1 つとして組織型 (変数名: Cell) が挙げられる。組織型は 4 水準 (Cell = 'adeno' 「腺癌」, 'small' 「小細胞癌」, 'large' 「大細胞癌」, 'squamous' 「扁平上皮癌」) のカテゴリカル変数である。本節では、組織型別の Kaplan-Meier プロットを作成することを考える。

### 2.1 グラフの出力範囲内へのリスク集合の出力

LIFETEST プロシジャでは、デフォルトではリスク集合の大きさは付与されないが、PLOTS = S のオプションとして、ATRISK を指定すれば、図 1 に示すような出力が得られる。その他のオプションの指定に関しては、大橋ら (2016) を参照されたい。

```
SAS プログラム proc lifetest data=VALung plots=s(atrisk atrisktick);  
                    time Time*Censor(1); strata Cell;  
                    run;
```

#### 出力結果

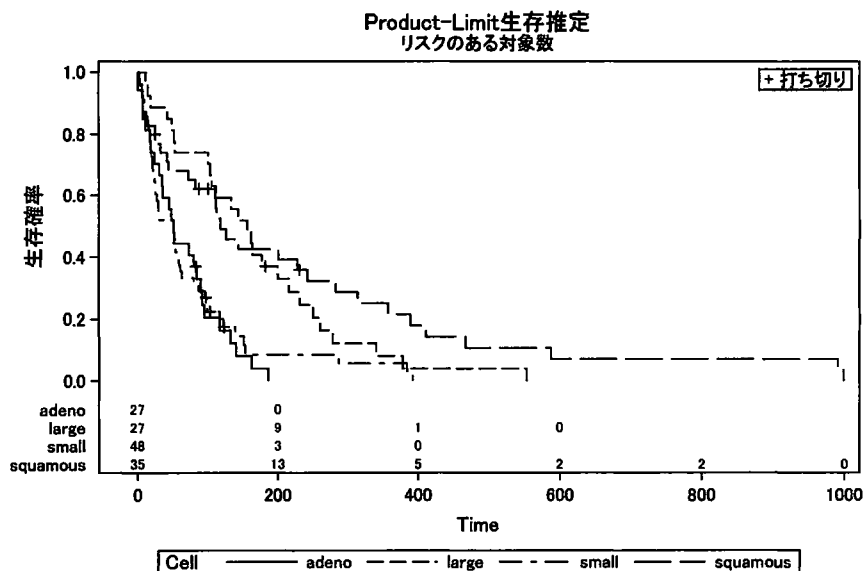


図 1: LIFETEST プロシジャによる Kaplan-Meier プロットの作成 1



```

SAS プログラム ods graphics;
ods output Survivalplot=Survivalplot;
proc lifetest data=VALung atrisk plots=s(atrisk);
  time Time*Censor(1); strata Cell;
run;
*-----*;
*-- データハンドリング (詳細は付録A参照) --*;
*-----*;
proc sgplot data=Survivalplot2 noautolegend;
  step x=Time y=Survival / group=StratumNum name='Survival';
  scatter x=Time y=Censored
    / group=StratumNum markerattrs=(symbol=plus);
  scatter x=tAtRisk y=StratumNum / markerchar=atrisk y2axis;
  keylegend 'Survival'
    / location=outside noborder position=bottom;
  yaxis offsetmin=0.20 values=(0 to 1 by 0.1) label='Survival';
  y2axis offsetmin=0.03 offsetmax=0.85
    min=1 max=4 label=' ' display=(noticks novalues);
  xaxis offsetmin=0.05
    values=(0 to &TimeMax by &TimeInterval) label='Time';
  format StratumNum Cellf.;
run;

```

出力結果

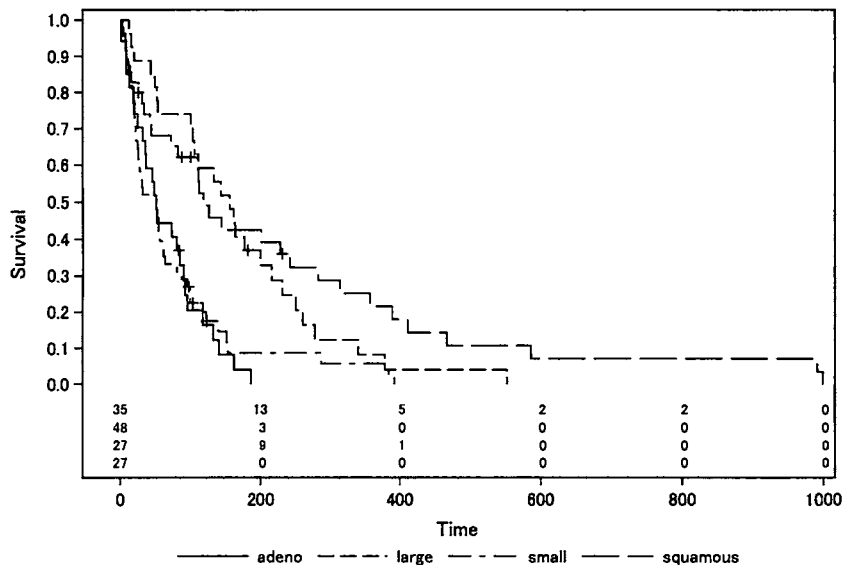


図 2: SGPLOT プロシジャによる Kaplan-Meier プロットの作成 1

しかし、図1の出力では、リスク集合の大きさが出力されていない時点および群が存在する。データセット VALung の特徴として、例数に対するイベント割合が高いため、後半の時点ではリスク集合の大きさが0となってしまう。リスク集合が0となると、0になった最初の時点では出力されるが、それ以降の時点では0は出力されない。本稿ではリスク集合の大きさが0の場合は0に統一する出力を考える。PLOTS=S のオプションで指定できない修飾を加える場合、生存関数の推定値を一度データセットに落として、グラフ関連のプロシジャで出力し直す必要がある。

図2はSGPLOT プロシジャを用いたプログラムおよびその出力結果を示しており、大橋ら (2016) 第2章で解説されている方法に加えて、Y2AXIS ステートメントに加えた上でYAXIS ステートメントのOFFSETMIN = を指定することで、グラフの出力範囲内にリスク集合の大きさを出力させるスペースを確保している。YAXIS および XAXIS ステートメントはグラフの縦軸および横軸の指定として、グラフの左側および下側の軸の設定を行っている一方、Y2AXIS および X2AXIS ステートメントではグラフの右側および上側の軸の設定を行っている。OFFSETMIN = と OFFSETMAX = を指定すれば、軸の下限と上限をそれぞれ指定できる。図2のプログラムでは、YAXIS ステートメントで OFFSETMIN = 0.20 を指定し、縦軸全体の20%の値からスタートさせている。縦軸全体は生存割合1であるため、生存割合20%に相当する位置が原点として描かれている。一方、右側の縦軸に関して、Y2AXIS ステートメントで OFFSETMIN = 0.03 を指定することで、横軸とリスク集合の値が重ならないように設定している。さらに、OFFSETMAX = 0.85 を指定することで、全体の85%の部分を出力範囲の上限としている。なお、ODS OUTPUT によって得られたデータセットに対して、グラフで出力する時点のリスク集合の大きさが0であった場合に0が出力されるよう、データハンドリングした上で、SGPLOT プロシジャを実行させる必要がある。データハンドリングの詳細は付録Aを参照されたい。

## 2.2 グラフの出力範囲外へのリスク集合の出力

図1および図2は、いずれもグラフの出力範囲内にリスク集合の大きさを追加した Kaplan-Meier プロットである。Kaplan-Meier プロットに関しては、学術論文においてリスク集合の大きさを Kaplan-Meier プロットの横軸の下段に示されることが多い。そのための方法として、SGRENDER プロシジャを用いて、TEMPLATE プロシジャで事前に定義したテンプレート、特に LAYOUT LATTICE ステートメントなどを活用することによって、リスク集合の大きさを Kaplan-Meier プロットの下段に示すことが可能である (魚住・浜田, 2011)。しかし、テンプレートの定義には、長文のプログラムを組まなければならない。そこで、V9.4 のSGPLOT プロシジャに追加されたステートメントを用いて、リスク集合の大きさを Kaplan-Meier プロットの下段に示す方法を考える。

図3はSGPLOT プロシジャを用いたプログラムおよびその出力結果を示している。V9.4から利用できる XAXISTABLE ステートメントを用いて、リスク集合の大きさを Kaplan-Meier プロットの下段に示している。XAXISTABLE ステートメントでは出力させたい値を表す変数 AtRisk を指定した上、オプションとしてリスク集合の大きさを出力させる横軸の値を持つ変数 tAtRisk を X = に指定し、群に該当する変数 StratumNum を CLASS = に指定している。共通の横軸を用いて、Kaplan-Meier プロットの出力とリスク集合の大きさの出力を行っているため、この場合 XAXISTABLE ステートメントを用いている。共通の縦軸を用いる場合は YAXISTABLE ステートメントを用いればよい。なお、XAXISTABLE ステートメントを用いて、グラフの範囲内にリスク集合の大きさを出力させたい場合、LOCATION = INSIDE をオプションとして指定すればよい。

図3では、SGPLOT プロシジャを用いてリスク集合の大きさを Kaplan-Meier プロットの下段に示したが、実は SAS/STAT 12.1 以降の LIFETEST プロシジャでは、オプション OUTSIDE を指定することで、リスク集合の大きさを Kaplan-Meier プロットの下段に示すことが可能である (図4)。なお、OUTSIDE(*k*) と指定することで、ODS GRAPHICS で構成される全体の出力範囲のうち、リスク集合の大きさを表示させる割合を指定できる。*k* のデフォルト値は、群の数の 0.035 倍である。図4の場合、4群に対してプロットを描いているため、全体のうち 14%の範囲にリスク集合の大きさを出力させている。

```
SAS プログラム proc sgplot data=Survivalplot2 noautolegend;
    step x=Time y=Survival / group=StratumNum name='Survival';
    scatter x=Time y=Censored
        / group=StratumNum markerattrs=(symbol=plus);
    xaxistable AtRisk / x=tAtRisk class=StratumNum;
    keylegend 'Survival'
        / location=outside noborder position=bottom;
    yaxis values=(0 to 1 by 0.1) label='Survival';
    xaxis values=(0 to &TimeMax by &TimeInterval) label='Time';
    format StratumNum Cellf.;
run;
```

出力結果

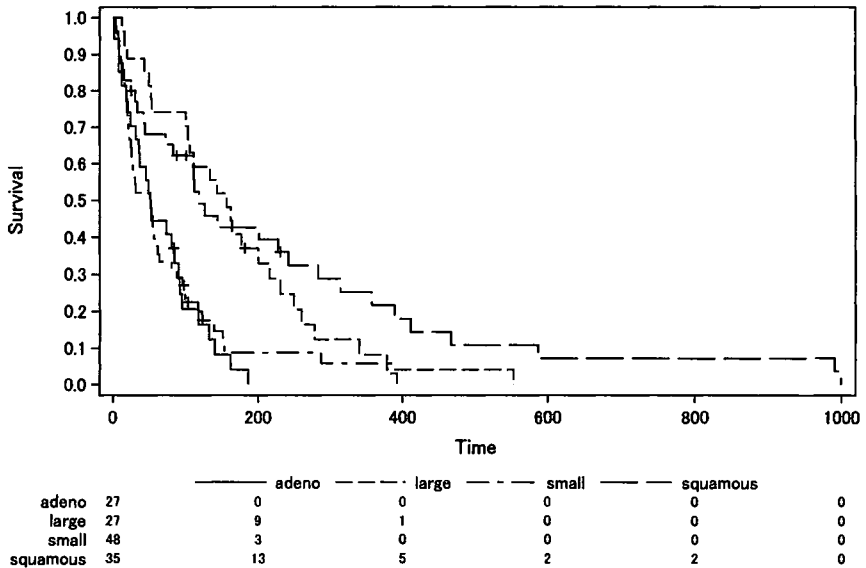


図3:SGPLOT プロシジャによる Kaplan-Meier プロットの作成 2

```
SAS プログラム proc lifetest data=VALung plots=s(atrisk atrisktick outside);
                    time Time*Censor(1); strata Cell;
run;
```

## 出力結果

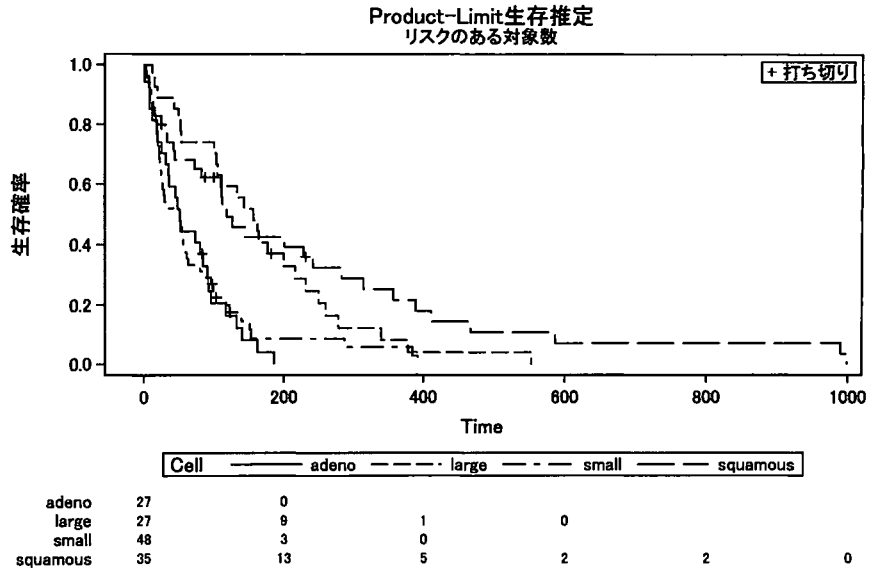


図 4: LIFETEST プロシジャによる Kaplan-Meier プロットの作成 2

### 3 Forest プロットの作成

医薬品開発では様々なプロットを用いてデータの可視化が行われるが、メタアナリシスやサブグループ解析の報告においては Forest プロットがよく用いられる。メタアナリシスであれば研究間の異質性を視覚的に確認するために、サブグループ解析であればサブグループ間でどのような傾向がみられるか確認するために用いられる。

Forest プロットにおいても、Kaplan-Meier プロットに対するリスク集合の大きさのように、実際の推定値およびその信頼区間などの修飾を加えることが多い。近年報告された臨床論文におけるサブグループ解析の結果を例に挙げると、Borghaei *et al.* (2015) はハザード比とその 95% 信頼区間に加え、各サブグループの例数も修飾して出力している。Motzer *et al.* (2015) は各群のイベント割合を修飾しており、点推定値のプロットの大きさについてもそれぞれのサブグループの例数に依存させて変化させている。Turner *et al.* (2015) は各サブグループにおける交互作用の p 値についても修飾させている。

SAS による Forest プロットの報告として、堀田 (2013) は、GPLOT プロシジャとその Annotation の機能を用いて、連続量データに対して Forest プロットを作成する方法を取りあげている。本稿では、Forest プロットに修飾する内容として、点推定値のプロットの大きさを例数に依存して変化させた上、各サブグループにおける例数、イベント割合、点推定値と信頼区間、交互作用の p 値を修飾して出力することとする。前節で用いたデータセット VALung における治療法 (Therapy2=1 「試験治療」、2 「標準治療」) の違いのサブグループ解析として、年齢カテゴリ (AgeC=1 「<65 歳」、2 「>=65 歳」)、組織型 (CellC=1 「腺癌」、2 「大細胞

癌」, 3「小細胞癌」, 4「扁平上皮癌」), 既往歴の有無 (Prior = 0「無」, 1「有」) をサブグループとした, 比例ハザードモデルによる解析を実施することを考える. 生存時間データに対するサブグループ解析であるため, 点推定値としてハザード比およびその 95%信頼区間を示すこととする. なお, データセット VALung に含まれている文字変数を数値変数として定義した変数が Therapy2, AgeC (Age から IF ステートメントで作成), CellC である.

表 1: Forest プロット作成時のデータセットに格納されている変数の概要

変数名	変数のタイプ	内容
Subgroup	文字	サブグループ名およびサブグループ内のカテゴリ名 [プロットに修飾]
Intpval	数値	交互作用の p 値 [プロットに修飾]
Type	数値	サブグループ名とサブグループ内のカテゴリ名の判別
N	数値	サブグループにおける例数
Grp1	文字	試験治療群におけるイベント数/例数 (%) [プロットに修飾]
Grp2	文字	標準治療群におけるイベント数/例数 (%) [プロットに修飾]
HazardRatio	数値	ハザード比
HRLowerCL	数値	ハザード比の 95%信頼下限
HRUpperCL	数値	ハザード比の 95%信頼上限
HR	文字	ハザード比 (95%信頼区間) [プロットに修飾]
Indent	数値	サブグループ名とサブグループ内のカテゴリ名の出力の 違いを表すインデントを設定
Forestval	数値	プロットの縦軸
Ref	数値	プロット行のうち, 色を変える行に対応する縦軸の値

SGPLOT プロシジャで Forest プロットを作成する場合, PHREG プロシジャを用いた比例ハザードモデルによる解析などを実施し, 表 1 の変数が含まれるデータセットを準備する. なお, PHREG プロシジャによるサブグループ解析から SGPLOT プロシジャで使用するデータセット作成までのプログラムは付録 B, 表 1 のデータセットのアウトプットは付録 C を参照されたい. 表 1 のデータを用いて, SGPLOT プロシジャで Forest プロットを作成するプログラムおよび出力結果を図 5 に示す. 図 5 では, BUBBLE ステートメントでハザード比の点推定値の大きさを●印で示し, 各サブグループにおける例数によって大きさを変更している. ただし, メタアナリシスで用いる Forest プロットの場合, 点推定値の大きさは各研究の例数の大きさとは限らず, 各研究における点推定値の分散の逆数とする場合もある (丹後, 2002). また, 横軸はハザード比であるので, XAXIS ステートメントで TYPE = LOG を指定することで対数スケールとしている. ただし, 縦軸の一番下においてサブグループでない Overall の結果が示されており, 縦軸に指定している変数 Forestval の値の大きさ

を逆順にしなければならない。このプロットを PowerPoint などに貼り付けて、ハザード比とその 95%信頼区間やサブグループ名などの修飾を加えて、図のファイルとして用いることが多いのではないだろうか。V9.4 から ODS PowerPoint も利用できるようになり (吉田ら, 2015), 手動で PowerPoint に貼り付けなくても, プログラムを実行させることで PowerPoint に出力できるような環境が整ってきているといえる。

しかし, 上記のような手順でなく, SAS ですべて操作して修飾を加えたプロットを作成したい場合, プロットの場合はどのような方法で行えばよいだろうか。

```
SAS プログラム proc sgplot data=forestdata noautolegend;
    refline ref / axis=y lineattrs=(thickness=30 color=cxf0f0f0);
    scatter y=forestval x=HazardRatio
        / errorbarattrs=(thickness=1 color=blue) markerattrs=(size=0)
        xerrorupper=HRUpperCL xerrorlower=HRLowerCL;
    bubble y=forestval x=HazardRatio size=N
        / bradiusmax=4.2 bradiusmin=0.2
        fillattrs=(color=blue) lineattrs=(color=blue);
    refline 1 / axis=x
        lineattrs=(pattern=shortdash) transparency=0.5;
    axis type=log minor min=0.1 max=5
        display=(nolabel) valueattrs=(size=7);
    yaxis display=none;
run;
```

出力結果

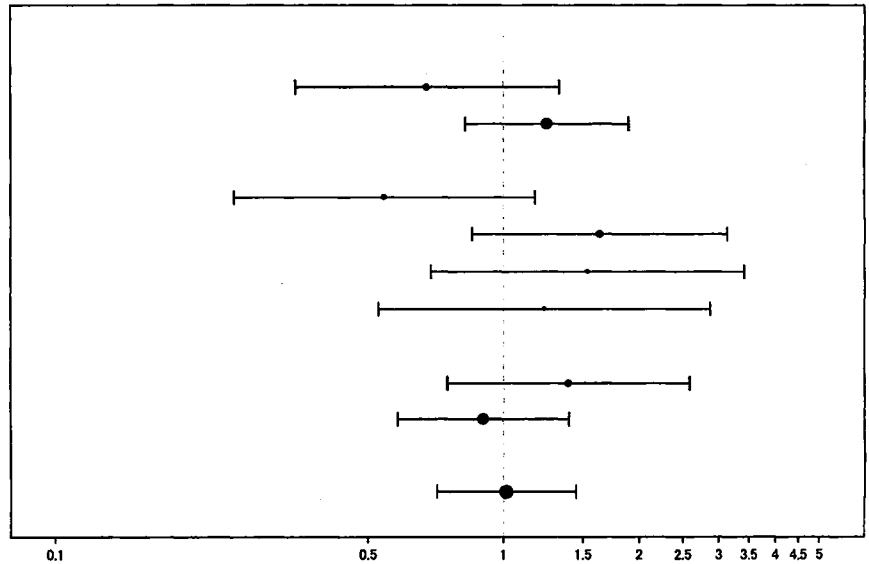


図 5: SGPLOT プロシジャによる Forest プロットの作成 1

Forest プロットに修飾を加えるためには、前節のように Y2AXIS ステートメントを用いた記述を図 5 のプログラムに加えれば作成できそうである。図 5 は V9.3 の SGPLOT プロシジャで作成可能であるが、V9.4 においてさらに多くのステートメントおよびオプションが追加されている。

図 6 は、V9.4 から利用できるようになった SGPLOT プロシジャの機能を用いて、修飾を加えた Forest プロットを作成するプログラムである。DATTRMAP=ATTRMAP として指定しているデータセットの詳細は、付録 D を参照されたい。図 5 までのプログラムに比べて長文になった上、出力結果も他のプロットよりも大きく示すことを意図して、プログラムと出力結果を別々に示している。

---

```
proc sgplot data=forestdata noautolegend
    dattrmap=attrmap nowall noborder nocycleattr;
    styleattrs axisextent=data;
    refline ref / axis=y lineattrs=(thickness=13 color=cxf0f0f0);
    scatter y=forestval x=HazardRatio
        / errorbarattrs=(thickness=1 color=blue)
        xerrorupper=HRUpperCL xerrorlower=HRLowerCL markerattrs=(size=0);
/* highlow y=forestval low=HRLowerCL high=HRUpperCL / lineattrs=(color=blue); */
    bubble y=forestval x=HazardRatio size=N
        / bradiusmax=4.2 bradiusmin=0.2
        fillattrs=(color=blue) lineattrs=(color=blue);
    scatter y=forestval x=HazardRatio / markerattrs=(size=0) x2axis;
    refline 1 / axis=x lineattrs=(pattern=shortdash) transparency=0.5;
    text x=x1 y=forestval text=text / position=center contributeoffsets=none;
    yaxistable subgroup / location=inside position=left
        textgroup=valtype labelattrs=(size=8) labeljustify=left
        titlejustify=left textgroupid=text indentweight=indent;
    yaxistable grp1 grp2 / location=inside position=left
        labelattrs=(size=8) valueattrs=(size=7)
        valuealign=center valuejustify=right;
    yaxistable HR intpval / location=inside position=right
        labelattrs=(size=8) valueattrs=(size=7) valuealign=center;
    yaxis reverse display=none offsetmin=0.1
        colorbands=odd colorbandsattrs=(transparency=1);
    xaxis type=log values=(0.2 0.5 2 5) minor display=(nolabel) valueattrs=(size=7);
    x2axis label='Hazard Ratio (95% CI)'
        display=(noline noticks novalues) labelattrs=(size=8);
run;
```

---

図 6 : SGPLOT プロシジャによる Forest プロットの作成 1 (V9.4 から実行可能)

表 2: Forest プロット作成時にデータセットに格納されている変数概要

ステートメント	役割	バージョンに関する注意
<b>STYLEATTRS</b>	<ul style="list-style-type: none"> <li>X 軸の範囲をデータに合わせて作成</li> </ul>	V9.4 以降で指定可能 (一部 TSIM3 以降でないとは指定できないオプションあり)
<b>REFLINE</b>	<ul style="list-style-type: none"> <li>色を変えるサブグループにおいて色の縦方向の幅を指定</li> <li>ハザード比 = 1 に対する参照線を出力</li> </ul>	-
<b>SCATTER</b>	<ul style="list-style-type: none"> <li>点推定値とその信頼区間のグラフを作成 (HIGHLOW ステートメントでも可能)</li> <li>グラフの上部における Hazard Ratio (95% CI) のラベルを修飾するために指定 (X2AXIS ステートメントとあわせて指定)</li> </ul>	-
<b>BUBBLE</b>	<ul style="list-style-type: none"> <li>点推定値を表す●印を出力 (サブグループにおける例数に依存して大きさを変化)</li> </ul>	-
<b>TEXT</b>	<ul style="list-style-type: none"> <li>グラフの下部 (&lt;-- Test Better や Standard Better --&gt;) の修飾</li> </ul>	V9.4 以降で指定可能
<b>YAXISTABLE</b>	<ul style="list-style-type: none"> <li>左側のサブグループ名の装飾</li> <li>左側の各群のイベント割合の装飾</li> <li>右側の点推定値とその信頼区間および交互作用の p 値の装飾</li> </ul>	V9.4 以降で指定可能 (一部 TSIM3 以降でないとは指定できないオプションあり)
<b>YAXIS</b>	<ul style="list-style-type: none"> <li>縦軸が変数 Forestval の値の大きさと逆順で出力</li> <li>左側のサブグループ名の装飾</li> </ul>	一部オプションは V9.4 以降で指定可能
<b>XAXIS</b>	<ul style="list-style-type: none"> <li>横軸を対数スケールで出力</li> </ul>	-
<b>X2AXIS</b>	<ul style="list-style-type: none"> <li>グラフの上部における Hazard Ratio (95% CI) のラベルを修飾する際の細い設定</li> </ul>	-

図 6 で指定している SGPLOT プロシジャのステートメントについて、表 2 にそれぞれの役割と V9.4 でのみ実行可能な注意点について示している。図 7 は、図 6 のプログラムの実行結果である。図 5 と異なり、縦軸において指定している値の大きさと逆順で出力されている。これは YAXIS ステートメントで REVERSE をオプション指定しているためである。また、参考までに、BUBBLE ステートメントの代わりに、HIGHLOW ステートメントを指定した上で作成したプロットを図 8 に示している。図 8 では、信頼下限・上限において縦方向のヒゲを装飾していない。Motzer *et al.* (2015) による報告で示したハザード比の信頼区間のプロットは、図 8 に近いといえる。図 7 および図 8 を見ると、学术论文に投稿する際の図として文句のないクオリティの図を作成できたといえる。前節の Kaplan-Meier プロットの作成では XAXISTABLE ステートメントを利用したが、図 7 および図 8 では YAXISTABLE ステートメントを利用している。YAXISTABLE ステート



メントのオプションとして、LOCATION = INSIDE と指定しているため、グラフの出力範囲内への修飾を行っていることになる。図6のSGPLOTプロシジャでは、YAXISステートメントにおいてDISPLAY = NONEを指定しているが、この記述を削除すれば、サブグループの項目名の左側に縦軸が出力され、グラフの出力範囲内への修飾であることを確認できる。

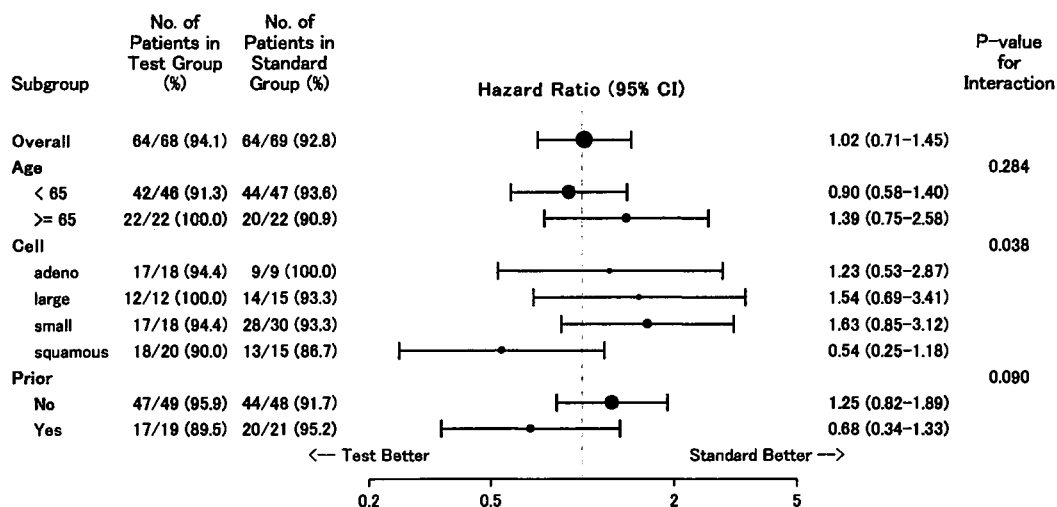


図7: SCATTER ステートメントを用いた Forest プロット

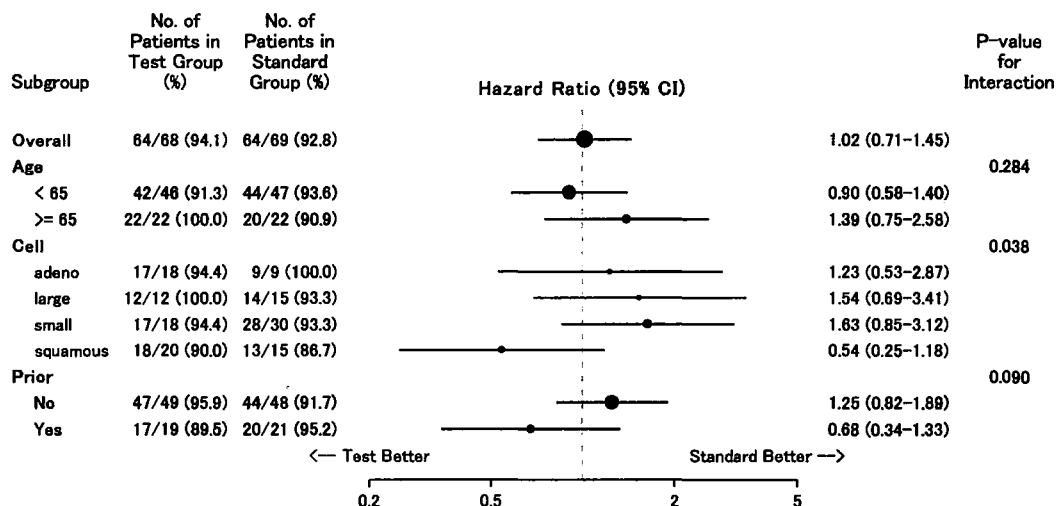


図8: HIGHLOW ステートメントを用いた Forest プロット

## 4 まとめ

本稿では、医薬品開発で用いられるよく用いられるグラフとして、Kaplan-Meier プロットと Forest プロットに焦点を当て、修飾を加えたプロットの作成方法を報告した。SAS による Kaplan-Meier プロットの作成に関してはいくつかの報告が行われているが、本稿では SAS/STAT12.1 以降、V9.4 以降で実施できる方法を示し、X2AXIS ステートメントの役割についても詳述した。なお、本稿で報告した Kaplan-Meier プロットの作成方法は、大橋ら (2016) の一部を引用している。また、Forest プロットの作成に関しては、Kaplan-Meier プロットと異なり SAS による実装報告が少ないトピックであったため、SGPLOT プロシジャを実行するにあたって、どのようなデータセットを用意しなければならないかについても詳述した。グラフの下段に数値出力を行う Kaplan-Meier プロットでは XAXISTABLE ステートメントによる実行手順を示したが、グラフの左右に数値出力を行う Forest プロットの作成では YAXISTABLE ステートメントを用いた。図 7 および図 8 のグラフは、見栄えとしてはグラフの出力範囲外への出力といえるが、YAXISTABLE ステートメントによるプログラム上は LOCATION = INSIDE を指定することで、グラフの出力範囲内への出力をしている。本稿で取りあげた Forest プロットの作成手順は、比例ハザードモデルによるハザード比の出力に限らず、ロジスティックモデルによるオッズ比の出力にも応用できる。なお、本稿で取りあげた生存時間データに対する Forest プロットの作成方法は、付録 B に SAS プログラムを詳細している。本稿の内容が国内外の多くの臨床試験の報告にお役に立てれば幸いである。

## 参考文献

- [1] Borghaei H, Paz-Ares L, Horn L, *et al.* Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *New England Journal of Medicine*. 373:1627–1639, 2015.
- [2] Hebbbar P. Lost in the Forest Plot? Follow the GTL AXISTABLE Road! *Proceedings of the SAS Global Forum*. SAS Institute Inc., Cary, NC, 2015. Available at <http://support.sas.com/resources/papers/proceedings15/SAS1748-2015.pdf>.
- [3] Matange S. Graphs are Easy with SAS® 9.4. *Proceedings of the SAS Global Forum*. SAS Institute Inc., Cary, NC, 2015. Available at <http://support.sas.com/resources/papers/proceedings15/SAS1780-2015.pdf>.
- [4] Motzer RJ, Escudier B, McDermott DF, *et al.* Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *New England Journal of Medicine*. 373:1803–1813, 2015.
- [5] SAS Institute Inc. *SAS/STAT(R) 12.1 User's Guide*. SAS Institute Inc., Cary, NC, 2012.
- [6] SAS Institute Inc. *SAS(R) 9.3 ODS Graphics: Procedures Guide* (3rd edn.). SAS Institute Inc., Cary, NC, 2012.
- [7] SAS Institute Inc. *SAS(R) 9.4 ODS Graphics: Procedures Guide* (4th edn.). SAS Institute Inc., Cary, NC, 2015.
- [8] Turner NC, Ro J, André F, *et al.* Palbociclib in Hormone-Receptor-Positive Advanced Breast Cancer. *New England Journal of Medicine*. 373:209–219, 2015.
- [9] 魚住龍史, 浜田知久馬. SG (Statistical Graphics) Procedures による Kaplan-Meier プロットの作成. SAS ユーザー総会 論文集 2011, 185–199.
- [10] 魚住龍史, 浜田知久馬. がん臨床試験における腫瘍縮小効果の検討に有用なグラフの作成 –SGPLOT プロシジャの最新機能を活用–. SAS ユーザー総会 論文集 2012, 151–165.

- [11] 大橋靖雄・浜田知久馬. 生存時間解析 -SAS による生物統計. 東京大学出版会, 1995.
- [12] 大橋靖雄・浜田知久馬・魚住龍史. 生存時間解析 応用編 -SAS による生物統計. 東京大学出版会, 2016.
- [13] 高浪洋平. SG プロシジャと GTL によるグラフの作成と ODS PDF による統合解析帳票の作成 ~TQT 試験における活用事例~. SAS ユーザー総会 論文集 2011, 201-219.
- [14] 高浪洋平. SAS と HTML アプリケーションによる CDISC ADaM 形式の解析用データセットを用いた有害事象の解析帳票・グラフ簡易作成ツールの開発事例. SAS ユーザー総会 論文集 2012, 185-205.
- [15] 丹後俊郎. メタ・アナリシス入門 -エビデンスの統合をめざす統計手法. 朝倉書店, 2002.
- [16] 豊泉樹一郎・財前政美・北西由武・都地昭夫. ODS GRAPHICS を用いた臨床試験データの可視化への挑戦. SAS ユーザー総会 論文集 2014, 307-323.
- [17] 長島健悟, 佐藤泰憲. Kaplan-Meier プロットに付加情報を追加するマクロの作成. SAS ユーザー総会 論文集 2010, 285-294.
- [18] 西本尚樹. データの可視化を加速する SAS/JMP のグラフ新機能 -ODS Graphics と JMP グラフビルダー-. SAS ユーザー総会 2013.
- [19] 平井隆幸・吉田早織・叶健・魚住龍史. ベクター形式を用いたグラフの作成と有用性. SAS ユーザー総会 論文集 2015, 303-310.
- [20] 堀田真一. Let's make Forest Plot by SAS. SAS ユーザー総会 論文集 2014, 675-679.
- [21] 吉田早織・平井隆幸・叶健・魚住龍史. ODS POWERPOINT の活用: SAS から Microsoft PowerPoint へのエクスポート. SAS ユーザー総会 論文集 2015, 295-302.

## 連絡先

E-mail : [uozumi@kuhp.kyoto-u.ac.jp](mailto:uozumi@kuhp.kyoto-u.ac.jp)

付録 A: 図 2 で省略したデータハンドリングの詳細

```

*=====*;
* リスク集合出力時間間隔 ;

%let TimeInterval=200;
*=====*;
* リスク集合の出力最大時間;

%let TimeMax=1000;
*=====*;
* フォーマット ;

proc format;
  value Cellf 1='adeno' 2='large' 3='small' 4='squamous';
run;
*=====*;
* データセットSurvivalplotからSurvivalplot2の作成 ;

data form;
do StratumNum=1 to 4; do time=0 to &TimeMax by &TimeInterval; output; end; end;
data atrisk0; merge form Survivalplot; by StratumNum time;
data Survivalplot2; set atrisk0;
  if AtRisk=. then do; AtRisk=0; tAtRisk=time; end;
run;

```

付録B: Forest プロット作成時のデータセット FORESTDATA の作成プログラム

```

*=====*;
* フォーマット ;

proc format;
  value Cellf 1='adeno' 2='large' 3='small' 4='squamous';
  value Priorf 10='Yes' 0='No'; value AgeCf 1='< 65' 2='>= 65';
  value Allf 1='Overall';
run;
*=====*;
* MACRO: ForestTTE ;
*=====*;
* サブグループ解析およびForestプロット用のデータセット作成
*=====*;
* 引数の説明 ;
* data : 解析対象データセット ;
* sub : サブグループを表すカテゴリ変数 ;
* varname : Forestプロットに示したいサブグループ名 ;
* format : サブグループのカテゴリの (事前に定義した) フォーマット名 ;
* grp : 群を表す変数 (grp = 2 を対照群とする) ;
* tte : 生存時間の評価項目を表す変数 ;
* censor : 打ち切りを表す変数 ;
* censorval : 打ち切り値 ;
* num : サブグループに付与した識別値 ;
* out : 出力データセット名 ;
*=====*;

%macro ForestTTE(data,sub,varname,format,grp,tte,censor,censorval,num,out) ;

proc sort data=&data out=data00; by &sub;
run;
*-----;
* 各群のイベント割合の計算 ;

ods listing close; ods output CensoredSummary=event00;
proc lifetest data=data00;
  time &tte*&censor(&censorval); strata &grp; by &sub;
run;
ods listing;

data event00;set event00; where &grp in(1,2);
  N=Total; Event=Failed; prop=Failed/Total;
  pevent=trim(left(put(cat(trim(left(Event)),"/",trim(left(N))," (",
    trim(left(put(round(prop*100,1e-1),8.1))),")"),20.)));
  keep &grp &sub pevent;
run;

proc transpose data=event00 out=event00; id &grp; by &sub; var pevent;
run;

data event00;set event00; rename _1=grp1 _2=grp2; drop _NAME_;
run;
*-----;
* 各サブグループにおける例数の計算 ;

ods listing close; ods output CensoredSummary=total00;
proc lifetest data=data00; time &tte*&censor(&censorval); by &sub;
run;
ods listing;

data total00;set total00; N=Total; Event=Failed; keep &sub N Event;
run;
*-----;
* ハザード比の計算 ;

ods listing close;
ods output ParameterEstimates=HR00;
proc phreg data=data00;
  class &grp(ref="2");

```

```

model &tte*& censor(& censorval) = &grp / rl ties=breslow; by &sub;
run;
ods listing;

data HR00;set HR00;
  where Parameter="&grp"; keep &sub HazardRatio HRLowerCL HRUpperCL HR;
  HR=trim(left(put(cat(trim(left(put(round(HazardRatio,1e-2),8.2))),
    " (",trim(left(put(round(HRLowerCL,1e-2),8.2))),"-",
    trim(left(put(round(HRUpperCL,1e-2),8.2))),")",20.))));
run;
-----;
* 交互作用のp値の計算 ;

ods listing close;
/* SAS/STAT14.1に対応, STAT/STAT13.2以前ではModelANOVAの代わりにType3を指定 */
ods output ModelANOVA=int00;
proc phreg data=data00;
  class &grp(ref="2") &sub;
  model &tte*& censor(& censorval) = &grp &sub &grp*&sub / rl ties=breslow;
run;
ods listing;

data int00; set int00; where Effect="&grp*&sub";
  intpval=trim(left(put(round(ProbChiSq,1e-3),pvalue8.3)));
  subgroup=&varname; subgroupc=&num; keep subgroup subgroupc intpval;
run;
-----;
* すべての計算結果を集約 ;

data surv00; merge total00 event00 HR00; by &sub; &grp=1;
data surv00;set surv00; where &sub^=.;
run;

data surv00;set surv00;
  subgroupc=&num; subgroup=put(&sub,&format); catc=&sub;
run;

data item00;set int00; valtype=1;
data cat00;set surv00; valtype=2; if &varname="Overall" then valtype=1;
data &out; length subgroup $ 50; set item00 cat00; drop &sub &grp;
run;

%mend ForestTTE;
*=====*;
* サブグループごとに解析実行 ;

%ForestTTE(VALung,all,"Overall",Allf.,Therapy2,Time,Censor,1,0,all);
%ForestTTE(VALung,AgeC,"Age",AgeCf.,Therapy2,Time,Censor,1,1,age);
%ForestTTE(VALung,CellC,"Cell",Cellf.,Therapy2,Time,Censor,1,2,cell);
%ForestTTE(VALung,Prior,"Prior",Priorf.,Therapy2,Time,Censor,1,3,prior);
*=====*;
* サブグループ解析結果の集約 ;

data forestdata00; length subgroup $ 50; set all age cell prior;
run;
*=====*;
* 全体の解析結果をマクロで実施した結果の調整 ;

data forestdata00;set forestdata00; if subgroupc=0 and N=. then delete;
run;
*=====*;
* インデントの設定 ;

data forestdata00;set forestdata00;
  forestval=_n_;
  if mod(subgroupc,2)^=0 then ref=forestval;
  indent=1; if subgroup="Overall" or intpval^="" then indent=0;

```

```

run;

ods listing close; ods output summary=forestvalmax;
proc means data=forestdata00 max; var forestval;
run;
ods listing;
*=====*;
* 横軸の説明テキストの設定 ;

data xaxis; length text $ 50; set forestvalmax;
forestval=forestval_Max+1; valtype=2;
xl=0.2; yl=forestval; text='<-- Test Better'; output;
xl=4; yl=forestval; text='Standard Better -->'; output;
run;

data forestdata;set forestdata00 xaxis;
run;
*=====*;
* 変数ラベルの指定 ;

data forestdata;set forestdata;
label grp1="No. of Patients in Test Group (%)"
grp2="No. of Patients in Standard Group (%)"
intpval="P-value for Interaction" subgroup="Subgroup" HR=" ";
run;
*=====*;
* 付録DのデータセットATTRMAP作成プログラム ;

data attrmap; length textweight $10;
id='text'; value='1'; textcolor='Black'; textsize=7; textweight='bold'; output;
id='text'; value='2'; textcolor='Black'; textsize=5; textweight='normal'; output;
run;

```

付録 C : Forest プロット作成時のデータセット FORESTDATA

Subgroup	Int pval	Type	N	Grp1	Grp2	Hazard Ratio	HR Lower CL	HR Upper CL	HR	Indent	Forest val	Ref
Overall		1	137	64/68 (94.1)	64/69 (92.8)	1.016	0.713	1.448	1.02 (0.71-1.45)	0	1	.
Age < 65	0.284	1	93	42/46 (91.3)	44/47 (93.6)	0.903	0.583	1.4	0.90 (0.58-1.40)	0	2	2
>= 65		2	44	22/22 (100.0)	20/22 (90.9)	1.393	0.751	2.582	1.39 (0.75-2.58)	1	4	4
Cell adeno	0.038	1	27	17/18 (94.4)	9/9 (100.0)	1.231	0.528	2.872	1.23 (0.53-2.87)	0	5	.
large		2	27	12/12 (100.0)	14/15 (93.3)	1.536	0.692	3.409	1.54 (0.69-3.41)	1	7	.
small		2	48	17/18 (94.4)	28/30 (93.3)	1.633	0.854	3.123	1.63 (0.85-3.12)	1	8	.
squamous		2	35	18/20 (90.0)	13/15 (86.7)	0.543	0.25	1.179	0.54 (0.25-1.18)	1	9	.
Prior No	0.09	1	97	47/49 (95.9)	44/48 (91.7)	1.248	0.822	1.894	1.25 (0.82-1.89)	0	10	10
Yes		2	40	17/19 (89.5)	20/21 (95.2)	0.676	0.344	1.33	0.68 (0.34-1.33)	1	12	12

付録 D : SGPLOT プロシジヤ (図 6) で用いるデータセット ATTRMAP

text	forestval Max	forestval	valtype	xl	yl
<-- Test Better	12	13	2	0.2	13
Standard Better -->	12	13	2	4	13

# 臨床試験におけるSGプロシジャを用いたGraphic Designの考察

○山崎 文寛<sup>1</sup>、清水 康平<sup>1</sup>、高浪 洋平<sup>1</sup>

<sup>1</sup>武田薬品工業株式会社 日本開発センター 生物統計室

Considerations on Good Graphic Design using SG Procedures in Clinical Trials

Fumihiko Yamasaki<sup>1</sup>, Kohei Shimizu<sup>1</sup>, Yohei Takanami<sup>1</sup>

<sup>1</sup>Takeda Pharmaceutical Company Limited

## 要旨

臨床試験で一般的に用いられてきた被験者の推移図や散布図等について、それらのグラフで表現しきれなかった情報を付加した新たなグラフを紹介し、SAS9.4のSGプロシジャを用いた作成方法を提案する。

キーワード：SGプロシジャ，Lasagna Plot（ラザニアプロット），Sunflower Plot

## 1 はじめに

臨床試験の統計解析業務において、グラフによるデータの可視化は非常に有用な分析ツールとなるが、一般的に用いられる被験者の推移図や散布図等から読み取れる情報には限りがある。一方、近年 SASをはじめとするソフトウェア技術の進歩により、数多くの種類のグラフが容易に、かつ高品質に作成できる環境が整いつつあり、Susan P.Dukeら(2015)は、適切にデザインされたグラフは強力なコミュニケーションツールとなる事、データに対する適切なグラフの選択が重要である事を主張するとともに、これまで表現しきれなかった情報を付加した新たなグラフの作成によるデータの分析方法を提案している。本稿では、それらのグラフを従来のグラフと比較・考察するとともに、SAS9.4のSGプロシジャによる作成方法及びカスタマイズ方法を報告する。

## 2 適切な Graphic Design について

グラフによるデータの可視化における課題の一つは、データに対して適切な Graphic Design を選択することである。CTSPedia (<https://www.ctspedia.org>)では、FDA・製薬企業・アカデミアのメンバーから構成される Safety Graphics Working Group によって推奨されている、適切な Graphic Design の選択方法(Choosing the Right Graph – What Is the Process?) やグラフを作成するサンプルのソースコード等が公開されている。本章では、その適切な Graphic Design の選択方法について概要をまとめる。なお、CTSPediaは、臨床研究者向けに、統計解析やグラフ等に関する知識・ツール・教育資料等を蓄積し、タイムリーで有用なアドバイスを提供するために設立されたものである。

## 2.1 適切な Graphic Design の原則

Safety Graphics Working Group は、図 2.1.1 に示すフローによって、適切な Graphic Design を選択することを推奨している。本フローには、解析する変数の性質（連続値、カテゴリ値等）、解析要件（経時データの解析、群間比較等）、グラフの表現方法（1 群で 1 図、複数群で 1 図等）、グラフの種類等に関する選択肢が設けられており、最終的にデータに対して適切な Graphic Design を選択することができるように構成されている。

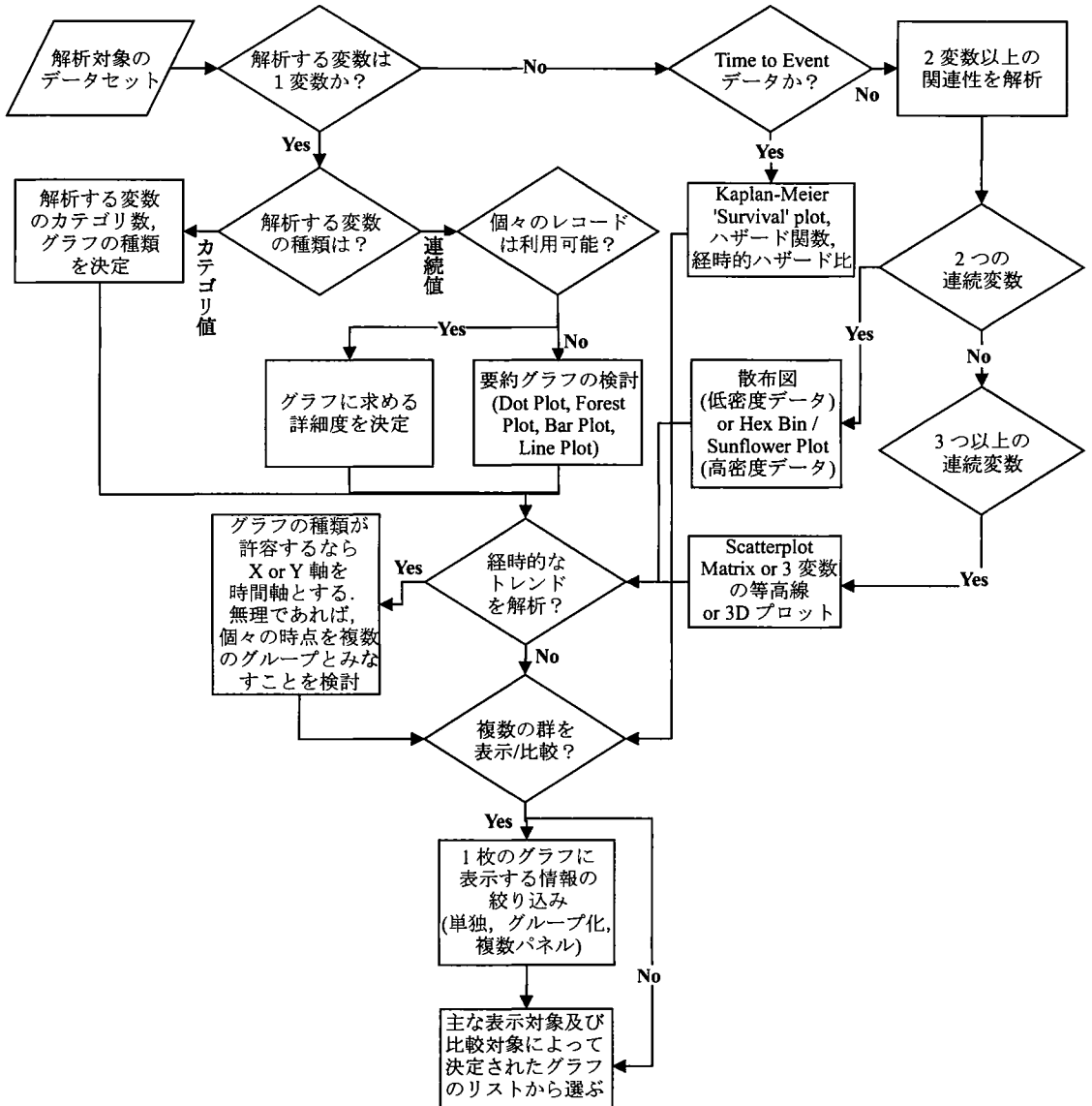


図 2.1.1 Graphic Design の選択するためのプロセス (Safety Graphics Working Group (2012) より抜粋)

一つの連続値データのグラフ作成において、個々のレコードが利用可能な状態であれば、図 2.1.2 に示



す詳細度 (Level of Detail) に従って、適切な Graphic Design を選択することができる。ここで言う詳細度とは、グラフ上でデータを示す線や記号等自身が示す情報量の多さのことであり、左に行くほど高く、右に行くほど低くなる。より多くの情報をグラフで示す場合は、詳細度の高い Graphic Design を選択する必要がある。

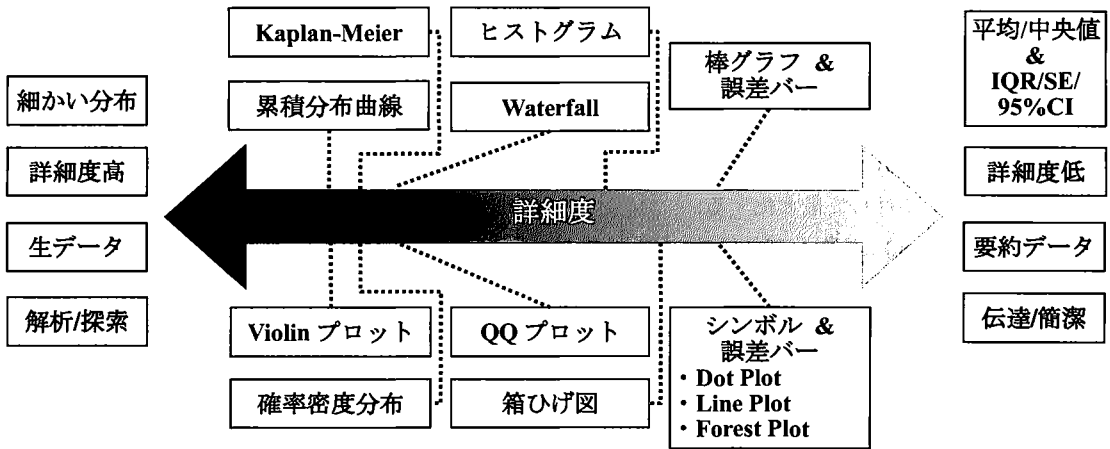


図 2.1.2 詳細度による Graphic Design の選択 (Susan P.Duke ら (2015) より抜粋)

1枚のグラフに示す情報は、群やサブグループの数によって、図2.1.3に示す方法に従って、適切な Graphic Design を選択することができる。その際、以下の2点のバランスによって選択する事が望ましい。

- ・ 比較したいグラフにおける要素をできるだけ近くに配置すること
- ・ 複数のパネルやグラフを使用することで、グラフの乱雑さを避けられること

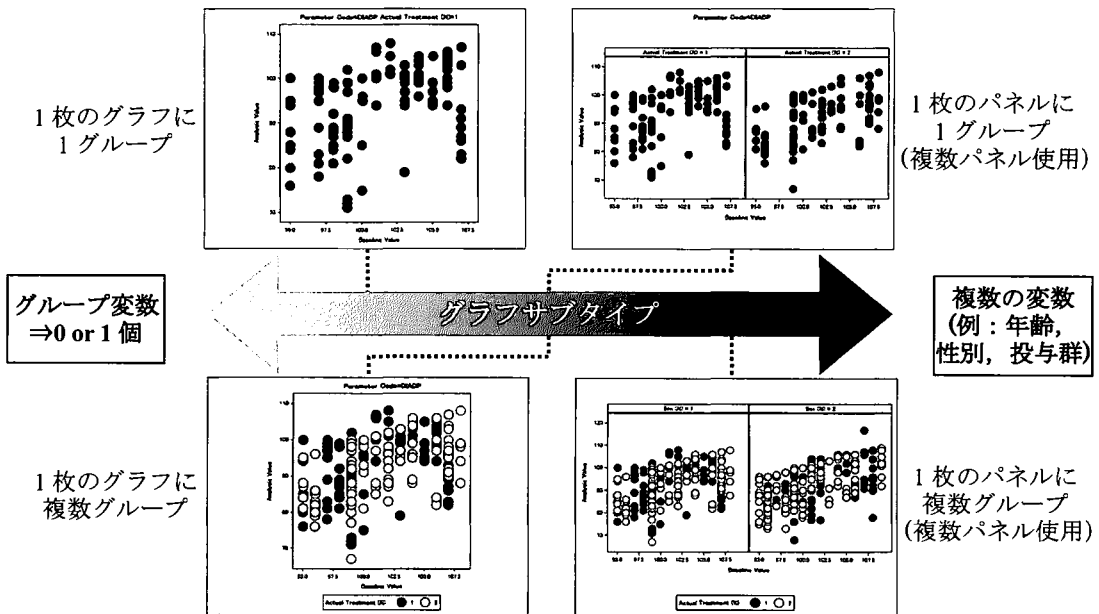


図 2.1.3 グラフサブタイプによる Graphic Design の選択 (Susan P.Duke ら (2015) より抜粋)

## 2.2 Graphic Design の推奨事項

Safety Graphics Working Group は、以下に示す「グラフ作成のための9つ最良の方法 (9 Best Practices for Making Graphs)」を CTSPedia で公開している。以下に主な内容を示す。原文については CTSPedia を参照されたい。

- ① 内容 (Content) : 説明文無しで意味する内容が表現され、明快・効果的・有益であるべき。
- ② 伝達 (Communication) : 伝達内容に対して本質的及び直感的であるべきで、情報の詰め込みすぎや、最も伝えたいことから逸らす表現は避ける。
- ③ 情報 (Information) : 不必要な情報は表示せず、必須情報の表示に充てるシンボル及び文字の割合を描画領域全体に対して最大化する。
- ④ 注釈 (Annotation) : 注釈及び凡例は、解釈を助け、最も伝えたいことから逸らさないように正しく配置し、読みやすいフォントを使用する。
- ⑤ 軸 (Axes) : 経時データに対して Visit 番号 (順序尺度) の代わりに時間 (連続尺度) を使用する等、興味のあるデータの特徴を示すのに適した軸スケールを検討する。また、ゼロを含まない場合はそれを明記し、比較するグラフ間で同じスケールとすることも重要。データの性質からグラフ形状が自動的に決まる場合はそれに従い、それ以外のグラフは横向きで、横方向の長さは縦方向よりも 50% 広くする。
- ⑥ 様式 (Styles) : 実葉は「A」、プラセボは「P」といったシンプルかつ直感的に解釈できるシンボルを選ぶ。スクリーンやポスター用には、太線や大きなシンボル・フォントを使用する。また、シンボル・線等は、すべて同一のものを指すべき。
- ⑦ 色 (Colors) : 情報を解釈する場合にだけカラーを使用する。その場合、対照的で明瞭な色を選ぶ (モノクロでは、イエローは白になり、レッド・グリーン・ブラウンは互いに区別できない)。なお、印刷等によりモノクロで見える可能性があれば、シンボルやラインスタイル等と同様に、その特徴を確保すべき。色の選択は、「COLORBREWER ([colorbrewer2.org](http://colorbrewer2.org))」が有用。
- ⑧ 技術 (Techniques) : データの関連性に従ってカテゴリを分類したり、一緒に表現すべき情報であれば一つのグラフで表現するといった、伝えたい内容を明確にするために確立された技術を使用する。
- ⑨ プロットの種類 (Types of plots) : データの分布を示すために最も適した種類のグラフを使用する。例えば、変数間の関連性を示す場合は散布図やラインプロット、第 3 の変数に関して、2 変数の関連性を示すには格子グラフを使用する。

## 3 SG プロシジャを用いた Graphics の提案

SG プロシジャは、高品質な統計グラフを作成するプロシジャである。複数のグラフを重ね合わせたり、並べて表示する場合に非常に有用である。SG プロシジャには以下の 5 種類があり、目的に応じて使い分けることができる。詳細については、高浪・舟尾 (2015) や舟尾 (2016) を参照されたい。なお、本稿では、上述した Safety Graphics Working Group が推奨する Graphic Design の中から、Lasagna Plot と Sunflower Plot を紹介するとともに、SG プロシジャによる作成方法を提案する。

表 3 SG プロシジャの種類

SG プロシジャ	概要
SGPLOT	様々な種類の 2 次元プロットを作成。複数の種類のグラフを重ね合わせて作成する

SGプロシジャ	概要
	ことが可能
SGPANEL	PANELBY ステートメントの指定に基づき、複数のグラフをパネル状に表示。SGPLOT で作成可能な、ほぼ全てのグラフをサポート
SGSCATTER	散布図の作成に特化したプロシジャで、変数の組合せごとに、複数の散布図を並べて1枚で作成
SGRENDER	Graph Template Language (GTL) でカスタマイズしたテンプレートを読み込んでグラフを作成
SGDESIGN	SGRENDER で読み込み可能な GTL を、ユーザがマウス操作でカスタマイズして作成

### 3.1 サンプルデータ

本項では、プラセボ対照の4群比較 (Placebo, ABC-123 10mg, ABC-123 20mg, ABC-123 30mg) で、1群あたり50例計200例、計5時点 (Baseline, Week2, Week4, Week6, Week8) から成る高血圧患者を対象とした仮定の臨床試験データから得られた、図 3.1.1 に示すバイタルサインの解析用データセット ADVS を用いて、Graphic Design の比較・検討を行う。本データセットは、CDISC ADaM BDS 形式に準拠しており、検査項目として、収縮期血圧、拡張期血圧等を含んでいる。また、成人における血圧値の分類 (表 3.1.2) を適用したカテゴリ変数 (AVALCAT1, BASECAT1) を含んでいる。なお、図 3.1.1 は一部のレコードを抜粋して表示している。

STUDYID	USUBID	TRTA	SAFFL	PARAM	PARAMCD	AVL	AVALCAT1	AVALCAT1N	BASE	BASECAT1	BASECAT1N	CHG	PCHG	AVIST	AVISTN	
1	ABC-123/001	ABC-123/001-0438	ABC-123 10mg	Y	Systolic Blood Pressure (mmHg)	SYSBP	154	140 to 159 mmHg	4	154	140 to 159 mmHg	4	0	0	Baseline	1
2	ABC-123/001	ABC-123/001-0438	ABC-123 10mg	Y	Systolic Blood Pressure (mmHg)	SYSBP	154	140 to 159 mmHg	4	154	140 to 159 mmHg	4	0	0	Week 2	2
3	ABC-123/001	ABC-123/001-0438	ABC-123 10mg	Y	Systolic Blood Pressure (mmHg)	SYSBP	145	140 to 159 mmHg	4	154	140 to 159 mmHg	4	-9	-5.8	Week 4	3
4	ABC-123/001	ABC-123/001-0438	ABC-123 10mg	Y	Systolic Blood Pressure (mmHg)	SYSBP	145	140 to 159 mmHg	4	154	140 to 159 mmHg	4	-9	-5.9	Week 6	4
5	ABC-123/001	ABC-123/001-0438	ABC-123 10mg	Y	Systolic Blood Pressure (mmHg)	SYSBP	132	130 to 139 mmHg	3	154	140 to 159 mmHg	4	-22	-14.3	Week 8	5
6	ABC-123/001	ABC-123/001-0438	ABC-123 10mg	Y	Diastolic Blood Pressure (mmHg)	DIABP	102	100 to 109 mmHg	5	102	100 to 109 mmHg	5	0	0	Baseline	1
7	ABC-123/001	ABC-123/001-0438	ABC-123 10mg	Y	Diastolic Blood Pressure (mmHg)	DIABP	97	90 to 99 mmHg	4	102	100 to 109 mmHg	5	-5	-4.9	Week 2	2
8	ABC-123/001	ABC-123/001-0438	ABC-123 10mg	Y	Diastolic Blood Pressure (mmHg)	DIABP	94	90 to 99 mmHg	4	102	100 to 109 mmHg	5	-8	-7.8	Week 4	3
9	ABC-123/001	ABC-123/001-0438	ABC-123 10mg	Y	Diastolic Blood Pressure (mmHg)	DIABP	98	90 to 99 mmHg	4	102	100 to 109 mmHg	5	-4	-3.9	Week 6	4
10	ABC-123/001	ABC-123/001-0438	ABC-123 10mg	Y	Diastolic Blood Pressure (mmHg)	DIABP	96	90 to 99 mmHg	4	102	100 to 109 mmHg	5	-6	-5.9	Week 8	5
11	ABC-123/001	ABC-123/001-0439	ABC-123 10mg	Y	Systolic Blood Pressure (mmHg)	SYSBP	152	140 to 159 mmHg	4	152	140 to 159 mmHg	4	0	0	Baseline	1
12	ABC-123/001	ABC-123/001-0439	ABC-123 10mg	Y	Systolic Blood Pressure (mmHg)	SYSBP	136	130 to 139 mmHg	3	152	140 to 159 mmHg	4	-16	-10.5	Week 2	2
13	ABC-123/001	ABC-123/001-0439	ABC-123 10mg	Y	Systolic Blood Pressure (mmHg)	SYSBP	133	130 to 139 mmHg	3	152	140 to 159 mmHg	4	-19	-12.5	Week 4	3
14	ABC-123/001	ABC-123/001-0439	ABC-123 10mg	Y	Systolic Blood Pressure (mmHg)	SYSBP	131	130 to 139 mmHg	3	152	140 to 159 mmHg	4	-21	-13.8	Week 6	4
15	ABC-123/001	ABC-123/001-0439	ABC-123 10mg	Y	Systolic Blood Pressure (mmHg)	SYSBP	131	130 to 139 mmHg	3	152	140 to 159 mmHg	4	-21	-13.8	Week 8	5
16	ABC-123/001	ABC-123/001-0439	ABC-123 10mg	Y	Diastolic Blood Pressure (mmHg)	DIABP	99	90 to 99 mmHg	4	99	90 to 99 mmHg	4	0	0	Baseline	1
17	ABC-123/001	ABC-123/001-0439	ABC-123 10mg	Y	Diastolic Blood Pressure (mmHg)	DIABP	89	85 to 89 mmHg	3	99	90 to 99 mmHg	4	-10	-10.1	Week 2	2
18	ABC-123/001	ABC-123/001-0439	ABC-123 10mg	Y	Diastolic Blood Pressure (mmHg)	DIABP	86	85 to 89 mmHg	3	99	90 to 99 mmHg	4	-13	-13.1	Week 4	3
19	ABC-123/001	ABC-123/001-0439	ABC-123 10mg	Y	Diastolic Blood Pressure (mmHg)	DIABP	91	90 to 99 mmHg	4	99	90 to 99 mmHg	4	-8	-8.1	Week 6	4
20	ABC-123/001	ABC-123/001-0439	ABC-123 10mg	Y	Diastolic Blood Pressure (mmHg)	DIABP	92	90 to 99 mmHg	4	99	90 to 99 mmHg	4	-7	-7.1	Week 8	5

図 3.1.1 バイタルサインのデータセット ADVS (CDISC ADaM 形式)

表 3.1.2 成人における血圧値の分類 (高血圧治療ガイドライン 2014 より抜粋)

検査項目	1. 至適血圧*	2. 正常血圧	3. 正常高値血圧	4. I 度高血圧	5. II 度高血圧	6. III 度高血圧
収縮期血圧 (SYSBP)	<120	120-129	130-139	140-159	160-179	>=180
拡張期血圧 (DIABP)	<80	80-84	85-89	90-99	100-109	>=110

\* : 収縮期血圧と拡張期血圧の and 条件

### 3.2 Spaghetti Plot と Lasagna Plot の比較検討

臨床試験で、各被験者の経時的なデータを単一グラフで示す場合は、Spaghetti Plot (スパゲッティプロ

ット)がよく使用される。しかし、被験者数が多くなるにつれて、線と線の重なりが多くなるため、外れ値の検出は可能だが、各被験者の推移を正確に読み取ることは困難となる。それを改善する手法として、Susan P. Dukeら(2015)によってLasagna Plot (ラザニアプロット)が提案されている。Spaghetti Plotがパスタ料理に似ていることからそのように呼ばれるのと同様に、Lasagna Plot もラザニア料理に似ていることからそのように呼ばれる。図 3.2.1 に拡張期血圧の Spaghetti Plot, 図 3.2.2 に拡張期血圧の Lasagna Plot を示した。

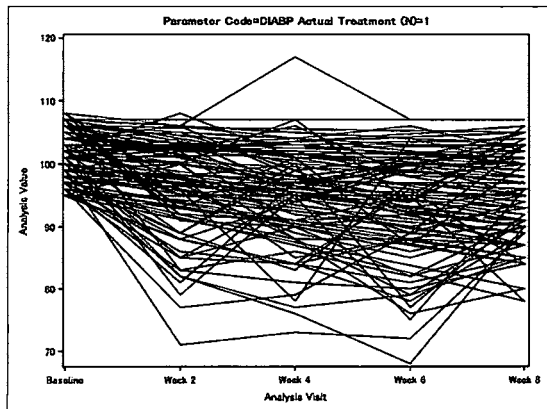


図 3.2.1 拡張期血圧の Spaghetti Plot

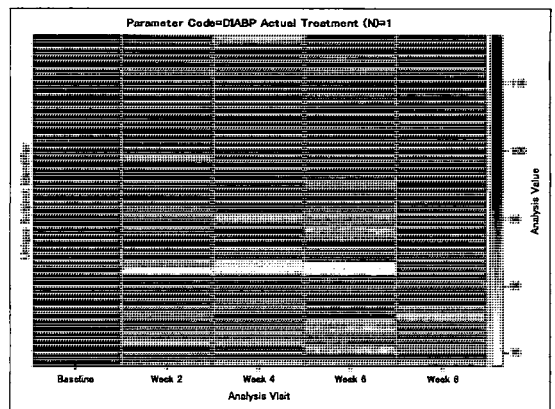


図 3.2.2 拡張期血圧の Lasagna Plot

図 3.2.1 は X 軸に時点, Y 軸に検査値を配置している。一方, 図 3.2.2 は X 軸に時点, Y 軸に被験者を配置しており, 各セルの色の濃淡で検査値の大小を示している。色は検査値が大きいほど濃く (黒に近づく), 小さいほど薄くなる (白に近づく) ように設定しており, グラフ右端に検査値と色の濃淡の対応を示す凡例を示している。被験者は Baseline の検査値の降順に配置しており, X 軸方向に見ていくと各被験者の推移が読み取り可能である。図 3.2.1 及び図 3.2.2 の作成に用いた拡張期血圧は, Week4 の最大値が 117 である。図 3.2.1 ではその最大値 117 が外れ値として読み取れるが, その前後の値は線の重なりによって読み取れない。図 3.2.2 では Week4 における最大値の具体的な値は読み取れないが, グラフ右端の凡例とグラフ中の色の濃さから, 上から 3 例目の被験者が該当し, ベースラインから高値で, Week4 で 117 を示した後, 元のレンジ (100~110) に戻ったこと等が読み取れる。なお, 本稿では, グレースケールを用いているが, SAS のデフォルトでは 3 色で値の大きい方から「赤色→白色→青色」と段階的に色が変化ようになっており, グレースケールよりも明確に値の大小を読み取れる。以下に, Spaghetti Plot と Lasagna Plot の比較結果を要約した。

表 3.2.1 Spaghetti Plot と Lasagna Plot の比較結果

	Spaghetti Plot	Lasagna Plot
被験者数	「少ない場合」向き	「多い場合」向き
データの種類	「連続値」向き	「連続値又はカテゴリ値」の両方
値の大小の判別	Y 軸の目盛及び線	色の濃淡 (変化) 例えば, サーマグラフィのように, 値の大きい方から「赤色→緑色→青色」のように変化させると分かりやすい

外れ値の 検出方法	集団から逸脱している線	色の濃淡 (変化) 例えば、値の大きい方を「赤色」、小さい方を「青色」としていた場合、他のセルと比べて際立って赤い場合や青い場合は、それが外れ値と確認できる
長所・短所	<ul style="list-style-type: none"> <li>・長所：被験者の数が少ない場合は、連続値であっても値を読み取ることが容易</li> <li>・短所：被験者の増加と共に線の重なりが増加するため、各被験者の推移が確認できない</li> </ul>	<ul style="list-style-type: none"> <li>・長所：被験者の数に関係なくプロットの重なりが発生しないため、各被験者の大よその推移が確認できる</li> <li>・短所：連続値の場合、値に応じて変化する色の細かい違いを読み取ることが難しいため、詳細な値が確認できない</li> </ul>

### 3.3 Lasagna Plot の作成方法

SGPLOT プロシジャにおいて、Lasagna Plot を作成する方法として、HEATMAPARM ステートメント又は HEATMAP ステートメントを利用する方法の 2 種類がある。両ステートメントは、3 変数を用いた 2 次元プロット (すなわち Heat map) を作成するもので、3 変数のうち 2 変数を用いて「複数の長方形から成る格子構造」をプロットし、残りの 1 変数によってその各長方形に色付け等を行うことで 3 変数の値を表現することができる。以下に、両ステートメントの基本構文及び主要なオプションを示す。

<基本構文>

```
HEATMAPARM X=<X 軸の変数> Y=<Y 軸の変数>
            COLORRESPONSE=<各セルの色付けに使用する変数>
            COLORMODEL=<各セルに適用する色の組合せ>;

HEATMAP    X=<X 軸の変数> Y=<Y 軸の変数>
            / COLORRESPONSE=<各セルの色付けに使用する変数>
            COLORMODEL=<各セルに適用する色の組合せ>;
```

<主なオプション>

<b>COLORRESPONSE</b> = <i>numeric-variable</i>	Heat map の各セルの色を、指定した <u>数値変数</u> の値に応じて変化させる
<b>COLORMODEL</b> = <i>color-ramp-style-element</i>   ( <i>color-list</i> )	<b>COLORRESPONSE</b> オプションと同時に使用する <u>色傾斜</u> (Color Ramp) 又は <u>色の組合せ</u> を指定 色傾斜は <b>ThreeColorRamp</b> , <b>ThreeColorAltRamp</b> , <b>TwoColorRamp</b> , <b>TwoColorAltRamp</b> が指定可能 具体的な色の組合せを指定する場合は、左から順に値の小さい順に適用したい色を列挙したうえで全体を括弧で囲む
<b>FILLATTRS</b> =( <i>TRANSPARENCY=number</i> )	Heat map のセル全体に対する、 <u>透過性を指定</u> (0~1 の値を指定し、0 が透過なし)
<b>OUTLINE</b>	Heat map のセル全体に対して、 <u>輪郭線を表示</u>
<b>OUTLINEATTRS</b> = <i>style-element</i>   <i>style-element (line-options)</i>   ( <i>line-options</i> )	Heat map のセル全体における <u>輪郭線の見え目</u> (色、線の種類・太さ) を指定
<b>NAME</b> ="text-string"	他のステートメントからの参照のために、本 Plot ステートメントに名前を割り当てる

<プログラムと出力例>

プログラム 3.3.1 は、色と検査値の関係の曖昧さを改善する方法として、表 3.1.2 の成人における血圧値の分類に従って作成したカテゴリ変数 AVALCAIN を用いている。実測値をそのままプロットするより

も色の段階を少なくすることで色間の区別を明確化しているが、カテゴリ化によって詳細な検査値の情報を失っているため、検査値の詳細さとグラフの見易さのバランスを考慮のうえ、選択されたい。なお、見易さの工夫として被験者のソート順は事前のソート処理により Baseline のカテゴリ値の降順としており、さらに、各被験者間の情報を明確に区別するために OUTLINE オプションを指定して各セルに輪郭線を引いている。この実行結果を図 3.3.1 に示す。プログラム 3.3.2 は、実測値を読み取れない欠点を解消する方法として示したが、Lasagna Plot 上に実測値そのものをシンボルとした Scatter Plot を重ね合わせるために、SCATTER ステートメントを同時に使用している。なお、グラフの解釈を助けるために、特に興味のある値（成人における血圧値分類がⅡ度高血圧以上に該当する値）のみを表示させた。これは出力したい値だけを格納した変数を事前に作成することで実現した。この実行結果を図 3.3.2 に示す。

```
proc sgplot data=<入力データセット>;
  format AVALCA1N <フォーマット名>;
  by PARAMCD TRTAN;
  heatmapparm x=AVISIT y=USUBJID colorresponse=AVALCA1N
    / name='a' colormodel=(WHITE BLACK) outline outlineattrs=(color=white);
  yaxis display=(NOVALUES NOTICKS);
  gradlegend 'a' / integer title = 'Diastolic Blood Pressure (mmHg)';
run;
```

プログラム 3.3.1 カテゴリ値を用いたグラフを作成

```
proc sgplot data=<入力データセット>;
  format AVALCA1N <フォーマット名>;
  by PARAMCD TRTAN;
  heatmapparm x=AVISIT y=USUBJID colorresponse=AVAL
    / name='a' colormodel=(WHITE BLACK) outline outlineattrs=(color=white);
  scatter x=AVISIT y=USUBJID / markerchar=AVAL2 markercharattrs=(color=white size=7);
  yaxis display=(NOVALUES NOTICKS);
  gradlegend 'a' / integer title = 'Diastolic Blood Pressure (mmHg)';
run;
```

プログラム 3.3.2 実測値を用いたグラフを作成

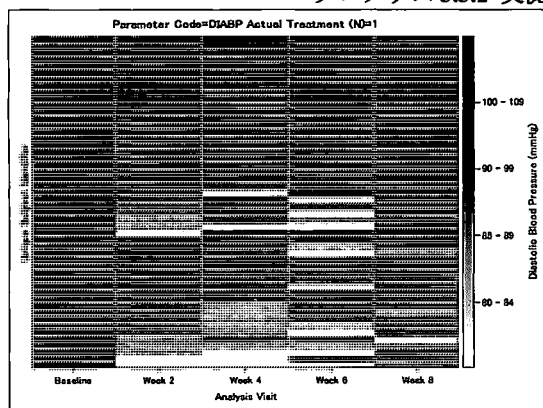


図 3.3.1 カテゴリ値を表示したグラフ

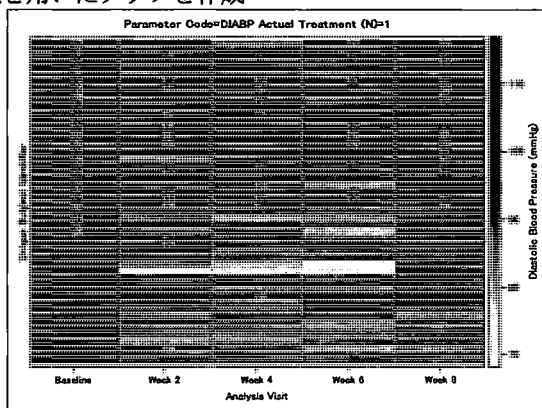
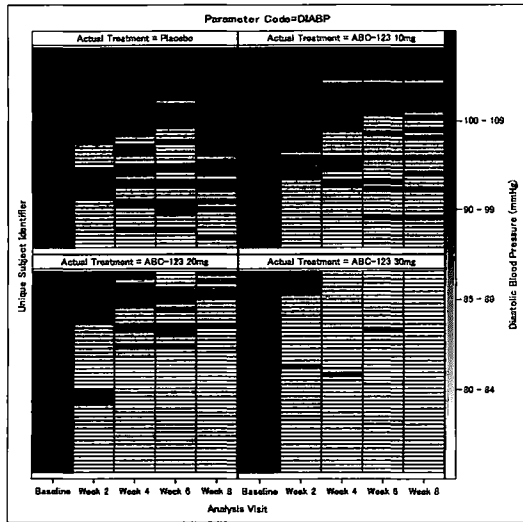


図 3.3.2 実測値を表示したグラフ

図 3.3.1 及び図 3.3.2 は、ある群における拡張期血圧をプロットしたものであったが、複数の群を比較したい場合には、プログラム 3.3.3 に示す SGPPANEL プロシジャを用いることで可能となる。パネルごとに分割したいグループ変数を PANELBY に、HEATMAPPARM ステートメントの"Y="には、各群における被験者の連番を格納した変数を指定する。これは、Y 軸を左右のグラフで共有しているためである。

この実行結果を図 3.3.3 に示す。詳細なプログラムは付録を参照されたい。



```
proc sgpanel data=<入力データセット>;
  by PARAMCD;
  format AVALCA1N <フォーマット名>;
  panelby TRTA;
  heatmapparm x=AVISIT
              y=<各群の被験者の連番>
              colorresponse=AVALCA1N / name='a'
              colormodel=(WHITE BLACK)
              outline ;
  rowaxis display=(NOVALUES NOTICKS);
  gradlegend 'a' / integer title
             = 'Diastolic Blood Pressure (mmHg)';
run;
```

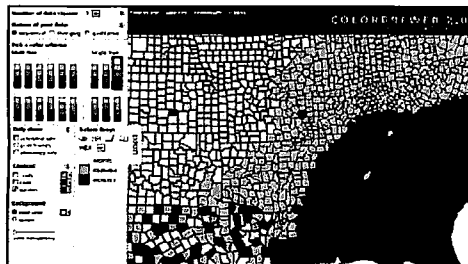
プログラム 3.3.3  
各群をパネル状に配置したグラフを作成

図 3.3.3 各群をパネル状に配置したグラフ

### 3.4 色の選択方法

Lasagna Plot は解釈しやすい色の組合せを選ぶことが重要である。本稿ではモノクロ印刷を考慮して白と黒の2色を選択したが、カラーが利用できる環境及び状況であれば赤や青といったカラーを積極的に利用したい。COLORBREWER ([colorbrewer2.org](http://colorbrewer2.org)) は適切な色の組合せの選択を容易にする。実際のサイトの画面を図 3.4 に示したが、画面上のオプションにより、色見本とともにカラーコードが表示される。画面左上から順に以下に記載したオプションを選択していくことで具体的に色が選択できる。

- ① 色の数：3～12色が選択可能で、データのカテゴリ数等に応じて選択
- ② 色の变化パターン：Sequential, Diverging, Qualitative の3パターンが選択可能で、それぞれ、薄い色から濃い色の变化、ある色から別の色への変化、全く別の色の組合せ
- ③ カラースキーム：Multi-hue (多色) 又は Single-hue (単色) が選択可能
- ④ 適切なカラースキームへの絞り込み：Colorblind Safe, Print Friendly, Photocopy Safe の3種類が選択可能で、それぞれ、色覚異常、デスクトップカラー、白黒印刷への対応
- ⑤ カラーコード：RGB (Red, Green, Blue (10進数)), HEX (RGBの16進数版), CMYK (Cyan, Magenta, Yellow, Key) の3種類が選択可能。このうち、HEXとCMYKがSASで指定可能で、HEXは主にディスプレイやデジタルカメラ、CMYKはプリンタや塗料で使用されるカラーコードとなっている。



選択した色は「COLORMODEL=(<カラーのリストを列挙>)」の形式で指定可能

- ・HEX：カラーコードの"#"を"CX"に置き換えて、「COLORMODEL=(CXf0f0f0 CXbdbdbd CX636363)」のように指定
- ・CMYK：カンマ区切りで表示されたカラーコードを各2桁に変換し、すべて繋げて8桁の数字にしたうえで、「COLORMODEL=(“00000006” “00000026” “00000061”)」のように指定

図 3.4 COLORBREWER

### 3.5 Scatter Plot と Sunflower Plot の比較検討

臨床試験で、各被験者の投与前後データや2変数の関連性を単一グラフで示す場合は、Scatter Plot（散布図）がよく使用される。しかし、被験者数が多くなるにつれて、シンボルとシンボルの重なりが多くなるため、被験者の密集具合を読み取ることは困難となる。それを改善する手法として、D.B. CARRら(1987)によってSunflower Plotが提案されている。花びらの数で被験者数を示し、それがひまわりの花のように見えることからそのように呼ばれる。図 3.5.1 に拡張期血圧の Scatter Plot、図 3.5.2 に拡張期血圧の Sunflower Plot を示した。

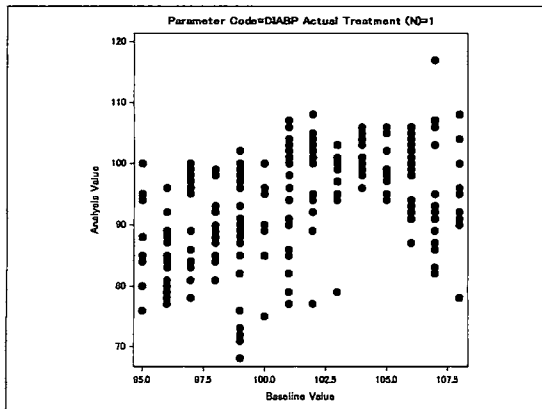


図 3.5.1 拡張期血圧の Scatter Plot

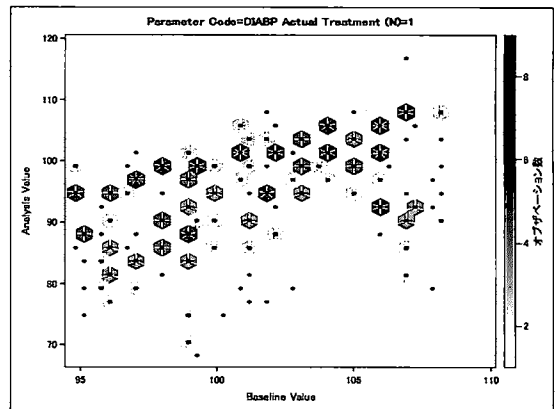


図 3.5.2 拡張期血圧の Sunflower Plot

いずれも X 軸に投与前値、Y 軸に Week 8 の値を配置している。図 3.5.2 はグラフ描画領域全体に六角形を敷き詰めたとうえで、その六角形の領域に含まれる値を取る被験者の数を、六角形に塗られた色の濃淡及び六角形の中心から等しい角度に出ている線で示している。例えば線が 3 本であれば被験者数は 3 である。六角形のピンは正方形のピンに比べて視覚的なバイアスが少なくとされているため、本稿では六角形のピンを用いている。なお、プロットの色にはグレースケールを用いているが、オレンジ色等を使用することで、グレースケールよりも明確に値の大小を読み取れるとともに、Sunflower のように見えるようになる。以下に、Scatter Plot と Sunflower Plot の比較結果を要約した。

表 3.5.1 Scatter Plot と Sunflower Plot の比較結果

	Scatter Plot	Sunflower Plot
被験者数	「少ない場合」向き	「少ない場合及び多い場合」の両方
データの種類	「連続値」向き	「連続値及びカテゴリ値」の両方
データの密度	判別不可	色の濃淡及び花びらの数
外れ値の検出	可能	可能
長所・短所	<ul style="list-style-type: none"> <li>長所：実測値をそのままプロットしているため、プロットの重なりが少ない場合は、比較的正確な値を読み取れる</li> <li>短所：連続値の場合、被験者の増加と共にプロットの重なりが発生するため、個々の値を読み取れない(カテゴリ値の場合は、被験者数に関係なくプロットの重なりが増大する場合がある)</li> </ul>	<ul style="list-style-type: none"> <li>長所：被験者の数に関係なく、プロットの重なりを色の濃淡と花びらの数によって表現できる</li> <li>短所：ピンの数によって視覚的なバイアスが発生する可能性がある(ピンの数を増やすと密度が低く見え、減らすと密度が高く見える)</li> </ul>



### 3.6 Sunflower Plot の作成方法

Sunflower Plot の作成には、SURVEYREG プロシジャ及び SGPLOT プロシジャの POLYGON ステートメントを利用する。SURVEYREG プロシジャは、線形モデルの当てはめ、回帰係数及びその共分散行列の算出を行ない、標本調査に基づくデータに対して回帰分析を行なうことができるプロシジャであり、PLOTS オプションを用いることで、各ビンの頂点の座標とその各ビンに含まれる被験者の数が算出できる。POLYGON ステートメントは、SGPLOT プロシジャ等で複数の多角形を描くことのできるステートメントであり、SURVEYREG プロシジャの実行結果を入力データとして多角形を作成できる。それらのプロシジャ及びステートメントの基本構文および主要なオプションを示す。

<基本構文>

```
PROC SURVEYREG DATA=<入力データセット>
    PLOT (NBINS=<ビンの数> WEIGHT=<グラフの種類>)=FIT (SHAPE=<シンボルの形状>) ;
    MODEL <Y 軸の変数> = <X 軸の変数> ;
RUN ;
```

```
POLYGON X=<X 軸の変数> Y=<Y 軸の変数> ID=<ID の変数> ;
```

<PROC SURVEYREG ステートメントの主なオプション>

PLOTS<(global-plot-options)> <=<plot-request <(plot-option)>>	2 変数による六角形又は長方形の Heatmap 上に、回帰直線を描く「PLOTS=all」, PLOTS (WEIGHT=HEATMAP)=FIT
<PLOT オプション> NBINS= <i>nbin1</i> < <i>nbin2</i> >	ビンの数を指定「10: 100 のビン, 10 20: 10×20 = 200 のビンが作成される」
<PLOT オプション> WEIGHT=BUBBLE   HEATMAP	レコード数に対して、領域が比例する Bubble Plot か、色の濃淡で示す Heat map のどちらかを指定
<FIT オプション> SHAPE=RECTANGULAR   HEXAGONAL	ヒートマップ領域を長方形と六角形のどちらにするかを指定 ※省略形の「REC」「HEX」の指定が可能

<POLYGON ステートメントの主なオプション>

COLORRESPONSE = <i>numeric-variable</i>	ビンの色を指定した数値変数の値に応じて変化させる
COLORMODEL = <i>color-ramp-style-element</i>   ( <i>color-list</i> )	COLORRESPONSE オプションと同時に使用する色の組合せを指定。色リストの直接指定も可能 ・3 色: ThreeColorRamp, ThreeColorAltRamp 2 色: TwoColorRamp, TwoColorAltRamp
FILLATTRS =(TRANSPARENCY= <i>number</i> )	ビンに対する、透過性を指定 (0~1 の値を指定し, 0 が透過なし)
OUTLINE	ビンに対して、輪郭線を表示
LINEATTRS = <i>style-element</i> < <i>options</i> >   ( <i>options</i> )	ビンの輪郭線の見たい目 (色, 線の種類・太さ) を指定
LABEL = <i>variable</i>	ビン上に表示するラベルが格納された変数を指定
NAME =" <i>text-string</i> "	他のステートメントからの参照のために、本 Plot ステートメントに名前を割り当てる

<プログラムと出力例>

プログラム 3.6.1 は、Sunflower Plot における各六角形のビンとその色の濃淡を表現するための、各ビンの頂点の座標と被験者の数を算出する。図 3.6.1(a)はその実行結果であり、hID は各ビンに付与された連番、XVar と YVar はビンの各頂点の座標、WVar は当該ビンの領域に含まれる値を持つ被験者数を示す。

このデータを仮に HexBin データと呼ぶ。次に各ビン上で被験者数を示す中心から出る直線をプロットするための Annotate データを作成する。本稿では図 3.6.1(b)に示すように六角形の内接円の円弧に向けて、中心から直線を引くこととした。X 軸と Y 軸のスケールが異なる場合は、正六角形及び真円ではないため、いずれかの方向に長い六角形及び楕円となることを考慮し、角度を等分するのではなく円弧を等分した位置に向けて線を引くこととした。また、六角形の中心の座標を  $(x_1, y_1)$ 、X 軸方向の最長半径を  $a$ 、Y 軸方向の最長半径を  $b$ 、X 軸に対して角度  $\theta$  の方向に直線を引いた場合、楕円の交点の座標は  $(x_1 + a \cos \theta, y_1 + b \sin \theta)$  となることを利用して、楕円の円周を被験者数に応じて等分した場合の座標を算出している。図 3.6.1(c)は、花序を示す各六角形の中心座標とそこから楕円の円周とに向けて引いた直線の始点・終点の座標を示す Annotate データである。詳細なプログラムは付録を参照されたい。

```
proc surveyreg data=<入力データセット> plots(nbins=30 weight=heatmap)=fit(shape=hex);
  model AVAL = BASE;
run;
```

プログラム 3.6.1 各六角形の頂点の座標と被験者の数を算出

	PARAMCD	TRTAN	hID	XVar	YVar	WVar
1	DIABP	1	1	107.87	80.73	1
2	DIABP	1	1	108.18	79.99	1
3	DIABP	1	1	108.18	78.52	1
4	DIABP	1	1	107.87	77.78	1
5	DIABP	1	1	107.55	78.52	1
6	DIABP	1	1	107.55	79.99	1
7	DIABP	1	2	106.91	82.93	2

図 3.6.1(a) HexBin データ

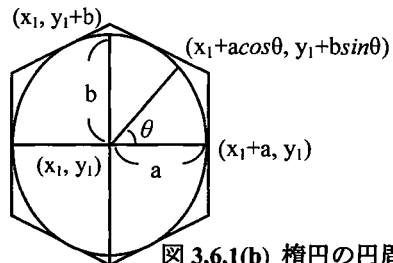


図 3.6.1(b) 楕円の円周上の座標

X1	Y1	X2	Y2	linecolor	function	drawspace	linethickness	fillcolor	width	height	display
95.830756203	97.056446773	95.829559832	98.409020754	BLACK	LINE	datavalue	0.7				
95.830756203	97.056446773	95.830158017	98.409026934	BLACK	LINE	datavalue	0.7				
107.82650803	57.229847583				OVAL	datavalue		BLACK	1	1	FILL
107.82650803	61.915329841				OVAL	datavalue		BLACK	1	1	FILL

図 3.6.1(c) Annotate データ

プログラム 3.6.2.1 は Sunflower Plot を作成するが、入力データセットには図 3.6.1(a)の HexBin データ、Annotate データセットには図 3.6.1(c)の Annotate データを指定する。この実行結果を図 3.6.2.1 に示す。また、プログラム 3.6.2.2 は花びらではなく被験者数を表示する。この実行結果を図 3.6.2.2 に示す。

```
proc sgplot data=<入力データセット> sganno=<Annotateデータセット>;
  polygon x=XVar y=YVar ID=hID / colorresponse=WVar fill
  colormodel=(cxffffff cxfeb24c cxf03b20);
run;
```

プログラム 3.6.2.1 被験者数を花びらで表現したグラフを作成

```
proc sgplot data=<入力データセット> sganno=<Annotateデータセット>;
  polygon x=XVar y=YVar ID=hID / colorresponse=WVar fill label=<被験者数を格納した変数>
  colormodel=(cxffffff cxfeb24c cxf03b20)
  labelattrs=(size=11 color=white weight=bold);
run;
```

プログラム 3.6.2.2 被験者数を数字で表現したグラフを作成

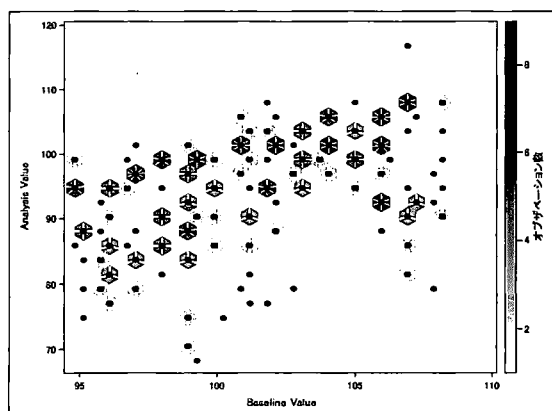


図 3.6.2.1 被験者数を花びらで表現したグラフ

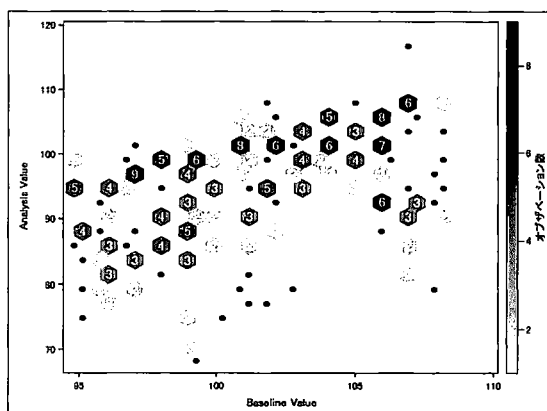


図 3.6.2.2 被験者数を数字で表現したグラフ

#### 4 まとめ

本稿では、Susan P.Duke ら(2015)によって提案された適切なグラフの選択方法を紹介した。そして、臨床試験でよく使われている Spaghetti Plot や Scatter Plot といったグラフを Lasagna Plot や Sunflower Plot といった付加情報を加えたグラフと比較・考察するとともに、SAS9.4 の SG プロシージャによる作成方法及びカスタマイズ方法を報告した。その中で、Lasagna Plot や Sunflower Plot はいずれもプロットの重なりがないため、個々の被験者の情報を比較的失うことなく集団の推移や分布を表現することが可能となることを示した。ただし、冒頭に述べたように、SAS をはじめとするソフトウェアの機能が年々向上している一方で、医薬品製造販売承認申請資料の作成においては、eCTD の規制もあり、依然としてモノクロ印刷を意識した記載をせざるを得ず、現状、これら最先端の技術をうまく活用できていない。本稿で紹介したグラフを総括報告書や eCTD 等で利用する場合は、さらなる議論が必要である。

連絡先

[fumihiro.yamasaki@takeda.com](mailto:fumihiro.yamasaki@takeda.com)

#### 参考文献

- Bruce J. Swihart et.al. (2010). Lasagna Plots A Saucy Alternative to Spaghetti Plots; *Epidemiology*. 2010 Sep; 21(5): 621–625
- Clinical Data Interchange Standards Consortium (2016). Analysis Data Model Implementation Guide Version 1.1
- Cynthia Brewer et.al. (2009). COLORBREWER2.0; Available at <http://colorbrewer2.org/> [Accessed 11 June 2016]
- D.B. CARR et.al. (1987). Scatterplot Matrix Techniques for Large N; *Journal of the American Statistical Association*; Vol. 82, No. 398 (Jun., 1987), 424-436
- David W. Scott (1988). A Note on Choice of Bivariate Histogram Bin Shape; *Journal of Official Statistics* Vol.4, No.1, 1988, p47-51

- ・ F. Bancken et.al. (2013). General Principles, Illustrations and Wiki Resources for Improving Statistical Graphs; International Meeting STATISTICAL METHODS IN BIOPHARMACY SFDS
- ・ Rick Wicklin (2014). How to create a hexagonal bin plot in SAS; Available at <http://blogs.sas.com/content/iml/2014/09/02/hexagonal-bin-plot.html> [Accessed 11 June 2016]
- ・ Safety Graphics Working Group (2012). Choosing the Right Graph – What is the Process?; Available at <https://www.ctspedia.org/do/view/CTSpedia/SelectRightGraph> [Accessed 11 June 2016]
- ・ SAS Institute Inc. SAS 9.4 ODS Graphics: Procedures Guide, Fifth Edition; Available at <https://support.sas.com/documentation/cdl/en/grstatproc/67909/HTML/default/viewer.htm> [Accessed 11 June 2016]
- ・ Susan P.Duke et.al. (2012). 9 Best Practices for Making Graphs; Available at <https://www.ctspedia.org/do/view/CTSpedia/BestPractices> [Accessed 11 June 2016]
- ・ Susan P.Duke et.al. (2015). Seeing is believing: good graphic design principles for medical research. *Statistics in Medicine* 2015; 34(22): 3040-59.
- ・ The Clinical Trials Safety Graphics Home Page; Available at <https://www.ctspedia.org/do/view/CTSpedia/StatGraphHome> [Accessed 11 June 2016]
- ・ William D. Dupont et.al. (2003). Density Distribution Sunflower Plots. *Journal of Statistical Software* 2003; 8(3)
- ・ 高浪洋平・舟尾暢男 (2015) 「統計解析ソフト『SAS』」(カットシステム)
- ・ 野中文雄 (2010) 「ActionScript 3.0 による三次元表現ガイドブック」(毎日コミュニケーションズ)
- ・ 舟尾暢男 (2016) 「もしも、SAS の sgplot と R の ggplot2 を比較したら…」(第1回ナニワデータサイエンス研究会)
- ・ 日本高血圧学会高血圧治療ガイドライン作成委員会 (2014) 「高血圧治療ガイドライン 2014」(ライフサイエンス出版)

## 付録

```

/* - Lasagna Plot - CDISC ADaM BDS形式のADVSデータセットを使用 */
proc sort data=OUTDATA.ADVS out=_ADVS_SORT;
  by PARAMCD TRTAN USUBJID AVISITN;
run;
proc transpose data=_ADVS_SORT out=_ADVS_TRANS prefix=V;
  by PARAMCD TRTAN USUBJID;
  var AVALCA1N;
  id AVISITN;
run;
data _ADVS_SORT_TRANS;
  merge _ADVS_SORT
        _ADVS_TRANS(rename=(V1=V1C V2=V2C V3=V3C V4=V4C V5=V5C));
  by PARAMCD TRTAN USUBJID;
run;
proc sort data=_ADVS_SORT_TRANS out=_ADVS_SORT_BY_VISIT_VALUE;
  by PARAMCD TRTAN V1C V2C V3C V4C V5C; *各時点のカテゴリ値昇順にソート;
run;
data _ADVS_SORT_BY_VISIT_VALUE;
  set _ADVS_SORT_BY_VISIT_VALUE;
  by PARAMCD TRTAN V1C V2C V3C V4C V5C;
  label _USUBJID = 'Unique Subject Identifier'; *パネル表示用の被験者連番;
  _USUBJID+1;
  if first.TRTAN then _USUBJID=1;
run;
proc sort data=_ADVS_SORT_BY_VISIT_VALUE out=ADVS;

```

```

by PARAMCD TRTAN _USUBJID USUBJID;
run;
proc format; *グラフ右端の凡例に表示するカテゴリ値のフォーマット;
value AVALCA1F
1='<80' 2='80 - 84' 3='85 - 89' 4='90 - 99' 5='100 - 109' 6='>=110';
run;
proc sgplot data=ADVS; *カテゴリ値を用いたLasagna Plot;
by PARAMCD TRTAN;
format AVALCA1N AVALCA1F.;
heatmapparm x=AVISIT y=USUBJID colorresponse=AVALCA1N / name='a' outline;
yaxis display=(NOVALUES NOTICKS);
gradlegend 'a' / integer title = 'Categories';
run;
proc sgpanel data=ADVS; *カテゴリ値を用いたLasagna Plot (パネル表示);
by PARAMCD;
format AVALCA1N AVALCA1F.;
panelby TRTAN;
heatmapparm x=AVISIT y=_USUBJID colorresponse=AVALCA1N / name='a' outline;
rowaxis display=(NOVALUES NOTICKS);
gradlegend 'a' / integer title = 'Categories';
run;

/* - Sunflower Plot - CDISC ADaM BDS 形式の ADVS データセットを使用 */
ods output fitplot=_HexMap(where=(not missing(hID))); *各六角形の頂点の座標;
proc surveyreg data=ADVS plots(nbins=30 weight=heatmap)=fit(shape=hex);
by PARAMCD TRTAN;
model AVAL = BASE;
run;
ods output summary=_HexCenter(keep=PARAMCD TRTAN hID Wvar Xvar_Mean Yvar_Mean);
proc means data=_HexMap mean; *各六角形の中心の座標;
by PARAMCD TRTAN hID Wvar;
var Xvar Yvar;
run;
data _Deformation_010;
set _HexMap;
by PARAMCD TRTAN hID;
_SEQ + 1;
if first.TRTAN then _SEQ = 1;
if _SEQ <= 6;
run;
ods output summary=_Deformation_020; *1個の六角形から歪みを計算;
proc means data=_Deformation_010 min max;
by PARAMCD TRTAN;
var Xvar Yvar;
run;
data _Deformation;
set _Deformation_020;
by PARAMCD TRTAN;
Diff_Line_X = Xvar_Max - Xvar_Min; *X軸の最大と最小の差;
Diff_Line_Y = Yvar_Max - Yvar_Min; *Y軸の最大と最小の差;
Fraction = 2 / sqrt(3); *本来のX軸に対するY軸の比;
Fraction_Act = Diff_Line_Y/Diff_Line_X; *実際のX軸に対するY軸の比;
Factor = sqrt(3) / 2 * Diff_Line_Y / Diff_Line_X; *正六角形からの変換係数;
if Fraction_Act = Fraction then do; *六角形に歪み無し;
Radius_X = Diff_Line_X / 2; *Y=0における六角形の内接円(楕円)の半径;
Radius_Y = Diff_Line_X / 2; *X=0における六角形の内接円(楕円)の半径;
end;
else if FRACTION_ACT > FRACTION then do; *六角形がY軸方向に伸長;
Radius_X = Diff_Line_X / 2;
Radius_Y = Diff_Line_X / 2 * FACTOR;
end;
else if FRACTION_ACT < FRACTION then do; *六角形がX軸方向に伸長;
Radius_X = DIFF_LINE_Y / 2 * sqrt(3) / 2 * FACTOR;
Radius_Y = DIFF_LINE_Y / 2 * sqrt(3) / 2;
end;
keep PARAMCD TRTAN Radius_X Radius_Y;
run;
data _PETAL_ANNO_010;
merge _HexCenter(rename=(Xvar_Mean=X1 Yvar_Mean=Y1))
_Deformation;

```

```

by PARAMCD TRTAN;
do Angle = 0 to 360 by 0.1; *六角形の内接円(楕円)の座標を0.1度刻みで導出;
  Angle = round(Angle, 0.1);
  X2 = Radius_X * sin(Angle * constant('pi') / 180) + X1; *X座標;
  Y2 = Radius_Y * cos(Angle * constant('pi') / 180) + Y1; *Y座標;
output;
end;
run;
proc sort data=_PETAL_ANNO_010 out=_PETAL_ANNO_020;
  by PARAMCD TRTAN hID WVar X1 Y1 Angle;
run;
data _PETAL_ANNO_021; *隣接する2点間の座標を同一レコードに持たせるための準備;
  set _PETAL_ANNO_020(rename=(Angle=Angle));
  by PARAMCD TRTAN hID WVar X1 Y1 Angle;
  X3 = X2; Y3 = Y2;
  Angle = round(Angle - 0.1, 0.1);
  keep PARAMCD TRTAN hID WVar X1 Y1 X3 Y3 Angle;
run;
data _PETAL_ANNO_030; *内接円(楕円)上の隣接する2点間の直線距離を導出;
  merge _PETAL_ANNO_020(in=d)
        _PETAL_ANNO_021;
  by PARAMCD TRTAN hID WVar X1 Y1 Angle;
  if d and X2^=. and Y2^=. and X3^=. and Y3^=. ;
  XY2_XY3 = sqrt((X3 - X2)**2 + (Y3 - Y2)**2);
run;
data _PETAL_ANNO_040; *内接円(楕円)上の各2点間の直線距離をレコードごとに累積;
  set _PETAL_ANNO_030;
  by PARAMCD TRTAN hID WVar X1 Y1 Angle;
  retain XY2_XY3_SUM;
  if first.hID then XY2_XY3_Sum = XY2_XY3;
  else XY2_XY3_Sum = XY2_XY3_Sum + XY2_XY3;
run;
data _PETAL_ANNO_041; *隣接2点間の距離と次の2点間の距離を同一レコードに持たせるための準備;
  set _PETAL_ANNO_040;
  by PARAMCD TRTAN hID WVar X1 Y1 Angle;
  Angle = round(Angle - 0.1, 0.1);
  XY2_XY3_Sum_Next = XY2_XY3_Sum;
  keep PARAMCD TRTAN hID WVar X1 Y1 Angle XY2_XY3_Sum_Next;
run;
data _PETAL_ANNO_050; *隣接する2点間の距離と次の2点間の距離を同一レコードに持たせる;
  merge _PETAL_ANNO_040(in=d)
        _PETAL_ANNO_041;
  by PARAMCD TRTAN hID WVar X1 Y1 Angle;
  if d;
  if missing(XY2_XY3_Sum_Next) then XY2_XY3_Sum_Next = XY2_XY3_Sum + 0.1;
run;
ods output summary=_PETAL_ANNO_060; *内接円(楕円)の円周(近似したもの)を導出;
proc means data=_PETAL_ANNO_050 sum;
  by PARAMCD TRTAN hID WVar X1 Y1;
  var XY2_XY3;
run;
data _PETAL_ANNO_070;
  set _PETAL_ANNO_060;
  by PARAMCD TRTAN hID WVar X1 Y1;
  Length_Per_WVar = XY2_XY3_Sum / WVar; *被験者数(=花びらの数)で内接円(楕円)の円周を割る;
  Length_Per_WVar_Sum = 0;
  do i = 1 to WVar by 1; *六角形に含まれる被験者数だけ繰り返す;
    Length_Per_WVar_Sum = Length_Per_WVar_Sum + Length_Per_WVar;
  output;
  end;
  keep PARAMCD TRTAN hID WVar X1 Y1 Length_Per_WVar_SUM;
run;
proc sql; *内接円(楕円)の円周を被験者数で等分した位置に最も近いXY座標を選択;
  create table _PETAL_ANNO_080 as
  select A.PARAMCD, A.TRTAN, A.hID, A.WVar, A.X1, A.Y1, A.X2, A.Y2, A.X3, A.Y3,
         XY2_XY3_Sum, XY2_XY3_Sum_Next, Length_Per_WVar_SUM, A.Angle
  from _PETAL_ANNO_050 as A,
       _PETAL_ANNO_070 as B
  where A.hID = B.hID and
        A.PARAMCD = B.PARAMCD and

```

```

A. TRTAN = B. TRTAN and
round(XY2_XY3_Sum , 0.000000000001) <=
round(Length_Per_WVar_SUM, 0.000000000001) <=
round(XY2_XY3_Sum_Next , 0.000000000001)
;
quit;
data _PETAL_ANNO; *花びら及び花序の描画用Annotateデータを作成;
set _PETAL_ANNO_080(in=a)
_HexCenter(in=b);
by PARAMCD TRTAN ;
if a and WVar>=2 then do; *被験者数を示す花びら(線);
linecolor='BLACK'; function='LINE'; drawspace='datavalue'; linethickness=2.0;
end;
else if b then do; *六角形の中心を示す花序(点);
X1=XVar_Mean; Y1=YVar_Mean;
fillcolor='BLACK'; function='OVAL'; drawspace='datavalue';
width=2.0; height=2.0; display='FILL';
end;
run;
proc sort data=_PETAL_ANNO out=_LOOP(keep=PARAMCD TRTAN) nodupkey;
by PARAMCD TRTAN;
run;
data _NULL_; *項目及び群の数だけループさせて、Sunflower Plotを作成;
set _LOOP;
call execute('proc sgplot data=_HexMap(where=(PARAMCD="" ||compress(PARAMCD)||" and TRTAN=' ||TRTAN||'))
sganno=_PETAL_ANNO(where=(PARAMCD="" ||compress(PARAMCD)||" and TRTAN=' ||TRTAN||'))');
call execute('polygon x=XVar y=YVar ID=hID
/ colorresponse=WVar fill colormodel=(white cxf03b20)');
call execute('run;');
run;
data _HexMap_Number; *花びらの代わりに被験者数を表示するためのAnnotateデータを作成;
set _HexMap;
WVar_Number=ifc(WVar=1, "", strip(put(WVar, best.)));
run;
data _NULL_; *項目及び群の数だけループさせて、Sunflower Plot (被験者数表示Ver.)を作成;
set _LOOP;
call execute('proc sgplot data=_HexMap_Number(where=(PARAMCD="" ||compress(PARAMCD)||" and
TRTAN=' ||TRTAN||'))
sganno=_PETAL_ANNO(where=(WVar=1 and PARAMCD="" ||compress(PARAMCD)||" and
TRTAN=' ||TRTAN||'))');
call execute('polygon x=XVar y=YVar ID=hID
/ colorresponse=WVar fill label=WVar_Number
labelattrs=(size=12 color=black weight=bold)
colormodel=(white cxf03b20)');
call execute('run;');
run;

```

# SASプロシジャを用いた生存時間データに対する例数設計の変革

○魚住 龍史<sup>1</sup> 矢田 真城<sup>2</sup> 浜田 知久馬<sup>3</sup>

<sup>1</sup>京都大学大学院医学研究科 医学統計生物情報学

<sup>2</sup>エイツーヘルスケア株式会社 開発戦略本部 生物統計部

<sup>3</sup>東京理科大学 工学部 情報工学科

Revolution in sample size calculation for time-to-event data using SAS procedures

Ryuji Uozumi<sup>1</sup>, Shinjo Yada<sup>2</sup>, and Chikuma Hamada<sup>3</sup>

<sup>1</sup>*Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine*

<sup>2</sup>*Biostatistics Department, Development Strategy Division, A2 Healthcare Corporation*

<sup>3</sup>*Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science*

## 要旨

POWER プロシジャは、様々な解析手法に対する例数設計や検出力を計算するために有用なプロシジャであり、TWOSAMPLESURVIVAL ステートメントを用いれば、生存時間解析のノンパラメトリック検定を前提とした例数設計を行うことができる。TWOSAMPLESURVIVAL ステートメントでは、Lakatos の方法が採用されており、生存時間解析の例数設計としてよく用いられている Freedman の方法や Schoenfeld の方法と比べて性能が良いことが報告されている。その一方、中間解析を含んだ群逐次デザインに用いられる SEQDESIGN プロシジャでは、ログランク検定を前提とした例数設計法として Schoenfeld の方法が採用されている。また、SAS/STAT 14.1 からの POWER プロシジャでは、新たに COXPH ステートメントが追加され、連続量データの共変量を含んだ多変量の比例ハザードモデルの解析に対する例数設計ができるようになった。本稿では、SAS で新たにできるようになった生存時間解析の例数設計に関する機能を取りあげ、それらの有用性について報告する。まず、Lakatos の方法から算出される必要例数以外の出力結果として、必要イベント数や単位時間あたりの登録例数がどのように算出されているか解説する。次に、POWER プロシジャの COXPH ステートメントで実施できるようになった例数設計法の特徴と有用性を報告する。最後に、患者の登録パターンが一樣でない場合の例数設計として、POWER プロシジャの代わりに、SEQDESIGN プロシジャで例数設計を行う方法を取りあげる。

キーワード：生存時間解析 例数設計 POWER TWOSAMPLESURVIVAL Lakatos COXPH Hsieh-Lavori SEQDESIGN Schoenfeld 登録パターン 切断指数分布



## 1 はじめに

医学研究では、ある基準の時刻から目的のイベントが起こるまでの時間を対象とした解析がよく行われる。このような解析手法を生存時間解析と呼び、LIFETEST, PHREG, LIFEREG といったプロシジャが 20 年以上用いられている (大橋・浜田, 1995)。近年では、SAS/STAT の機能拡張により、区間打ち切りデータに関するプロシジャとして、SAS/STAT 13.1 から ICLIFETEST, SAS/STAT 13.2 から ICPHREG が追加された (SAS Institute Inc., 2013, 2014)。さらに、SAS/STAT 14.1 からの LIFETEST プロシジャでは、EVENTCODE = オプションが追加され、競合リスクの解析が実施できるようになった (SAS Institute Inc., 2015)。

このように生存時間解析に関連したプロシジャが拡張されている一方、近年の医学研究では、研究目的を明確にするために、試験の必要例数を計画段階で事前に見積もらなければならない。これを例数設計と呼び、生存時間解析の例数設計を行うためには、POWER プロシジャにおける TWOSAMPLESURVIVAL ステートメントが有用である (浜田・安藤, 2005)。生存時間解析の例数設計法として、Freedman の方法 (Freedman, 1982) と Schoenfeld の方法 (Schoenfeld, 1981) がよく用いられているが、POWER プロシジャで採用されている Lakatos の方法 (Lakatos, 1988) の方が良い性能であることが報告されている (Lakatos and Lan, 1992; 中西ら, 2006; 水澤・浜田, 2008; 魚住ら, 2009)。また、SAS/STAT 14.1 からの POWER プロシジャでは、新たに COXPH ステートメントが追加され、連続量データの共変量を含んだ多変量の比例ハザードモデルの解析に対する例数設計ができるようになった (SAS Institute Inc., 2015)。

本稿では、SAS で新たにできるようになった生存時間解析の例数設計に関連した機能を取りあげる。まず、POWER プロシジャの TWOSAMPLESURVIVAL ステートメントによる Lakatos の方法から算出される必要イベント数や登録例数の算出について取りあげ、その有用性を示す。次に、POWER プロシジャの COXPH ステートメントで実施できるようになった例数設計法について解説し、TWOSAMPLESURVIVAL ステートメントと比較した場合の有用性を示す。最後に、POWER プロシジャの代わりに、SEQDESIGN プロシジャで生存時間解析の例数設計を行う有用性について取りあげる。

## 2 生存時間解析における例数設計法

生存時間解析においてはイベント数が情報量となるため、必要例数を求めるためには、まずは必要イベント数を見積もる必要がある。必要イベント数を求めた上で、

$$\text{必要例数} = \text{必要イベント数} / 2 \text{ 群を併せたイベントの割合} \quad (1)$$

を計算すればよい。

生存時間解析における必要イベント数は、Cox の比例ハザードモデルによるハザード比 (hazard ratio) を用いて算出される。比例ハザードモデルは、共変量ベクトル  $\mathbf{x}$  で条件付けたある患者集団のハザード関数  $h(t|\mathbf{x})$  が基準患者集団のハザード関数  $h_0(t)$  の定数倍になるモデルである (Cox, 1972)。ここで、回帰パラメータ  $\omega$ 、単変量  $x$  に対する比例ハザードモデル

$$\log[h(t|x)/h_0(t)] = \omega x \quad (2)$$

を考えると、ハザード比は

$$\frac{h(t|x=1)}{h(t|x=0)} = \exp(\omega)$$

となる。ただし、 $t$  は時間を表す。

## 2.1 Freedman の方法と Schoenfeld の方法

生存時間解析の例数設計法として、Freedman の方法 (Freedman, 1982) と Schoenfeld の方法 (Schoenfeld, 1981) がよく用いられている。それぞれの方法で算出される必要イベント数を  $D_F$ ,  $D_S$  とする。検定の有意水準を両側  $\alpha$ , 検出力の名義水準を  $1 - \beta$  とすると、必要イベント数はそれぞれ次の式で求まる。

$$D_F = \frac{(z_{\alpha/2} + z_{\beta})^2 (w \cdot HR + 1)^2}{w(HR - 1)^2} \quad (3)$$

$$D_S = \frac{(1 + w)^2 (z_{\alpha/2} + z_{\beta})^2}{w \{\log(HR)\}^2} \quad (4)$$

ただし、 $z_{\kappa}$  を標準正規分布の上側  $100 \cdot \kappa$  %点、 $HR$  をハザード比、 $w$  を群間の例数の割付比 (対照群 : 試験群 = 1 :  $w$ ) とする。いずれの式においても 2 群を併せた必要イベント数を表しており、1 群あたりの必要イベント数を求めるためには、それぞれの式から求めた  $D_F$  あるいは  $D_S$  を 0.5 倍すればよい。(3), (4) 式のいずれも、生存時間分布に指数分布を仮定して導出される。

これらの方法は、テイラー展開により 2 次式で近似すると一致するため、帰無仮説  $H_0: HR = 1$  の近傍では、両式はほぼ等しくなる (浜田・藤井, 2003)。また、 $w = 1$  の場合、Freedman の方法と比べて、Schoenfeld の方法の方が必要イベント数は必ず少なく済む。なお、いずれの方法においても、ハザード比を用いて例数設計を行っているが、ノンパラメトリックなログランク検定の例数設計とほぼ等価である。

ここで、試験群 ( $E$ ) と対照群 ( $C$ ) で構成される 2 群間比較試験を想定し、 $w = 1$  の場合で考える。このとき、必要例数を求めるために必要な条件となる、(1) 式の分母として示した 2 群を併せたイベントの割合について、各群における最終生存割合  $\pi_C$ ,  $\pi_E$  を用いて、

$$2 \text{ 群を併せたイベントの割合} = \frac{(1 - \pi_C) + (1 - \pi_E)}{2}$$

と見積もることができる。しかし、実際の臨床試験では、患者は逐次的に登録され、一度にすべての患者が組み入れられるわけではないため、登録期間を考慮する必要があるといえる。臨床試験の登録期間を  $A$  年、フォローアップ期間を  $F$  年として、範囲  $0 \sim A$  を確率密度  $1/A$  で一様分布に従って登録されると仮定すると、群  $g \in \{C, E\}$  における試験期間中の期待イベント割合は

$$P_g = 1 - \frac{1}{A} \int_0^A S_g(A + F - t) dt$$

となる。ただし、 $S_g(A + F - t)$  は時点  $(A + F - t)$  における患者の生存関数である。生存時間分布が指数分布に従うと仮定すると、

$$P_g = 1 - \frac{1}{A} \int_0^A \exp\{-\lambda_g \cdot (A + F - t)\} dt = 1 - \frac{\exp(-\lambda_g F) \{1 - \exp(-\lambda_g A)\}}{A \lambda_g} \quad (5)$$

となる。ただし、 $\lambda_g$  は群  $g \in \{C, E\}$  におけるハザードを表す指数分布の母数である。よって、2 群を併せた期待イベント割合  $(P_C + P_E)/2$  を用いて、必要例数を見積もることができる。

## 2.2 SEQDESIGN プロシジャによる Schoenfeld の方法

2.1 節の (3), (4) 式に基づく Freedman の方法および Schoenfeld の方法による必要イベント数、さらに必要例数の計算には、DATA ステップによる実装が容易にできる (大橋ら, 2016)。3 節で述べる POWER プロシジャ

ャによる例数設計のように、SASのプロシジャで出力させたい場合、Schoenfeldの方法であれば SEQDESIGN プロシジャを活用することで、必要イベント数と必要例数を算出することができる。

SEQDESIGN プロシジャは、本来群逐次デザインの検討を行うために有用なプロシジャである (浜田, 2010)。SAMPLESIZE ステートメントを指定することで、群逐次デザインの例数設計も行うことができる。

SEQDESIGN プロシジャでは、DESIGN ステートメントにおいて、群逐次デザインにおけるステージの数 (すなわち、中間解析の回数) を指定する。NSTAGE=1 とすれば、中間解析なしの固定デザインに相当する。

プログラム 1 は、検定の有意水準を両側  $\alpha$ 、検出力の名義水準を  $1-\beta$  のデザインを想定している。ALT=TWOSIDED で両側検定を指定しており、片側検定の場合には LOWER あるいは UPPER を指定する。DESIGN ステートメントにおいて NSTAGE=1 を指定した下で、SAMPLESIZE ステートメントで MODEL=TWOSAMPLESURVIVAL を指定すれば、2 群の生存時間データに対してログランク検定を実施することを前提とした、Schoenfeld の方法に基づく必要イベント数および必要例数を算出できる。なお、SAMPLESIZE ステートメントのオプションとして、NULLHAZARD= で対照群に対するハザードを指定し、ACCTIME= で登録期間、FOLTIME= でフォローアップ期間、WEIGHT= で群間の割付比 (対照群の例数 : 試験群の例数 = 1:w) を指定している。

プログラム 1: SEQDESIGN プロシジャによる Schoenfeld の方法に基づく例数設計

```
proc seqdesign altref=-log(HR);
  design nstages=1 beta= $\beta$  alpha= $\alpha$  alt=twosided;
  samplesize model=twosamplesurvival
              (nullhazard= $\lambda_C$ 
              ACCTIME=A FOLTIME=F
              weight=w);
run;
```

#### ・数値例

生存時間データに対してログランク検定による群間比較を行うことを想定し、有意水準を 5% (両側)、検出力の名義水準を  $1-\beta=0.80$ 、対照群 (C) における 5 年生存割合を 65%、試験群 (E) における 5 年生存割合を 80% (フォローアップ期間はすべての個体で等しく 5 年とする)、割付比として対照群の例数 : 試験群の例数 = 1:1 の条件の下、例数設計を行う場合を考える。各群の 5 年生存割合より、生存時間分布に指数分布を仮定すると、各群のハザードはそれぞれ 0.0862, 0.0446 となる。このとき、ハザード比  $HR$  は 0.518、対数ハザード比は  $\log(HR)=-0.6578$  となる。プログラム 2 は、上記の条件の下、SEQDESIGN プロシジャで Schoenfeld の例数設計を行うためのプログラムである。

プログラム 2 を実行させると、必要イベント数は 72.56 となり、必要例数は 227.61 例となる。なお、プログラム 2 をそのまま実行すると、対数ハザード比などの値における丸めの誤差より、必要イベント数および必要例数は上述した値とならない。付録 A に詳細なプログラムを示しており、DATA ステップで (4) 式を計算した値と一致することを確認できる。

なお、SAMPLESIZE ステートメントによって、必要イベント数と必要例数に加えて、Information 「情報量」が出力されるが、これは対数ハザード比の分散の逆数であり、

$$I = D_s \times \frac{w}{(1+w)^2}$$

で算出される値である。

プログラム 2: 数値例に対する Schoenfeld の方法に基づく例数設計

```
proc seqdesign altref=0.6578;
  design nstages=1 beta=0.20 alpha=0.05 alt=twosided;
  samplesize model=twosamplesurvival
              (nullhazard=0.0862
              ACCTIME=2 FOLTIME=5
              weight=1);
run;
```

### 3 POWER プロシジャによる例数設計

POWER プロシジャにおける生存時間解析の例数設計に関するステートメントとして、TWOSAMPLESURVIVAL と COXPH が挙げられる。TWOSAMPLESURVIVAL ステートメントでは、Lakatos の方法に基づき例数設計が実施される (Lakatos, 1988)。その一方、COXPH ステートメントは SAS/STAT 14.1 から追加された機能であり、Hsieh-Lavori の方法に基づく例数設計法である (Hsieh and Lavori, 2000)。

#### 3.1 TWOSAMPLESURVIVAL ステートメント

POWER プロシジャの TWOSAMPLESURVIVAL ステートメントにおいて採用されている Lakatos の方法では、観察期間を小区間に分けて区間ごとに期待イベント数を算出する。Freedman の方法や Schoenfeld の方法では、群間のハザードあるいは対数ハザードに差がないという帰無仮説を検定する一方、Lakatos の方法では区間ごとに群間のハザードに差がないという帰無仮説を検定する。

TWOSAMPLE ステートメントによる例数設計として、比例ハザードモデルを仮定した方法を考える。試験群 ( $E$ ) と対照群 ( $C$ ) で構成される 2 群間比較試験を想定し、時点  $t^*$  における群  $g \in \{C, E\}$  の生存割合  $S_g(t^*)$  から、ハザード比

$$HR = \frac{\log S_E(t^*)}{\log S_C(t^*)}$$

を推定し、例数設計を行う場合を考える。プログラム 3 は、群間の割合比が対照群 : 試験群 = 1 :  $w$  の下、POWER プロシジャによるログランク検定に基づく例数設計法である。CURVE オプションで両群の生存割合を指定しているが、片方の群のみ指定した上でハザード比 HAZARDRATIO = を指定してもよい。なお、各群の生存割合として、1 時点のみ指定しているので、比例ハザード性を仮定した例数設計が行われる。

さらに、バージョン 9.3 からの POWER プロシジャでは、検出力の名義水準を指定した下で、必要例数の他にも、表 1 に示す値も算出できるようになった (SAS Institute Inc., 2011)。例えば、必要イベント数を算出するためには、プログラム 3 における NTOTAL = の代わりに EVENTSTOTAL = を指定すればよい。

表 1: 必要例数以外に算出できる値

TWOSAMPLESURVIVAL ステートメントにおける指定	算出される数値の内容
EVENTSTOTAL   EVENTTOTAL   EETOTAL =	期待イベント数
ACCRUALRATETOTAL   ARTOTAL =	単位時間あたりの 2 群の登録例数
ACCRUALRATEPERGROUP   ACCRUALRATEPERG   ARPERGROUP   ARPERG =	単位時間あたりの 1 群あたりの登録例数

プログラム 3 のように、指数分布を想定した例数設計を行う場合、Lakatos の方法による必要イベント数  $D_L$  は、必要例数  $N_L$ 、群  $g \in \{C, E\}$  における試験期間中の期待イベント割合  $P_C, P_E$ 、および群間の割付比 (対照群の例数 : 試験群の例数 = 1 :  $w$ ) を用いて、

$$D_L = N_L \times \frac{P_C + wP_E}{1 + w}$$

と算出される。詳細は付録 B を参照されたい。

プログラム 3: POWER プロシジャによる比例ハザード性を仮定した例数設計

```
proc power;
  twosamplesurvival test=logrank
  curve("Control") = (t*):(SC)
  curve("Experimental") = (t*):(SE)
  groupsurvival = "Control" | "Experimental"
  groupweights = (1 w)
  accrualtime = A followuptime = F
  ntotal = .
  power = 1-β;
run;
```

・数値例

プログラム 4 は、2.2 節の数値例に対して、比例ハザード性を仮定した Lakatos の方法に基づく例数設計を行うためのプログラムである。必要イベント数は 74.35、必要例数は 233.23 例となる。1 群あたりの必要例数を算出したい場合は、NTOTAL = の代わりに NPERGROUP = を指定すればよい。NTOTAL = や NPERGROUP = では、小数点以下を切り上げた例数が出力されるため、プログラム 4 の右側に示した必要例数を求めるためのプログラムでは NFRACCTOTAL を指定し、小数点以下を切り上げる前の例数も出力している。NTOTAL = の代わりに、ACCRUALRATETOTAL = を指定すると、1 年あたりの登録例数が  $N_L/A = 233.23/2 = 116.62$  と出力される。

プログラム 4: 数値例に対する比例ハザード性を仮定した Lakatos の方法に基づく例数設計

必要イベント数	必要例数
<pre> proc power;   twosamplesurvival test=logrank   curve("Control") = (5):(0.65)   curve("Experimental") = (5):(0.80)   groupsurvival = "Control"                       "Experimental"   groupweights = (1 1)   accrualtime = 2   followuptime = 5   eventstotal = .   power = 0.8; run; </pre>	<pre> proc power;   twosamplesurvival test=logrank   curve("Control") = (5):(0.65)   curve("Experimental") = (5):(0.80)   groupsurvival = "Control"                       "Experimental"   groupweights = (1 1)   accrualtime = 2   followuptime = 5   ntotal = . nfractional   power = 0.8; run; </pre>

### 3.2 COXPH ステートメント

POWER プロシジャにおける生存時間解析の例数設計に関するステートメントとしては、以前は TWOSAMPLESURVIVAL のみであった。しかし、SAS/STAT 14.1 から新たに COXPH ステートメントが追加された。COXPH ステートメントでは、Hsieh-Lavori の方法に基づく例数設計が行われ、多変量の比例ハザードモデルの説明変数として、連続変数を加えた場合の変数間の相関を考慮した例数設計を実施できる。

#### 3.2.1 Hsieh-Lavori の方法

2 節の (2) 式に示した、単変量に対する Cox の比例ハザードモデルによるスコア検定統計量は次のようになる。

$$S^2 = \frac{\sum_{i \in D} (x_i - e_i)^2}{\sum_{i \in D} v_i}$$

ただし、 $D$  はイベント発現データの集合を表す。さらに、リスク集合を  $R$  として、

$$e_i = \frac{\sum_{j \in R} x_j}{n_i},$$

$$v_i = \frac{\sum_{j \in R} (x_j - e_i)^2}{n_i}$$

とする。

ここで、

$$\eta_i = \frac{\sum_{j \in R} x_j \exp(ax_j)}{\sum_{j \in R} \exp(ax_j)}$$

とすると、スコア検定統計量の分子は、

$$\sum_{i \in D} (x_i - e_i) = \sum_{i \in D} (x_i - \eta_i) + \sum_{i \in D} (\eta_i - e_i)$$

と分解できる。このとき、第1項は漸近的に

$$\sum_{i \in D} (x_i - \eta_i) \sim N(0, \sum_{i \in D} \sum_{j \in R} (x_j - \eta_i)^2 / n_i)$$

であり、 $\omega \rightarrow 0$ のとき、

$$\eta_i \rightarrow \frac{\sum_{j \in R} x_j}{n_i} = e_i,$$

$$\sum_{i \in D} \sum_{j \in R} \frac{(x_j - \eta_i)^2}{n_i} \rightarrow \sum_{i \in D} \sum_{j \in R} \frac{(x_j - e_i)^2}{n_i} = \sum_{i \in D} v_i$$

となるため、

$$\sum_{i \in D} (x_i - \eta_i) \sim N(0, \sum_{i \in D} v_i)$$

となる。

また、 $\omega = 0$ の周りでテイラー展開による1次近似を行うと、

$$\begin{aligned} \eta_i &\approx \frac{\sum_{j \in R} x_j}{n_i} + \omega \cdot \frac{\left\{ \sum_{j \in R} x_j^2 \exp(ax_j) \right\} \left\{ \sum_{j \in R} \exp(ax_j) \right\} - \left\{ \sum_{j \in R} x_j \exp(ax_j) \right\}^2}{\left\{ \sum_{j \in R} \exp(ax_j) \right\}^2} \\ &= e_i + \omega \cdot v_i \end{aligned}$$

となるため、

$$\sum_{i \in D} (\eta_i - e_i) \approx \omega \sum_{i \in D} v_i$$

となる。

以上より、 $D$ に含まれるデータ数が $D_{HL}$ とすると、

$$S = \frac{\omega \sum_{i \in D} v_i}{\sqrt{\sum_{i \in D} v_i}} = \omega \sqrt{\sum_{i \in D} v_i} = \omega \sqrt{\sigma^2 D_{HL}}$$

と表すことができる。ただし、 $\sigma^2$ は単変量 $x$ の分散である。ここで、推定すべき回帰パラメータ $\omega = \log(HR)$

として、必要イベント数は

$$D_{HL} = \frac{(z_{\alpha/2} + z_{\beta})^2}{\sigma^2 \{\log(HR)\}^2} \quad (6)$$

と表すことができる。(4)式のSchoenfeldの方法と同様に、 $\omega = 0$ における漸近式である。(4)式に比べて保守的な例数設計の式であるといえる。

次に、多変量 $\mathbf{x} = (x_1 \ x_2 \ \dots \ x_k)$ に対するCoxの比例ハザードモデル

$$\log[h(t|\mathbf{x})/h_0(t)] = \omega'_1 x_1 + \omega'_2 x_2 + \dots + \omega'_k x_k \quad (7)$$

を考えると、パラメータの推定値の分散を用いて得られるVIF (variance inflation factor) は、

$$VIF = \frac{V_k[\hat{\omega}'_1]}{V_1[\hat{\omega}]} = \frac{1}{1 - R^2}$$

となる。ただし、 $V_k[\hat{\omega}_1]$ は(7)式の $k$ 変量のモデルから得られる $\hat{\omega}_1$ の分散、 $V_1[\hat{\omega}]$ は(2)式の単変量のモデルから得られる $\hat{\omega}$ の分散、 $R^2$ は $x_2, x_3, \dots, x_k$ に対する $x_1$ の回帰式から得られる寄与率である。検出力を名義水準に保つために、多変量の比例ハザードモデルによる解析に対する必要イベント数は

$$D'_{HL} = D_{HL} \cdot VIF = \frac{D_{HL}}{1 - R^2} \quad (8)$$

となる。

(6), (8)式に基づき計算される必要イベント数を用いて、(1)式から必要例数を見積もることができる。

### 3.2.2 COXPH ステートメントによる Hsieh-Lavori の方法に基づく例数設計

プログラム5は、POWER プロシジャのCOXPH ステートメントを用いて、Hsieh-Lavoriの方法による例数設計を行うためのプログラムである。必要イベント数の計算式である(6)式は、(3), (4)式と異なり、 $\sigma^2$ を指定しなければならない。さらに、必要イベント数の計算には影響しないが、寄与率 $R^2$ も指定することができ、 $R^2$ が大きいほど必要例数は多くなる。また、必要例数を求める場合には、EVENTPROB = でイベント割合を指定する必要がある。

プログラム5: POWER プロシジャによる Hsieh-Lavori の方法に基づく例数設計

必要イベント数	必要例数
<pre>proc power;   coxreg     hazardratio = HR     rsquare = R<sup>2</sup>     stddev = SD     eventstotal = .     power = 1-β; run;</pre>	<pre>proc power;   coxreg     hazardratio = HR     rsquare = R<sup>2</sup>     stddev = SD     ntotal = .     eventprob = イベント割合     power = 1-β; run;</pre>

\*SAS/STAT 14.1 より実行可能

#### ・数値例

COXPH ステートメントによる例数設計の数値例として、SAS/STAT PHREG プロシジャのマニュアルで取り扱われている、多発性骨髄腫のデータを用いる。このデータは、アルキル化剤で65人の患者を治療する試験から得られた死亡または打ち切りまでの生存時間データであり、このうち48人がイベント発生、17人が打ち切りであった。生存時間の共変量として、診断時の尿素窒素を表すBUNの対数をとった値(LogBUN)、白血球数を表すWBCの対数をとった値、ヘモグロビン値、血清カルシウム値などの9つが取りあげられており、9つの共変量のうち予後因子を特定することに関心がある。

ここで、LogBUNが興味のある共変量であると仮定すると、他の8つの共変量との関連をあることが考えられる。このとき、他の8つの共変量に対するLogBUNの回帰式から得られる寄与率 $R^2$ は0.1839、VIFで示すと1.225となり、LogBUNの標準偏差は0.3126となる。



多発性骨髄腫のデータの LogBUN に関心がある場合、有意水準 5% (片側)、検出力の名義水準 80% の下、対数ハザード比 1 の差を検出できるような例数設計を行うにあたって、まず必要イベント数を求めると、(6) 式より 64 例と算出される。(1) 式より、必要イベント数から必要例数を求めると、イベント割合が  $1 - 17/65 = 0.738$  なので、 $64/0.738 = 86.67$  となる。よって、(8) 式より、多変量の比例ハザードモデルによる解析に対する必要例数は、86.67 を  $VIF = 1.225$  倍して、106.17 となる。

プログラム 6 は、これらの計算を POWER プロシジャで実施するためのプログラムである。 $R^2$  を考慮して計算すると、必要イベント数は 78、必要例数は 105 と計算される。なお、 $R^2 = 0$  とすれば、必要イベント数が 64、必要例数が 86 と計算されることを確認できる。

プログラム 6: 数値例に対する Hsieh-Lavori の方法に基づく例数設計

必要イベント数	必要例数
<pre>proc power;   coxreg     alpha = 0.05     sides = 1     hazardratio = 2.718     rsquare = 0.1837     stddev = 0.3126     eventstotal = .     nfrac     power = .8; run;</pre>	<pre>proc power;   coxreg     alpha = 0.05     sides = 1     hazardratio = 2.718     rsquare = 0.1837     stddev = 0.3126     ntotal = .     nfrac     eventprob = 0.7385     power = .8; run;</pre>

\*SAS/STAT 14.1 より実行可能

プログラム 6 は、SAS/STAT14.1 から実行可能である。なお、日本語版の SAS/STAT 14.1 から、POWER プロシジャの出力結果は一部日本語で出力される。出力結果の詳細は付録 E を参照されたい。

#### 4 登録パターンが一様でない場合の例数設計

3.1 節の POWER プロシジャにおける TWOSAMPLESURVIVAL ステートメントでは、脱落までの時間が指数分布に従うと仮定して、例数設計を行うことが可能である (大橋ら, 2016)。しかし、登録パターンについては、登録期間中に一様に患者が登録されると仮定した場合のみで、脱落までの時間のように指数分布を仮定した検討を行うことはできない。

しかし、2.2 節で取りあげた SEQDESIGN プロシジャを用いれば、期間内で一様に登録されるパターン以外に、切断指数 (truncated exponential) 分布に従って登録されると仮定することができる。切断指数分布の確率密度関数および累積分布関数はそれぞれ

$$g(t) = \frac{\gamma \exp(-\gamma t)}{1 - \exp(-\gamma A)},$$

$$G(t) = \frac{1 - \exp(-\gamma t)}{1 - \exp(-\gamma A)},$$

となる ( $0 \leq t \leq A$ ). ただし,  $\gamma$  は形状パラメータである ( $\gamma \neq 0$ ). 切断指数分布の累積分布関数の形状は図 1 のようになり (作成プログラムは付録 F を参照), 点線で示した一様分布と比較すると,  $\gamma < 0$  のときは凸形の登録パターンとなり, その一方  $\gamma > 0$  のときは凹形の登録パターンとなる.

一様分布に従って登録される場合の期待イベント割合を表す (5) 式に対して, 登録パターンが切断指数分布に従う場合の期待イベント割合は,

$$P_g' = \int_F^{A+F} g(A+F-t) \{1 - \exp(-\lambda_g t)\} dt = 1 + \frac{\gamma \exp\{-\lambda_g(A+F)\} [1 - \exp\{(\lambda_g + \gamma)A\}]}{(1 - \exp(-\gamma A))(\lambda_g - \gamma)} \quad (9)$$

となる (Lachin and Foulkes, 1986). よって,  $w = 1$  の場合で考えると, 2.1 節と同様の手順, すなわち 2 群を併せた期待イベント割合  $(P_C' + P_E')/2$  を用いて, 必要例数を見積もることができる.

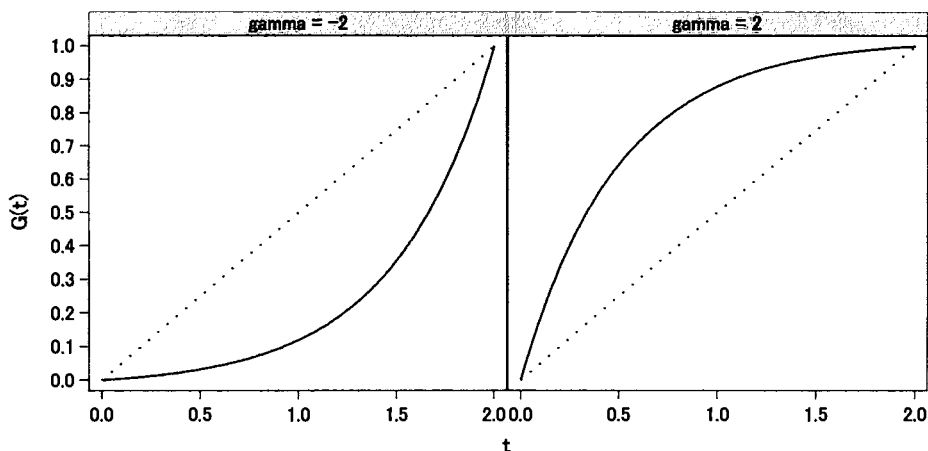


図 1: 切断指数分布の累積分布関数

#### 4.1 SEQDESIGN プロシジャによる登録パターンが一様でない場合の例数設計

2.2 節に示した方法に基づき SEQDESIGN プロシジャを活用すれば, POWER プロシジャでは検討できなかった登録パターンが一様でない場合の例数設計を行うことができる. プログラム 7 は, SEQDESIGN プロシジャを用いて, 登録パターンが切断指数分布に従うと仮定した下で, 必要イベント数および必要例数を算出するためのプログラムである. 2.2 節同様, SEQDESIGN プロシジャの例数設計は Schoenfeld の方法に基づき実施され, 登録パターンの考慮についてのみ Lachin-Foulkes の方法に基づいている.

プログラム 7 では, SEQDESIGN プロシジャの SAMPLESIZE ステートメントにおいて, ACCRUAL = EXP(PARM = GAMMA') を指定することで, 登録パターンにパラメータ  $\gamma'$  の切断指数分布を仮定することができる. 注意すべき点として, ACCRUAL = で指定する GAMMA' は, (9) 式の  $\gamma$  を直接指定するのではなく,  $\gamma' = A\gamma$  を指定する. 例えば, 登録期間として ACCTIME = 2 を指定しており, 図 1 で考えた  $\gamma = -2$  の切断

指数分布を仮定したい場合には、`PARM=-4` と指定すればよい。なお、`SAMPLESIZE` ステートメントにおいて `ACCRUAL=` を指定しなければ、デフォルトである `ACCRUAL=UNIFORM` が適用される。

プログラム 7: SEQDESIGN プロシジャによる登録パターンが切断指数分布に従うと仮定した例数設計

```

proc seqdesign altref=-log(HR);
  design nstages=1 beta= $\beta$  alpha= $\alpha$  alt=twosided;
  samplesize model=twosamplesurvival
              (nullhazard= $\lambda_c$ 
              ACCTIME=A FOLTIME=F
              accrual=exp(parm= $\gamma$ )
              weight=w);
run;

```

・ 数値例

登録パターンが一様でない場合の数値例として、2.2 節の数値例と同様の条件であるが、図 1 のパラメータが  $\gamma = -2, 2$  の切断指数分布に登録パターンが従う場合を考える。

プログラム 8: 数値例に対する登録パターンが切断指数分布に従う場合の例数設計

凸形の登録パターンの場合 ( $\gamma = -2$ )	凹形の登録パターンの場合 ( $\gamma = 2$ )
<pre> <b>proc seqdesign</b> altref=0.6578;   design nstages=1           beta=0.20 alpha=0.05           alt=twosided;   samplesize model=twosamplesurvival               (nullhazard=0.08616               ACCTIME=2 FOLTIME=5               <b>accrual=exp(parm=-4)</b>               weight=1); <b>run;</b> </pre>	<pre> <b>proc seqdesign</b> altref=0.6578;   design nstages=1           beta=0.20 alpha=0.05           alt=twosided;   samplesize model=twosamplesurvival               (nullhazard=0.08616               ACCTIME=2 FOLTIME=5               <b>accrual=exp(parm=4)</b>               weight=1); <b>run;</b> </pre>

プログラム 2 による実行結果より、必要イベント数は 72.56、必要例数は 227.61 例であった。SEQDESIGN プロシジャによる登録パターンが一様でない場合の例数設計では、(4) 式に基づき必要イベント数を求め、(9) 式から得られる期待イベント割合から、(1) 式に基づき必要例数が算出される。そのため、必要イベント数は登録パターンを変動させても同様の値が得られる。しかし、必要例数に関して、登録パターンが切断指数分布に従う場合、プログラム 8 より、 $\gamma = -2$  の場合は 245.46 例に増大し、 $\gamma = 2$  の場合は 212.42 例に減少することを確認できる。

## 6 まとめ

本稿では、生存時間解析の例数設計に関連したプロシジャとして、POWER プロシジャ、SEQDESIGN プロシジャによる例数設計について取りあげた。

POWER プロシジャによる Lakatos の方法に基づく例数設計では、これまで例数のみしか出力できず、イベント数については算出できなかった。しかし、SAS/STAT 9.3 からイベント数についても算出され、例数設計法として性能が優れている Lakatos の方法を実務に適用する機会が増えることが予想される。

SAS/STAT 14.1 以降の POWER プロシジャでは、Lakatos の方法に加えて、Hsieh-Lavori の方法に基づく例数設計もできるようになった。がんを対象とした第 III 相試験のように、生存時間データを主要評価項目とした 2 群間比較試験の例数設計では、Lakatos の方法が有用である。しかし、疫学研究や非ランダム化比較試験では、生存時間に対する共変量として連続変数を加える場合があり、興味のある共変量が他の共変量と独立でないことが考えられる。このとき、Hsieh-Lavori の方法に基づき、VIF を考慮して例数設計を行うことが可能である。

また、POWER プロシジャでは、登録期間中に患者が一様に登録されると仮定しており、登録パターンが一様でない場合の検討を行うことができない。そこで、POWER プロシジャの代わりに、SEQDESIGN プロシジャで例数設計を行うことは有用であり、登録パターンが切断指数分布に従うと仮定した下で、Schoenfeld の方法に基づく例数設計を行うことができることを示した。Schoenfeld の方法は Lakatos の方法より性能が劣ってしまうが、生存時間解析の例数設計法として最もよく用いられている方法であり (Barthel et al, 2006), 日本語訳が出版されている生存時間解析のテキストにおいても取りあげられている (Collett, 2003; Hosmer et al, 2008). そのため、実務上、登録パターンが一様でない場合の必要イベント数、必要例数を検討するにあたって大変有用といえる。

比例ハザード性が成り立たない場合、比例ハザードモデルを仮定しない方法として、生存時間分布に区分直線モデルを仮定することで、精度良く例数設計を実施できる。POWER プロシジャにより、生存時間分布に区分直線モデルを仮定した例数設計を行うためには、CURVE オプションで 2 時点以上の生存割合を指定すればよい。比例ハザードモデルを仮定せずに、生存時間分布に区分直線モデルを仮定する方法については、魚住ら (2009), 大橋ら (2016) を参照されたい。

なお、POWER プロシジャによる Hsieh-Lavori の方法、SEQDESIGN プロシジャによる登録パターンが一様でない場合の例数設計に関しては、DATA ステップによる実装も付録に示しており (付録 C, G), プロシジャによる実行結果 (付録 D, H) と一致することを確認できる。

## 参考文献

- [1] Barthel FMS, Babiker A, Royston P, Parmar MK. Evaluation of sample size and power for multiarm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. *Statistics in Medicine*. 25:2521–2542, 2006.
- [2] Collett D. *Modelling Survival Data in Medical Research* (2nd edn.). Chapman and Hall/CRC, Boca Raton, FL, 2003. 宮岡悦良(監訳), グラクソ・スミスクライン株式会社パイオメディカルデータサイエンス部: 安藤英一・今井由希子・遠藤輝・兼本典明・張方紅・寺尾工・橋本浩史・中間剛介(訳). 医薬統計のための生存時間データ解析. 共立出版, 2013.

- [3] Cox DR. Regression models and life-tables (with Discussion). *Journal of the Royal Statistical Society, Series B.* 34:187–220, 1972.
- [4] Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine.* 1:121–129, 1982.
- [5] Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions, Vol.1* (2nd edn.). John Wiley and Sons, New York, NY, 1994.
- [6] Hosmer, D. W. Jr., Lemeshow, S. and May, S. *Applied Survival Analysis: Regression Modeling of Time to Event Data* (2nd edn.). John Wiley and Sons, Hoboken, NJ, 2008. 五所正彦(監訳), 佐藤泰憲・竹内久朗・長島健悟・中水流嘉臣・平川晃弘・松永信人・山田雅之(訳). 生存時間解析入門 (原書第2版). 東京大学出版会, 2014.
- [7] Hsieh FY, Lavori PW. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials.* 21:552–560, 2000.
- [8] Lachin JM, Foulkes MA. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics.* 42:507–519, 1986.
- [9] Lakatos E. Sample size based on the log-rank statistic in complex trials using the log-rank test. *Biometrics.* 44:229–241, 1988.
- [10] Lakatos E, Lan KKG. A comparison of sample size methods for the logrank statistic. *Statistics in Medicine.* 11:179–191, 1992.
- [11] SAS Institute Inc. *SAS/STAT(R) 9.3 User's Guide.* SAS Institute Inc., Cary, NC, 2011.
- [12] SAS Institute Inc. *SAS/STAT(R) 13.1 User's Guide.* SAS Institute Inc., Cary, NC, 2013.
- [13] SAS Institute Inc. *SAS/STAT(R) 13.2 User's Guide.* SAS Institute Inc., Cary, NC, 2014.
- [14] SAS Institute Inc. *SAS/STAT(R) 14.1 User's Guide.* SAS Institute Inc., Cary, NC, 2015.
- [15] Schoenfeld D. The asymptotic properties of nonparametric tests for comparing survival distribution. *Biometrika.* 68:316–319, 1981.
- [16] 魚住龍史・水澤純基・浜田知久馬. 生存時間解析における Lakatos の症例数設計法の有用性の評価. SAS ユーザー総会論文集, 143–152, 2009.
- [17] 魚住龍史・森田智視. 生存時間解析における三種の神器. 呼吸, 34: 1083–1089, 2015.
- [18] 大橋靖雄・浜田知久馬. 生存時間解析 –SAS による生物統計. 東京大学出版会, 1995.
- [19] 大橋靖雄・浜田知久馬・魚住龍史. 生存時間解析 応用編 –SAS による生物統計. 東京大学出版会, 2016.
- [20] 張方紅・寺尾工. 非劣性試験の症例数設計方法の紹介 –生存時間データの場合–. SAS ユーザー総会論文集, 87–100, 2010.
- [21] 張方紅. SAS による生存時間解析の実務. SAS ユーザー総会論文集, 167–184, 2011.
- [22] 中西豊支・五所正彦・菅波秀規. POWER プロシジャを用いた生存時間解析における症例数設定方法の統計学的一考察. SAS Forum ユーザー会論文集, 161–169, 2006.
- [23] 浜田知久馬・藤井陽介. 生存時間解析の症例数設計. 日本 SAS ユーザー会 論文集, 73–98, 2003.
- [24] 浜田知久馬・安藤英一. POWER プロシジャによる症例数設計. SAS Forum ユーザー会 論文集, 127–151, 2005.
- [25] 浜田知久馬. SAS による中間解析のデザインと解析. SAS ユーザー総会 論文集, 111–179, 2010.
- [26] 水澤純基・浜田知久馬. 生存時間解析における症例数設計方法の性能比較. SAS ユーザー総会 論文集, 19–28, 2008.

## 連絡先

E-mail : [uozumi@kuhp.kyoto-u.ac.jp](mailto:uozumi@kuhp.kyoto-u.ac.jp)

### 付録 A : SEQDESIGN プロシジャによる Schoenfeld の方法に基づく例数設計の確認プログラム

```
*=====*;
* 追加の条件設定 ;

%let alpha=0.05; %let beta=0.2; %let power=%sysevalf(1-&beta); %let w=1;
%let Sc=0.65; %let Se=0.80;
%let HR=%sysevalf(%sysfunc(log(&Se))/%sysfunc(log(&Sc)));
%let logHR=%sysevalf(-%sysfunc(log(&HR)));
%let accrual=2; %let followup=5;
%let hc=%sysevalf(-%sysfunc(log(&Sc))/&followup);
%let he=%sysevalf(-%sysfunc(log(&Se))/&followup);
*=====*;
* SEQDESIGNプロシジャによるSchoenfeldの方法に基づく例数設計 ;

proc seqdesign altref=&logHR;
  design nstages=1 beta=&beta alpha=&alpha alt=twosided;
  samplesize model=twosamplesurvival
    ( nullhazard=&hc weight=&w
      ACCTIME=&accrual FOLTIME=&followup accrual=uniform );
run;
*=====*;
* DATAステップによるSEQDESIGNプロシジャの計算結果の確認 ;

data sample;
  alpha=&alpha; beta=&beta; w=&w;
  za=probit(1-alpha/2); zb=probit(1-beta); Sc=&Sc; Se=&Se; HR=log(Se)/log(Sc);
  accrual=&accrual; followup=&followup; total=accrual+followup; hc=&hc; he=&he;
  Pc=1-exp(-hc*followup)*(1-(exp(-hc*accrual)))/(accrual*hc);
  Pe=1-exp(-he*followup)*(1-(exp(-he*accrual)))/(accrual*he);
  DtotalS=(1+w)**2*(za+zb)**2/(w*log(HR)**2);
  NtotalS=DtotalS*(1+w)/(Pc+w*Pe);
run;
```

### 付録 B : DATA ステップによる Lakatos の方法に基づく必要イベント数の計算プログラム

```
*=====*;
* Lakatosの方法に基づく必要例数 ;

ods output Output=lakatos;
proc power;
  twosamplesurvival test=logrank
  curve("Control") = (5):(0.65) curve("Experimental") = (5):(0.80)
  groupsurvival = "Control" | "Experimental"
  groupweights = (1 1) accrualtime = 2 followuptime = 5
  ntotal = . nfractional power = 0.8;
run;
*=====*;
* DATAステップによるLakatosの方法に基づく必要例数から必要イベント数の計算 ;

data lakatos; set lakatos;
  Sc=0.65; Se=0.80; HR=log(Se)/log(Sc);
  hc=-log(Sc)/FollowUpTime; he=-log(Se)/FollowUpTime;
  Pc=1-exp(-hc*FollowUpTime)*(1-(exp(-hc*AccrualTime)))/(AccrualTime*hc);
  Pe=1-exp(-he*FollowUpTime)*(1-(exp(-he*AccrualTime)))/(AccrualTime*he);
  Dtotal=FractionalN*(Weight1*Pc+Weight2*Pe)/(Weight1+Weight2);
run;
*=====*;
* Lakatosの方法に基づく必要イベント数の確認 ;
```

```

proc power;
  twosamplesurvival test=logrank
  curve("Control") = (5):(0.65) curve("Experimental") = (5):(0.80)
  groupsurvival = "Control" | "Experimental"
  groupweights = (1 1) accrualtime = 2 followuptime = 5
  eventstotal = . power = 0.8;
run;

```

付録 C: DATA ステップによる Hsieh-Lavori の方法に基づく必要イベント数の計算プログラム

```

*====*;
* 変数LogBUNの標準偏差の確認 ;

proc means data=Myeloma;
  var logBUN;
run;
*====*;
* VIFの確認 ;

proc reg data=Myeloma;
  model Time = LogBUN HGB Platelet Age LogWBC Frac LogPBM Protein SCalc / vif;
run;quit;
*====*;
* 寄与率R2の確認 ;

proc reg data=Myeloma;
  model LogBUN = HGB Platelet Age LogWBC Frac LogPBM Protein SCalc;
run;quit;
*====*;
* 条件の指定 ;

%let alpha=0.05; %let beta=0.2; %let power=%sysevalf(1-&beta); %let w=1;
%let logHR=1; %let HR=%sysfunc(exp(&logHR));
%let sigma=0.3126; %let VIF=1.225; %let rs=%sysevalf(1-1/&VIF);
%let n0=65; %let c0=17; %let eventp0=%sysevalf(1-&c0/&n0);
*====*;
* DATAステップによるHsieh-Lavoriの方法に基づく例数設計 ;

data sample;
  alpha=&alpha; beta=&beta; w=&w; za=probit(1-alpha); zb=probit(1-beta);
  logHR=&logHR; HR=exp(logHR);
  sigma=&sigma; VIF=&VIF; rsquare=1-1/VIF; pi=1-&c0/&n0;
  DtotalHL=(za+zb)**2/(sigma*logHR)**2; NtotalHL=DtotalHL/pi;
  NtotalHL_VIF=NtotalHL*VIF; DtotalHL_VIF=DtotalHL*VIF;
run;

```

付録 D: COXPH ステートメントによる Hsieh-Lavori の方法の計算プログラム

```

*====*;
* 必要イベント数 - DtotalHL, DtotalHL_VIF に一致 ;

proc power;
  coxreg
  alpha = &alpha sides = 1 hazardratio = &HR rsquare = 0,&rs stddev = &sigma
  eventstotal = . nfrac power = &power;
run;
*====*;
* 必要例数 - NtotalHL, NtotalHL_VIF に一致 ;

proc power;
  coxreg
  alpha = &alpha sides = 1
  hazardratio = &HR rsquare = 0,&rs stddev = &sigma
  ntotal = . nfrac eventprob = &eventp0 power = &power;
run;

```

付録 E: COXPH ステートメント (付録 D) による Hsieh-Lavori の方法の計算結果

● 必要イベント数

計算された Ceiling Event Total				
インデックス	R-square	Fractional Event Total	実際の検出力	Ceiling Event Total
1	0.000	63.268887	0.804	64
2	0.184	77.504387	0.802	78

● 必要例数

計算された上限の総標本サイズ				
インデックス	R-square	小数点の総標本サイズ	実際の検出力	上限の総標本サイズ
1	0.000	85.676618	0.801	86
2	0.184	104.953858	0.800	105

付録 F: 切断指数分布 (図 1) の作成プログラム

```

=====
* 切断指数分布に従う登録パターン (A = 2) ;
data trunexp;
a=2;
do gamma=-2,2; do t=0 to 2 by 0.01;
  g=(1-exp(-gamma*t))/(1-exp(-gamma*a)); y=0.5*t; output;
end; end;
run;
=====
* 登録パターン別にグラフ化 ;
proc sgpanel data=trunexp noautolegend;
panelby gamma / columns=2;
series x=t y=g / lineattrs=(pattern=1 color=black);
series x=t y=y / lineattrs=(pattern=35 color=black);
rowaxis values=(0 to 1 by 0.1) label="G(t)";
colaxis values=(0 to 2 by 0.5) label="t";
run;

```

付録 G: DATA ステップによる登録パターンを考慮した例数設計プログラム

```

=====
* 付録Aの条件に加えて以下を設定 ;
%let gamma=-2; %let gamma seqdesign=%sysevalf(&accrual*&gamma);
=====
* DATAステップによる登録パターンを考慮したSchoenfeldの方法に基づく例数設計 ;
data sample;
alpha=&alpha; beta=&beta; w=&w; za=probit(1-alpha/2); zb=probit(1-beta);
gamma=&gamma; Sc=&Sc; Se=&Se; HR=log(Se)/log(Sc);
accrual=&accrual; followup=&followup; total=accrual+followup; hc=&hc; he=&he;
PcUni=1-exp(-hc*followup)*(1-(exp(-hc*accrual)))/(accrual*hc);
PeUni=1-exp(-he*followup)*(1-(exp(-he*accrual)))/(accrual*he);
PcExp=1+(gamma*exp(-hc*total)*(1-exp((hc-gamma)*accrual)))/
  (((1-exp(-gamma*accrual))*(hc-gamma)));
PeExp=1+(gamma*exp(-he*total)*(1-exp((he-gamma)*accrual)))/
  (((1-exp(-gamma*accrual))*(he-gamma)));
DttotalS=(1+w)**2*(za+zb)**2/(w*log(HR)**2);
NtotalS_Uni=DttotalS*(1+w)/(PcUni+w*PeUni);
NtotalS_Exp=DttotalS*(1+w)/(PcExp+w*PeExp);
run;

```



付録 H: SEQDESIGN プロシジャによる登録パターンを考慮した例数設計プログラム

```
*=====*;
* 登録パターンが一樣の場合の必要例数の確認 ;

proc seqdesign altref=&logHR;
  design nstages=1 beta=&beta alpha=&alpha alt=twosided;
  samplesize model=twosamplesurvival
    ( nullhazard=&hc weight=&w
      ACCTIME=&accrual FOLTIME=&followup accrual=uniform );
run;
*=====*;
* 登録パターンが一樣でない場合の必要例数の確認 ;

proc seqdesign altref=&logHR;
  design nstages=1 beta=&beta alpha=&alpha alt=twosided;
  samplesize model=twosamplesurvival
    ( nullhazard=&hc weight=&w
      ACCTIME=&accrual FOLTIME=&followup
      accrual=exp(parm=&gamma_seqdesign) );
run;
```

# 生存時間データに対するベイズ流例数設計

○矢田 真城<sup>1</sup> 魚住 龍史<sup>2</sup> 浜田 知久馬<sup>3</sup>

<sup>1</sup> エイツーヘルスケア株式会社 開発戦略本部 生物統計部

<sup>2</sup> 京都大学大学院医学研究科 医学統計生物情報学

<sup>3</sup> 東京理科大学 工学部 情報工学科

Bayesian sample size calculation for survival analysis

Shinjo Yada<sup>1</sup>, Ryuji Uozumi<sup>2</sup>, and Chikuma Hamada<sup>3</sup>

<sup>1</sup> *Biostatistics Department, Development Strategy Division, A2 Healthcare Corporation*

<sup>2</sup> *Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine*

<sup>3</sup> *Department of Information and Computer Technology, Tokyo University of Science*

## 要旨

臨床試験では、試験のデザインや目的に従い試験結果の精度を保証するために、必要と考えられる例数を設定する。必要例数を解析的に導出することが難しい場合、シミュレーションにより必要例数を見積もる方法が用いられる。シミュレーションによるベイズ流例数設計では、“もし…としたら”という状況の下で仮想的に生成されたデータを解析し、その結果がモデル判定基準を満たすような例数を、当該臨床試験に必要な例数と見積もる。このとき、検定の有意水準、検出力、効果の差といった概念は、事後分布に関連付けて提示することができる。本稿では、シミュレーションによるベイズ流例数設計として、indifference zone を用いた方法について紹介する。これは、効果の差を表すモデルパラメータの信用区間と indifference zone との位置関係から必要例数を見積もる方法である。シミュレーションの実行手順に関してまとめ、適用例として生存時間データにおける例数設計をとりあげ、対応する SAS プログラムと簡単なシミュレーション結果を示す。

キーワード： indifference zone, design prior, analysis prior

## 1. はじめに

臨床試験における例数設計は、試験デザインの重要な一つの要素である（上坂，2006）．統計学的な例数設計には、精度に基づく例数設計と検出力に基づく例数設計とがあるが、臨床試験では検出力に基づく例数設計がよく用いられている．検定に用いる帰無仮説及び対立仮説を設定し、検定の有意水準  $\alpha$ 、検出力  $1-\beta$ 、予想される群間での効果の差  $\Delta$  とそのばらつきを与え、適切な検定方法を選択することで、必要例数を算出することができる．ただし、必要例数を解析的に算出するためには、対立仮説のもとでの検定統計量の分布を知る必要がある．より複雑なモデルを想定した場合、対立仮説のもとでの検定統計量が従う分布の導出が難解になり、必要例数を明示できないことがありえる．このような場合、シミュレーションに基づき必要例数を見積もる方法がある（大橋等，2016）．

シミュレーションに基づくベイズ流の例数設計では、必要例数  $N$  を仮に定めた後、1) 予め設定した事前分布に基づきモデルパラメータを生成させる、2) 生成させたモデルパラメータを用いてデータを発生させる、3) 発生させたデータを用いてベイズ流に解析する、4) 得られた事後分布から予め設定したモデル性能基準（model performance criteria）と照らし合わせる、という手順をとり、モデル性能基準を満たしていなければ、 $N$  を変えて1)から4)を再度実行する（Wang and Gelfand, 2002）．このとき、検定の有意水準、検出力、群間での効果の差といった概念は事後分布に関連付けて定義される．そこで本稿では、事後分布に基づくベイズ流例数設計の1つとして、indifference zone による方法を取りあげ、シミュレーションによる例数設計の手順を示し、検定の有意水準、検出力との類似性をまとめる．

本稿の構成は以下のとおりである．第2節でindifference zone を用いたベイズ流例数設計について説明する．第3節にて、indifference zone を用いたベイズ流例数設計として、生存時間解析における適用例を具体的に紹介する．最後に第4節で本稿のまとめを行う．

## 2. ベイズ流例数設計

### 2.1. Indifference zone

本稿では、対照群に対する試験群の有効性を示すことを目的とした、並行2群比較試験における例数設計を対象とする．”indifference zone” を用いたベイズ流例数設計において、基本となるアイデアは、検定の際に設定する帰無仮説「 $H_0: \Delta = 0$ 」をある範囲をもった区間  $\Delta \in [\delta_L, \delta_U]$  に置き換えることにある．ここに、 $\Delta$  は対照群（control）に対する試験群（treatment）の2群間での効果の差、 $\delta_U$  は試験群がこの値を超えると対照群に比べて臨床的に優れているといえる限界値であり、 $\delta_L$  はこの値を下回ると試験群が対照群よりも臨床的に劣ると考えられる閾値である．このindifference zone  $[\delta_L, \delta_U]$  と2群間での効果の差  $\Delta$  の95%信用区間との位置関係から、試験群と対照群との優劣を結論付けることが可能となる．効果の差  $\Delta$  が大きいほど対照群に対する試験群の優越性が示されている状態であるとき、 $\Delta$  の95%信用区間に基づく判定基準（decision rule）は図2.1のようになる（Spiegelhalter et al., 1994）．”accept treatment”（試験群を採択）、”reject control”（対照群を棄却）、”equivalence”（同等）、”reject treatment”（試験群を棄却）、”accept control”（対照群を採択）、”no decision”（判定不能）のいずれかの判定が下される．

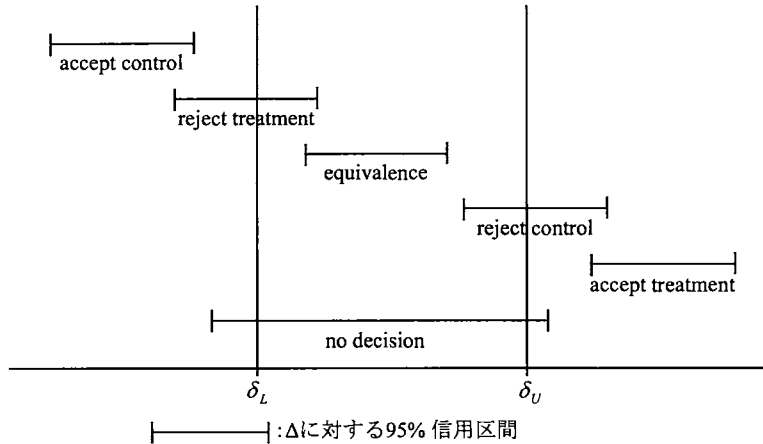


図 2.1 : 効果の差  $\Delta$  の 95%信用区間に基づく indifference zone と判定基準との関係

## 2.2. シミュレーションによる例数設計

例数設計の大きな流れとしては次のようになる。

- 手順 1 データに対してあてはめるモデルを想定し, 2 群間での効果の差  $\Delta$ , indifference zone  $[\delta_L, \delta_U]$  及びモデルパラメータの事前分布を設定する。
- 手順 2 事前分布に基づき  $N$  人分の被験者の観測値を仮想的に生成する
- 手順 3 手順 2 で生成された仮想的な観測値を用いてベイズ流にパラメータを更新し,  $\Delta$  の 95%信用区間を推定する。
- 手順 4  $\Delta$  の 95%信用区間と indifference zone  $[\delta_L, \delta_U]$  との位置関係から, 図 2.1 に従い 6 つの判定基準のいずれに該当するかを判定する。これが 1 回のシミュレーションによって得られる判定結果となる。
- 手順 5 手順 2 から手順 4 までを数千から数万回繰り返すことで, 判定基準ごとの経験的確率を計算することができ, 例数  $N$  に対する検出力などを推定することができる。

ベイズ流アプローチでは, 解析事前分布 (analysis prior あるいは fitting prior) とデザイン事前分布 (design prior あるいは sampling prior) という 2 種類の事前分布を設定し, 例数設計を行う。解析事前分布は, 事前情報をモデル化し (ときにヒストリカルデータに代表される情報をモデル化し), 事後分布を得るために用いられるべきであり, デザイン事前分布はパラメータの不確実性をモデル化するために用いられるべきである (Brutti et al., 2008)。

解析事前分布は, 文字どおり集積されたデータに対しモデルをあてはめる, 即ち解析するために用いることを想定し設定される事前分布である (Wang and Gelfand, 2002)。たとえば対照群の治療効果をパラメータ  $\theta_1$  で, 試験薬の治療効果をパラメータ  $\theta_2$  で表すモデルを考える。対照群については試験開始前にかんがりの情報を得ているが, 試験薬については対照群ほどの豊富な情報はもちあわせていないかもしれない。このように, 各モデルパラメータに対する事前分布の検討においては, それまでの経験則やヒストリカルデータに対する解析結果のレビューから始めるべきであり, それが解析事前分布を設定するときに役立つ (Berry et al, 2011)。一方, デザイン事前分布は "もし…だったとしたら" という見解に基づき設定される事前分布である (Wang and Gelfand, 2002)。即ち, 試験をデザインするときに用いられる事前分布であり, 過去に行われた試験の結果や文献調査により収集された情報をもとに検討される (Berry et al, 2011)。

上記手順でいえば、デザイン事前分布は手順2で、解析事前分布は手順3で、それぞれ使用する。手順2において、デザイン事前分布に基づきモデルパラメータのもとで*i*番目の被験者 ( $i = 1, 2, \dots, N$ ) に対する観測値を仮想的に生成させる。Type I エラーに完全に対応する "Bayesian analogs" を求めるためには、デザイン事前分布に従いパラメータを固定させて観測値を生成させるが、よりベイズ流に行うのであれば、デザイン事前分布からサンプリングされたモデルパラメータを用いて観測値を生成させることも考えられる。続く手順3にてパラメータを更新し  $\Delta$  の95%信用区間を推定する際に、解析事前分布を使用する。

頻度論におけるシミュレーションによる例数設計では、帰無仮説及び想定する対立仮説を決め、対立仮説のもとで仮想的にデータを生成させる。生成させたデータに対し治療効果に差があるかの仮説検定を行い、有意水準  $\alpha$  で検定が有意となるか判定する。この過程をシミュレーション回数  $N_{rep}$  繰り返し、有意になる割合を調べればこれが検出力の推定値となる。これに対し、ベイズ流のシミュレーションによる例数設計では、デザイン事前分布に基づき仮想的にデータを生成させる。生成されたデータに対し解析事前分布を用いてベイズ流に解析し、モデル性能基準と照らし合わせる。この過程をシミュレーション回数  $N_{rep}$  繰り返し、検出力を推定する。indifference zone を用いたベイズ流例数設計では、"モデル性能基準と照らし合わせる" ところが、 $\Delta$  と indifference zone との位置関係から6つの判定基準のいずれに該当するかを判定するステップに該当する。

### 3. 生存時間解析における適用例の紹介

#### 3.1. 試験デザインと解析方法

適用例として、生存時間解析における例数設計を紹介する (Berry et al, 2011)。対照群の5年生存率が0.65、試験群の5年生存率が0.80、対照群の例数：試験群の例数 = 1:1 の条件の下で、例数設計を行う場合を考える。生存時間分布に形状パラメータ  $\gamma$ 、尺度パラメータ  $\eta$  のワイブル分布

$$f(t) = \frac{\gamma}{\eta^\gamma} t^{\gamma-1} \exp\left(-\left(\frac{t}{\eta}\right)^\gamma\right) \quad (t > 0 \text{ のとき；それ以外は全て } 0) \quad (3.1)$$

を仮定する。ここで

$$\lambda = -\gamma \log \eta$$

とおくとき、生存関数  $S(t)$  及びハザード関数  $h(t)$  は  $\lambda$  を用いて  $S(t) = \exp(-\exp(\lambda)t^\gamma)$ 、 $h(t) = \gamma \exp(\lambda)t^{\gamma-1}$  となる。ここで治療群を表す共変量を  $x$  ( $x=0$  のとき対照群、 $x=1$  のとき試験群) として、共変量をモデルに組み込むために  $\lambda = -(\beta_0 + \beta_1 x)$  とおくと、対照群に対する試験群のハザード比は

$$HR = \frac{\gamma \exp(-\beta_0 - \beta_1) t^{\gamma-1}}{\gamma \exp(-\beta_0) t^{\gamma-1}} = e^{-\beta_1} \quad (3.2)$$

となり、 $\beta_1 > 0$  のとき  $HR < 1$  となり試験群のほうが対照群よりも死亡のリスクが低いことを表す。共変量をモデルに組み込む際、 $\lambda = \beta_0 + \beta_1 x$  とせず  $\lambda = -(\beta_0 + \beta_1 x)$  とおいたのは、そうすることで  $\beta_1$  の値が大きいほど試験群のほうが対照群よりも死亡のリスクが低い、つまり試験群のほうが対照群に比べて臨床上に望ましい状態であることを表すため、2群の効果の差を表す指標として  $\beta_1$  をとれば、 $\beta_1$  の95%信用区間と6つの判定基準との関係に図2.1がそのまま適用できるためである ( $\lambda = \beta_0 + \beta_1 x$  とおくと、 $\beta_1$  が大きいほど試験群のほうが対照群に比べて劣る状態を表すため、図2.1において treatment と control を反転させねばならなくなる)。対照群における時点  $t$  での生存率を  $S_1(t)$ 、試験群における時点  $t$  での生存率を  $S_2(t)$  と表すと、上記モデルのも

とでは

$$S_1(t) = \exp(-\exp(-\beta_0)t^\gamma)$$

$$S_2(t) = \exp(-\exp(-\beta_0 - \beta_1)t^\gamma)$$

により

$$\beta_0 = -\log\left(\frac{-\log S_1(t)}{t^\gamma}\right) \quad (3.3)$$

$$\beta_1 = \log\left(\frac{-\log S_1(t)}{t^\gamma}\right) - \log\left(\frac{-\log S_2(t)}{t^\gamma}\right) \quad (3.4)$$

である。

いま、 $i$  番目の被験者における、イベント発現時間あるいは打ち切り時間を  $t_i$ 、群を表す変数を  $x_{1i}$  (対照群であれば  $x_{1i} = 0$ 、試験群であれば  $x_{1i} = 1$ )、打ち切りを表す変数を  $v_i$  (イベント発現のとき  $v_i = 1$ 、打ち切りのとき  $v_i = 0$ ) と表すとき ( $i = 1, 2, \dots, n$ )、被験者の生存時間が互いに独立であるとして、 $n$  人分のデータ  $D$  が与えられたもとの尤度は、 $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ 、 $\mathbf{x}_i = (1, x_{1i})^T$  として

$$L(\boldsymbol{\beta}, \gamma | D) = \gamma^d \exp\left[\sum_{i=1}^n \left\{v_i(\mathbf{x}_i^T \boldsymbol{\beta}) + v_i(\gamma - 1) \log t_i - t_i^\gamma \exp(\mathbf{x}_i^T \boldsymbol{\beta})\right\}\right] \quad (3.5)$$

となる。ここに

$$d = \sum_{i=1}^n v_i$$

とおいた。モデルパラメータ  $\boldsymbol{\beta}$  の事前分布として平均ベクトル  $\boldsymbol{\mu}_0$ 、分散共分散行列  $\boldsymbol{\Sigma}_0$  の 2 変量正規分布を、ワイブル分布の形状パラメータ  $\gamma$  の事前分布として平均  $\alpha_0 / \kappa_0$ 、分散  $\alpha_0 / \kappa_0^2$  のガンマ分布をそれぞれとりあげると、 $\boldsymbol{\beta}$  と  $\gamma$  の同時事後分布は

$$p(\boldsymbol{\beta}, \gamma | D) \propto \gamma^{\alpha_0 + d - 1} \exp\left[\sum_{i=1}^n \left\{v_i(\mathbf{x}_i^T \boldsymbol{\beta}) + v_i(\gamma - 1) \log t_i - t_i^\gamma \exp(\mathbf{x}_i^T \boldsymbol{\beta})\right\} - \kappa_0 \gamma - (\boldsymbol{\beta} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) / 2\right]$$

で与えられる (Ibrahim et al., 2001)。 $\boldsymbol{\beta}$  の事後分布は閉じた形式をもたず解析的に求められないため、MCMC 法 (Markov Chain Monte Carlo method) を用いて事後分布を導出する。

以降では簡単のため、ワイブル分布(3.1)の形状パラメータが固定値  $\gamma = 2$  をとる、即ち生存時間分布がレイリー分布に従うものとする (Leemis and McQueston, 2008)。あてはめるモデルを決めたら、そのモデルに基づいて、対照群と試験群の効果の差を表すパラメータと、対応する indifference zone  $[\delta_L, \delta_U]$  を設定することになる。対照群と試験群の効果の差を表すパラメータとして  $\beta_1$  をとりあげると、 $\beta_1 < 0$  のとき  $HR > 1$  となり試験群のほうが対照群よりも死亡のリスクが高くなることから、indifference zone の下限値  $\delta_L$  を 0 とおく。対照群の 5 年生存率が 0.65、試験群の 5 年生存率が 0.80 のとき、(3.4)より

$$\beta_1 = \log(-\log S_1(5)/5^2) - \log(-\log S_2(5)/5^2) = \log(-\log(0.65)/5^2) - \log(-\log(0.80)/5^2) = 0.658$$

であることから、indifference zone の上限値  $\delta_U$  を 0.658 とおく。

ベイズ流アプローチによる例数設計では、前述のとおり、モデルパラメータの事前分布を設定しなければならない。Speigelhalter et al. (1994) は、事前分布の設定に関して、"reference prior"、"clinical prior" (臨床的事前分布)、"skeptical prior" (懐疑的事前分布)、"enthusiastic prior" (楽観的事前分布) の 4 種類から構成される "community of priors" の利用を推奨している。reference prior は「無情報」を客観的に表現した事前分布であり、いわゆる無情報事前分布 (non-informative prior) である。skeptical prior は試験薬に対して懐疑的な見解な立場での事前分布であり、より保守的な事前分布といえる。これに対して、enthusiastic prior は試験薬が対照

薬に比べて優れているとの信念に基づいた事前分布であり、臨床試験担当者にとって楽観的な事前分布といえる。本稿では、モデルパラメータ  $\beta_1$  の懐疑的事前分布として平均 0 の正規分布を用いる。事前分布として用いる正規分布の分散は、 $P(\beta_1 > \delta_U) = 0.05$  となるよう  $0.40^2$  と設定する。楽観的事前分布として平均 0.658、分散  $0.40^2$  の正規分布を、無情報事前分布として平均 0、分散  $10^4$  の正規分布をそれぞれ用いる。モデルパラメータ  $\beta_0$  は対照群の治療効果を規定する局外パラメータである。そこで、 $\beta_1$  が 3 つの事前分布いずれの場合においても平均  $-\log(-\log S_1(5)/5^2) = 4.061$ 、分散  $0.20^2$  の正規分布を用いる。以上の設定のもとで、3.2 の手順 2 から手順 5 を実行することで、Type I エラー、検出力に対する "Bayesian analogs" を求めることも可能となる。 $\beta_1$  のデザイン事前分布として懐疑的事前分布  $N(0, 0.40^2)$  を用いたとき、対照群を却下する経験的確率がベイズ流 Type I エラー (Bayesian Type I error) の推定値となる。また、 $\beta_1$  のデザイン事前分布として楽観的事前分布  $N(0.658, 0.40^2)$  を用いたとき、対照群を却下する経験的確率がベイズ流 Type II エラー (Bayesian Type II error) の推定値であり、試験群を却下する経験的確率がベイズ流検出力 (Bayesian power) の推定値となる。

### 3.2. SAS プログラムコード

プログラム 3.1 は、上記事前分布に基づき 1 群あたり 140 人の被験者の仮想的な観測値を生成させるための SAS プログラムの一例である。なお、プログラムの作成にあたっては、Minnesota 大学公衆衛生大学院の web サイト (<http://www.biostat.umn.edu/~brad/software/BRugs/>) にて公開されている BUGS コードを参考にした。少し補足すると、 $\beta_1$  に対してはデザイン事前分布として懐疑的事前分布、楽観的事前分布を記載し、楽観的事前分布が採用されるよう、懐疑的事前分布の設定に関するコードをコメントアウトしている。プログラム 3.1 を実行することにより生成されるデータセット SIM では、イベント発現または打ち切りの時間 (年) を表す変数が TIME、群を表す変数が X (対照群であれば 0、試験群であれば 1)、打ち切りを表す変数が CENSOR (打ち切りであれば 0、イベント発現であれば 1) である。レイリー分布に従う乱数 T を生成させ、 $T > 5$  であれば打ち切り例とみなして CENSOR に 0 を代入し、TIME に 5 と代入する。逆に  $T \leq 5$  であればイベント発現例とみなして CENSOR に 1 を代入し、TIME に T を代入している。

プログラム 3.1 : 擬似乱数による生存時間データの生成

```
data SIM ;
  gamma = 2 ;
  EST1 = 0.65 ; EST2 = 0.80 ;
  call streaminit(1234) ;
  do i = 1 to 140 ;
    do treatment = 1 to 2 ;
      x = treatment-1 ;
      beta0_mean = -log(-log(EST1))/(5**gamma) ;
      beta0 = RAND('NORMAL',beta0_mean,0.20) ;
      /*optimistic design prior*/
      beta1_mean = log(-log(EST1))/(5**gamma)-log(-log(EST2))/(5**gamma) ;
      beta1 = RAND('NORMAL',beta1_mean,0.40) ;
      /*skeptical design prior*/
```

```

/* beta1 = RAND('NORMAL',0,0.40);*/
lambda = -(beta0+beta1*x);
t = RAND('WEIBULL', gamma,(1/exp(lambda))**(1/gamma));

output;
end;
end;
run;
data SIM;set SIM;
keep i x time censor;
if round(t,1e-8)>5 then do; censor = 0; time = 5; end;
else if round(t,1e-8)<=5 then do; censor = 1; time = t; end;
run;

```

デザイン事前分布に基づき  $N$  人分の仮想的な観測値が生成できれば、それらを用いてベイズ流に解析し、2 群の効果の差  $\Delta$  (今回は  $\beta_1$  が該当する) の 95%信用区間を推定する。プログラム 3.2 は、プログラム 3.1 で生成された 1 組の仮想的な観測値を用いて  $\beta_1$  の 95%信用区間を推定するための SAS プログラムの一例である。 $\beta_1$  に対する事前分布として、無情報事前分布、懐疑的事前分布、楽観的事前分布を記載し、無情報事前分布が解析事前分布として採用されるよう、他の 2 つの事前分布の設定に関するコードをコメントアウトした。(3.5)より対数尤度関数が

$$\log L = \sum_{i=1}^n \left\{ v_i \log \gamma - (\beta_0 + \beta_1 x_i) v_i + v_i (\gamma - 1) \log t_i - t_i^\gamma \exp(-(\beta_0 + \beta_1 x_i)) \right\}$$

で表されることを用い、MCMC プロシジャにて対数尤度関数を直接記述している。得られた 10000 組のモンテカルロ標本を用いて  $\beta_1$  の 95%信用区間を推定した。今回のように、生存時間分布に標準的な分布をあてはめるのであれば、対数尤度関数を直接記述する以外に、SAS 関数 LOGPDF 及び LOGSDF を用いて対数尤度を記述することも可能である。その他 MCMC プロシジャの詳細については、SAS Institute Inc. (2015) を参照されたい。

プログラム 3.2 : MCMC プロシジャによる生存時間データの解析

```

proc mcmc data = SIM outpost = OUT nbi = 1000 nmc = 10000 seed = 1234;
gamma = 2; EST1 = 0.65; EST2 = 0.80;
beta0_mean = -log(-log(EST1)/(5**gamma));
parms beta0 beta1 0;
prior beta0~normal(beta0_mean,sd = 0.2);
/*optimistic fitting prior*/
/* beta1_mean = log(-log(EST1)/(5**gamma))-log(-log(EST2)/(5**gamma)); */
/* prior beta1~normal(beta1_mean,sd = 0.40); */
/*skeptical fitting prior*/
/* prior beta1~normal(0,sd = 0.40); */

```



```

/*reference fitting prior*/
  prior beta1~normal(0,sd = 1e+2);
  lambda = -(beta0+beta1*x);
  llike = censor*(log(gamma) + (gamma-1)*log(time) + lambda)
          - exp(lambda)*(time**gamma);
  model general(llike);
run;
proc univariate data = OUT noprint;
  var beta1; output out = OUT pctlpre = P_ pctlpts=2.5 97.5;
run;

```

2 群の効果の差  $\beta_1$  の 95%信用区間が推定されれば, indifference zone  $[\delta_L, \delta_U]$  との位置関係から, 図 2.1 に従い 6 つの判定基準のいずれに該当するかを判定することになる. プログラム 3.3 はその一例である. 変数 D1, D2, ..., D6 は, 対照群を採択, 試験群を棄却, 同等, 対照群を棄却, 試験群を採択, 判定不能に対する判定フラグであり 1 のとき該当, それ以外全て 0 をとる. 以上のステップをシミュレーション回数  $N_{rep}$  繰り返して得られる変数 D1, D2, ..., D6 の平均値が, 各判定基準の経験的確率となる.

プログラム 3.3 : 信用区間に基づく判定基準フラグの導出

```

data OUT; set OUT;
  gamma = 2; EST1 = 0.65; EST2 = 0.80;
  deltaL = 0;
  deltaU = log(-log(EST1)/(5**gamma))-log(-log(EST2)/(5**gamma));
  array a1[6] d1-d6;
  do i = 1 to 6;
    a1[i] = 0;
    if P_97_5 < deltaL then a1[1] = 1;
    else if p_97_5 < deltaU and p_97_5 > deltaL and p_2_5 < deltaL then a1[2] = 1;
    else if p_2_5 > deltaL and p_97_5 < deltaU then a1[3] = 1;
    else if p_2_5 > deltaL and p_2_5 < deltaU and p_97_5 > deltaU then a1[4] = 1;
    else if p_2_5 > deltaU then a1[5] = 1;
    else if p_2_5 < deltaL and p_97_5 > deltaU then a1[6] = 1;
  end;
run;

```

### 3.3. シミュレーション結果

表 3.1 は、デザイン事前分布に楽観的事前分布を、解析事前分布に無情報事前分布を、それぞれ用いたときのシミュレーション結果をまとめたものである。シミュレーション回数は 1000 回とした。デザイン事前分布に楽観的事前分布を用いる、即ち、楽観的事前分布に基づいて仮想データを生成するため、対照群に対する試験群の真のハザード比は(3.2)より 0.518 であり、「試験群を棄却」あるいは「対照群を採択」と判定された経験的確率を合算した値がベイズ流検出力の推定値となる。表 3.1 より、1 群あたりの例数  $N=80$  のときには、「試験群を棄却」と判定される割合は 0.63 であり、「試験群を棄却」よりもさらに強い「対照群を採択」と判定される割合も 0.01 となる。「試験群を棄却」あるいは「対照群を採択」と判定された割合  $0.63 + 0.01 = 0.64$  が、このシミュレーション結果から推定されるベイズ流検出力となる。デザイン事前分布に楽観的事前分布を設定し、対照群に比べて試験群が有効な状況であるにも関わらず、「対照群を棄却」と判定された割合は 0.02 であり、よってベイズ流 Type II エラーは 0.02 と推定される。例数の増加に伴ってベイズ流検出力も上がり、 $N=80$  のとき 0.64、 $N=100$  のとき 0.77、 $N=120$  のとき 0.82、 $N=140$  のとき 0.87、 $N=160$  のとき 0.90 となった。また、 $N=80$  のときは「判定不能」と判定された割合は 0.34 だったが、例数の増加に伴いその割合は減少していき、 $N=160$  のときには 0.07 であった。 $N$  が大きくなるにつれて  $\Delta$  の信用区間の幅が狭くなり、その結果、判定不能以外の判定結果となる割合が高くなっている。

表 3.1：無情報事前分布を解析事前分布として用いたときの 6 つの判定基準に対する経験的確率

$N$	Accept control	Reject treatment	Equivalence	Reject control	Accept treatment	No decision
80	0.00	0.02	0.00	0.63	0.01	0.34
100	0.00	0.02	0.00	0.76	0.01	0.22
120	0.00	0.02	0.00	0.81	0.01	0.16
140	0.00	0.03	0.00	0.86	0.01	0.11
160	0.00	0.03	0.00	0.89	0.01	0.07

表 3.2 は解析事前分布に懐疑的事前分布を用いたときの、表 3.3 は解析事前分布に楽観的事前分布を用いたときの、シミュレーション結果の要約である。シミュレーション条件は、解析事前分布の設定以外、全て表 3.1 のときと同じである。ベイズ流検出力は、 $N=80$  のとき懐疑的事前分布では 0.49、楽観的事前分布では 0.89、 $N=120$  のとき懐疑的事前分布では 0.64、楽観的事前分布では 0.95、 $N=140$  のとき懐疑的事前分布では 0.80、楽観的事前分布では 0.96、 $N=160$  のとき懐疑的事前分布では 0.86、楽観的事前分布では 0.97 となった。 $N=80, 100, 120, 140, 160$  いずれの場合も、3 つの事前分布の中では、楽観的事前分布を用いたときのベイズ流検出力が高く、懐疑的事前分布を用いたときのベイズ流検出力が低かった。事前分布の設定の仕方が必要な例数が異なることが確認できる。

表 3.2：懐疑的事前分布を解析事前分布として用いたときの 6 つの判定基準に対する経験的確率

$N$	Accept control	Reject treatment	Equivalence	Reject control	Accept treatment	No decision
80	0.00	0.11	0.00	0.49	0.00	0.40
100	0.00	0.12	0.00	0.64	0.00	0.25
120	0.00	0.11	0.00	0.73	0.00	0.17
140	0.00	0.10	0.00	0.80	0.00	0.10
160	0.00	0.10	0.00	0.86	0.00	0.04

表 3.3 : 楽観的事前分布を解析事前分布として用いたときの 6 つの判定基準に対する経験的確率

$N$	Accept control	Reject treatment	Equivalence	Reject control	Accept treatment	No decision
80	0.00	0.01	0.00	0.89	0.00	0.11
100	0.00	0.01	0.00	0.92	0.00	0.07
120	0.00	0.02	0.00	0.95	0.00	0.03
140	0.00	0.01	0.00	0.96	0.00	0.03
160	0.00	0.02	0.00	0.97	0.00	0.01

## 4. まとめ

本稿では、シミュレーションによるベイズ流例数設計として *indifference zone* を用いた方法についてまとめ、生存時間解析における適用例を紹介した。適用例では、一度に全患者が登録され、登録された患者のフォローアップ期間は全て等しい状況を想定している。しかし実際には、患者の登録は逐次的であり、フォローアップ期間も全ての患者で等しいとは限らない。より現実の臨床試験で起こりうる状況に応じた例数設計が求められる（魚住等, 2016）。

全ての試験に共通する困難な問題は、例数の計算に用いる効果と変動の大きさの設定であり、このための考え方には、a) 臨床的に意味のある最小の差を定めること、b) 他の試験データによるエフェクトサイズの推定、のように 2 通りある（上坂, 2006）。*indifference zone* を用いた方法では、臨床的に意味のある最小の差を定めることが、*indifference zone* の設定に対応しているといえる。b) の考え方は、先行する試験の結果を利用することが可能な場合に有用である。その際、今回紹介した方法であれば、先行する試験のデータをベイズ流に解析し、得られたモデルパラメータの事後分布を用いてデザイン事前分布を設定し、当該試験における必要例数を検討することができる。効果と変動の大きさの設定にあたり、参考可能な試験は存在するが試験薬に関する情報はない、文献上での情報であるため、詳細なデータは入手できず、試験の結果しかわからない状況もおこりうる。Hobbs and Carlin (2008) は、このような場合への適用例もあわせて紹介している。ヒストリカルデータを、そのデータがもつ情報量に応じて新規試験におけるデザインと解析内容に柔軟に組み入れることが可能なことが、ベイズ流例数設計の大きな魅力ではないだろうか。

## 参考文献

- [1] Berry, S.M., Carlin, B.P., Lee, J.J. and Müller, P. (2011). *Bayesian Adaptive Methods for Clinical Trials*. New York: Chapman & Hall/CRC.
- [2] Brutti, P., De Santis, F. and Gubbiotti, S. (2008). Robust Bayesian sample size determination in clinical trials. *Statistics in Medicine*. **27**(13):2290-2306.
- [3] Hobbs, B.P. and Carlin, B.P. (2008). Practical Bayesian design and analysis for drug and device clinical trials. *Journal Biopharmaceutical Statistics*. **18**:54-80.
- [4] Ibrahim, J.G., Chen, M.H., Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York.
- [5] Leemis, L.M. and McQueston, J.T. (2008). Univariate Distribution Relationships. *The American Statistician*. **62**: 45-53.
- [6] SAS Institute Inc. (2015). *SAS/STAT(R) 14.1 User's Guide*. Cary, NC, USA: SAS Institute Inc.

- [7] Spiegelhalter, D. J., Freedman, L. S., Parmar, M. K. B. (1994). Bayesian approaches to randomized trials (with discussion). *Journal of Royal Statistical Society: Series A*. **157**:357-416.
- [8] Wang, F. and Gelfand, A. E. (2002). A Simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science*. **17**(2):195-208.
- [9] 上坂浩之(2006). 医薬品開発のための臨床試験の計画と解析. 朝倉書店.
- [10] 魚住龍史・矢田真城・浜田知久馬 (2016). SAS プロシジャを用いた生存時間データに対する例数設計の变革. SAS ユーザー総会論文集.
- [11] 大橋靖雄・浜田知久馬 (1995). 生存時間解析 -SASによる生物統計. 東京大学出版会.
- [12] 大橋靖雄・浜田知久馬・魚住龍史 (2016). 生存時間解析 応用編 -SASによる生物統計. 東京大学出版会.
- [13] 中村剛 (2001). Cox 比例ハザードモデル. 朝倉書店.

## 連絡先

E-mail: [yada-s@a2healthcare.com](mailto:yada-s@a2healthcare.com)

生物統計・医薬関連／SASシステム

# SASにおける文芸的プログラミングへの試み

○吹谷 芳博<sup>1</sup>

(<sup>1</sup>株式会社エスアールディ)

## Attempt to Literate Programming Using SAS

Yoshihiro Fukiya

SRD Co., Ltd.

## 要旨

プログラミング開発においてプログラム仕様書等の文書作成は重要とされてきている。その主な理由として、プログラム品質の担保や文書を作成することでプログラム開発時の相互理解を高めるなどがあげられる。しかし、その一方で文書作成はソースコードとは分けて作成するため、その分の負担をかけていることも事実である。

そこで著者はこの問題の解決方法の一つとして「文芸的プログラミング」を SAS に適用させることを検討した。本発表は R 言語のパッケージである「knitr」を利用して、文芸的プログラミングを実装した事例を紹介する。この事例を通して実装方法や実装したことによる利点・欠点等の検討事項について報告する。

キーワード：文芸的プログラミング， markdown， knitr， LaTeX

## 1. はじめに

### 1-1 背景と目的

医薬品業界に限らず、プログラミング開発においてソースコードの作成・運用の効率化については議論がなされるところである。著者はソースコードの運用を効率化するために GitLab を用いたバージョン管理システムについて紹介を昨年 SAS ユーザー総会 2015 で報告した。しかし、前回の発表では運用周りを中心とした内容であり、プログラムの作成の部分における効率化についてあまり焦点にしていなかった。今回の試みとしてソースコードとその関連文書の作成の効率化にスポットを当てた内容を報告する。

プログラムの作成に関してはプログラム仕様書やその他の関連文書の作成も必要となるため、具体的にはプログラム仕様書などの関連文書の作成は重要とされてきている。その理由として、プログラム品質の担保するため、プログラム開発チームソースコードの共有理解を深めることがあげられる。しかし、現場のプログラマーとしてはソースコードと関連文書の整合性を保ちつつ、短期間での作成はとても負担がかかり、また運用面でも同様のことが言える。そこで Knuth<sup>3)</sup> (TeX の作者) が提唱したプログラミング方法論の一つである「文芸的プログラミング」はこれらの問題を解決する一つのアイデアである著者は考えた。この文芸的プログラミングとはソースコードと文書を併記したメタソースを記述し、そこからソースコードと文書をそれぞれ生成させることで、情報の一体性を高めている。

R 言語は文芸的プログラミングを比較的好く使われている言語のひとつである。R 言語はこれまで Sweave<sup>5)</sup>(2002)や knitr<sup>6)</sup> (2012)などが文芸的プログラミングを実装するパッケージが開発されてきた。現在、R 言語の文芸的プログラミングといえば knitr と言われるほど多くの人に使用されてきている。その理由として Sweave と knitr の大きな違いは出力形式の豊富さ、メタソースの記載しやすさと拡張性の高さである。具体的には Sweave は html 形式や pdf 形式への出力が主な形式であるが、knitr は Sweave ができる出力形式にプラスして word 形式やウェブスライド形式など多岐にわたる。また、メタソースの文書部分は Sweave が LaTeX 形式で記載するのに対して knitr は LaTeX 形式よりも理解しやすい markdown 形式を使って記載している。最後の違いとして knitr は R 言語だけでなく、python や ruby などの言語の文芸的プログラミングにも対応しており、注目もされている。一方、SAS における文芸的プログラミングについては SASweave<sup>4)</sup>(2007)や StatRep<sup>1)</sup>(2012)が開発されている。しかし、主な機能として Sweave とほぼ同様な機能であり、また通常の業務では pdf 形式より word 形式を使用することが多いため、SAS で knitr のようなツールがないか検討を行っていた。

これらの経緯から著者は knitr の拡張機能で SAS の文芸的プログラミングが可能ではないかと考え調査したところ、Doug<sup>2)</sup>が R markdown と SAS を用いて html 形式への変換を行っている事例を見つけ、これを word 形式で実用的なレベルまでに拡張できないか検討を行った。具体的には文章変換後に数式の表示、リストの表示及び図の表示の3つを評価ポイントとし、html 形式と word 形式を比較し、その結果について考察も行った。また、変換時にうまくいかなかった点については改善策を行い、改善後の結果についても考察を行う。

## 1-2 用語説明

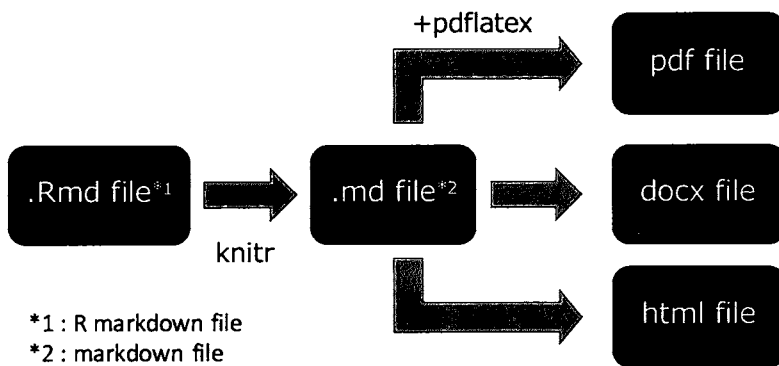
- R : オープンソースの統計解析プログラムツール
- LaTeX : テキストベースの組版処理システム
- Rstudio : R の統合開発環境 (オープンソース)
- Pandoc : 文書変換ツールで `haskell` というプログラムで作成されている。  
(Rstudio インストール時に `pandoc` もインストールされるため、セットアップ不要。)
- markdown (.md) : 文書を記述するための軽量マークアップ言語のひとつ。
- R markdown (.Rmd) : markdown を R 用に拡張させたファイル。
- chunk : R markdown でソースコードの実行や結果の表示などを制御するために設定するもの。

## 2.変換フローの概要と実行方法について

### 2.1 変換フローの概要

#### 2.1.1 変換フロー説明

knitr は R markdown 形式を変換することで、多様な形式に出力することができる (Figure 1)。具体的には knitr は chunk で指定されたソースコードを実行しつつ、R markdown 形式から markdown 形式に変換し、その後 pandoc による変換が行われる。Rstudio の GUI に knitr を実行するボタンが実装されているが今回そのボタンを使用せずにコードで出力形式や指定を細かく調整できる `rmarkdown` パッケージを使って実行を行った。



S

Figure 1. 文書変換フローの模式図

### 2.2.1 R markdown の構造と chunk の指定について (R と SAS の違い)

◇ R markdown の構造は以下の大きく 3 つに分かれる (Figure 2) .

- 1) 文書プロパティや出力形式の調節を記載するメタ情報 (yaml 形式)
- 2) 文書部分 (markdown 形式)
- 3) ソースコードを記載する部分

この構造は SAS を用いた時も大きく変わらず、変更する部分は code block に SAS コードを記載することと、このあとに説明する chunk 内にあるオプションの設定である。

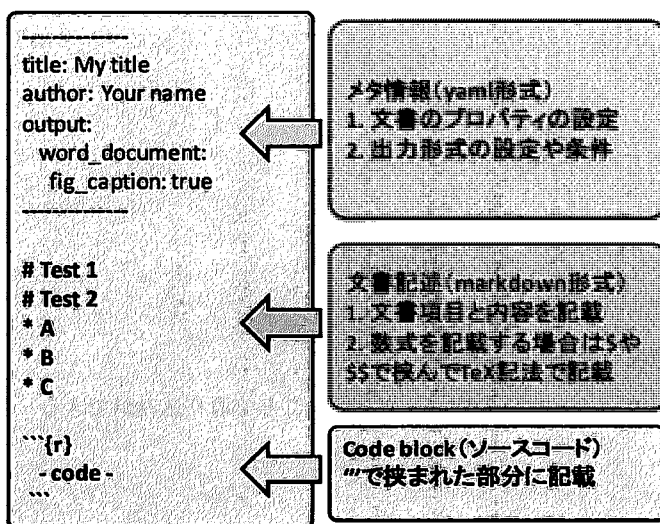


Figure 2. R markdown 構成の模式図



## ◇ chunk の指定

chunk の部分にプログラム実行の有無や結果表示の制御するパラメータがあり、必要に応じて調整する (Figure 3) . 何も変更がないときはすべてデフォルトで動作するため、注意が必要である. SAS の場合は実行ファイル (.exe) 先とコマンドプロンプトなどで実行する際の log 出力及び lst 出力先の設定も必要であるため, Figure 3.で記載した設定が必要である. "opts\_chunk\$set"変数を事前に設定すれば, code block どの path を設定しなくても動作するようになる.

```
# R言語の時のChunkの使い方.
```{r}
-r code -
```

# SASの時のChunkの使い方.
```{r, message=FALSE}
saspath <- "C:/.../sas.exe"
sasopts <- "-nosplash -log 'C:/.../' -print 'C:/.../' -ls 75"
opts_chunk$set(engine="sas", engine.path=saspath, engine.opts=sasopts,
comment="", results="hide")
...

```{r}
-sas code -
```
```

Figure 3. chunk の使い方の模式図

## 2.2 実行方法

### 2.2.1 実行環境の整備

実行環境については以下のソフトとパッケージが必要となる.

1. R
2. Rstudio
3. Library (knitr, rmarkdown)

この順番にインストールを行えば, 実行環境の設定は完了する.

パッケージのインストールの場合は"install.packages('rmarkdown')"と示すことで取り込みができる.

## 2.2.2 文書変換の評価

Doug が公開している"SAS using R Markdown"を基に以下の3つについて word 形式で出力できるように実装し、評価を行った。比較対象として Doug が行っている html 形式への出力も行った。出力内容については「実用 SAS 生物統計ハンドブック」<sup>8)</sup>の生存時間解析のページを参考にした。また、word 形式で出力した場合にうまく行かなかった部分については原因も探り、改善策の検討も行った。

- 1) 数式 (TeX 記法)
- 2) リストの表示
- 3) 図の表示

## 3.評価 (結果)

### 3.1 文書変換の評価

実装結果については Table 1.に示した。

Table 1. 従来の方法で word 形式に出力結果

|             | html 形式 | word 形式 |
|-------------|---------|---------|
| 数式 (Tex 記法) | ○       | ○*1     |
| リストの表示      | ○       | △       |
| 図の表示        | ○       | ×       |

\*1 : office のバージョンにより文字化けが発生 (office 2007)。

今回の結果では Doug が示した方法で word 形式の出力に適応させると数式以外はそのままで方法では適応が難しいということが示された。個別に結果を見るとリストの表示についてはリストのみの表示をする設定にすれば問題なく list は読み込まれて code block 内で表示されるが、図も出力する設定の場合は設定の仕方によってはうまく取り込まれなかったため表示がされなかった。また、図についても同様に自動的に取り込まれなかった。

### 3.2 問題点と改善策

Doug が示した方法は全体の出力結果を html 形式として出力し、文書部分の html 形式に出力結果の html 形式を取り込むことで表示をしていたこと原因であった。word 形式では html 形式を取り込んで表示できないため、以下の方法で対応を行った。

リストの表示の改善方法

1. chunk 内のリスト出力先の設定をフォルダではなく、特定のファイル名にする。
2. 表示したい部分に新たに chunk を記載し、リストを取り込む設定を行う。

例) ````\{r eval=FALSE, code=readLines("C:/.../rmdput.lst")}\`````

## 図の表示の改善方法

1. 図の出力を直下のフォルダではなく、特定のフォルダ先を指定するようにする。
2. 図を表示させたい部分に markdown 形式の記載で画像を読み込む設定を行う。

例) ![結果](C:/...../Kaplan-Meier.png "Kaplan Meier")

以上の方法を行うことにより html 形式で表示した内容とほとんど変わらない形で表示することができた。詳細の結果については画像で APPENDEX に表示する。

## 4. 結論と考察

### 4.1. 文書変換の評価についての考察

knitr を用いた文芸プログラミングを SAS でも可能であることを示し、出力形式についても html だけでなく実用性のある word への変換も可能であると示した。word 形式への変換が実用的に使えるとなれば SAS macro の仕様書等を効率よく作成できるのではないかと考えている。

今回は PC SAS で行っているため、UNIX SAS での検証は行っていない。しかし、UNIX 環境に同様の環境を作ることによって実行は可能だと思われる。今後は word 形式以外にもウェブスライド形式や pdf 形式への変換等も行い検証していきたい。

### 4.2 今後の検討課題

今回の試みは今後の SAS における文芸的プログラミングの発展やこの手法を応用した Reproducible Research<sup>7)</sup>への貢献にもつながったのではないかと考える。今後の文芸的プログラミングが広まるためには SAS university edition でも実装可能であることが望まれるが、今回のバッチを使う方法は SAS university edition はバッチ実行できないため適用できない。SAS university edition でも今回のような文書変換フローが実装できればと考えている。また、SAS において pandoc を使用した文書変換だと syntax highlighting が対応していないため、対応できるようにしたいと考えている。

## 5.参考文献

- 1) Arnold, T and Kuhfeld, WF (2012) “Using SAS and LaTeX to Create Documents with Reproducible.”
- 2) Doug, H (2015) “SAS using R markdown”  
URL:<http://www.ssc.wisc.edu/~hemken/SASworkshops/SASWindows/Markdown/SASmarkdown.html>
- 3) Knuth, Donald E (1984) “Literate programming.” The Computer Journal (1984) 27 (2): 97-111.
- 4) Lenth RV, Højsgaard S (2007) “SASweave: Literate Programming Using SAS.” Journal of Statistical Software, 19(8).  
15 Page 16.
- 5) Leisch, F (2002) “Sweave: Dynamic Generation of Statistical Reports Using Literate Data Analysis.” Compstat pp  
575-580
- 6) Xie, Y (2012) “An Introduction to knitr.”
- 7) Xie, Y (2014) “knitr: a comprehensive tool for reproducible research in R.” Implement Reprod Res.
- 8) 浜田 知久馬 (2005) “実用 SAS 生物統計ハンドブック” サイエンス社, p172-177

## 6.APPENDIX

Figure 4. Run.R (rmarkdown による文書変換プログラム)

```
1 library(knitr)
2 library(rmarkdown)
3 render("rmd2word.Rmd", "word_document") # MS Wordレポート
```

Figure 5. rmd2word.Rmd (1/2)

```
1 ▾ ---
2 title: 生存時間解析 (SAS)
3 author: "@Yoshihiro Fukiya"
4 output:
5   word_document:
6     fig_width: 2
7     fig_height: 1.5
8     fig_caption: true
9     pandoc_args: [
10      "-s", "--mathml"
11    ]
12 ▸ ---
13
14 ▾ ### 【統計手法】
15 ▾ #### Kaplan-Meier法
16
17 $$ \hat{S}(t) = (1 - d_{1/n_1}) \times (1 - d_{2/n_2}) \times \dots \times \prod_{t_i < t} \left\{ 1 - \frac{d_i}{n_i} \right\} $$
18
19 * $t_i$: イベント (死亡) が起こった時点.
20 * $d_i$: 時点 $t_i$ での死亡数.
21 * $n_i$: 時点 $t_i$ 直前のリスク集合の大きさ.
22
23 ▾ ##### 生存関数の標準誤差はGreenwoodの公式により計算される.
24
25 $$ SE\left(\hat{S}(t)\right) \approx \hat{S}(t) \cdot \sqrt{\sum_{t_i < t} \frac{d_i}{n_i} \cdot (n_i - d_i)} $$
26
27 ▾ ##### 時点ごとのそれぞれの推定値 $\hat{S}(t)$ に対する正規近似した $1 - \alpha / 2$ 信頼区間.
28
29 $$ \text{信頼上限: } UCL = \hat{S}(t) + z_{\alpha/2} \cdot SE\left(\hat{S}(t)\right) $$
30 $$ \text{信頼下限: } LCL = \hat{S}(t) - z_{\alpha/2} \cdot SE\left(\hat{S}(t)\right) $$
31
32 ▾ ### 【データセットの構造】
33 * データセット名: KM
34 * 群分け変数: TRT
35   - 対照薬群(0)
36   - 治療薬群(1)
37 * 時間変数: TIME
38 * 打ち切り変数: CENSOR
39   - 打ち切り(0)
40   - イベント(1)
41
```

Figure 6 rmd2word.Rmd (2/2)

```

41
42 - ### 【SASプログラム】
43 - ```{r, message=FALSE, echo=FALSE}
44 saspath <- "C:/Program Files/SAS/SASFoundation/9.2/sas.exe"
45 sasopts <- "-nosplash -log 'C:/Users/y-fukiya/Documents/Markdown/rmdlog.log' -print
46 'C:/Users/y-fukiya/Documents/Markdown/rmdput.lst' -ls 75"
47 opts_chunk$set(engine="sas", engine.path=saspath, engine.opts=sasopts, comment="", results="hide")
48
49 - ```{r data1}
50 * ----- 出力設定;
51 options linesize=80 pagesize=66 nodate;
52
53 * ----- フォーマット作成;
54 proc format;
55     value TRTF      0='CONTROL'
56                   1='NEWDRUG2';
57     value CENSORF  0='CENSORED'
58                   1='EVENT';
59 run;
60
61 * ----- サンプルデータ作成;
62 data KM;
63     do TRT=0,1;
64         do i=1 to 16;
65             input TIME CENSOR @@;output;
66         end;end;
67     format TRT TRTF. CENSOR CENSORF.;
68     keep TRT TIME CENSOR;
69 cards;
70
71
72     4 1 5 1 7 1 7 1 12 1 14 0 17 1 20 1
73     22 1 31 1 33 0 34 1 43 1 46 0 50 1 52 1
74     9 1 12 1 14 0 20 1 20 1 25 1 26 1 31 0
75     34 0 41 1 44 0 48 1 55 1 60 1 61 0 77 0
76 ;
77 run;
78
79 proc print data=KM;
80 run;
81
82 * ----- 出力設定;
83 ods graphics / reset=index imagename="Kaplan-Meier" imagefmt=png ;
84 ods html close;
85 ods listing gpath="C:/Users/y-fukiya/Documents/Markdown";
86
87 ods select survivalplot;
88 proc lifetest data=KM plots=s;
89     time TIME*CENSOR(0);
90     strata TRT;
91 run;
92
93 ods listing close;
94 ods graphics / reset=all;
95 ods listing;
96 - ```
97
98 - ```{r eval=FALSE, code=readLines("C:/Users/y-fukiya/Documents/Markdown/rmdput.lst")}
99 - ```
100
101 ! [proc lifetestの結果](C:/Users/y-fukiya/Documents/Markdown/Kaplan-Meier.png "Kaplan Meier")

```

## 生存時間解析 (SAS)

@Yoshihiro Fukiya

### 【統計手法】

#### Kaplan-Meier 法

$$\hat{S}(t) = (1 - d_1/n_1) \times (1 - d_2/n_2) \times \dots = \prod_{t_i < t} (1 - d_i/n_i)$$

- $t_i$  : イベント (死亡) が起こった時点.
- $d_i$  : 時点  $t_i$  での死亡数.
- $n_i$  : 時点  $t_i$  直前のリスク集合の大きさ.

生存関数の標準誤差は Greenwood の公式により計算される.

$$SE(\hat{S}(t)) \approx \hat{S}(t) \cdot \sqrt{\sum_{t_i < t} \frac{d_i}{n_i \cdot (n_i - d_i)}}$$

時点ごとのそれぞれの推定値  $\hat{S}(t)$  に対する正規近似した  $1 - \alpha/2$  信頼区間.

$$\text{信頼上限 : } UCL = \hat{S}(t) + z_{\alpha/2} \cdot SE(\hat{S}(t))$$

$$\text{信頼下限 : } LCL = \hat{S}(t) - z_{\alpha/2} \cdot SE(\hat{S}(t))$$

### 【データセットの構造】

- データセット名 : KM
- 群分け変数 : TRT
  - 対照薬群(0)
  - 治療薬群(1)
- 時間変数 : TIME
- 打ち切り変数 : CENSOR
  - 打ち切り(0)
  - イベント(1)

Figure 8. rmd2word.docx (2/4)

```
【SAS プログラム】
* ----- 出力設定;
options linesize=80 pagesize=66 nodate;

* ----- フォーマット作成;
proc format;
  value TRTF      0='CONTROL'
                 1='NEWDRUG2';
  value CENSORF  0='CENSORED'
                 1='EVENT';
run;

* ----- サンプルデータ作成;
data KM;
  do TRT=0,1;
    do i=1 to 16;
      input TIME CENSOR @@;output;
    end;end;
  format TRT TRTF. CENSOR CENSORF.;
  keep TRT TIME CENSOR;
cards;

  4 1 5 1 7 1 7 1 12 1 14 0 17 1 20 1
  22 1 31 1 33 0 34 1 43 1 46 0 50 1 52 1
  9 1 12 1 14 0 20 1 20 1 25 1 26 1 31 0
  34 0 41 1 44 0 48 1 55 1 60 1 61 0 77 0
;
run;

proc print data=KM;
run;

* ----- 出力設定;
ods graphics / reset=index imagename="Kaplan-Meier" imagefmt=png ;
ods html close;
ods listing gpath="C:/Users/y-fukiya/Documents/Markdown";

ods select survivalplot;
proc lifetest data=KM plots=s;
  time TIME*CENSOR(0);
  strata TRT;
run;

ods listing close;
```

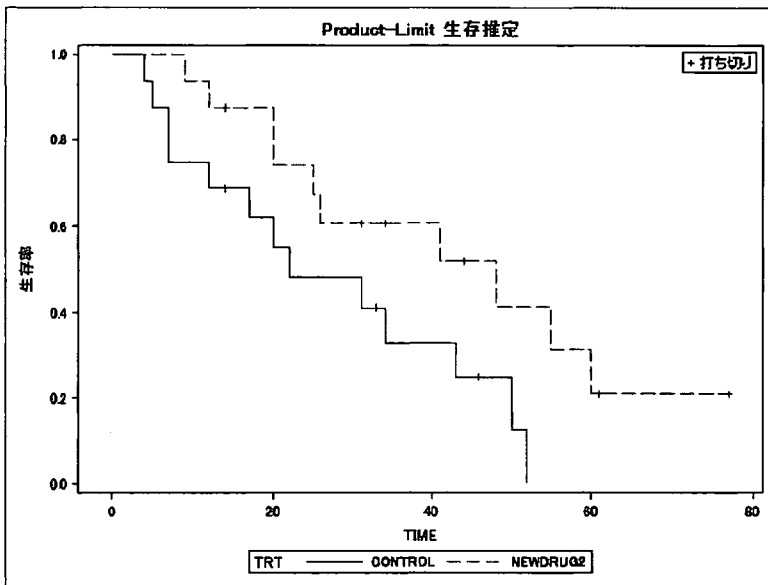


Figure 9. rmd2word.docx (3/4)

```
ods graphics / reset=all;  
ods listing;
```

| OBS | TRT      | TIME | CENSOR   |
|-----|----------|------|----------|
| 1   | CONTROL  | 4    | EVENT    |
| 2   | CONTROL  | 5    | EVENT    |
| 3   | CONTROL  | 7    | EVENT    |
| 4   | CONTROL  | 7    | EVENT    |
| 5   | CONTROL  | 12   | EVENT    |
| 6   | CONTROL  | 14   | CENSORED |
| 7   | CONTROL  | 17   | EVENT    |
| 8   | CONTROL  | 20   | EVENT    |
| 9   | CONTROL  | 22   | EVENT    |
| 10  | CONTROL  | 31   | EVENT    |
| 11  | CONTROL  | 33   | CENSORED |
| 12  | CONTROL  | 34   | EVENT    |
| 13  | CONTROL  | 43   | EVENT    |
| 14  | CONTROL  | 46   | CENSORED |
| 15  | CONTROL  | 50   | EVENT    |
| 16  | CONTROL  | 52   | EVENT    |
| 17  | NEWDRUG2 | 9    | EVENT    |
| 18  | NEWDRUG2 | 12   | EVENT    |
| 19  | NEWDRUG2 | 14   | CENSORED |
| 20  | NEWDRUG2 | 20   | EVENT    |
| 21  | NEWDRUG2 | 20   | EVENT    |
| 22  | NEWDRUG2 | 25   | EVENT    |
| 23  | NEWDRUG2 | 26   | EVENT    |
| 24  | NEWDRUG2 | 31   | CENSORED |
| 25  | NEWDRUG2 | 34   | CENSORED |
| 26  | NEWDRUG2 | 41   | EVENT    |
| 27  | NEWDRUG2 | 44   | CENSORED |
| 28  | NEWDRUG2 | 48   | EVENT    |
| 29  | NEWDRUG2 | 55   | EVENT    |
| 30  | NEWDRUG2 | 60   | EVENT    |
| 31  | NEWDRUG2 | 61   | CENSORED |
| 32  | NEWDRUG2 | 77   | CENSORED |

Figure 10. rmd2word.docx (4/4)



*proc lifetest* の結果

## 医療・健診データに対するSAS Visual Analytics/Visual Statistics の適用可能性の検討

○大津 洋<sup>1</sup>、大橋 靖雄<sup>2</sup>

(<sup>1</sup>国立国際医療研究センター臨床研究センターデータサイエンス部、

<sup>2</sup>中央大学理工学部人間総合理工学科)

医療ビックデータの時代となり、疾患予防や健康指導・疾患と治療の予後調査などの場にデータを強力に活用していこうという動きが強まってきている。

医療データの臨床研究活用においては、使うデータを予め決めることで、その集団における基礎統計量を用いることで、推定することが主であった。しかしながら、長寿高齢化が最も進むわが国において、ひとつの疾患で死に至るということではなく、多くのエピソードによって構成されることが増え、そのことが、「Cardio-oncology」や「糖尿病とうつ病」といった新たな研究領域を生むことになっている。また、医療情報を提供することになる個人においても、集団の評価も重要であるが、「自分の場合ではどうなのか？」という似た個人での情報をフィードバックし、最適化されたサービス・特徴づけを期待する傾向にある。

このように、我々はできるだけ多くのデータを取り扱いながら、集団と個人にバランスよく結果をフィードバックさせて、迅速な意思決定に活かせることが必要になる。特に、自治体による健康指導というのは、基礎自治体という単位で、長期にデータを活用することで、より適切な指導に活かせる可能性が高い分野だと考える。

近年のBIツールにおいては、上記のことが柔軟に対応できることを明記しているものが多いが、提供側の情報量が多く、実際に行っている側のニーズに対応できるものであるのか？ということについて、十分な情報量がないことから、その適用可能性を試行することとした。

実際に自治体などで利用する可能性がある人達にニーズ調査を行った結果を、SAS Visual Statistics/Visual Analysis (以下、SAS VA/VS)にどこまで反映できるかを検討し、必要な情報に到達できるかということで、活用可能性を評価することとした。

### 現状

ニーズ調査を反映したシステムが SAS VA/VSにどこまで対応できたかは、現在進行中のため、当日のプレゼンテーションにて発表する。

チュートリアル

SASによる生存時間分布の予測「Death Noteの統計学」

○浜田知久馬 魚住龍史  
(東京理科大学)

医薬研究では生存時間解析のモデルとして比例ハザードモデルが標準的に用いられる。この方法では共変量の効果をハザード比を用いて表すが、予後を評価する上では特定の共変量を持つ患者の生存時間分布を予測することがより重要である。比例ハザードモデルで解析を行うためのPHREGプロシジャではBASELINE文を用いて、セミパラメトリックモデルの枠組みの中で、任意の時点における生存率を予測することが可能である。LIFETESTプロシジャではKaplan-Meier法に基づいて、ノンパラメトリックに生存関数を推定することができる。更にLIFEREGプロシジャでは、ワイブル分布、対数ロジスティック分布等のパラメトリックな生存時間分布を仮定して、生存時間を予測することができる。また分位点回帰を行うQUANTREGプロシジャを打ち切りがある場合に拡張したQUANTLIFEプロシジャがバージョン9.4から新たに追加された。このプロシジャを利用するとメディアン等の分位点と共変量の関係をモデル化することが出来る。

本発表では、チュートリアルとしてLIFETESTプロシジャ、PHREGプロシジャ、LIFEREGプロシジャ、QUANTLIFEプロシジャを用いた生存関数の予測について解説する。

表1 生存時間分布の予測のためのプロシジャの比較

| プロシジャ     | 焦点       | モデル            | アプローチ | 共変量   | 打ち切り          |
|-----------|----------|----------------|-------|-------|---------------|
| LIFETEST  | 生存関数     |                | ノンパラ  | 層別    | 右側            |
| PHREG     | ハザード関数   | 比例ハザードモデル      | セミパラ  | モデル調整 | 右側            |
| LIFEREG   | 生存時間     | 加速モデル          | パラ    | モデル調整 | 右側,<br>左側, 区間 |
| QUANTLIFE | 生存時間の分位点 | 加速モデル<br>加法モデル | セミパラ* | モデル調整 | 右側            |

\*QUANTLIFE：生存時間分布に特定の分布を仮定しないが、共変量で%点の関係をモデル化

# データハンドリングにおけるSQLプロシジャの利活用

## —PROC SQL入門—

古隅 弘樹

(兵庫県立大学 経済学部)

Effective Use of SQL Procedure for Data Handling —Introduction to Proc SQL—

Hiroki Furuzumi

School of Economics, University of Hyogo

### 要旨

データセットの作成や整備といったデータハンドリングに関わる部分でSQLプロシジャは威力を発揮する。データハンドリングを主にデータステップで行ってきた利用者にとってはプログラミングの幅が増えることになり、リレーショナル・データベースのシステム開発経験者にとっては馴染みのあるSQL命令がSASシステムにおいても使用できるメリットは大きい。SASシステムにおけるSQLの実装がProc SQLであり、その使用例を交えた初心者向けのチュートリアルとして報告をおこなう。

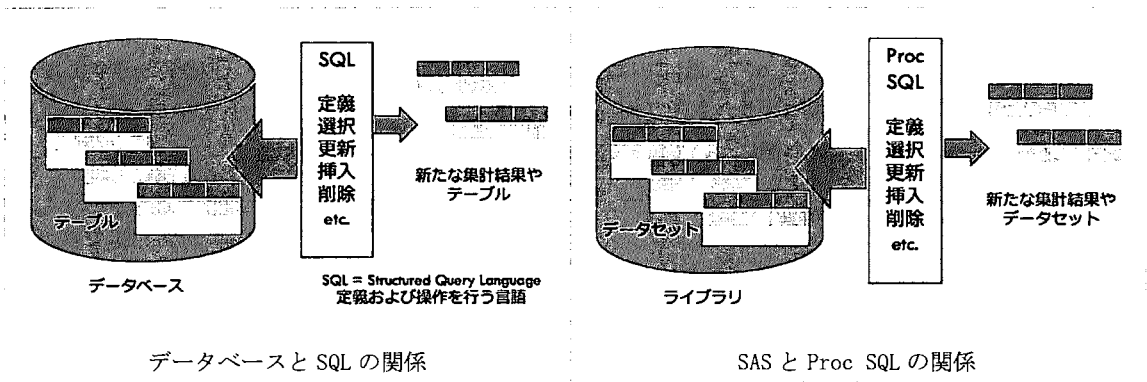
キーワード： データハンドリング、Proc SQL、データステップ、リレーショナル・データベース

### 1. はじめに

リレーショナル・データベースに関してシステム開発の経験がある者にとって、SQLはとても身近な言語である。SASシステムにおけるデータ管理は、データベースと同様に取り扱うことができ、SASでSQLが利用できるのはその恩恵によるものである。分析処理を行うには、まずしっかりとデータの整備が不可欠であり、そのためのデータハンドリングにおいて、データステップのみならず、SQLを利用できるメリットは大きい。今回の報告では、SQLの利用に不慣れな利用者向けのチュートリアルとして、Proc SQLの使い方の基礎から、応用に向けての便利な機能を少し紹介したい。

### 2. データベースとSQL

データベースではデータの定義や操作のための命令体系が必要になる。前者をDDL (Data Definition Language: データ定義言語)、後者をDML (Data Manipulation Language: データ操作言語)と呼ぶが、リレーショナル・データベースにおいてこれらの両方を実装した標準言語がSQL (Structured Query Language)である。



データベースの中にデータを格納したテーブルが複数あり、それらの編集や、集計、新たなテーブルを作成するといった操作を行うのが SQL である。SAS システムの場合では、ライブラリの中にデータセットがあるイメージになる。ライブラリはデータセットの保存先であり、ファイルシステムにおけるディレクトリ（フォルダ）に当たる。

関連用語の対応（「SAS 9.4 SQL プロシジャ：ユーザーガイド」より）

| SAS      | SQL  | データ処理 |
|----------|------|-------|
| データファイル  | テーブル | ファイル  |
| オブザベーション | 行    | レコード  |
| 変数       | 列    | フィールド |

厳密には SAS データセットには、SAS データファイルと SAS データビューの 2 種類があり、SAS データファイルとは、データ値に加え、データに関連づけられたディスクリプタ情報（変数の型や長さ、エンジン名など）を含んだものであるが、ここでは特に区別せず単に「データセット」と呼ぶ。

### 3. Proc SQL の基礎

Proc SQL は、Base SAS に含まれる SQL の実装であり、レポートや要約統計量の生成、データセットに対する操作、などを行うことができることから、データステップや他のプロシジャの代替としても使用できる。まずは、Proc SQL の基本的な使い方からはじめよう。

#### 3.1. Proc SQL の使い方

Proc SQL は対話型インターフェイスになっており、実行中は SQL 文を逐次実行できる。つまり、Proc SQL 文を実行すれば、Quit 文を実行して終了しない限り、SQL 形式で記述する Select 文や Create 文などのデータベースに対する操作命令であるクエリ（Query：問い合わせ）を繰り返し実行することができる。

```

【Proc SQL の基本構造】
proc sql [ options ] ;
  select 文;
  create 文;
  insert 文;
  ...
quit;

```

### 3.2. SELECT 文の基本構文

SQL における基本は SELECT 文と言っても過言では無い。その名のとおり、テーブルから列や行を選択して出力するための命令であり、選択クエリとも呼ばれる。なお、Select 句で指定する出力列と From 句で指定するテーブル（データセット）の参照については省略できないが、それ以外は省略できる。

|                                                                                                                                                                                                                                  |                                                                                                                                               |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| <b>【SELECT 文の基本構文】</b><br>Select 列 1 [, 列 2, ...]<br>From テーブル 1 [, テーブル 2, ...]<br>[Where 条件 1 [ and/or 条件 2... ] ]<br>[Group by 列 1 [, 列 2, ... ] ]<br>[Having 条件 1 [ and/or 条件 2... ] ]<br>[Order by 列 1 [, 列 2, ... ] ]<br>; | <b>【注 釈】</b><br>Select 出力列（変数）の指定<br>From 参照するデータセット<br>[Where 出力条件 ]<br>[Group by グループ化変数 ]<br>[Having グループ出力条件 ]<br>[Order by 並べ替え順序 ]<br>; |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|

### 3.3. 出力列の指定

Select 句では、From 句で指定したテーブルから出力する列名を指定する。複数の列を出力する場合は、出力順にコンマ区切りで指定する。なお、「\*」はテーブル内の全ての列を意味する記号であり、テーブルにおける格納順に列を出力する。

|                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>【記述例】 出力列の指定</b> <ul style="list-style-type: none"><li>・テーブル（dataset1）から全ての列を格納順に出力する<br/>Select * from dataset1;</li><li>・テーブル（dataset1）から特定の列（var1, var3）だけを出力する<br/>Select var1, var3 from dataset1;</li><li>・テーブル（dataset1）から特定の列（var1, var3, var2）を指定順に出力する<br/>Select var1, var3, var2 from dataset1;</li></ul> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

### 3.4. 出力列の別名および属性指定

出力列の指定において、元の列名とは異なる列名（列エイリアス: column-alias）を指定できる。as キーワードを使用し、列エイリアスは SAS の命名規則に従う必要がある。また、列に対して属性を指定できる。指定可能な属性は、ラベル（label=）、出力書式（format=）、入力書式（informat=）、長さ（length=）である。属性指定は、列エイリアス指定がある場合にはその後ろに、複数の属性を指定する場合は空白区切りで指定する。なお、label=' #' と指定した列は出力結果で列見出しが空欄になる。Format の指定については、「SAS 9.4 出力形式と入力形式リファレンス」等の関連資料を参照されたい。

|                                                                                                                                                                                                                                                    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>【記述例】 出力列の別名および属性指定</b> <ul style="list-style-type: none"><li>・列の別名指定（as エイリアス）<br/>Select var1 as als1, var2 as als2 from dataset1;</li><li>・列の属性指定（ラベル名やフォーマット指定など）<br/>Select var1 label='NAME' format=\$40. , var2 from dataset1;</li></ul> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

### 3.5. 計算式による列の追加

From 句で参照するデータセットの列を加工して新しい列を追加指定できる。出力列の指定において計算式や関数式を記述することができる。この場合は列名が無いので、必要があれば列エイリアスを指定する。



**【記述例】 計算式による列の追加**

- 人口 (pop) と面積 (area) から人口密度 (density) を作成する  
 Select pop, area, pop / area as density format=comma8. from dataset1;
- 姓 (sei) と名 (mei) を空白で結合して氏名 (simei) を作成する  
 Select sei, mei, catx(' ', sei, mei) as simei from dataset1;  
 /\* catx() 関数では結合に際して末尾の空白文字は除去される \*/

単純な計算式ではなく、条件によって値を変えたい場合には、Case 式を用いて処理を定義することができる。これによりデータステップにおける Select-When-Otherwise 構文と同様の処理が可能である。

**【記述例】 Case 式を用いた条件による値指定**

- 得点 (SCORE) に応じた評価 (EVAL) を作成  
 Select id, score,  
   Case  
     When 0 <= score < 60 then 'D'  
     When 60 <= score < 70 then 'C'  
     When 70 <= score < 80 then 'B'  
     When 80 <= score < 90 then 'A'  
     When 90 <= score <=100 then 'A'  
     Else 'Error'  
   End as eval length=5  
 From dataset1  
 ;

**【出力例】**

| ID | SCORE | EVAL  |
|----|-------|-------|
| 1  | 60    | C     |
| 2  | 95    | A+    |
| 3  | 30    | D     |
| 4  | .     | Error |

Case に続けてオペランド (列名) を指定する場合は、when に続けて値を指定する。なお、条件は等式に限定され、比較演算子などの他の演算子は使えない。

**【記述例】 Case-Operand 式を用いた条件による値指定**

- 性別コード (SEX\_CD) に応じた名称 (SEX\_MF) を作成  
 Select sex\_cd,  
   Case sex\_cd  
     When 1 then 'MALE'  
     When 2 then 'FEMALE'  
     Else 'NONE'  
   End as sex\_mf length=5  
 From dataset1  
 ;

**【出力例】**

| SEX_CD | SEX_MF |
|--------|--------|
| 1      | MALE   |
| 2      | FEMALE |
| 3      | NONE   |
| .      | NONE   |

**[参考] 欠損値を置換して出力**

引数の最初に現れる欠損値でない値を返す関数 Coalesce() を出力列の指定で用いることで、欠損値の出力を置換処理できる。なお、引数が全て欠損値の場合は欠損値が返されるので、欠損値を出力したくない場合は、引数の最後に欠損値ではない値を指定する。

**【記述例】 欠損値の置換出力**

- 列 var2 (文字列) にある欠損値を文字列 'N.A.' に置換して出力  
 Select var1, coalesce(var2, 'N.A.') as var2 from dataset1;
- 同じことは case 式でもできる  
 Select var1, case when var2 is missing then 'N.A.' else var2 end as var2 from dataset1;

### 3.6. 結果の並べ替え

Order by 句を用いることで、指定した列の値を基準にして出力結果をソートできる。列名の後ろに asc (昇順: ascending) または desc (降順: descending) を指定できるが、省略時は昇順である。複数の列を基準にソートしたい場合は、優先順位の高い列からコンマ区切りで指定する。Order by 句では列出力の指定に含まれない列を指定できる (並べ替えに使用した列を出力しなくてよい)。なお、欠損値は昇順の場合は先頭に表示される。また、Proc SQL の sortseq= オプションで言語やエンコーディングを指定することもできる。

| <p><b>【記述例】 order by 句</b></p> <ul style="list-style-type: none"> <li>データを身長 (height) の降順で並べ替え、同じ身長なら体重 (weight) の昇順で並べ替える</li> </ul> <pre>Select id, height, weight   from dataset1   order by height desc, weight ;</pre> | <p><b>【出力例】</b></p> <table border="1"> <thead> <tr> <th>ID</th> <th>HEIGHT</th> <th>WEIGHT</th> </tr> </thead> <tbody> <tr> <td>3</td> <td>175</td> <td>80</td> </tr> <tr> <td>2</td> <td>160</td> <td>70</td> </tr> <tr> <td>1</td> <td>160</td> <td>75</td> </tr> <tr> <td>4</td> <td>155</td> <td>60</td> </tr> </tbody> </table> | ID     | HEIGHT | WEIGHT | 3 | 175 | 80 | 2 | 160 | 70 | 1 | 160 | 75 | 4 | 155 | 60 |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|--------|--------|---|-----|----|---|-----|----|---|-----|----|---|-----|----|
| ID                                                                                                                                                                                                                          | HEIGHT                                                                                                                                                                                                                                                                                                                               | WEIGHT |        |        |   |     |    |   |     |    |   |     |    |   |     |    |
| 3                                                                                                                                                                                                                           | 175                                                                                                                                                                                                                                                                                                                                  | 80     |        |        |   |     |    |   |     |    |   |     |    |   |     |    |
| 2                                                                                                                                                                                                                           | 160                                                                                                                                                                                                                                                                                                                                  | 70     |        |        |   |     |    |   |     |    |   |     |    |   |     |    |
| 1                                                                                                                                                                                                                           | 160                                                                                                                                                                                                                                                                                                                                  | 75     |        |        |   |     |    |   |     |    |   |     |    |   |     |    |
| 4                                                                                                                                                                                                                           | 155                                                                                                                                                                                                                                                                                                                                  | 60     |        |        |   |     |    |   |     |    |   |     |    |   |     |    |

### 3.7. 出力条件の指定

Where 句を用いることで結果出力のための条件を指定できる。条件指定では比較演算子 (=, <, <=, <>, >, >=) 等を利用でき、複数の条件は論理演算子 (and, or) で結合できる。

| <p><b>【記述例】 where 句</b></p> <ul style="list-style-type: none"> <li>身長が 160 以上 (height&gt;=160) の男性 (sex='MALE') の行を出力</li> </ul> <pre>Select id, height, sex   from dataset1   where height&gt;=160 and sex='MALE' ;</pre> | <p><b>【出力例】</b></p> <table border="1"> <thead> <tr> <th>ID</th> <th>HEIGHT</th> <th>SEX</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>175</td> <td>MALE</td> </tr> <tr> <td>3</td> <td>160</td> <td>MALE</td> </tr> <tr> <td>8</td> <td>166</td> <td>MALE</td> </tr> <tr> <td>9</td> <td>180</td> <td>MALE</td> </tr> </tbody> </table> | ID   | HEIGHT | SEX | 1 | 175 | MALE | 3 | 160 | MALE | 8 | 166 | MALE | 9 | 180 | MALE |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|--------|-----|---|-----|------|---|-----|------|---|-----|------|---|-----|------|
| ID                                                                                                                                                                                                                         | HEIGHT                                                                                                                                                                                                                                                                                                                                    | SEX  |        |     |   |     |      |   |     |      |   |     |      |   |     |      |
| 1                                                                                                                                                                                                                          | 175                                                                                                                                                                                                                                                                                                                                       | MALE |        |     |   |     |      |   |     |      |   |     |      |   |     |      |
| 3                                                                                                                                                                                                                          | 160                                                                                                                                                                                                                                                                                                                                       | MALE |        |     |   |     |      |   |     |      |   |     |      |   |     |      |
| 8                                                                                                                                                                                                                          | 166                                                                                                                                                                                                                                                                                                                                       | MALE |        |     |   |     |      |   |     |      |   |     |      |   |     |      |
| 9                                                                                                                                                                                                                          | 180                                                                                                                                                                                                                                                                                                                                       | MALE |        |     |   |     |      |   |     |      |   |     |      |   |     |      |

Where 句では比較演算子に加えてさまざまな演算子が利用できる。これらすべての演算子は、NOT 演算子を付加することで否定条件も作成できる (not in, not like, is not missing, など)。また、Like 演算子ではワイルドカード文字が使える、パーセント記号 (%) は任意の文字列に一致し、アンダーライン (\_) は任意の 1 文字に一致する。文字列の比較で大文字・小文字の混在がある場合は、UPCASE() または UPPER() 関数で大文字に変換し、条件指定も含め大文字に統一して比較するなどの工夫をする。

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>【記述例】 where 句で使える便利な演算子</b></p> <ul style="list-style-type: none"> <li>得点 (score) が 60 から 100 の間<br/>where score between 60 and 100</li> <li>サイズ (size) が L か LL か 3L のいずれか<br/>where size in ('L', 'LL', '3L')</li> <li>メッセージ (msg) に「error」を含む<br/>where msg contains 'error'</li> <li>名前 (name) が A で始まるか、E で終わる 5 文字か<br/>where upper(name) like 'A%' or upper(name) like '____E'</li> <li>得点 (score) が欠損値<br/>where score is missing /* is null でも可 */</li> </ul> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

SQL 形式で記述された Select 文などの命令をクエリというが、クエリの中でカッコ ( ) で括って記述されたクエリのことをサブクエリという。Where 句ではサブクエリを用いて条件指定を記述できる演算子 (any, all, exists) が利用できる。なお、これらの演算子も前に NOT 演算子を付加して否定条件を作成できる。

**【記述例】 where 句におけるサブクエリ**

- dataset2 の得点 (score) の少なくとも 1 つを上回っている  
where score > any (select score from dataset2)
- dataset3 の得点 (score) の全てを上回っている  
where score > all (select score from dataset3)
- サブクエリの実行結果が 1 行以上存在する  
where exists (select \* from dataset4 where ... )

**[参考] 切り捨て文字列 (Truncated String) の比較演算子**

文字列の比較において相互の文字列の長さの違いが問題になることがある。その場合は切り捨て文字列の比較演算子を用いるとよい。これらの演算子では、比較を実行する前に、長い文字列が短い文字列の長さと同じになるように切り捨てが行われる。この切り捨て処理は Proc SQL 内部で実行され、オペランド (被演算部) に指定した列の内容が切り捨て文字列で上書きされることはない。

切り捨て文字列の比較演算子 (「SAS 9.4 SQL プロシジャガイド」より)

| 演算子 | 定義                | 記述例                           |
|-----|-------------------|-------------------------------|
| EQT | 切り捨て文字列に等しい       | where Name eqt 'Aust';        |
| GTT | 切り捨て文字列より大きい      | where Name gtt 'Bah';         |
| LTT | 切り捨て文字列より小さい      | where Name ltt 'An';          |
| GET | 切り捨て文字列と等しいかより大きい | where Country get 'United A'; |
| LET | 切り捨て文字列と等しいかより小さい | where Lastname let 'Smith';   |
| NET | 切り捨て文字列と等しくない     | where Style net 'TWO';        |

**3.8. データの要約**

複数行の値を用いて集計することができる。出力列の指定において合計や平均などを計算する関数を指定することで複数行の値を 1 つの値に要約することから要約関数と呼ばれる。計算式を指定する場合と同様に列名がないので、必要に応じて列エイリアスを指定する。

Proc SQL における主な要約関数一覧 (「SAS 9.4 SQL プロシジャ: ユーザーガイド」より)

| 要約関数              | 定義                        | 要約関数   | 定義                                      |
|-------------------|---------------------------|--------|-----------------------------------------|
| AVG, MEAN         | 値の平均                      | RANGE  | 値の範囲                                    |
| COUNT,<br>FREQ, N | 非欠損値の数                    | STD    | 標準偏差                                    |
|                   |                           | STDERR | 平均の標準誤差                                 |
| CSS               | 修正平方和                     | SUM    | 値の合計                                    |
| CV                | 変動係数 (パーセント)              | SUMWGT | WEIGHT 変数値の合計                           |
| MAX               | 最大値                       | T      | 母集団の平均値がゼロに等しいという仮説を検定するためのスチューデントの t 値 |
| MIN               | 最小値                       |        |                                         |
| NMISS             | 欠損値の数                     | USS    | 無修正平方和                                  |
| PRT               | より大きなスチューデントの t の絶対値を得る確率 | VAR    | 分散                                      |

| <b>【記述例】要約関数</b><br>得点 (score) が欠損でない人数、<br>合計、平均を出力する<br><pre> Select   count(score) as count,   sum(score) as sum,   avg(score) as avg from dataset1; </pre> | <b>【dataset1】</b><br><table border="1"> <thead> <tr> <th>ID</th> <th>SCORE</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>60</td> </tr> <tr> <td>2</td> <td>70</td> </tr> <tr> <td>3</td> <td>.</td> </tr> <tr> <td>4</td> <td>50</td> </tr> </tbody> </table> | ID    | SCORE | 1 | 60 | 2 | 70 | 3 | . | 4 | 50 | <b>【出力結果】</b><br><table border="1"> <thead> <tr> <th>COUNT</th> <th>SUM</th> <th>AVG</th> </tr> </thead> <tbody> <tr> <td>3</td> <td>180</td> <td>60</td> </tr> </tbody> </table> | COUNT | SUM | AVG | 3 | 180 | 60 |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-------|---|----|---|----|---|---|---|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-----|-----|---|-----|----|
|                                                                                                                                                                | ID                                                                                                                                                                                                                                                              | SCORE |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |
|                                                                                                                                                                | 1                                                                                                                                                                                                                                                               | 60    |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |
|                                                                                                                                                                | 2                                                                                                                                                                                                                                                               | 70    |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |
|                                                                                                                                                                | 3                                                                                                                                                                                                                                                               | .     |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |
| 4                                                                                                                                                              | 50                                                                                                                                                                                                                                                              |       |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |
| COUNT                                                                                                                                                          | SUM                                                                                                                                                                                                                                                             | AVG   |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |
| 3                                                                                                                                                              | 180                                                                                                                                                                                                                                                             | 60    |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |
|                                                                                                                                                                |                                                                                                                                                                                                                                                                 |       |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |
|                                                                                                                                                                |                                                                                                                                                                                                                                                                 |       |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |
|                                                                                                                                                                |                                                                                                                                                                                                                                                                 |       |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |
|                                                                                                                                                                |                                                                                                                                                                                                                                                                 |       |       |   |    |   |    |   |   |   |    |                                                                                                                                                                                   |       |     |     |   |     |    |

### 3.9. データのグループ化

要約関数を Group by 句と併用することで、グループ毎にデータを要約することができる。グループ化の基準となる列 (グループ化変数) を指定し、複数の列を用いる場合はコンマ区切りで優先順位の高い順に指定する。なお、要約関数を使わないグループ化では、group by 句が order by 句に変換されて実行される。また、グループ化変数に欠損値があると、欠損値を持つ行でグループ化されるので注意が必要である。

| <b>【記述例】 Group by 句</b><br>世帯番号 (HHID) でグループ化して、<br>世帯員 (HMID) の数を出力<br><pre> Select hhid, count(hmid) as count   from dataset1 group by hhid ;  Select hhid, count(*) as count   from dataset1 group by hhid ; /* 欠損値を含む*/ </pre> | <b>【dataset1】</b><br><table border="1"> <thead> <tr> <th>HHID</th> <th>HMID</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1</td> </tr> <tr> <td>1</td> <td>2</td> </tr> <tr> <td>1</td> <td>3</td> </tr> <tr> <td>2</td> <td>1</td> </tr> <tr> <td>2</td> <td>2</td> </tr> </tbody> </table> | HHID | HMID | 1 | 1 | 1 | 2 | 1 | 3 | 2 | 1 | 2 | 2 | <b>【出力結果】</b><br><table border="1"> <thead> <tr> <th>HHID</th> <th>COUNT</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>3</td> </tr> <tr> <td>2</td> <td>2</td> </tr> </tbody> </table> | HHID | COUNT | 1 | 3 | 2 | 2 |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------|---|---|---|---|---|---|---|---|---|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|-------|---|---|---|---|
|                                                                                                                                                                                                                                    | HHID                                                                                                                                                                                                                                                                                           | HMID |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
|                                                                                                                                                                                                                                    | 1                                                                                                                                                                                                                                                                                              | 1    |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
|                                                                                                                                                                                                                                    | 1                                                                                                                                                                                                                                                                                              | 2    |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
|                                                                                                                                                                                                                                    | 1                                                                                                                                                                                                                                                                                              | 3    |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
| 2                                                                                                                                                                                                                                  | 1                                                                                                                                                                                                                                                                                              |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
| 2                                                                                                                                                                                                                                  | 2                                                                                                                                                                                                                                                                                              |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
| HHID                                                                                                                                                                                                                               | COUNT                                                                                                                                                                                                                                                                                          |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
| 1                                                                                                                                                                                                                                  | 3                                                                                                                                                                                                                                                                                              |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
| 2                                                                                                                                                                                                                                  | 2                                                                                                                                                                                                                                                                                              |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
|                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
|                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
|                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |
|                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                          |      |       |   |   |   |   |

グループ化されたクエリの結果に対して出力条件を付加したい場合は Having 句を用いる。

| <b>【記述例】 Having 句</b><br>世帯番号 (HHID) でグループ化した世帯員<br>(HMID) の数が 3 以上を出力<br><pre> Select hhid, count(hmid) as count   from dataset1   group by hhid   having count(*)&gt;=3 ; </pre> | <b>【dataset1】</b><br><table border="1"> <thead> <tr> <th>HHID</th> <th>HMID</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1</td> </tr> <tr> <td>1</td> <td>2</td> </tr> <tr> <td>1</td> <td>3</td> </tr> <tr> <td>2</td> <td>1</td> </tr> <tr> <td>2</td> <td>2</td> </tr> </tbody> </table> | HHID | HMID | 1 | 1 | 1 | 2 | 1 | 3 | 2 | 1 | 2 | 2 | <b>【出力結果】</b><br><table border="1"> <thead> <tr> <th>HHID</th> <th>COUNT</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>3</td> </tr> </tbody> </table> | HHID | COUNT | 1 | 3 |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------|---|---|---|---|---|---|---|---|---|---|---------------------------------------------------------------------------------------------------------------------------------------------------------|------|-------|---|---|
|                                                                                                                                                                                    | HHID                                                                                                                                                                                                                                                                                           | HMID |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
|                                                                                                                                                                                    | 1                                                                                                                                                                                                                                                                                              | 1    |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
|                                                                                                                                                                                    | 1                                                                                                                                                                                                                                                                                              | 2    |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
|                                                                                                                                                                                    | 1                                                                                                                                                                                                                                                                                              | 3    |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
| 2                                                                                                                                                                                  | 1                                                                                                                                                                                                                                                                                              |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
| 2                                                                                                                                                                                  | 2                                                                                                                                                                                                                                                                                              |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
| HHID                                                                                                                                                                               | COUNT                                                                                                                                                                                                                                                                                          |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
| 1                                                                                                                                                                                  | 3                                                                                                                                                                                                                                                                                              |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
|                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
|                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
|                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |
|                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                |      |      |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                         |      |       |   |   |

## 4. データの水平結合 (merge)

複数のデータセットにおける共通の情報をキーとして、レコード (オブザベーション) 単位で列方向に結合することを水平結合 (merge) という。データステップで処理する場合は merge 文と by 文を使う。異なるデータセットを水平 (列) 方向に結合する方法にはいくつかの種類があり、Proc SQL ではデカルト積 (cross join)、内部結合 (inner join)、外部結合 (outer join)、自然結合 (natural join) などをサポートしている。

### 4.1. デカルト積 (Cross Join)

結合に使用するキーを指定せず、対象のデータセットの各行について全ての組み合わせが出力される。

|                                                                                                                                                |                   |              |                   |              |               |              |            |              |
|------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|--------------|-------------------|--------------|---------------|--------------|------------|--------------|
| <b>【記述例】デカルト積 (Cross Join)</b><br>Select *<br>from dataset1, dataset2;<br><br>Select *<br>from dataset1<br>cross join dataset2;<br>/* 結果は同じ */ | <b>【dataset1】</b> |              | <b>【dataset2】</b> |              | <b>【出力結果】</b> |              |            |              |
|                                                                                                                                                | <b>ID1</b>        | <b>DATA1</b> | <b>ID2</b>        | <b>DATA2</b> | <b>ID1</b>    | <b>DATA1</b> | <b>ID2</b> | <b>DATA2</b> |
|                                                                                                                                                | 1                 | A            | 3                 | X            | 1             | A            | 3          | X            |
|                                                                                                                                                | 2                 | B            | 4                 | Y            | 1             | A            | 4          | Y            |
|                                                                                                                                                |                   |              |                   |              | 2             | B            | 3          | X            |
|                                                                                                                                                |                   |              |                   | 2            | B             | 4            | Y          |              |

#### 4.2. 内部結合 (Inner Join)

結合対象のデータセットにおいて、指定した条件に合致する組み合わせが出力される。From 句でテーブルエイリアス (table-alias) を利用すると列指定や条件指定を記述しやすい。

|                                                                                                                                                                                                                                         |                   |              |                   |              |               |              |              |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|--------------|-------------------|--------------|---------------|--------------|--------------|
| <b>【記述例】内部結合 (Inner Join)</b><br>Select a.id, data1, data2<br>from dataset1 as a, dataset2 as b<br>where a.id = b.id ;<br><br>Select a.id, data1, data<br>from dataset1 as a<br>inner join dataset2 as b<br>on a.id = b.id ; /* 結果は同じ*/ | <b>【dataset1】</b> |              | <b>【dataset2】</b> |              | <b>【出力結果】</b> |              |              |
|                                                                                                                                                                                                                                         | <b>ID</b>         | <b>DATA1</b> | <b>ID</b>         | <b>DATA2</b> | <b>ID</b>     | <b>DATA1</b> | <b>DATA2</b> |
|                                                                                                                                                                                                                                         | 1                 | A            | 1                 | X            | 1             | A            | X            |
|                                                                                                                                                                                                                                         | 2                 | B            | 2                 | Y            | 2             | B            | Y            |
|                                                                                                                                                                                                                                         |                   |              | 3                 | Z            |               |              |              |

#### 4.3. 左外部結合 (Left Outer Join)

左側 (from 句) の全行と、右側 (left join 句) の条件 (on 句) が一致した行との組み合わせが出力される。不一致の場合、右側の列データは欠損となる。

|                                                                                                                                           |                   |              |                   |              |               |              |              |
|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------|--------------|-------------------|--------------|---------------|--------------|--------------|
| <b>【記述例】左外部結合 (Left Outer Join)</b><br>Select a.id, a.data1, b.data2<br>from dataset1 as a<br>left join dataset2 as b<br>on a.id = b.id ; | <b>【dataset1】</b> |              | <b>【dataset2】</b> |              | <b>【出力結果】</b> |              |              |
|                                                                                                                                           | <b>ID</b>         | <b>DATA1</b> | <b>ID</b>         | <b>DATA2</b> | <b>ID</b>     | <b>DATA1</b> | <b>DATA2</b> |
|                                                                                                                                           | 1                 | A            | 1                 | X            | 1             | A            | X            |
|                                                                                                                                           | 2                 | B            | 2                 | Y            | 2             | B            | Y            |
|                                                                                                                                           | 4                 | D            | 3                 | Z            | 4             | D            | .            |

#### 4.4. 右外部結合 (Right Outer Join)

右側 (right join 句) の全行と、左側 (from 句) の条件 (on 句) が一致した行との組み合わせが出力される。不一致の場合、左側の列データは欠損となる。

|                                                                                                                                             |                   |              |                   |              |               |              |              |
|---------------------------------------------------------------------------------------------------------------------------------------------|-------------------|--------------|-------------------|--------------|---------------|--------------|--------------|
| <b>【記述例】右外部結合 (Right Outer Join)</b><br>Select a.id, a.data1, b.data2<br>from dataset1 as a<br>right join dataset2 as b<br>on a.id = b.id ; | <b>【dataset1】</b> |              | <b>【dataset2】</b> |              | <b>【出力結果】</b> |              |              |
|                                                                                                                                             | <b>ID</b>         | <b>DATA1</b> | <b>ID</b>         | <b>DATA2</b> | <b>ID</b>     | <b>DATA1</b> | <b>DATA2</b> |
|                                                                                                                                             | 1                 | A            | 1                 | X            | 1             | A            | X            |
|                                                                                                                                             | 2                 | B            | 2                 | Y            | 2             | B            | Y            |
|                                                                                                                                             | 4                 | D            | 3                 | Z            | 4             | .            | Z            |

#### 4.5. 完全外部結合 (Full Outer Join)

左側 (from 句) の全行と右側 (full join 句) の全行の組み合わせが出力される。双方で条件 (on 句) が一致した行に加えて、一方にしかない行も含まれる。一致しなかった場合、それぞれの列データは欠損となる。

| 【記述例】完全外部結合 (Full Outer Join)                                                                             | 【dataet1】                                                                                                                                                                                  | 【dataet2】 | 【出力結果】 |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
|-----------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|--------|---|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-------|---|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-------|-------|---|---|---|---|---|---|---|---|---|---|---|---|
| <pre>Select a. id, a. data1, b. data2 from dataset1 as a full join dataset2 as b on a. id = b. id ;</pre> | <table border="1"> <thead> <tr><th>ID</th><th>DATA1</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td></tr> <tr><td>2</td><td>B</td></tr> <tr><td>4</td><td>D</td></tr> </tbody> </table> | ID        | DATA1  | 1 | A | 2 | B | 4 | D | <table border="1"> <thead> <tr><th>ID</th><th>DATA2</th></tr> </thead> <tbody> <tr><td>1</td><td>X</td></tr> <tr><td>2</td><td>Y</td></tr> <tr><td>3</td><td>Z</td></tr> </tbody> </table> | ID | DATA2 | 1 | X | 2 | Y | 3 | Z | <table border="1"> <thead> <tr><th>ID</th><th>DATA1</th><th>DATA2</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td><td>X</td></tr> <tr><td>2</td><td>B</td><td>Y</td></tr> <tr><td>3</td><td>.</td><td>Z</td></tr> <tr><td>4</td><td>D</td><td>.</td></tr> </tbody> </table> | ID | DATA1 | DATA2 | 1 | A | X | 2 | B | Y | 3 | . | Z | 4 | D | . |
| ID                                                                                                        | DATA1                                                                                                                                                                                      |           |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 1                                                                                                         | A                                                                                                                                                                                          |           |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 2                                                                                                         | B                                                                                                                                                                                          |           |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 4                                                                                                         | D                                                                                                                                                                                          |           |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| ID                                                                                                        | DATA2                                                                                                                                                                                      |           |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 1                                                                                                         | X                                                                                                                                                                                          |           |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 2                                                                                                         | Y                                                                                                                                                                                          |           |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 3                                                                                                         | Z                                                                                                                                                                                          |           |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| ID                                                                                                        | DATA1                                                                                                                                                                                      | DATA2     |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 1                                                                                                         | A                                                                                                                                                                                          | X         |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 2                                                                                                         | B                                                                                                                                                                                          | Y         |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 3                                                                                                         | .                                                                                                                                                                                          | Z         |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 4                                                                                                         | D                                                                                                                                                                                          | .         |        |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |

#### 【参考】欠損値 (NULL, Missing) の影響

PROC SQL では欠損値 (NULL, MISSING) の特殊な扱いに注意が必要である。1つは、昇順でのソート時には欠損値が先頭になり、数値の大小や、文字コードの前後に依存しないことである。もう1つは、結合 (マージ) においても同じ型 (文字 or 数値) の欠損値と接合することであり、SQL の実装では欠損値を結合に使わないのが一般的である。これは、SAS における他のプロシジャやデータステップでの処理との整合性によるものである。

| 【例】キーに欠損値を含む内部結合                                                                                    | 【dataet1】                                                                                                                                                                                                                | 【dataet2】 | 【出力結果】 |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
|-----------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|--------|---|---|---|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-------|---|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-------|-------|---|---|---|---|---|---|---|---|---|---|---|---|
| <pre>Select a. id, a. data1, b. data2 from dataset1 as a, dataset2 as b where a. id = b. id ;</pre> | <table border="1"> <thead> <tr><th>ID</th><th>DATA1</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td></tr> <tr><td>2</td><td>B</td></tr> <tr><td>.</td><td>C</td></tr> <tr><td>.</td><td>D</td></tr> </tbody> </table> | ID        | DATA1  | 1 | A | 2 | B | . | C | . | D | <table border="1"> <thead> <tr><th>ID</th><th>DATA2</th></tr> </thead> <tbody> <tr><td>1</td><td>X</td></tr> <tr><td>2</td><td>Y</td></tr> <tr><td>.</td><td>Z</td></tr> </tbody> </table> | ID | DATA2 | 1 | X | 2 | Y | . | Z | <table border="1"> <thead> <tr><th>ID</th><th>DATA1</th><th>DATA2</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td><td>X</td></tr> <tr><td>2</td><td>B</td><td>Y</td></tr> <tr><td>.</td><td>C</td><td>Z</td></tr> <tr><td>.</td><td>D</td><td>Z</td></tr> </tbody> </table> | ID | DATA1 | DATA2 | 1 | A | X | 2 | B | Y | . | C | Z | . | D | Z |
| ID                                                                                                  | DATA1                                                                                                                                                                                                                    |           |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 1                                                                                                   | A                                                                                                                                                                                                                        |           |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 2                                                                                                   | B                                                                                                                                                                                                                        |           |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| .                                                                                                   | C                                                                                                                                                                                                                        |           |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| .                                                                                                   | D                                                                                                                                                                                                                        |           |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| ID                                                                                                  | DATA2                                                                                                                                                                                                                    |           |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 1                                                                                                   | X                                                                                                                                                                                                                        |           |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 2                                                                                                   | Y                                                                                                                                                                                                                        |           |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| .                                                                                                   | Z                                                                                                                                                                                                                        |           |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| ID                                                                                                  | DATA1                                                                                                                                                                                                                    | DATA2     |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 1                                                                                                   | A                                                                                                                                                                                                                        | X         |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| 2                                                                                                   | B                                                                                                                                                                                                                        | Y         |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| .                                                                                                   | C                                                                                                                                                                                                                        | Z         |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |
| .                                                                                                   | D                                                                                                                                                                                                                        | Z         |        |   |   |   |   |   |   |   |   |                                                                                                                                                                                            |    |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                |    |       |       |   |   |   |   |   |   |   |   |   |   |   |   |

#### 【参考】ユニークデータの抽出と重複データの処理

水平統合 (マージ) では、緩い結合条件を指定するとレコードの重複結合 (1対多、多対多) が生じやすい。マージ後のレコード件数や ID 番号のユニーク性をチェックするように心がけよう。

#### ユニークな行のみ出力する (Distinct キーワードの利用)

- ・姓 (sei) と名 (mei) のユニークな組み合わせの一覧を出力する  

```
Select distinct sei, mei from dataset1; /* 重複行を出力しない */
```
- ・ユニークな姓 (sei) の数を数える

```
Select count(distinct sei) as count from dataset1;
```

#### グループ化と集約関数 count() を使用した重複行の検出

- ・姓 (sei) と名 (mei) の重複 (同姓同名) があれば、重複数 (count) と併せて出力する  

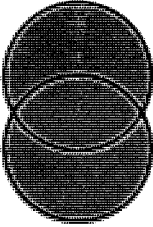
```
Select sei, mei, count(*) as count from dataset1 group by sei, mei having count(*)>1;
```

### 5. データの垂直結合

複数のデータセットを垂直方向に結合することもできる。SQL では集合 (セット) 演算子を用いる。基本的には列の位置に基づいて (列名と無関係に) データセットを垂直に結合する。

## 5.1. 和集合 (UNION)

両方のクエリ結果を統合し、重複行を削除する。それぞれに追加行を持つデータを統合する際に便利である。

| 【例】和集合 (UNION 演算子)       |  | 【dataet1】 |       | 【dataet2】 |       | 【出力結果】 |       |
|--------------------------|-----------------------------------------------------------------------------------|-----------|-------|-----------|-------|--------|-------|
|                          |                                                                                   | ID1       | DATA1 | ID2       | DATA2 | ID1    | DATA1 |
| Select * from dataset1   |                                                                                   | 1         | A     | 1         | A     | 1      | A     |
| Union                    |                                                                                   | 2         | B     | 2         | Y     | 2      | B     |
| Select * from dataset2 ; |                                                                                   | 4         | D     | 3         | Z     | 4      | D     |
|                          |                                                                                   |           |       |           |       | 2      | Y     |
|                          |                                                                                   |           |       |           |       | 3      | Z     |

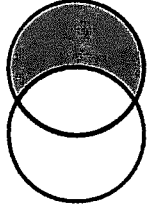
### [参考] ALL キーワード

和集合 (UNION) に ALL キーワードをつけると、両方のクエリ結果から重複を含む行を作成する。ALL キーワードをつけることで重複行の除去が行われなくなる。重複を含めたデータを作成する場合に使用する。

| 【例】UNION 演算子<br>+ All キーワード | 【dataet1】 |       | 【dataet2】 |       | 【出力結果】 |       |
|-----------------------------|-----------|-------|-----------|-------|--------|-------|
|                             | ID1       | DATA1 | ID2       | DATA2 | ID1    | DATA1 |
| Select * from dataset1      | 1         | A     | 1         | A     | 1      | A     |
| Union <u>All</u>            | 2         | B     | 2         | Y     | 2      | B     |
| Select * from dataset2 ;    | 4         | D     | 3         | Z     | 4      | D     |
|                             |           |       |           |       | 1      | A     |
|                             |           |       |           |       | 2      | Y     |
|                             |           |       |           |       | 3      | Z     |

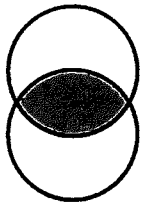
## 5.2. 差集合 (EXCEPT)

最初のクエリ結果にのみ含まれる行を生成する。例えば、新データから旧データとの重複情報を差し引けば、挿入・更新された情報を抽出することができる。

| 【例】差集合 (Except 演算子)      |  | 【dataet1】 |       | 【dataet2】 |       | 【出力結果】 |       |
|--------------------------|-------------------------------------------------------------------------------------|-----------|-------|-----------|-------|--------|-------|
|                          |                                                                                     | ID1       | DATA1 | ID1       | DATA2 | ID1    | DATA1 |
| Select * from dataset1   |                                                                                     | 1         | A     | 1         | A     | 2      | B     |
| Except                   |                                                                                     | 2         | B     | 2         | Y     | 4      | D     |
| Select * from dataset2 ; |                                                                                     | 4         | D     | 3         | Z     |        |       |

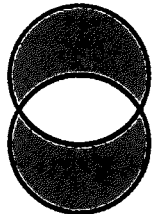
## 5.3. 積集合 (INTERSECT)

2つのクエリ結果に共通する行を生成する。2つのデータで変わっていない情報を抽出できる。新旧の調査名簿で継続調査の対象や存続情報の抽出に利用できる。

| <b>【例】積集合 (Intersect 演算子)</b><br><br>Select * from dataset1<br>Intersect<br>Select * from dataset2 ; |  | <b>【dataset1】</b><br><table border="1"> <thead> <tr><th>ID1</th><th>DATA1</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td></tr> <tr><td>2</td><td>B</td></tr> <tr><td>4</td><td>D</td></tr> </tbody> </table> | ID1   | DATA1 | 1 | A | 2 | B | 4 | D | <b>【dataset2】</b><br><table border="1"> <thead> <tr><th>ID2</th><th>DATA2</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td></tr> <tr><td>2</td><td>Y</td></tr> <tr><td>3</td><td>Z</td></tr> </tbody> </table> | ID2 | DATA2 | 1 | A | 2 | Y | 3 | Z | <b>【出力結果】</b><br><table border="1"> <thead> <tr><th>ID1</th><th>DATA1</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td></tr> </tbody> </table> | ID1 | DATA1 | 1 | A |
|------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-------|---|---|---|---|---|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-------|---|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------|-----|-------|---|---|
|                                                                                                      |                                                                                   | ID1                                                                                                                                                                                                              | DATA1 |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                  |     |       |   |   |
|                                                                                                      |                                                                                   | 1                                                                                                                                                                                                                | A     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                  |     |       |   |   |
|                                                                                                      |                                                                                   | 2                                                                                                                                                                                                                | B     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                  |     |       |   |   |
| 4                                                                                                    | D                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                  |     |       |   |   |
| ID2                                                                                                  | DATA2                                                                             |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                  |     |       |   |   |
| 1                                                                                                    | A                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                  |     |       |   |   |
| 2                                                                                                    | Y                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                  |     |       |   |   |
| 3                                                                                                    | Z                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                  |     |       |   |   |
| ID1                                                                                                  | DATA1                                                                             |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                  |     |       |   |   |
| 1                                                                                                    | A                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                  |     |       |   |   |

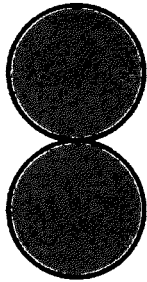
#### 5.4. 排他的和集合 (EXCLUSIVE UNION)

一方のクエリ結果にのみ含まれる行を作成する。管理バージョンの異なるデータを照合して、それぞれで追加・変更された情報を抽出できる。相互の差集合に対して和集合を取れば良い。

| <b>【例】排他的和集合</b><br><br>(select * from dataset1<br>except<br>select * from dataset2)<br>union<br>(select * from dataset2<br>except<br>select * from dataset1); |  | <b>【dataset1】</b><br><table border="1"> <thead> <tr><th>ID1</th><th>DATA1</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td></tr> <tr><td>2</td><td>B</td></tr> <tr><td>4</td><td>D</td></tr> </tbody> </table> | ID1   | DATA1 | 1 | A | 2 | B | 4 | D | <b>【dataset2】</b><br><table border="1"> <thead> <tr><th>ID2</th><th>DATA2</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td></tr> <tr><td>2</td><td>Y</td></tr> <tr><td>3</td><td>Z</td></tr> </tbody> </table> | ID2 | DATA2 | 1 | A | 2 | Y | 3 | Z | <b>【出力結果】</b><br><table border="1"> <thead> <tr><th>ID1</th><th>DATA1</th></tr> </thead> <tbody> <tr><td>2</td><td>B</td></tr> <tr><td>4</td><td>D</td></tr> <tr><td>2</td><td>Y</td></tr> <tr><td>3</td><td>Z</td></tr> </tbody> </table> | ID1 | DATA1 | 2 | B | 4 | D | 2 | Y | 3 | Z |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-------|---|---|---|---|---|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-------|---|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-------|---|---|---|---|---|---|---|---|
|                                                                                                                                                                |                                                                                   | ID1                                                                                                                                                                                                              | DATA1 |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
|                                                                                                                                                                |                                                                                   | 1                                                                                                                                                                                                                | A     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
|                                                                                                                                                                |                                                                                   | 2                                                                                                                                                                                                                | B     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
| 4                                                                                                                                                              | D                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
| ID2                                                                                                                                                            | DATA2                                                                             |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
| 1                                                                                                                                                              | A                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
| 2                                                                                                                                                              | Y                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
| 3                                                                                                                                                              | Z                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
| ID1                                                                                                                                                            | DATA1                                                                             |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
| 2                                                                                                                                                              | B                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
| 4                                                                                                                                                              | D                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
| 2                                                                                                                                                              | Y                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |
| 3                                                                                                                                                              | Z                                                                                 |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                            |     |       |   |   |   |   |   |   |   |   |

#### 5.5. 連結 (OUTER UNION)

列を統合せずにクエリの結果を連結する。列の順番では連結されず、別の列として垂直に連結される。

| <b>【例】連結</b><br><br>(OUTER UNION 演算子)<br><br>Select * from dataset1<br>outer union<br>Select * from dataset2; |  | <b>【dataset1】</b><br><table border="1"> <thead> <tr><th>ID1</th><th>DATA1</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td></tr> <tr><td>2</td><td>B</td></tr> <tr><td>4</td><td>D</td></tr> </tbody> </table> | ID1   | DATA1 | 1 | A | 2 | B | 4 | D | <b>【dataset2】</b><br><table border="1"> <thead> <tr><th>ID2</th><th>DATA2</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td></tr> <tr><td>2</td><td>Y</td></tr> <tr><td>3</td><td>Z</td></tr> </tbody> </table> | ID2 | DATA2 | 1 | A | 2 | Y | 3 | Z | <b>【出力結果】</b><br><table border="1"> <thead> <tr><th>ID1</th><th>DATA1</th><th>ID2</th><th>DATA2</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td><td>.</td><td>.</td></tr> <tr><td>2</td><td>B</td><td>.</td><td>.</td></tr> <tr><td>4</td><td>D</td><td>.</td><td>.</td></tr> <tr><td>.</td><td>.</td><td>1</td><td>A</td></tr> <tr><td>.</td><td>.</td><td>2</td><td>Y</td></tr> <tr><td>.</td><td>.</td><td>3</td><td>Z</td></tr> </tbody> </table> | ID1 | DATA1 | ID2 | DATA2 | 1 | A | . | . | 2 | B | . | . | 4 | D | . | . | . | . | 1 | A | . | . | 2 | Y | . | . | 3 | Z |
|---------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-------|---|---|---|---|---|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-------|---|---|---|---|---|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-------|-----|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|                                                                                                               |                                                                                    | ID1                                                                                                                                                                                                              | DATA1 |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|                                                                                                               |                                                                                    | 1                                                                                                                                                                                                                | A     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|                                                                                                               |                                                                                    | 2                                                                                                                                                                                                                | B     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 4                                                                                                             | D                                                                                  |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| ID2                                                                                                           | DATA2                                                                              |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 1                                                                                                             | A                                                                                  |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 2                                                                                                             | Y                                                                                  |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 3                                                                                                             | Z                                                                                  |                                                                                                                                                                                                                  |       |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| ID1                                                                                                           | DATA1                                                                              | ID2                                                                                                                                                                                                              | DATA2 |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 1                                                                                                             | A                                                                                  | .                                                                                                                                                                                                                | .     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 2                                                                                                             | B                                                                                  | .                                                                                                                                                                                                                | .     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 4                                                                                                             | D                                                                                  | .                                                                                                                                                                                                                | .     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| .                                                                                                             | .                                                                                  | 1                                                                                                                                                                                                                | A     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| .                                                                                                             | .                                                                                  | 2                                                                                                                                                                                                                | Y     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| .                                                                                                             | .                                                                                  | 3                                                                                                                                                                                                                | Z     |       |   |   |   |   |   |   |                                                                                                                                                                                                                  |     |       |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |       |     |       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

[参考] CORRESPONDING (CORR) キーワード

OUTER UNION で CORR キーワードを使用すると、クエリの結果の同名の列を統合し、列名が異なるものを統合せずに連結する。ID は重複しないが列定義が異なるようなデータを統合する場合に便利である。

| <b>【例】OUTER UNION 演算子<br/>+ CORR キーワード</b><br><br>Select * from dataset1<br>outer union <u>corr</u><br>Select * from dataset2; | <b>【dataset1】</b><br><table border="1"> <thead> <tr><th>ID</th><th>DATA1</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td></tr> <tr><td>2</td><td>B</td></tr> </tbody> </table> | ID    | DATA1 | 1 | A | 2 | B | <b>【dataset2】</b><br><table border="1"> <thead> <tr><th>ID</th><th>DATA2</th></tr> </thead> <tbody> <tr><td>11</td><td>X</td></tr> <tr><td>12</td><td>Y</td></tr> </tbody> </table> | ID | DATA2 | 11 | X | 12 | Y | <b>【出力結果】</b><br><table border="1"> <thead> <tr><th>ID</th><th>DATA1</th><th>DATA2</th></tr> </thead> <tbody> <tr><td>1</td><td>A</td><td>.</td></tr> <tr><td>2</td><td>B</td><td>.</td></tr> <tr><td>11</td><td>.</td><td>X</td></tr> <tr><td>12</td><td>.</td><td>Y</td></tr> </tbody> </table> | ID | DATA1 | DATA2 | 1 | A | . | 2 | B | . | 11 | . | X | 12 | . | Y |
|--------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|-------|---|---|---|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-------|----|---|----|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-------|-------|---|---|---|---|---|---|----|---|---|----|---|---|
|                                                                                                                                | ID                                                                                                                                                                                | DATA1 |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |
|                                                                                                                                | 1                                                                                                                                                                                 | A     |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |
|                                                                                                                                | 2                                                                                                                                                                                 | B     |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |
| ID                                                                                                                             | DATA2                                                                                                                                                                             |       |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |
| 11                                                                                                                             | X                                                                                                                                                                                 |       |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |
| 12                                                                                                                             | Y                                                                                                                                                                                 |       |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |
| ID                                                                                                                             | DATA1                                                                                                                                                                             | DATA2 |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |
| 1                                                                                                                              | A                                                                                                                                                                                 | .     |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |
| 2                                                                                                                              | B                                                                                                                                                                                 | .     |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |
| 11                                                                                                                             | .                                                                                                                                                                                 | X     |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |
| 12                                                                                                                             | .                                                                                                                                                                                 | Y     |       |   |   |   |   |                                                                                                                                                                                     |    |       |    |   |    |   |                                                                                                                                                                                                                                                                                                   |    |       |       |   |   |   |   |   |   |    |   |   |    |   |   |



## 6. テーブルの編集

ここまでは、Select 文を用いた行や列の選択や結合などの操作が中心であったが、Select 文以外の SQL 命令、とりわけ、テーブルの編集に関する命令について取り上げる。

### 6.1. テーブルの作成

Create Table 文を利用することで、新規テーブルを作成できる。新規テーブルにおける列定義を指定したり、既存テーブルの列定義をコピーしたり、選択クエリ (Select 文) の結果を新規テーブルに保存したりする方法がある。

#### 【記述例】 Create Table 文

- ・新規テーブル (dataset1) を作成し、3つの列: id (数値)、name (20文字)、age (数値) を定義する。  
`Create table dataset1 ( id num, name char(20), age num) ;`
- ・新規テーブル (dataset2) に dataset1 の列定義をコピーする。行はコピーしない。  
`Create table dataset2 like dataset1 ;`
- ・新規テーブル (dataset3) に select 文の実行結果を (dataset1 の内容) 保存する。  
`Create table dataset3 as Select * from dataset1 ;`

### 6.2. テーブルへの行挿入

Insert 文を使用することで、既存のテーブル (データセット) に対して行挿入を行うことができる。

#### 【記述例】 Insert 文

- ・データセット (dataset1) の指定列に対応する値を2行挿入する  
`Insert into dataset1 (id, name, age) values (5, 'Tanaka', 38) values (6, 'Suzuki', 40) ;`
- ・データセット (dataset1) の格納順に値を挿入する  
`Insert into dataset1 set id=5, name='Tanaka', age=38 set id=6, name='Suzuki', age=40 ;`
- ・データセット (dataset1) の格納順に値を挿入する  
`Insert into dataset1 values (5, 'Tanaka', 38) values (6, 'Suzuki', 40) ;`
- ・データセット (dataset1) の指定列に対応する値を dataset2 から抽出して挿入する  
`Insert into dataset1 (id, name, age) select id, name, age from dataset2 ;`

### 6.3. テーブルのデータ更新

Update 文を使用することで、既存のテーブル (データセット) の内容を更新できる。その際、更新対象となる行を Where 句で条件指定し、Set 句で更新内容を指定する。サブクエリを使うことで更新用データを参照して更新することもできる。データステップにおける update 文を用いる場合も同様のことができる。

#### 【記述例】 Update 文

- ・価格 (price) を税込み金額で更新する  
`Update dataset1 set price = price * 1.08 where country = 'JP' ;`
- ・更新用データ (dataset2) を参照して dataset1 の内容を更新する  
`Update dataset1 as a  
set data = (select data from dataset2 as b where a.id=b.id)  
where exists (select * from dataset2 as b where a.id=b.id) ;  
/* where 句を忘れると他の行が欠損値で更新されるので要注意 */`

### 6.4. テーブルの列定義の変更

Alter Table 文を用いることで、既存テーブルの列に対して追加、変更、削除の操作を行うことができる。

**【記述例】 Alter Table 文**

- dataset1 に列 var2 を追加し、文字型 (10 バイト)、出力書式\$10. とする  
Alter table dataset1 add var2 char(10) label='variable2' format=\$10. ;
- dataset1 の列 var2 を変更し、文字型 (20 バイト)、出力書式\$20. とする  
Alter table dataset1 modify var2 char(20) format=\$20. ;
- dataset1 から列 var2 を削除する  
Alter table dataset1 drop var2 ;

## 6.5. その他 削除

Drop や Delete 文を用いることで、テーブルや行などを削除することができる。

**【記述例】 さまざまな削除**

- テーブル (dataset1) の削除  
Drop table dataset1 ;
- テーブル (dataset1) からインデックス (インデックス名 : idx) の削除  
Drop index idx from dataset1 ;
- ビュー (dataview1) の削除  
Drop view dataview1 ;
- テーブル (dataset1) の行の削除: A で始まる名前(name) の行  
Delete from dataset1 where upper(name) like 'A%' ;

## 7. 便利な機能

Proc SQL には他にも様々な機能があるが、その便利な機能の一部について紹介する。また、Proc SQL の実装では、一般的なデータベースには無い、SAS システムとの親和性を高めるための独自拡張も行われており、これらの利用によって、SAS システムの強力な機能と連携して処理を行うことができる。

### 7.1. SQL ビュー (View)

SQL ビューとは、SQL で記述されたクエリが格納されたデータセットの一種である。Data ステップや Proc ステップから参照された際にクエリ (SQL) が実行され、実体の情報を参照できるようになる。データファイルのように実データを保持しないため記憶容量を節減できる。内容に対する更新や削除などの操作は、ビューにおいてクエリが参照している元のデータセットにも反映される。Proc SQL では作成に Create View 文を使用する。なお、Select 文等の From 句内でネストされたクエリ式をインラインビュー (Inline View) と呼び、外側のクエリがデータの選択に使用する一時テーブルを Proc SQL が内部生成して処理が行われる。当該クエリの実行が終われば一時テーブルは削除されるので、余分なプログラムステップを省き、不要なテーブルの増加を防ぐことができるメリットがある。詳細は、「SAS 9.4 SQL プロシジャ : ユーザーガイド」および「SAS 9.4 言語リファレンス」等の関連資料を参照されたい。

**【記述例】 View および Inline View**

- dataset1 を参照する SQL ビュー (dataview1) を作成する  
Create view dataview1 as select \* from dataset1 ;
- From 句内でのインラインビューの使用例  
Select a. id, a. name, b. addr  
from (select id, name from dataset1) as a,  
      (select id, addr from dataset2) as b  
where a. id=b. id ;

## 7.2. 索引 (Index)

テーブルの各行 (データセットのオブザベーション) に高速かつ効率的にアクセスするための情報を索引 (Index) と呼び、索引が作られていれば、他のデータステップやプロシジャにおける by 文でも使用され、ソートが不要になる。また、WHERE 句の処理を最適化、グループ処理を容易にすることができ、SQL クエリのみならず、SAS システムにおける効率的な処理の実行に貢献する。索引の種類には、1 つのキー変数 (列) の値のみを使用した単一インデックスと、2 つ以上のキー変数の値を使用した複合インデックスがあり、いずれも Create Index 文を使用して作成できる。索引については、データステップにおけるデータセットオプション (Index=) や、Proc DATASETS の MODIFY 文と INDEX CREATE 文でも作成できる。詳細は「SAS 9.4 言語リファレンス」や「Base SAS 9.4 プロシジャガイド」等の関連資料を参照されたい。

### 【記述例】索引 (Index)

- dataset1 の列 id に単一インデックス id を作成する  
Create index id on dataset1 (id) ;
- dataset1 の列 area および addr に複合インデックス places を作成する  
Create index places on dataset1 (area, addr) ;

## 7.3. データセットオプション

Proc SQL におけるテーブル (データセット) の指定部分では、ほとんどのデータセットオプション (KEEP= DROP= RENAME= OBS= WHERE= など) が利用可能である。データセットオプションの詳細は、「SAS 9.4 データセットオプション: リファレンス」等の関連資料を参照されたい。

### 【記述例】データセットオプション

- dataset2 の処理対象を先頭 15 行とし、dataset1 の列名を id から hhid に変更する  
Create table dataset1 (rename=(id=hhid)) as select \* from dataset2 (obs=15) ;

## 7.4. マクロ機能

データの内容や状況に応じてプログラムを自動生成するなど、高度な利用を想定する場合にはマクロ機能の利用は不可欠である。Proc SQL においても、Select 文における Into 句を用いて実行結果をマクロ変数へ出力したり、クエリの実行状況を格納した自動マクロ変数 (SQLOBS, SQLRC, など) を利用したりすることができる。マクロ機能の詳細は、「SAS 9.4 SQL プロシジャ: ユーザーガイド」および「SAS 9.4 マクロ言語: リファレンス」等の関連資料を参照されたい。

### 【記述例】INTO 句を使用したマクロ変数への出力

- クエリ結果の先頭行の値を格納するマクロ変数を作成する  
Select id, name, addr into :ids, :names, :addrs from dataset1 ;  
%put &ids &names &addrs ;
- 複数行の結果 (99 件分) を格納するマクロ変数を作成する  
Select id, name into :ids1-ids99, :names1-:names99 from dataset1 ;
- 値を連結してマクロ変数へ格納する  
Select name into :names separated by ',' from dataset1 ;  
/\* into 句の末尾 (from の前) に NOTRIM を指定すると空白を除去しない \*/

## 8. さいごに

データハンドリング、とりわけ、データセットの作成や整備に関わる部分で、Proc SQL は威力を発揮する。Proc SQL は多機能であるが万能ではないので、データステップや他のプロシジャの得意分野とうまく組み合わせた SAS プログラムの作成が不可欠である。

## 謝 辞

SAS ユーザー総会の世話人であり恩師でもある周防節雄・兵庫県立大学名誉教授にはさまざまな助言をいただいた。また、今回の報告は、先だって開催された九州地区 SAS ユーザー会（社会科学 SAS 分科会）における報告をベースにしたものであるが、そこでの参加者からも有用なご意見をいただいた。この場を借りて御礼申し上げたい。

## 参考文献

- ・古隅弘樹（2011）「企業データを統合するための名寄せ処理技法」 SAS ユーザー総会 アカデミア/テクノロジー&ソリューション 報告（2011年7月29日、於：神戸国際会議場）

## 参考資料

日本語版 SAS オンラインヘルプ（最終アクセス：2016.6.17）

<http://www.sas.com/offices/asiapacific/japan/service/onlinehelp.html>

- ・ SAS 9.4 SQL プロシジャ：ユーザーガイド
- ・ SAS 9.4 言語リファレンス：解説編 第3版
- ・ Base SAS 9.4 プロシジャガイド 第4版
- ・ SAS 9.4 出力形式と入力形式リファレンス
- ・ SAS 9.4 マクロ言語：リファレンス 第4版
- ・ SAS 9.4 データセットオプション：リファレンス 第2版

# JMPによる第4回 Let's データ分析の規定課題の解析

高橋 行雄  
BioStat 研究所(株)

JMP Manipulation to Solve the Compulsory Exercises  
for the 4th Micro Data Competition in SAS Japan Users Forum 2016

Yukio Takahashi  
BioStat Research Co.,Ltd.

**要旨：** 第35回 SAS ユーザー総会(2016年)にて、(独)統計センターの擬似マイクロデータを題材に第4回目の「Let's データ分析 ミクロデータ分析コンテスト」を開催した。残念ながら(独)統計センターからの擬似マイクロデータの提供が2016年3月末で中止されたことにより、本年が最終年となる。そこで、このコンテストの規定問題の出題者として、統計ソフト JMP を用いたマイクロデータの統計解析に関するチュートリアルとして、ボリュームが大きい社会調査データに対する統計解析の進め方、実際に JMP ではどのように行うのか、JMP の操作の基本にも触れつつ出題の意図も含めて解説を行うこととした。(独)統計センターから多くのマイクロデータが公開されるようになったことは喜ばしいことである。それらを用いて新たなチャレンジを行うためには、教科書的なアプローチを超えた様々な創意工夫が必要となるので、擬似マイクロデータを例にして、どのようにアプローチしたらよいか、具体的に例示した。

**キーワード：** 全国消費実態調査、擬似マイクロデータ、集計乗率、10万世帯比、分位点、JMP

## 1. はじめに

国の統計調査の集計結果については、冊子あるいは電子的に公開されており自由に利用することができるが、その元の個票データは本来その目的である統計作成以外の目的のために利用・提供できないとの縛りがあった。最近の統計法の改正を受けて、統計の研究や教育など公益に資するために使用される場合に限り、(独)統計センターが個人や企業が特定できない形に加工した匿名データの提供依頼を申出ることによって入手できるようになった。全国消費実態調査(全消)は、1989年、1994年、1999年、2004年の4回分の匿名データが公開されており、それらのデータボリュームは、カンマ区切りの固定長データで、それぞれ1,000メガバイトと膨大である。

これらの大規模な公開データを用いた統計解析を行う前のトレーニングとして「擬似マイクロデータ」を用いた「Let's データ分析 ミクロデータ分析コンテスト」を企画した<sup>1)</sup>。用いた擬似マイクロデータは、全国消費実態調査(全消)の2004年の結果を用いて、変数名間の相関係数を考慮して元のデータの特性が失われなように人工的に生成されたデータであり教育現場で自由に使い、マイクロデータ統計解析の教材となるように作成された<sup>2)</sup>。

とはいえ、擬似マイクロデータのデータは 32,027 件、197 変数でありファイルの容量は 32K×200×8 バイト=48,000 k バイト=48 メガバイトとなる。現代の標準的な PC の Excel でもデータを表示するためにも 20 秒弱とかかるが、近年、基本的なデータ操作が可能となったことは、喜ばしい。Excel は、現代の算盤としての地位は確立してはいるものの、自由自在にマイクロデータの統計解析を行うための機能は残念ながら持っていない。








長年にわたり多くの統計ソフトを筆者は、必要に応じて使い分けてきた。その中で、筆者の PC 上で生き残り日々用いて使用頻度が高い JMP を用いて、マイクロデータの統計解析を実践するための、基本的な操作手順を示す。また、規定問題をどのように意図で出題したのか、どのように作成したかを丁寧に解説しつつ、今後マイクロデータの統計解析を試みる人達の道標になることを願っている。出題した規定問題の作成に際して、全く独立に SAS を用いた検証<sup>3)</sup>が共同企画者の周防によって行なわれているので、目的は同じでもアプローチは異なることを理解しつつ、新たなチャレンジを期待したい。

擬似マイクロデータを用いた SAS でのプログラミング形式での文献<sup>4)</sup>、匿名データを用いた JMP での GUI 形式での文献<sup>5)</sup>に、全く異なるアプローチが示されている。同じベンダーから提供されている統計ソフトなので SAS で作成された SAS データセットを JMP で直接読み込み、GUI により各種のグラフを作成できるし、逆に JMP で作成された JMP データセットを SAS データセットとして保存でき、そのまま SAS で利用できる。SAS は、文献で提案されているあらゆる統計解析を実施できることを目指して日々拡張が続けているのに対して、JMP は SAS の標準的な統計解析をベースとして SAS を凌駕するグラフィカルな結果の表示とカスタマイズ機能が日々増強され提供されている。

## 2. 解析用データの作成

(独)統計センターからダウンロードした擬似マイクロデータは、5,000 行ごとの ZIP 圧縮ファイルであり表示 1 に示すファイルが含まれている。これらのファイルを開くために、(独)統計センターから別途パスワードを入手した。

表示 1 ZIP ファイル giji-microdata1 の内容

| 名前                                                                                                                      | 圧縮サイズ    | サイズ       | パスワード保護 |
|-------------------------------------------------------------------------------------------------------------------------|----------|-----------|---------|
|  GJI_2004zensho_dataset(00001~05000) | 4,319 KB | 14,532 KB | 有       |
|  GJI_2004zensho_guide                | 254 KB   | 273 KB    | 有       |
|  GJI_2004zensho_layout_code          | 56 KB    | 268 KB    | 有       |
|  GJI_2004zensho_questionnaire        | 11 KB    | 43 KB     | 有       |
|  GJI_2004zensho_summary_table        | 12 KB    | 41 KB     | 有       |
|  GJI_2004zensho_vname                | 1 KB     | 2 KB      | 有       |
|  資料一覧                                | 96 KB    | 103 KB    | 有       |

データは、カンマ区切り (CSV) 形式であるので、Excel で開くと表示 2 に示すように、データのみのファイルであり、7 ファイルに分割されている。

表示 2 GIJI\_2004zensho\_dataset(00001~05000).csv の最初の 10 行 21 変数

|    | A | B | C | D | E | F | G | H | I | J | K | L  | M | N  | O       | P       | Q      | R      | S      | T | U |
|----|---|---|---|---|---|---|---|---|---|---|---|----|---|----|---------|---------|--------|--------|--------|---|---|
| 1  | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | VV | 1 | 19 | 2873.03 | 370447  | 215047 | 211011 | 194776 | 0 | 0 |
| 2  | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | VV | 1 | 15 | 5055.37 | 1134350 | 591755 | 580650 | 535975 | 0 | 0 |
| 3  | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | VV | 1 | 15 | 3337.98 | 754758  | 431302 | 423208 | 390646 | 0 | 0 |
| 4  | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 5 | VV | 1 | 21 | 3264.17 | 459079  | 299831 | 299831 | 299831 | 0 | 0 |
| 5  | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 5 | VV | 1 | 9  | 4282.7  | 477208  | 280484 | 280484 | 280484 | 0 | 0 |
| 6  | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 5 | VV | 1 | 9  | 4009.38 | 427095  | 311915 | 311915 | 311915 | 0 | 0 |
| 7  | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 2 | VV | 3 | 17 | 5853.02 | 649739  | 339553 | 295731 | 295731 | 0 | 0 |
| 8  | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 2 | VV | 3 | 10 | 2207.22 | 568988  | 280663 | 244441 | 244441 | 0 | 0 |
| 9  | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 2 | VV | 3 | 11 | 2248.31 | 454048  | 221531 | 192940 | 192940 | 0 | 0 |
| 10 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 2 | VV | 3 | 14 | 3947.41 | 453148  | 285956 | 249051 | 249051 | 0 | 0 |

JMP で CSV 形式のファイルを読み込むと、変数名は、列 1, 列 2, ..., 列 197 と自動的に割り付けられる。全体で 7 ファイルを別々に読み込み、それらを JMP の【テーブル・連結】で行方向に結合し、32,027 人分の JMP ファイル (サイズは 47MB) を作成する。変数名が通し番号のままでは、迅速な解析も行えず、結果の出力も使いものにならない。何らかの意味を持つ変数名を付与する必要がある。変数名一覧は、GIJI\_2004zensho\_vname.csv にあり、表示 3 に示すように通し番号が振られており役に立たない。

表示 3 変数名リスト

|               |          |          |
|---------------|----------|----------|
| SetaiKubun    | Youto001 | Youto170 |
| SetaiJinin    | Youto002 | Youto171 |
| ShuugyouJinin | Youto003 | Youto172 |
| Kouzou        | Youto004 | Youto173 |
| Tatekata      | Youto005 | Youto174 |
| Shoyuu        | Youto006 | Youto175 |
| S1_Sex        | Youto007 | Youto176 |
| S1_Age        | Youto008 | Youto177 |
| S1_Shuugyou   | Youto009 | Youto178 |
| S1_Kigyoubun  | Youto010 | Youto179 |
| S1_Kigyoubun  | Youto011 | Youto180 |
| S1_Sangyou    | Youto012 | Youto181 |
| S1_Shokugyou  | Youto013 | Youto182 |
| Weight        | Youto014 | Youto183 |

そこで、表示 4 に示した GIJI\_2004zensho\_layout\_code.xls のレイアウトシートから、Excel の手作業で不要な行 (2, 3, 4, 7, 8, 9, ..., 526) を削除し、項目名の入っている行 (1, 5, 6, 8, ..., 527) を拾い出す。

表示 5 に示すように、行番号、階層、項目名、変数名を残し他は削除する。新たな通し変数名として、接頭語 V001, V002, ..., V0014, Y001, Y002, ..., Y183 を作成し、これに項目名を Concatenate 関数で結合し、合成\_変数名とする。

表示 4 擬似マイクロデータのレイアウトシート

| 行番号 | 項目名        | 階層 | 位置 | バイト数 | 配置 | 型 | 種別 | 変数名           | 符号  | 符号内容                | 備考 |
|-----|------------|----|----|------|----|---|----|---------------|-----|---------------------|----|
| 1   | 世帯区分       | 1  | 1  | 1    |    |   | 1  | SetaiKubun    | 1   | 勤労                  |    |
| 2   |            |    |    |      |    |   |    |               | 2   | 勤労以外                |    |
| 3   |            |    |    |      |    |   |    |               | 3   | 無職                  |    |
| 4   |            | 1  | 2  | 1    |    |   |    |               | .   | 項目の区切り              |    |
| 5   | 世帯票        | 1  |    |      |    |   |    |               |     |                     |    |
| 6   | 世帯人員       | 2  | 3  | 2    | 2  | 1 | 2  | SetaiJinin    | Δ2~ | 2人~                 |    |
| 7   |            | 2  | 5  | 1    |    |   |    |               | .   | 項目の区切り              |    |
| 8   | 有業人員       | 2  | 6  | 2    | 2  | 1 | 3  | ShuugyouJinin | Δ1~ | 1人~                 |    |
| 9   |            |    |    |      |    |   |    |               | VV  | 不詳                  |    |
| 10  |            | 2  | 8  | 1    |    |   |    |               | .   | 項目の区切り              |    |
| 11  | 現住居等に関する事項 | 2  |    |      |    |   |    |               |     |                     |    |
| 12  | 住居の構造      | 3  | 9  | 1    |    |   | 3  | Kouzou        | 1   | 木造                  |    |
| 13  |            |    |    |      |    |   |    |               | 2   | 防火木造                |    |
| 14  |            |    |    |      |    |   |    |               | 3   | 鉄骨・鉄筋コンクリート造        |    |
| 15  |            |    |    |      |    |   |    |               | 4   | その他（ブロック造り、レンガ造りなど） |    |
| 16  |            |    |    |      |    |   |    |               | V   | 不詳                  |    |
| 17  |            | 3  | 10 | 1    |    |   |    |               | .   | 項目の区切り              |    |

:

|     |      |   |      |    |   |   |  |          |          |        |  |
|-----|------|---|------|----|---|---|--|----------|----------|--------|--|
| 523 | 財産購入 | 4 | 2928 | 15 | 2 | 1 |  | Youto181 | ΔΔΔΔΔΔΔΔ | 円      |  |
| 524 |      | 4 | 2943 | 1  |   |   |  |          | 0.00000~ | 項目の区切り |  |
| 525 | その他  | 4 | 2944 | 15 | 2 | 1 |  | Youto182 | ΔΔΔΔΔΔΔΔ | 円      |  |
| 526 |      | 4 | 2959 | 1  |   |   |  |          | 0.00000~ | 項目の区切り |  |
| 527 | 繰越金  | 3 | 2960 | 15 | 2 | 1 |  | Youto183 | ΔΔΔΔΔΔΔΔ | 円      |  |
|     |      |   |      |    |   |   |  |          | 0.00000~ |        |  |

表示 5 変数名テーブル

| 行番号 | 階層 | 項目名   | 変数名           | 接頭語  | 合成_変数名     | JMP列番号 |
|-----|----|-------|---------------|------|------------|--------|
| 1   | 1  | 世帯区分  | SetaiKubun    | V001 | V001_世帯区分  | 列1     |
| 6   | 2  | 世帯人員  | SetaiJinin    | v002 | v002_世帯人員  | 列2     |
| 8   | 2  | 有業人員  | ShuugyouJinin | V003 | V003_有業人員  | 列3     |
| 12  | 3  | 住居の構造 | Kouzou        | V004 | V004_住居の構造 | 列4     |

:

|     |   |      |          |      |           |      |
|-----|---|------|----------|------|-----------|------|
| 523 | 4 | 財産購入 | Youto181 | Y181 | Y181_財産購入 | 列195 |
| 525 | 4 | その他  | Youto182 | Y182 | Y182_その他  | 列196 |
| 527 | 3 | 繰越金  | Youto183 | Y183 | Y183_繰越金  | 列197 |

結合した JMP ファイルを CSV 形式で保存し、表示 6 に示すように Excel の「行列を入れ替える」機能を用いて値のみを貼り付け保存し、このファイルを JMP で読み込む。

表示 6 Excel の 1 行目へ合成\_変数名の貼り付け

| V001_世帯区分 | 転置<br>張付<br>➔ | <table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> <th>H</th> <th>I</th> <th>J</th> <th>K</th> <th>L</th> <th>M</th> <th>N</th> <th>O</th> <th>GO</th> </tr> </thead> <tbody> <tr> <td></td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列</td> <td>列15</td> <td>列197</td> </tr> <tr> <td>1</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> <td>10</td> <td>11</td> <td>12</td> <td>13</td> <td>14</td> <td></td> <td></td> </tr> <tr> <td>2</td> <td>1</td> <td>2</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>2</td> <td>1</td> <td>1</td> <td>3</td> <td>VV</td> <td>1</td> <td>19</td> <td>2873.03</td> <td>21590.0402</td> </tr> <tr> <td>3</td> <td>1</td> <td>2</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>2</td> <td>1</td> <td>1</td> <td>3</td> <td>VV</td> <td>1</td> <td>15</td> <td>5055.37</td> <td>47930.0771</td> </tr> <tr> <td>4</td> <td>1</td> <td>2</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>2</td> <td>1</td> <td>1</td> <td>3</td> <td>VV</td> <td>1</td> <td>15</td> <td>3337.98</td> <td>59806.9563</td> </tr> <tr> <td>5</td> <td>1</td> <td>2</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>2</td> <td>1</td> <td>1</td> <td>5</td> <td>VV</td> <td>1</td> <td>21</td> <td>3264.17</td> <td>23930.8901</td> </tr> <tr> <td>32028</td> <td>1</td> <td>3</td> <td>1</td> <td>3</td> <td>5</td> <td>7</td> <td>1</td> <td>4</td> <td>1</td> <td>1</td> <td>5</td> <td>5</td> <td>3</td> <td>23</td> <td>13592.7</td> <td>55098.2714</td> </tr> </tbody> </table> |   | A | B | C | D | E | F | G | H | I  | J  | K  | L  | M  | N       | O          | GO |  | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列15 | 列197 | 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |  |  | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | VV | 1 | 19 | 2873.03 | 21590.0402 | 3 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | VV | 1 | 15 | 5055.37 | 47930.0771 | 4 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 3 | VV | 1 | 15 | 3337.98 | 59806.9563 | 5 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 5 | VV | 1 | 21 | 3264.17 | 23930.8901 | 32028 | 1 | 3 | 1 | 3 | 5 | 7 | 1 | 4 | 1 | 1 | 5 | 5 | 3 | 23 | 13592.7 | 55098.2714 |
|-----------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|---------|------------|----|--|---|---|---|---|---|---|---|---|---|---|---|---|---|---|-----|------|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|--|--|---|---|---|---|---|---|---|---|---|---|---|---|----|---|----|---------|------------|---|---|---|---|---|---|---|---|---|---|---|---|----|---|----|---------|------------|---|---|---|---|---|---|---|---|---|---|---|---|----|---|----|---------|------------|---|---|---|---|---|---|---|---|---|---|---|---|----|---|----|---------|------------|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---------|------------|
|           |               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | A | B | C | D | E | F | G | H | I | J  | K  | L  | M  | N  | O       | GO         |    |  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |      |   |   |   |   |   |   |   |   |   |   |    |    |    |    |    |  |  |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |       |   |   |   |   |   |   |   |   |   |   |   |   |   |    |         |            |
|           |               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列 | 列  | 列  | 列  | 列  | 列  | 列15     | 列197       |    |  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |      |   |   |   |   |   |   |   |   |   |   |    |    |    |    |    |  |  |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |       |   |   |   |   |   |   |   |   |   |   |   |   |   |    |         |            |
| 1         |               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |         |            |    |  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |      |   |   |   |   |   |   |   |   |   |   |    |    |    |    |    |  |  |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |       |   |   |   |   |   |   |   |   |   |   |   |   |   |    |         |            |
| 2         |               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1  | 3  | VV | 1  | 19 | 2873.03 | 21590.0402 |    |  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |      |   |   |   |   |   |   |   |   |   |   |    |    |    |    |    |  |  |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |       |   |   |   |   |   |   |   |   |   |   |   |   |   |    |         |            |
| 3         |               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1  | 3  | VV | 1  | 15 | 5055.37 | 47930.0771 |    |  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |      |   |   |   |   |   |   |   |   |   |   |    |    |    |    |    |  |  |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |       |   |   |   |   |   |   |   |   |   |   |   |   |   |    |         |            |
| 4         |               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1  | 3  | VV | 1  | 15 | 3337.98 | 59806.9563 |    |  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |      |   |   |   |   |   |   |   |   |   |   |    |    |    |    |    |  |  |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |       |   |   |   |   |   |   |   |   |   |   |   |   |   |    |         |            |
| 5         |               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1  | 5  | VV | 1  | 21 | 3264.17 | 23930.8901 |    |  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |      |   |   |   |   |   |   |   |   |   |   |    |    |    |    |    |  |  |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |       |   |   |   |   |   |   |   |   |   |   |   |   |   |    |         |            |
| 32028     |               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 1 | 3 | 1 | 3 | 5 | 7 | 1 | 4 | 1 | 1  | 5  | 5  | 3  | 23 | 13592.7 | 55098.2714 |    |  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |      |   |   |   |   |   |   |   |   |   |   |    |    |    |    |    |  |  |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |   |   |   |   |   |   |   |   |   |   |   |   |    |   |    |         |            |       |   |   |   |   |   |   |   |   |   |   |   |   |   |    |         |            |



変数名が名義尺度の場合に、数値などにコード化されているので、元のカテゴリ名を集計表に反映するために JMP では、【値ラベル】の機能を用いて対応することができるが、手作業での対応となるので、元のデータそのものを一括して入れ替えることにする。そのためには、表示 7 に示すように、全ての合成\_変数名に対して符号と符号内容を一括した JMP ファイルとしておく。

表示 7 符号テーブル

| 行番号 | 合成_変数名         | 変数名         | 符号 | 符号データ             |
|-----|----------------|-------------|----|-------------------|
| 26  | 40 V007_性別     | S1_Sex      | 1  | 1_男               |
| 27  | 41 V007_性別     | S2_Sex      | 2  | 2_女               |
| 28  | 43 V008_年齢5歳階級 | S1_Age      | 1  | 01_24歳未満          |
| ⋮   |                |             |    |                   |
| 58  | 77 V012_産業符号   | S1_Sangyou  | 1  | 01_農業             |
| 59  | 78 V012_産業符号   | S2_Sangyou  | 2  | 02_林業             |
| 60  | 79 V012_産業符号   | S3_Sangyou  | 3  | 03_漁業             |
| ⋮   |                |             |    |                   |
| 75  | 94 V012_産業符号   | S18_Sangyou | 18 | 18_公務(他に分類されないもの) |
| 76  | 95 V012_産業符号   | S19_Sangyou | 19 | 19_その他(非就業を含む)    |
| 77  | 96 V012_産業符号   | S20_Sangyou | VV | VV_不詳             |

合成\_変数名の JMP ファイルに表示 8 に示すように通し番号の列「No」を追加して、「V001\_世帯区分」から「V013\_職業符号」13 変数の変数タイプを「文字」に統一し、行ごとに「列の積み重ね」で、縦方向に展開する。積み重ねたデータに対し、符号テーブルを【結合 (Join)】で符号データを追加する。

表示 8 列を積み重ねたデータに符号データを結合

| No     | 合成_変数           | データ | 符号データa       | 符号データb       |
|--------|-----------------|-----|--------------|--------------|
| 1      | 1 V001_世帯区分     | 1   | 1_勤労         | 1_勤労         |
| 2      | 1 V002_世帯票      | 2   |              | 2            |
| 3      | 1 V003_有業人員     | 1   |              | 1            |
| 4      | 1 V004_住居の構造    | 1   | 1_木造         | 1_木造         |
| 5      | 1 V005_住居の建て方   | 1   | 1_一戸建        | 1_一戸建        |
| 6      | 1 V006_住居の所有関係  | 1   | 1_持ち家(世帯員名義) | 1_持ち家(世帯員名義) |
| 7      | 1 V007_性別       | 1   | 1_男          | 1_男          |
| 8      | 1 V008_年齢5歳階級   | 2   | 02_25~29歳    | 02_25~29歳    |
| 9      | 1 V009_就業・非就業   | 1   | 1_就業         | 1_就業         |
| 10     | 1 V010_企業区分     | 1   | 1_民営         | 1_民営         |
| 11     | 1 V011_企業規模     | 3   | 3_30~499人    | 3_30~499人    |
| 12     | 1 V012_産業符号     | VV  | VV_不詳        | VV_不詳        |
| 13     | 1 V013_職業符号     | 1   | 01_常用労務作業者   | 01_常用労務作業者   |
| 14     | 2 V001_世帯区分     | 1   | 1_勤労         | 1_勤労         |
| 15     | 2 V002_世帯票      | 2   |              | 2            |
| 416349 | 32027 V011_企業規模 | V   | V_不詳         | V_不詳         |
| 416350 | 32027 V012_産業符号 | VV  | VV_不詳        | VV_不詳        |
| 416351 | 32027 V013_職業符号 | VV  | VV_不詳        | VV_不詳        |

このようにして得られた 416,351 行の符号データを、元の 32,027 行に【列の分割】で戻す。この符号データを元に戻すためには、【結合 (Join)】で行っても良いし、置き換える列を削除して【連

結】してもよい。表示9に規定問題の作成に関連する変数を選択した結果を示す。

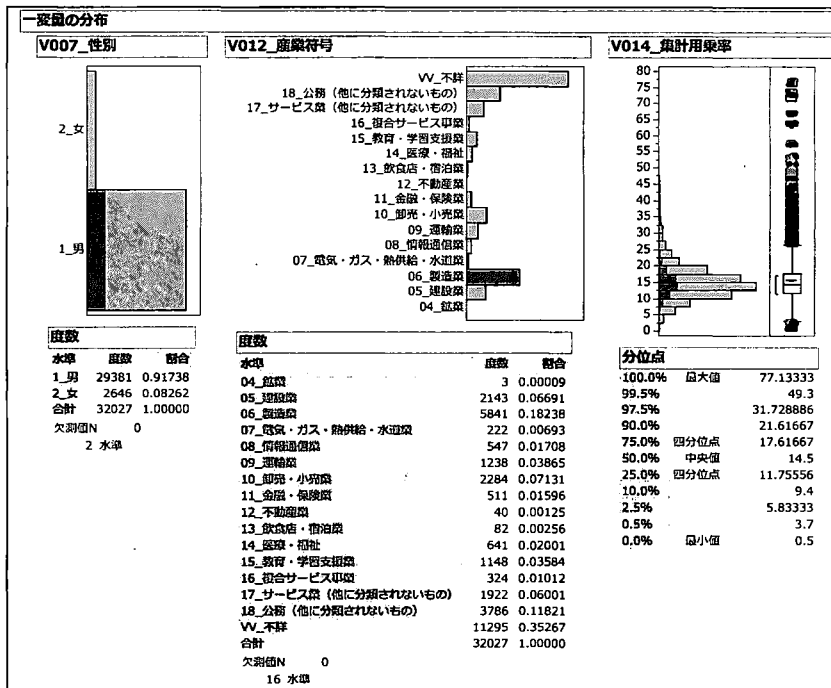
表示 9 全変数から規定問題を作成するためのサブセット

| No    | V007_性別  | V012_産業符号  | V014_集計用乗率 | V014_10万分比 | Y004_経常収入 | Y038_食料 | Y004_経常5分位 | Y038_食料5分位 | Y038_食料log10 |
|-------|----------|------------|------------|------------|-----------|---------|------------|------------|--------------|
| 1     | 1_1男     | VV_不詳      | 18.9000    | 3.8146     | 211011    | 37942   | 1          | 1          | 4.5791       |
| 2     | 2_1男     | VV_不詳      | 15.4333    | 3.1149     | 580650    | 68063   | 4          | 3          | 4.8329       |
| 3     | 3_1男     | VV_不詳      | 15.4333    | 3.1149     | 423208    | 67993   | 3          | 3          | 4.8325       |
| 3948  | 3948_2女  | 05_建設業     | 17.8444    | 3.6016     | 282118    | 55288   | 1          | 2          | 4.7426       |
| 3949  | 3949_1男  | 10_卸売・小... | 14.9667    | 3.0207     | 242525    | 46790   | 1          | 1          | 4.6702       |
| 3950  | 3950_1男  | 10_卸売・小... | 12.2333    | 2.4691     | 284500    | 47359   | 1          | 1          | 4.6754       |
| 3951  | 3951_1男  | 10_卸売・小... | 14.9667    | 3.0207     | 342233    | 47752   | 2          | 1          | 4.6790       |
| 3952  | 3952_1男  | 06_製造業     | 15.5000    | 3.1284     | 639642    | 46342   | 5          | 1          | 4.6660       |
| 32024 | 32024_1男 | 05_建設業     | 12.1500    | 2.4522     | 495115    | 123418  | 3          | 5          | 5.0914       |
| 32025 | 32025_1男 | VV_不詳      | 10.9000    | 2.2000     | 204448    | 47891   | 1          | 1          | 4.6803       |
| 32026 | 32026_1男 | VV_不詳      | 15.6000    | 3.1486     | 174699    | 114082  | 1          | 5          | 5.0572       |
| 32027 | 32027_1男 | VV_不詳      | 6.2000     | 1.2513     | 105224    | 45739   | 1          | 1          | 4.6603       |

### 3. JMPによる解析

JMPによる解析作業は、【一変量の分布】から始めて、表示10に示すようにデータの特徴を把握すると共に、一部のヒストグラム「V012\_産業符号=06\_製造業」を選択することにより、他の変数での選択したデータの分布がハイライトされている。

表示 10 JMPの「一変量の分布」によるヒストグラム



## 規定課題 1) 産業符号・性別 クロス表

規定問題の表1は、産業符号別世帯主の性別の単純クロス表及び10万世帯比でのクロス表である。単純クロス表の作成の基本は、【二変量の関係】で10万世帯あたりのクロス表を作成するために「集計乗率」を用いる。集計乗率の合計は、495464.76574なので、 $100,000/495464.7657=0.201830699001665$ をそれぞれの集計乗率に掛ければよい。ただし、それらの合計は完全に100,000ではなく、99999.95とわずかに小さい。10万世帯比のクロス表の作成には、【二変量の関係】で、「V014\_10万分比」を【度数】として用いればよい。結果を表1aに示す。

表1a 【二変量の関係】を使ったクロス表

単純クロス表

| 分割表                  |       |      |       |
|----------------------|-------|------|-------|
| V007_性別              |       |      |       |
| 度数                   | 1_男   | 2_女  | 合計    |
| 04_鉱業                | 3     | 0    | 3     |
| 05_建設業               | 2134  | 9    | 2143  |
| 06_製造業               | 5668  | 173  | 5841  |
| 07_電気・ガス・熱供給・水道業     | 222   | 0    | 222   |
| 08_情報通信業             | 547   | 0    | 547   |
| 09_運輸業               | 1235  | 3    | 1238  |
| 10_卸売・小売業            | 2083  | 201  | 2284  |
| 11_金融・保険業            | 472   | 39   | 511   |
| 12_不動産業              | 40    | 0    | 40    |
| 13_飲食店・宿泊業           | 50    | 32   | 82    |
| 14_医療・福祉             | 371   | 270  | 641   |
| 15_教育・学習支援業          | 1059  | 89   | 1148  |
| 16_複合サービス事業          | 321   | 3    | 324   |
| 17_サービス業（他に分類されないもの） | 1816  | 106  | 1922  |
| 18_公務（他に分類されないもの）    | 3656  | 130  | 3786  |
| VV_不詳                | 9704  | 1591 | 11295 |
| 合計                   | 29381 | 2646 | 32027 |

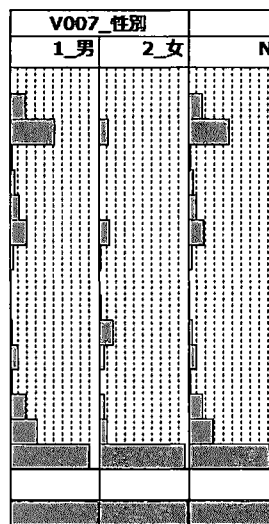
10万世帯比でのクロス表

| 分割表                  |       |      |        |
|----------------------|-------|------|--------|
| V007_性別              |       |      |        |
| 度数                   | 1_男   | 2_女  | 合計     |
| 04_鉱業                | 7     | 0    | 7      |
| 05_建設業               | 6346  | 31   | 6377   |
| 06_製造業               | 17837 | 510  | 18347  |
| 07_電気・ガス・熱供給・水道業     | 697   | 0    | 697    |
| 08_情報通信業             | 2020  | 0    | 2020   |
| 09_運輸業               | 3852  | 6    | 3858   |
| 10_卸売・小売業            | 6644  | 589  | 7233   |
| 11_金融・保険業            | 1447  | 104  | 1551   |
| 12_不動産業              | 144   | 0    | 144    |
| 13_飲食店・宿泊業           | 158   | 107  | 265    |
| 14_医療・福祉             | 1022  | 810  | 1832   |
| 15_教育・学習支援業          | 3110  | 269  | 3379   |
| 16_複合サービス事業          | 981   | 11   | 992    |
| 17_サービス業（他に分類されないもの） | 5842  | 326  | 6167   |
| 18_公務（他に分類されないもの）    | 10691 | 430  | 11121  |
| VV_不詳                | 31204 | 4806 | 36011  |
| 合計                   | 92000 | 8000 | 100000 |

表1b 【表の作成】による10万世帯比でのクロス表とチャート

表の作成

| V012_産業符号            | V007_性別 |      | N      |
|----------------------|---------|------|--------|
|                      | 1_男     | 2_女  |        |
| 04_鉱業                | 7       | 0    | 7      |
| 05_建設業               | 6346    | 31   | 6377   |
| 06_製造業               | 17837   | 510  | 18347  |
| 07_電気・ガス・熱供給・水道業     | 697     | 0    | 697    |
| 08_情報通信業             | 2020    | 0    | 2020   |
| 09_運輸業               | 3852    | 6    | 3858   |
| 10_卸売・小売業            | 6644    | 589  | 7233   |
| 11_金融・保険業            | 1447    | 104  | 1551   |
| 12_不動産業              | 144     | 0    | 144    |
| 13_飲食店・宿泊業           | 158     | 107  | 265    |
| 14_医療・福祉             | 1022    | 810  | 1832   |
| 15_教育・学習支援業          | 3110    | 269  | 3379   |
| 16_複合サービス事業          | 981     | 11   | 992    |
| 17_サービス業（他に分類されないもの） | 5842    | 326  | 6167   |
| 18_公務（他に分類されないもの）    | 10691   | 430  | 11121  |
| VV_不詳                | 31204   | 4806 | 36011  |
| N                    |         |      |        |
| T                    | 92000   | 8000 | 100000 |



JMP でクロス表の作成は、【要約】による方法、あるいは【表の作成】によっても作成することができる。ただし、縦横の合計を求めるためには、若干の追加作業が必要となる。【二変量の関係】で別々に出力した結果を規定課題で設定したように併合表を JMP だけで実現できないので、【表の作成】で単純クロス表および表 1b に示した 10 万世帯比でのクロス表を別々に作成し、結果を JMP ファイルに出力し、その結果を Excel に貼り付けて表 1c に示すように、縦方向の合計を出すことにする。

Excel の機能で縦方向の合計を算出する場合に、10 万世帯比の場合には、クロス表内のセルの度数を縦方向に足し合わせた場合には、それぞれのセルの表示が四捨五入されて表示されているので、表示されている数字を足し合わせても合計が 100,000 とはならないことに注意が必要である。

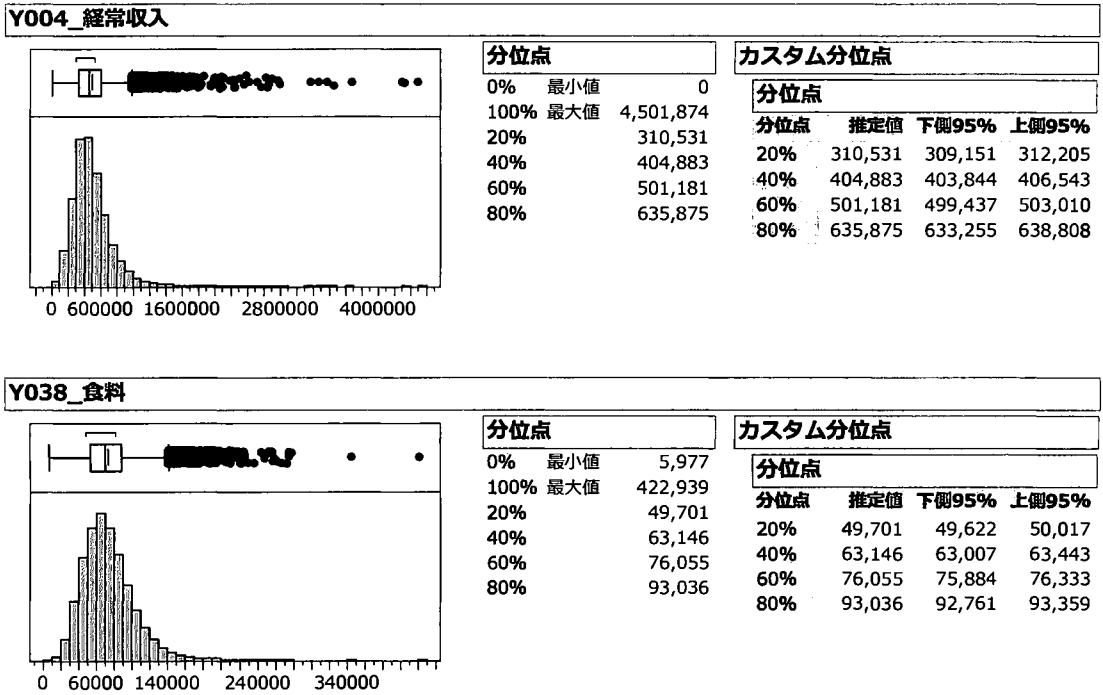
表 1c JMP で基本集計を行い Excel で整形した結果

| 産業符号              | a) 単純クロス表 |       |        | b) 10万世帯比でのクロス表 |       |         |
|-------------------|-----------|-------|--------|-----------------|-------|---------|
|                   | 男         | 女     | 合計     | 男               | 女     | 合計      |
| 鉱業                | 3         | 0     | 3      | 7               | 0     | 7       |
| 建設業               | 2,134     | 9     | 2,143  | 6,346           | 31    | 6,377   |
| 製造業               | 5,668     | 173   | 5,841  | 17,837          | 510   | 18,347  |
| 電気・ガス・熱供給・水道業     | 222       | 0     | 222    | 697             | 0     | 697     |
| 情報通信業             | 547       | 0     | 547    | 2,020           | 0     | 2,020   |
| 運輸業               | 1,235     | 3     | 1,238  | 3,852           | 6     | 3,858   |
| 卸売・小売業            | 2,083     | 201   | 2,284  | 6,644           | 589   | 7,233   |
| 金融・保険業            | 472       | 39    | 511    | 1,447           | 104   | 1,551   |
| 不動産業              | 40        | 0     | 40     | 144             | 0     | 144     |
| 飲食店・宿泊業           | 50        | 32    | 82     | 158             | 107   | 265     |
| 医療・福祉             | 371       | 270   | 641    | 1,022           | 810   | 1,832   |
| 教育・学習支援業          | 1,059     | 89    | 1,148  | 3,110           | 269   | 3,379   |
| 複合サービス事業          | 321       | 3     | 324    | 981             | 11    | 992     |
| サービス業(他に分類されないもの) | 1,816     | 106   | 1,922  | 5,842           | 326   | 6,167   |
| 公務(他に分類されないもの)    | 3,656     | 130   | 3,786  | 10,691          | 430   | 11,121  |
| 不詳                | 9,704     | 1,591 | 11,295 | 31,204          | 4,806 | 36,011  |
| 合計                | 29,381    | 2,646 | 32,027 | 92,000          | 8,000 | 100,000 |

## 規定課題 2) 経常収入および食料費

各種の統計調査の結果の表示で 10 階級区分が標準的に用いられている。探索的な統計解析を行うためには煩雑であり、また統計量として 4 分位点が標準的ではあるが、中央値前後で 2 区分されてしまう。そこで、中央値前後を 1 区分とするために 5 分位に着目した。JMP の【一変量の分布】での標準出力結果に対して【表示オプション・(分位点の間隔の設定, 任意の分位点)】の設定で表 2a に示すように求めることができる。課題では集計乗率を使うように指示しているが、10 万世帯比を用いても全く同じ結果となる。

表 2a JMP で求めた 5 分位点



JMP で求めた 5 分位点を JMP データとして出力し、Excel のシートに張りつけた結果を表 2b に示す。

表 2b Excel シートに貼り付けた結果

|     | 経常収入    | 食料費    |
|-----|---------|--------|
| 分位点 | 推定値     | 推定値    |
| 20% | 310,531 | 49,701 |
| 40% | 404,883 | 63,146 |
| 60% | 501,181 | 76,055 |
| 80% | 635,875 | 93,036 |

### 規定課題 3) 男女別・経常収入および食料費の 5 分位分布

経常収入および食料費を 5 分位数によってカテゴリ化する方法は、集計乗率を考慮しないのであれば、列内の数値に対する順位付けの Col Rank 関数および整数化する Ceiling 関数を用いて

$$\text{経常収支 5 分位} = \text{Ceiling}(\text{Col Rank}(Y004\_経常収入))$$

のように数値計算で求めることができる。しかしながら、10 万世帯比とした場合には、データ数を 10 万世帯倍とするような重みつき順位を用いる必要があり、簡単な関数計算での求め方は、思いつ

かない。そこで、最も素朴な If 関数で実現することとした。この結果は、表示 9 にすでに示してある。

経常収支 5 分位=

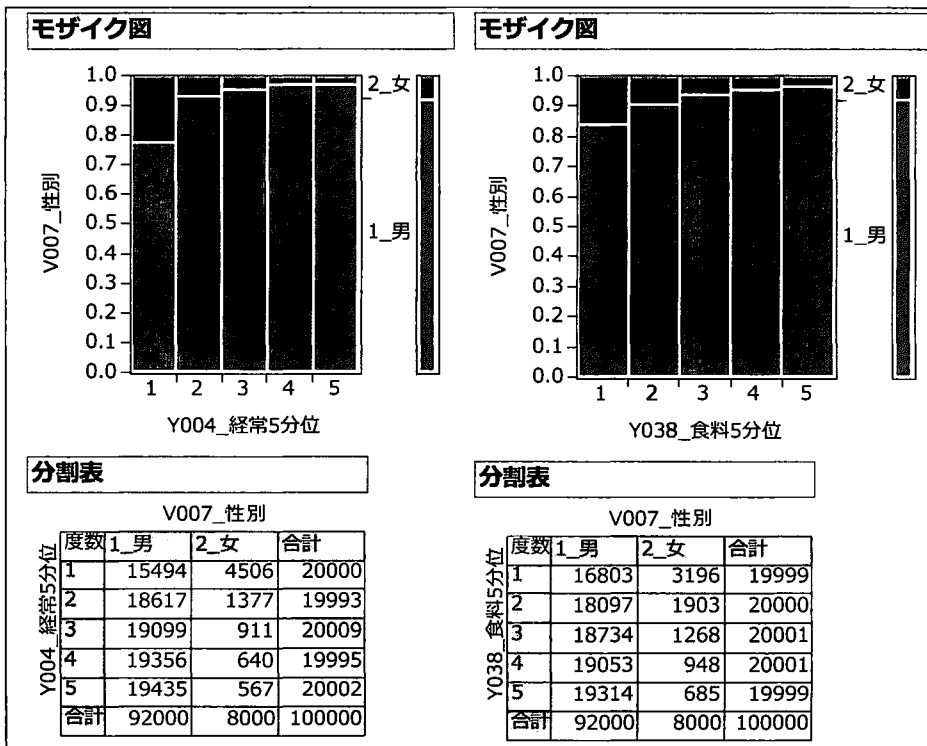
|                     |     |
|---------------------|-----|
| Y004_経常収入 < 310531  | ⇒ 1 |
| Y004_経常収入 < 404883  | ⇒ 2 |
| Y004_経常収入 < 501181  | ⇒ 3 |
| Y004_経常収入 < 635875  | ⇒ 4 |
| Y004_経常収入 < 9999999 | ⇒ 5 |
| else                | ⇒ . |

食料費 5 分位=

|                   |     |
|-------------------|-----|
| Y038_食料 < 49701   | ⇒ 1 |
| Y038_食料 < 63146   | ⇒ 2 |
| Y038_食料 < 76055   | ⇒ 3 |
| Y038_食料 < 93036   | ⇒ 4 |
| Y038_食料 < 9999999 | ⇒ 5 |
| else              | ⇒ . |

性別の経常収入 5 分位および食料費 5 分位の別々のクロス表なので、「二変量の関係」を用い帯グラフも同時に示す。

表 3a 【二変量の関係】によるクロス表および縦棒帯グラフ



この課題は、10万世帯比での性別は共通な2つのクロス表を併合する課題である。【二変量の関係】では、別々のクロス表であるが、【表の作成】を用いれば、表 3b に示すように複合クロス表を作成することができる。ただし、10万世帯比を度数としたので集計結果は、整数でなく実数で表示

されるので、【形式の変更】で小数点以下を0とした固定小数点表示を選び、形式的に整数表示とした。

表 3a 【表の作成】による複合クロス表

| V007_性別    |       |      |        |
|------------|-------|------|--------|
| Y004_経常5分位 | 1_男   | 2_女  | すべて    |
| 1          | 15494 | 4506 | 20000  |
| 2          | 18617 | 1377 | 19993  |
| 3          | 19099 | 911  | 20009  |
| 4          | 19356 | 640  | 19995  |
| 5          | 19435 | 567  | 20002  |
| すべて        | 92000 | 8000 | 100000 |

| V007_性別    |       |      |        |
|------------|-------|------|--------|
| Y004_経常5分位 | 1_男   | 2_女  | すべて    |
| 1          | 15494 | 4506 | 20000  |
| 2          | 18617 | 1377 | 19993  |
| 3          | 19099 | 911  | 20009  |
| 4          | 19356 | 640  | 19995  |
| 5          | 19435 | 567  | 20002  |
| すべて        | 92000 | 8000 | 100000 |

| V007_性別    |       |      |        |
|------------|-------|------|--------|
| Y038_食料5分位 | 1_男   | 2_女  | すべて    |
| 1          | 16803 | 3196 | 19999  |
| 2          | 18097 | 1903 | 20000  |
| 3          | 18734 | 1268 | 20001  |
| 4          | 19053 | 948  | 20001  |
| 5          | 19314 | 685  | 19999  |
| すべて        | 92000 | 8000 | 100000 |

【二変量の関係】で作成されたクロス表を JMP ファイルに出力した結果には、縦横の計がカットされているので、Excel の合計の関数を使って計算する必要がある。【表の作成】の場合には、「すべて」も含まれるので、そのまま張付けられる。

表 3c Excel にデータを取り込み縦横の計を計算

| 経常収入 |           |        |       |         | 食料費 |          |        |       |         |
|------|-----------|--------|-------|---------|-----|----------|--------|-------|---------|
| 5分位  | 5分位数      | 男      | 女     | 計       | 5分位 | 5分位数     | 男      | 女     | 計       |
| 1    | 310,531未満 | 15,494 | 4,506 | 20,000  | 1   | 49,701未満 | 16,803 | 3,196 | 19,999  |
| 2    | 310,531以上 | 18,617 | 1,377 | 19,993  | 2   | 49,701以上 | 18,097 | 1,903 | 20,000  |
| 3    | 404,883以上 | 19,099 | 911   | 20,009  | 3   | 63,146以上 | 18,734 | 1,268 | 20,001  |
| 4    | 501,181以上 | 19,356 | 640   | 19,995  | 4   | 76,055以上 | 19,053 | 948   | 20,001  |
| 5    | 635,875以上 | 19,435 | 567   | 20,002  | 5   | 93,036以上 | 19,314 | 685   | 19,999  |
|      | 計         | 92,000 | 8,000 | 100,000 |     | 計        | 92,000 | 8,000 | 100,000 |

食料費の5分位数に出題ミスがあり横計を一部訂正した

#### 規定課題 4) 男女別・経常収入・食料費 についての3重クロス表

JMP での多重クロス表は、【表の作成】が基本である。10万世帯比を度数としたので、3重クロス表は、整数でなく実数で表示されるので、【形式の変更】で小数点以下を0とした固定小数点表示を選び、形式的に整数表示とした結果を表 4a に示す。さらに、性別ごとの計、全体の計を入れるために【集計統計量の追加】を選択する。

【表の作成】に引き続き、【データテーブルに出力】を実行し、結果を JMP テーブルに出力し、必要な部分を選択コピーし、表 4b に示すように Excel シートに貼り付ける。

表 4a 【表の作成】による3重クロス表の作成

| 表の作成    |            | Y038_食料5分位 |       |       |       |       |        |  |
|---------|------------|------------|-------|-------|-------|-------|--------|--|
| V007_性別 | Y004_経常5分位 | 1          | 2     | 3     | 4     | 5     | すべて    |  |
| 1_男     | 1          | 5827       | 3974  | 2798  | 1834  | 1060  | 15494  |  |
|         | 2          | 5130       | 4501  | 3978  | 2996  | 2012  | 18617  |  |
|         | 3          | 2911       | 4301  | 4395  | 4294  | 3198  | 19099  |  |
|         | 4          | 1906       | 3233  | 4082  | 4903  | 5231  | 19356  |  |
|         | 5          | 1029       | 2087  | 3481  | 5026  | 7813  | 19435  |  |
|         | すべて        | 16803      | 18097 | 18734 | 19053 | 19314 | 92000  |  |
| 2_女     | 1          | 2462       | 994   | 565   | 314   | 172   | 4506   |  |
|         | 2          | 371        | 399   | 235   | 211   | 161   | 1377   |  |
|         | 3          | 243        | 225   | 216   | 164   | 63    | 911    |  |
|         | 4          | 68         | 134   | 156   | 143   | 140   | 640    |  |
|         | 5          | 53         | 152   | 96    | 117   | 149   | 567    |  |
|         | すべて        | 3196       | 1903  | 1268  | 948   | 685   | 8000   |  |
| すべて     | すべて        | 19999      | 20000 | 20001 | 20001 | 19999 | 100000 |  |

表 4b Excelによる3重クロス表の整形

|     | 経常収入   | 食料(5分位) |        |        |        |         | 計      |
|-----|--------|---------|--------|--------|--------|---------|--------|
|     | 5分位    | 1       | 2      | 3      | 4      | 5       |        |
| 男   | 1      | 5,827   | 3,974  | 2,798  | 1,834  | 1,060   | 15,494 |
|     | 2      | 5,130   | 4,501  | 3,978  | 2,996  | 2,012   | 18,617 |
|     | 3      | 2,911   | 4,301  | 4,395  | 4,294  | 3,198   | 19,099 |
|     | 4      | 1,906   | 3,233  | 4,082  | 4,903  | 5,231   | 19,356 |
|     | 5      | 1,029   | 2,087  | 3,481  | 5,026  | 7,813   | 19,435 |
|     | 計      | 16,803  | 18,097 | 18,734 | 19,053 | 19,314  | 92,000 |
| 女   | 1      | 2,462   | 994    | 565    | 314    | 172     | 4,506  |
|     | 2      | 371     | 399    | 235    | 211    | 161     | 1,377  |
|     | 3      | 243     | 225    | 216    | 164    | 63      | 911    |
|     | 4      | 68      | 134    | 156    | 143    | 140     | 640    |
|     | 5      | 53      | 152    | 96     | 117    | 149     | 567    |
|     | 計      | 3,196   | 1,903  | 1,268  | 948    | 685     | 8,000  |
| 男女計 | 19,999 | 20,000  | 20,001 | 20,001 | 19,999 | 100,000 |        |

規定課題 5) 男女別・経常収入別に食料費を常用対数変換した平均値の比較

性別の経常収入5分位についての10万世帯比のクロス表は、すでに規定課題3)作成した結果に一致する。食料費の常用対数変換は、Log10関数を用いる。幾何平均の計算は、別途行わなければならないので、【要約】で対応することにする。

【要約】では、10万世帯比を【度数】に、性別および経常収支5分位を【グループ化】に、常用対数(食料費)を統計量欄から【N】および【平均】を選択して、【統計量】に入れる。結果はJMPデータとして得られるので、幾何平均は、



幾何平均 =  $\sqrt[10]{\text{平均}(Y038\_食料\log10)}$

で別途計算する。性別の統計量は計算されていないので、【要約】によって性別のみで計算し、表 5a に示すように手作業で JMP テーブルに挿入する。この結果をコピーし、Excel シートに張りつけた結果を表 5b に示す。

表 5a JMP による幾何平均

| GUI 全変数のサブセット2の要約 (V007_性別、Y004_経常5分位) - JMP                            |         |            |       |                 |                  |        |        |        |
|-------------------------------------------------------------------------|---------|------------|-------|-----------------|------------------|--------|--------|--------|
| ファイル(F) 編集(E) テーブル(T) 行(R) 列(C) 実験計画 (DOE)(D) 分析(A) グラフ(G) ツール(O) 表示(S) |         |            |       |                 |                  |        |        |        |
| GUI 全変数のサブ...                                                           |         |            |       |                 |                  |        |        |        |
| ソース                                                                     |         |            |       |                 |                  |        |        |        |
| 重ね合わせプロット                                                               |         |            |       |                 |                  |        |        |        |
|                                                                         | V007_性別 | Y004_経常5分位 | 行数    | N(Y038_食料log10) | 平均(Y038_食料log10) | 幾何平均   |        |        |
|                                                                         | 1 1_男   | 1          | 5017  | 15,494          | 4.7389           | 54,814 |        |        |
|                                                                         | 2 1_男   | 2          | 5924  | 18,617          | 4.7832           | 60,704 |        |        |
|                                                                         | 3 1_男   | 3          | 6084  | 19,099          | 4.8368           | 68,675 |        |        |
|                                                                         | 4 1_男   | 4          | 6161  | 19,356          | 4.8826           | 76,311 |        |        |
|                                                                         | 5 1_男   | 5          | 6195  | 19,435          | 4.9335           | 85,793 |        |        |
|                                                                         | 6 1_男   | •          | 29381 | 92,000          | 4.8395           | 69,107 |        |        |
|                                                                         | 7 2_女   | 1          | 1546  | 4,506           | 4.6713           | 46,913 |        |        |
|                                                                         | 8 2_女   | 2          | 424   | 1,377           | 4.7811           | 60,405 |        |        |
|                                                                         | 9 2_女   | 3          | 289   | 911             | 4.7813           | 60,439 |        |        |
|                                                                         | 10 2_女  | 4          | 209   | 640             | 4.8617           | 72,730 |        |        |
| すべての行                                                                   | 12      | 10         | 2_女   | 4               | 209              | 640    | 4.8617 | 72,730 |
| 選択されている行                                                                | 2       | 11         | 2_女   | 5               | 178              | 567    | 4.8774 | 75,407 |
| 除外されている行                                                                | 0       | 12         | 2_女   | •               | 2646             | 8,000  | 4.7325 | 54,018 |
| 表示しない行                                                                  | 0       |            |       |                 |                  |        |        |        |
| ラベルのついた行                                                                | 0       |            |       |                 |                  |        |        |        |

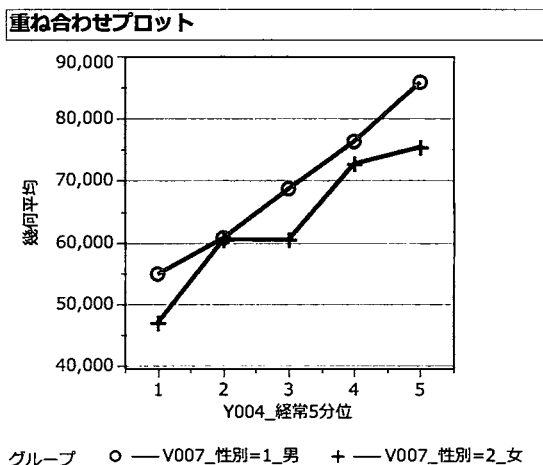
表 5b JMP データの Excel シートへの貼り付け

|   | 経常収入 |  | 食料     |        |        |
|---|------|--|--------|--------|--------|
|   | 5分位  |  | 世帯数    | 常用対数平均 | 幾何平均   |
| 男 | 1    |  | 15,494 | 4.7389 | 54,814 |
|   | 2    |  | 18,617 | 4.7832 | 60,704 |
|   | 3    |  | 19,099 | 4.8368 | 68,675 |
|   | 4    |  | 19,356 | 4.8826 | 76,311 |
|   | 5    |  | 19,435 | 4.9335 | 85,793 |
|   | 計    |  | 92,000 | 4.8395 | 69,107 |
| 女 | 1    |  | 4,506  | 4.6713 | 46,913 |
|   | 2    |  | 1,377  | 4.7811 | 60,405 |
|   | 3    |  | 911    | 4.7813 | 60,439 |
|   | 4    |  | 640    | 4.8617 | 72,730 |
|   | 5    |  | 567    | 4.8774 | 75,407 |
|   | 計    |  | 8,000  | 4.7325 | 54,018 |

幾何平均の折れ線グラフは、【重ね合わせプロット】を用いる。幾何平均を【Y】に、経常5分位を【X】に、性別を【グループ変数】にいれ、出力されたデフォルト画面を図表 5c に示すよう

に GUI 機能で調整する。このスクリプトは、図表 5d に示すように JMP テーブルに貼り付けてあり、何時でも再現できる。

図表 5c JMP による幾何平均の折れ線グラフ



図表 5d JMP による折れ線グラフ作成のための自動保存されたスクリプト

```

重ね合わせプロット
重ね合わせプロット(
  X(:Y004_経常5分位),
  Y(:幾何平均),
  グループ変数(:V007_性別),
  グループの重ね合わせ,
  Y軸[1] << {最小値(39982), 最大値(94457), 目盛り間隔(10000),
  補助目盛り(1), Label Row Nesting(1), ラベル行(
  {Show Major Grid(1), Show Minor Grid(1)}}
)},
  点をつなぐ(1),
  レポートへ送信(
    ディスバッチ(
      {},
      "106",
      ScaleBox(2),
      {最小値(39982), 最大値(94457), 目盛り間隔(10000),
      補助目盛り(1), Label Row Nesting(1), ラベル行(
      {Show Major Grid(1), Show Minor Grid(1)}}
    )
  ),
  ディスバッチ(
    {},
    "Overlay Plot Graph",
    FrameBox(3),
    {フレームサイズ(259, 254), マーカーサイズ(4), 線の幅のスケール(3)}
  ),
  ディスバッチ({}, "", LegendBox, {タイトルの指定("グループ")})
)

```

#### 4. 考察

擬似マイクロデータを用いたこれまでの「マイクロデータ分析コンテスト」の優秀論文は、SAS ユーザー総会の論文集にすべて収録され公開されており、SAS あるいは JMP の使用経験のある人達には容易に理解できると思われるが、他の統計ソフト Stata で経済データの統計解析<sup>6,7)</sup>を実施している人たちには、平易ではないと思われる。そこで、SAS および JMP の使用歴のない人達を念頭にしたチュートリアルを、第 4 回目の規定問題を題材に行うことにした。

擬似マイクロデータは、先頭行に変数名が付いていない CSV 形式であるので、JMP で内容の把握しやすい変数名を GUI ベースで付ける必要がある。規定問題の範囲内の限定された変数ならば、手作業で変数名を付け替えることは容易ではあるが、更なる探索的な解析を行う場合には推奨することはできない。そこで、全 197 変数を一括した変数名の付け替え方法を示した。

名義変数のカテゴリ名は、できる限り整数でコーディングし、統計ソフトのプログラムで符号(コード)と符号名(カテゴリ名)の対応付けを一括して行うのは、SAS などのプログラミング言語系の統計ソフトの一般的な対応である。GUI 系の JMP では、符号と符号名を手作業で変更入力する必要があり、きわめて煩雑であり推奨できない。元の符号名自体がデータとして入力されていれば、JMP ではそのまま各種の解析が問題なくできるので、全ての変数の符号名を事前に Excel で整形し、JMP ファイルとして取り込み、リレーショナルデータベース的な GUI 操作で、一括して置き換える方法を示した。

このように統計解析を行うために利用者が煩雑な前処理を必要とすることは、“教育的”ではあるものの擬似マイクロデータ使って統計教育を推進しようとする者にとって阻害要因である。目的とする集計表が明示されていて、Fortran あるいは Cobol などのプログラミング言語を使ってプログラミングする場合には、表示 4 で示したレイアウトシートが不可欠であることは承知している。しかしながら、多くの統計ソフトが普及し、現在、Fortran あるいは Cobol などのプログラミング言語を使って統計解析を行っている人達の存在を確認できない。

擬似マイクロデータのレイアウトシートに「位置」と「バイト数」が示され、その中に「項目の区切り」としてカンマ「,」もあることから、古典的な大型コンピュータ時代の流れをくんだ固定長レコードと Excel でも操作できる CSV 形式となっていることに苦肉の努力がなされている。区切り記号が入ってなければ、JMP および Excel でデータを扱うことは不可能で、筆者も慣れ親しんできた SAS で固定長のデータと認識したプログラミングが統計解析に不可欠となる。さいわいハイブリッドな固定長 CSV 形式でのデータ提供であったので、JMP での統計解析が可能となった。

コード化された名義尺度データに対しても古典的な「レイアウトシート」ではなく、表示 5 で示したような「変数名テーブル」があれば、速やかに解析のためのデータファイルを生成できる。また、表示 7 で示した「符号テーブル」があれば、リレーショナル DB 的な操作で、データセット内の符号を符号名にすることは、本稿で示したように容易である。

第 4 回目の規定問題は、集計乗率から推定される 10 万世帯比での集計・解析を主体にした。集計乗率は実数で提供されており、各種の統計量の算出に際しては、集計乗率を「重み」として使うものと文献<sup>7)</sup>での記述により思い込みをしていた。これは、誤解であることが、マイクロデータの解析の経験をつむ事により実感してきた。SAS および JMP には、集計乗率を「重み」として使うことも、実数ではあるが「度数」として扱うことも可能であり、「度数」として設定することが適

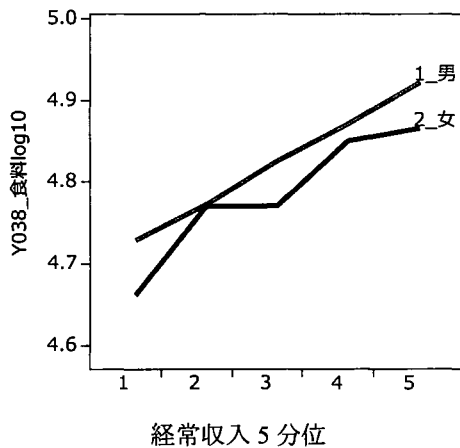
切であることを規定問題の作成を通して確認した。

政府系のマイクロデータの統計解析の基本は、集計乗率から推定する10万人比（10万世帯比）であることを再認識し規定問題の中心とした。また、政府系の統計資料では、10階級区分が標準的に用いられているが、探索的な統計解析を行うためには重すぎるので、10万世帯比に対する5階級区分とし規定問題に取り込んだ。

最後の規定問題5)では、表示11に示した不揃いの二元配置分散分析を念頭に常用対数変換および幾何平均なども盛り込んだ問題としたのであるが、共同主催者から経済統計の分野では一般的ではないので、規定問題としてはいかがなものかとのコメントがあり、設問を丁寧に書き直して最終的には、2元配置分散分析での交互作用の解釈をするための折れ線グラフの作成ができるように段階的な設問に変更した。

表示11 「度数」を10万世帯比をした二元配置分散分析の交互作用解析プロット

| 効果の検定              |        |     |           |          |            |  |
|--------------------|--------|-----|-----------|----------|------------|--|
| 要因                 | パラメータ数 | 自由度 | 平方和       | F値       | p値(Prob>F) |  |
| V007_性別            | 1      | 1   | 7.23392   | 312.8156 | <.0001*    |  |
| Y004_経常5分位         | 4      | 4   | 130.22158 | 1407.790 | <.0001*    |  |
| V007_性別*Y004_経常5分位 | 4      | 4   | 4.57338   | 49.4416  | <.0001*    |  |



表示12および13にしめした10万世帯比および集計乗率を「重み」とした分散分析は、誤差分散の不一致などにより微妙に異なり完全に一致しない。どのように使い分けるのか、何が適切なのかは、今後の検討課題である。一般的には、重みは該当するデータの分散の逆数とすることが知られているが、マイクロデータの「重み」一般的な定義とは異なり、本質的には「度数」であり、どのように正当化するのかの検討も今後の課題である。

表示 12 「重み」を 10 万世帯比した二元配置分散分析の交互作用解析

| 効果の検定              |        |     |           |          |            |  |
|--------------------|--------|-----|-----------|----------|------------|--|
| 要因                 | パラメータ数 | 自由度 | 平方和       | F値       | p値(Prob>F) |  |
| V007_性別            | 1      | 1   | 7.23392   | 100.1642 | <.0001*    |  |
| Y004_経常5分位         | 4      | 4   | 130.22158 | 450.7773 | <.0001*    |  |
| V007_性別*Y004_経常5分位 | 4      | 4   | 4.57338   | 15.8313  | <.0001*    |  |

平方和は「度数」の場合と同じだが、誤差の平均平方が大きいため F 値が小さい

表示 13 「重み」を集計乗率とした二元配置分散分析の交互作用解析

| 効果の検定              |        |     |           |          |            |  |
|--------------------|--------|-----|-----------|----------|------------|--|
| 要因                 | パラメータ数 | 自由度 | 平方和       | F値       | p値(Prob>F) |  |
| V007_性別            | 1      | 1   | 35.84154  | 100.1642 | <.0001*    |  |
| Y004_経常5分位         | 4      | 4   | 645.20237 | 450.7773 | <.0001*    |  |
| V007_性別*Y004_経常5分位 | 4      | 4   | 22.65950  | 15.8313  | <.0001*    |  |

平方和は異なるが、誤差の平均平方も大きくなるため、F 値は 10 万世帯比を「重み」にした場合と同じ

## 文 献

- 1) 高橋行雄, 周防節雄(2013), 統計センター提供の教育用擬似マイクロデータを用いた SAS/JMP によるデータ分析コンテストの概要, [http://www.nstac.go.jp/services/pdf/131122\\_4-1.pdf](http://www.nstac.go.jp/services/pdf/131122_4-1.pdf) .
- 2) 独立行政法人統計センター(2012), 教育用擬似マイクロデータの開発とその利用～平成 16 年全国消費実態調査を例として～, <http://www.nstac.go.jp/services/pdf/sankousiryoku2407.pdf>.
- 3) 周防節雄(2016), SAS ユーザー総会 2016 における「Let's データ分析 第 4 回マイクロデータ分析コンテスト」の規定課題の SAS プログラム解説, SAS ユーザー総会 2016 論文集掲載予定.
- 4) 周防節雄(2015), 全国消費実態調査の匿名データとその符号表から自動的に SAS のデータセット, 変数ラベルと変数フォーマットを作成する SAS プログラム, SAS ユーザー総会 2015 論文集, 257-278.
- 5) 高橋行雄(2015), 統計センターの匿名データ 13 万件を用いた統計解析の実践, SAS ユーザー総会 2015 論文集, 145-164.
- 6) 筒井淳也, 平井裕久, 水谷正明ら(2011), Stata で計量経済学入門 第 2 版, ミネルヴァ書房.
- 7) 二木美苗(2013), 「全国消費実態調査 匿名データの整理概要 -kuma\_do 及び変数対応表の活用方法を中心に」, <http://www.econ.kobe-u.ac.jp/kuma/satellite/pdf/308.pdf> .
- 8) 松田芳郎, 伴金美, 美添泰人編著(2000), ミクロ統計の集計解析と技法, 日本評論社.

# SAS ユーザー総会 2016 における 「Let's データ分析 第4回マイクロデータ分析コンテスト」の 規定課題の SAS プログラム解説

周防 節雄  
兵庫県立大学・名誉教授

Exemplary SAS Programs to Solve the Compulsory Exercises  
for the 4<sup>th</sup> Micro Data Competition in SAS Japan Users Forum 2016

Setsuo Suoh  
Professor Emeritus of the University of Hyogo

## 要旨

SAS ユーザー総会 2016 における「Let's データ分析 第4回マイクロデータ分析コンテスト」の規定課題について、出題者の立場から、模範解答の SAS プログラムを解説する。

**キーワード** 全国消費実態調査、教育用擬似マイクロデータ、proc freq、proc tabulate

## 1. はじめに

政府機関や地方公共団体が実査する統計調査から得られた統計調査データは、一般的に公的統計と言われる。これらは、全数調査もあれば、標本調査もある。いずれにしても、我々一般ユーザーの利用に提供される際は、その全サンプルが入手できる訳ではなく、部分的に抽出されたマイクロデータが提供される。その際、元の母集団が推定できるように、各レコードには集計用乗率(復元乗率とも言う)が付与されている。

独立行政法人統計センターは、平成 16 年の全国消費実態調査のマイクロデータから教育目的で擬似マイクロデータを作成して、公的マイクロデータの利用実習用に提供してきた。SAS ユーザー総会では、2013 年からこの教育用擬似マイクロデータ(以後、「擬似マイクロデータ」と略称する)を使用して、毎年、データコンペを実施してきた、今年で 4 回目を数える。このコンペは、規定課題(付録 5 参照)と自由課題から成り、規定課題は参加者全員に課せられる。

今年の規定課題では集計用乗率を考慮したプログラミングが要求される。本稿では、この規定課題のための SAS プログラムを解説する。

## 2. 教育用擬似マイクロデータの概要

統計センターから提供される擬似マイクロデータは7個の CSV ファイルに分割されている。これらの CSV ファイルにはメタ情報がないので、これだけでは使えないが、この他にメタ情報として「符号表」(表1)というエクセル形式のファイルがあり、これがメタ情報として CSV ファイルに対応している<sup>1</sup>。その中にある変数名を使って SAS のデータセットに変換した。ただし、変数名の先頭部に「Youto」が付いている変数は、全て「Youto」を「Y」に rename して使い易くした。できあがった SAS データセット

<sup>1</sup> このメタ情報を使って、CSV ファイルから SAS データセットへの自動変換については、(周防 2014a,b)で詳細に解説している。

は世帯を単位とする 32,027 オブザベーションから成る。付録 2 に全ての変数を示す<sup>2</sup>。規定問題では、黒く塗りつぶした4つの変数だけを使えばよい。後で製表の際に必要なになるので、「性別」と「産業符号」の変数用に format を定義しておく(表2右)。この proc format は符号表を使って SAS プログラムから自動作成された(周防 2014a,b)。

表 1 符号表

| 行番号 | 項目名        | 階層 | 位置   | バイト数 | 配置 | 型 | 種別 | 変数名           | 符号               | 符号内容                |
|-----|------------|----|------|------|----|---|----|---------------|------------------|---------------------|
| 1   | 世帯区分       | 1  | 1    | 1    |    |   |    | SetaiKubun    | 1                | 勤労                  |
| 2   |            |    |      |      |    |   |    |               | 2                | 勤労以外                |
| 3   |            |    |      |      |    |   |    |               | 3                | 無職                  |
| 4   |            | 1  | 2    | 1    |    |   |    |               | ,                | 項目の区切り              |
| 5   | 世帯員        | 1  |      |      |    |   |    |               |                  |                     |
| 6   | 世帯人員       | 2  | 3    | 2    | 2  | 1 |    | SetaiJinin    | Δ2~              | 2人~                 |
| 7   |            | 2  | 5    | 1    |    |   |    |               | ,                | 項目の区切り              |
| 8   | 有業人員       | 2  | 6    | 2    | 2  | 1 |    | ShuugyouJinin | Δ1~              | 1人~                 |
| 9   |            |    |      |      |    |   |    |               | VV               | 不詳                  |
| 10  |            | 2  | 8    | 1    |    |   |    |               | ,                | 項目の区切り              |
| 11  | 現住居等に関する事項 | 2  |      |      |    |   |    |               |                  |                     |
| 12  | 住居の構造      | 3  | 9    | 1    |    |   |    | Kouzou        | 1                | 木造                  |
| 13  |            |    |      |      |    |   |    |               | 2                | 防火木造                |
| 14  |            |    |      |      |    |   |    |               | 3                | 鉄骨・鉄筋コンクリート造        |
| 15  |            |    |      |      |    |   |    |               | 4                | その他(ブロック造り、レンガ造りなど) |
| 16  |            |    |      |      |    |   |    |               | V                | 不詳                  |
| 17  |            | 3  | 10   | 1    |    |   |    |               | ,                | 項目の区切り              |
| 中略  |            |    |      |      |    |   |    |               |                  |                     |
| 523 | 財産購入       | 4  | 2928 | 15   | 2  | 1 |    | Yuto181       | ΔΔΔΔΔΔΔΔ0.00000~ | 円                   |
| 524 |            | 4  | 2943 | 1    |    |   |    |               | ,                | 項目の区切り              |
| 525 | その他        | 4  | 2944 | 15   | 2  | 1 |    | Yuto182       | ΔΔΔΔΔΔΔΔ0.00000~ | 円                   |
| 526 |            | 4  | 2959 | 1    |    |   |    |               | ,                | 項目の区切り              |
| 527 | 繰越金        | 3  | 2960 | 15   | 2  | 1 |    | Yuto183       | ΔΔΔΔΔΔΔΔ0.00000~ | 円                   |

表2 変数S1\_Sangyuu

| 元の値 | 数値型 | 符号内容              |
|-----|-----|-------------------|
| 文字型 | に変換 |                   |
| Δ1  | 1   | 農業                |
| Δ2  | 2   | 林業                |
| Δ3  | 3   | 漁業                |
| Δ4  | 4   | 鉱業                |
| Δ5  | 5   | 建設業               |
| Δ6  | 6   | 製造業               |
| Δ7  | 7   | 電気・ガス・熱供給・水道業     |
| Δ8  | 8   | 情報通信業             |
| Δ9  | 9   | 運輸業               |
| 10  | 10  | 卸売・小売業            |
| 11  | 11  | 金融・保険業            |
| 12  | 12  | 不動産業              |
| 13  | 13  | 飲食店・宿泊業           |
| 14  | 14  | 医療・福祉             |
| 15  | 15  | 教育・学習支援業          |
| 16  | 16  | 複合サービス事業          |
| 17  | 17  | サービス業(他に分類されないもの) |
| 18  | 18  | 公務(他に分類されないもの)    |
| 19  | 19  | その他(非就業を含む)       |
| VV  | .   | 不詳                |

(符号表から作成)

```
proc format;
value xS1_Sangyuu
1="農業"
2="林業"
3="漁業"
4="鉱業"
5="建設業"
6="製造業"
7="電気・ガス・熱供給・水道業"
8="情報通信業"
9="運輸業"
10="卸売・小売業"
11="金融・保険業"
12="不動産業"
13="飲食店・宿泊業"
14="医療・福祉"
15="教育・学習支援業"
16="複合サービス事業"
17="サービス業(他に分類されないもの)"
18="公務(他に分類されないもの)"
19="その他(非就業を含む)"
.= "不詳";

value xS1_Sex
1="男"
2="女";
注:SASプログラム「giji_format.sas」から一部抽出
run;
```

<sup>2</sup> この変数一覧表は付録 1 のプログラムにある proc datasets からの出力リストを、保存の物理的順序に並べ替えたものである。このプログラムは、オブザベーション数が多いデータセットの場合に、便利なツールである。

### 3. 集計用乗率

ここでは、集計用乗率(変数名:weight)について簡単に解説する。32,027 オブザベーションの変数weightの値を全て合計すると、495464.76574になる。因みに、最初のオブザベーションのweightの値は18.9、二番目のオブザベーションでは15.43333である。これは、全体の母集団の中でそれぞれ0.0000381460(=18.9÷495464.76574)、0.0000311492(=15.43333÷495464.76574)の比重があるということである。つまり、提供された擬似マイクロデータでは、共に1件分のレコードであるが、母集団を推定する時は、件数が変わるということになる。この割り算の値に100を掛ければ、全体を100オブザベーションとした時のオブザベーション数になる。

従って、全体を100万オブザベーションとすると、それぞれ、オブザベーションが38.1460個(=0.0000381460×100万)、31.1492個(=0.0000311492×100万)あることになる。今回の規定問題の中には100万世帯比で計算する課題があるが、その時は、この集計用乗率を使用しなければならない。

### 4. 規定問題のプログラム解説

模範解答のSASプログラムは付録3に掲載している。説明用に各行には連番号を付している。プログラムから結果viewer画面(付録4)に出力するときは、必ずtitle文を付して、プログラムのどの箇所からの出力かがすぐに分かるように配慮している。以下の解説では、該当するプログラムの行番号を「#3」のように先頭で明示している。

#3: %include している SAS プログラム「giji\_format.sas」のうち、規定問題に関する format 文だけを抜き出して、表2(右)に示している。

#6: 教育用擬似マイクロデータの SAS データセット zensho は、付録2に示す変数から成り、オブザベーションは32,027件ある。ここで規定問題に必要な変数(付録2の黒色のセルに白字)だけでテンポラリーデータセット kadai を作成する。変数名もわかりやすい名前に rename した。

#9: 集計用乗率の変数 weight の合計を計算して、変数 allWGT に保存。#38～39 で使う。

#17: proc freq を使って、課題の表1の「a)単純クロス表」完成。

#27: proc freq の tables 文のオプション out=table1 で、「乗率あり」のクロス表を出力すると同時に、データセット table にその結果を保存する。変数 COUNT と PERCENT は SAS が自動作成する。

#34: そのデータセット table の中味を表示。その際、sum 文で 変数 COUNT と PERCENT の合計値も表示させて、この変数の意味を確認。

#38: データセット kadai の最後のオブザベーションの変数 allWGT の値を、%put 文でマクロ変数 allWGT に格納する。これ以後、この値は&allWGT でこのプログラム中で使える。#40 で確認。

#42: データセット kadai に元々ある集計用乗率 weight を使って、10万比に換算した乗率 wgt10man を計算して、データセット newWGT に追加・保存。wgt10man の累計 allwgt10man も計算しておき、最後のオブザベーションの時に LOG 画面に「allwgt10man=100000」を出力して、確認。

#48: データセット newWGT を使って、10万比の乗率で度数分布表を作成。

#56: 上で作成した度数分布表と同じ表を、オプション FORMAT=COMMA6. を使って、小数点なしで再計算。課題の表1の「b)10万世帯比でのクロス表」完成。

#66: proc sgplot を使って、乗率を考慮した経常収支のヒストグラムを作図。

#71: proc sgplot を使って、乗率を考慮した食料費のヒストグラムを作図。



#80: proc univariate を使って、データセット newWGT から集計用乗率を使って、経常収入と食料費の 5 分位点を求め、新しいデータセット fivePCTL を作成。変数 income20～income80、food20～food80 の 8 変数で、オブザベーションは 1 個。weight 文を使わなければヒストグラムは自動的に出力されるが、weight 文を使うと出力されないで、先に、proc sgplot でヒストグラムを描いた。

#90: proc transpose を使って、1オブザベーションのデータセット fivePCTL を縦に並べ替え、新しいデータセット finishTable2 を作成。

#94: データセット finishTable2 を加工して課題の表 2 の原型になるデータセット TB2SEIRI を作成。

#102: データセット TB2SEIRI を、経常収入と食料費で別々のデータセット income と food に分割。

#111: データセット income と food を横並びにマージして、課題の表 2 が完成。

#119: 数値変数 income20 の値をマクロ変数 income20 に格納する方法を確認。

#120: %put 文を実行すると LOG 画面に“310531.0652”と表示。これは経常収入の 20%分位点。

#122: 先に作成したデータセット fivePCTL の 8 個の変数を四捨五入して小数点以下を削除する。

#129: options macrogen を使うと、この直後の SAS マクロの展開結果を LOG 画面に表示してくれるので、マクロが正しく展開されたかを確認できる。

#130: データセット fivePCTL の 8 個の変数 income20～income80、food20～food80 の値を同じ名前のマクロ変数に格納。

#150: 経常収入と食料費の 5 分位用の format を、上で設定したマクロ変数を使って定義するマクロ format を作成。

#168: マクロ format を実行。以後、フォーマット income. と food. が使用可能。

#170: 乗率 wgt10man を使って経常収入の 5 分位毎の頻度表を出力。

#176: 乗率 wgt10man を使って食料費の 5 分位毎の頻度表を出力。課題の表 3 が完成。

#183: 課題の表 4 の右端の「計」欄を 10 万比の乗率で計算するための工夫。変数 X の値を 1 に設定することで、これが乗率に応じて膨れる。

#185: proc tabulate で、先に定義したフォーマットを使えば、5 分位毎の分布が計算できる。課題の表 4 完成。

#198: データセット newWGT を使って、食料費の常用対数を計算、データセット table5 を作成。

#202: proc tabulate でデータセット table5 を使って、課題の表 5 のうち、幾何平均欄以外の部分が完成。10 万比の乗率を使っているが、SUMWGT を使うのがミソ。この後で幾何平均を計算するために、この結果表をデータセット result5 に保存しておく。

#217: 直ぐ上で出力したデータセット result5 を性別にあらかじめソートしておく。

#224: データセット result5 を使って、幾何平均を計算し、データセット final\_result5 を作成。

#229: proc print を使って、データセット final\_result5 を LABEL 文併用すれば、課題の表 5 が完成。

#236: データセット final\_result5 を使って折れ線グラフを描くのに、「計」の列が不要なので削除して、データセット plot を作成。

#243: データセット plot を使って、proc sgplot で折れ線グラフを作図。男女で折れ線の色の区別はできるが、線種や線の太さの指定がうまく効かないので、モノクロで印刷すると、男女が識別できないのが懸案事項として残る。

## 5. まとめ

公的統計のマイクロデータを分析するときは、常に集計用乗率を意識する必要がある。つまり、入手したデータは全数ではないので、母集団を推定するのに必要な情報となる。今回の規定問題では、proc tabulate を使うケースが多いが、weight 文を使うと、ヒストグラムが出力されないで、別途方

法を考えなければならない。今回、SAS9.4から使えるようになった proc sgplot を使ったが、当初、その存在を知らなかった。

新機能は絶えず気をつけていないとその存在すら知らないことになる。一昔前までは、SAS のバージョンが上がる度に印刷物のマニュアルを購入していたが、いつのまにか冊数が増えてきたので止めた。今では、大抵のことは、ネット上で必要な情報が得られるが、それでも分からない時は、知り合いのSASのヘビーユーザーに質問して教えてもらうことにしている。2011年に神戸で開催された SAS ユーザー総会で数独を解くSASプログラムを報告したが、その英語版を翌年に米国フロリダで開催された SAS Global Forum 2012 で報告をした。この時、地元のユーザーから proc FCMP が SAS9.3 から使えるようになり、自作の SAS サブルーティンや SAS 関数が作れることを初めて聞いた。このプロシジャーについては、昨年の東大での SAS ユーザー総会でチュートリアルをした。日本のユーザーでも SAS ユーザー総会でこの機能を使っている報告を見つけたので、しっかりと新機能について目を光らせているユーザーもいるのだなと感心した。

実を言うと、今回のデータコンペの規定問題を自分で出題していながら、ヒストグラムや折れ線グラフを描くの到手こずった。普段から使っていれば何でもない難易度なのだが、自分では SAS でグラフを描くことは余りないので、大変勉強になった。何人かの知人の知恵を借りた。SAS のヘビーユーザーになる近道は、自分の周りにヘビーユーザーの輪を作ることと心得た。

## 参考文献

1. 周防節雄 (2014a) 政府統計マイクロデータの符号表から SAS 変数のラベルとフォーマットを自動生成する SAS プログラムの作成方法、『SAS ユーザー総会 2014 論文集』, pp831-844
2. 周防節雄 (2014b) 教育用擬似マイクロデータの符号表から SAS データセット用の変数ラベルと変数フォーマットを自動作成する SAS プログラム、「公的統計のマイクロデータ等を用いた研究の新展開」研究集会、<http://mighty.gk.u-hyogo.ac.jp/confidential/suoh20141121.html>、統計数理研究所

## 謝辞

全国消費実態調査の教育用擬似マイクロデータの利用に便宜を図って頂いた(独)統計センターに感謝の意を表します。本稿は学術振興会科学研究費(課題番号 24530232: 研究代表者・周防節雄)による研究成果の一部である。記して謝意に代えたい。

## 付録 1 SAS データセットの変数一覧表を出力するプログラム

```
/*-----' lookdata.sas' -----*/
/* Utility program to display all information in a specified sas dataset*/
/*-----*/
* Specify the following macro variables:
* Only specified number of observations are displayed in Result Viewer window;
/*lib name(libref or 1st-level name) */ %let libname= suoh;
/*definition of lib name[drive name:%directory name] */
%let lib_what= F:%全消費全消費擬似マイクロデータ%zensho_permanent;
/*SAS dataset name(no extension) */ %let dsname= zensho;
/*Number of observations shown in Result Viewer */ %let obs_no= 1;
libname &libname "&lib_what";
options ls=74 ps=60 nocenter nonumber;
proc datasets library=&libname; contents data=&dsname; run; quit;
proc print data=&libname..&dsname(obs=&obs_no);
title
"SAS dataset name: &lib_what..&dsname (first &obs_no. observations only)";
run;
```

## 付録2 変数一覧表

| #  | 変数              | 変数ラベル            | #   | 変数   | 変数ラベル        | #   | 変数   | 変数ラベル                                   |
|----|-----------------|------------------|-----|------|--------------|-----|------|-----------------------------------------|
| 1  | SetaiKubun      | 世帯区分             | 67  | Y53  | 牛乳           | 133 | Y119 | 健康保持用摂取品                                |
| 2  | SetaiJinin      | 世帯人員             | 68  | Y54  | 乳製品          | 134 | Y120 | 保健医療用品・器具                               |
| 3  | ShuugyouJinin   | 有業人員             | 69  | Y55  | 卵            | 135 | Y121 | 保健医療サービス                                |
| 4  | Kouzou          | 住居の構造            | 70  | Y56  | 野菜・海藻        | 136 | Y122 | 交通・通信                                   |
| 5  | Tatekata        | 住居の建て方           | 71  | Y57  | 生鮮野菜         | 137 | Y123 | 交通                                      |
| 6  | Shoyuu          | 住居の所有関係          | 72  | Y58  | 乾物・海藻        | 138 | Y124 | 自動車等関係費                                 |
| 7  | S1_Sex          | 性別               | 73  | Y59  | 大豆加工品        | 139 | Y125 | 自動車等購入                                  |
| 8  | S1_Age          | 年齢5歳階級           | 74  | Y60  | 他の野菜・海藻加工品   | 140 | Y126 | 自転車購入                                   |
| 9  | S1_Shugyou      | 就業・非就業の別         | 75  | Y61  | 果物           | 141 | Y127 | 自動車等維持                                  |
| 10 | S1_KigyokuKubun | 企業区分             | 76  | Y62  | 生鮮果物         | 142 | Y128 | 通信                                      |
| 11 | S1_KigyokuKibo  | 企業規模             | 77  | Y63  | 果物加工品        | 143 | Y129 | 教育                                      |
| 12 | S1_Sangyou      | 産業符号             | 78  | Y64  | 油脂・調味料       | 144 | Y130 | 授業料等                                    |
| 13 | S1_Shokugyou    | 職業符号             | 79  | Y65  | 油脂           | 145 | Y131 | 教科書・学習参考教材                              |
| 14 | Weight          | 集計用乗率            | 80  | Y66  | 調味料          | 148 | Y132 | 補習教育                                    |
| 15 | Y1              | 年間収入             | 81  | Y67  | 菓子類          | 147 | Y133 | 娯楽                                      |
| 18 | Y2              | 収入総額             | 82  | Y68  | 調理食品         | 148 | Y134 | 娯楽用耐久財                                  |
| 17 | Y3              | 実収入              | 83  | Y69  | 主食の調理食品      | 149 | Y135 | 娯楽用品                                    |
| 18 | Y4              | 経常収入             | 84  | Y70  | 他の調理食品       | 150 | Y136 | 書籍・他の印刷物                                |
| 19 | Y5              | 勤め先収入            | 85  | Y71  | 飲料           | 151 | Y137 | 娯楽サービス                                  |
| 20 | Y6              | 事業・内職収入          | 88  | Y72  | 茶類           | 152 | Y138 | 宿泊料                                     |
| 21 | Y7              | 農林漁業収入           | 87  | Y73  | コーヒ-・ココア     | 153 | Y139 | バンク旅行費                                  |
| 22 | Y8              | 家賃収入             | 88  | Y74  | 他の飲料         | 154 | Y140 | 月謝類                                     |
| 23 | Y9              | 他の事業収入           | 89  | Y75  | 酒類           | 155 | Y141 | 他の娯楽サービス                                |
| 24 | Y10             | 内職収入             | 80  | Y76  | 外食           | 156 | Y142 | その他の消費支出                                |
| 25 | Y11             | 本業以外の勤め先・事業・内職収入 | 91  | Y77  | 一般外食         | 157 | Y143 | 雑費                                      |
| 28 | Y12             | 他の経常収入           | 92  | Y78  | 学校給食         | 158 | Y144 | 理美容サービス                                 |
| 27 | Y13             | 財産収入             | 93  | Y79  | 住居           | 159 | Y145 | 理美容用品                                   |
| 28 | Y14             | 社会保険給付           | 94  | Y80  | 家賃地代         | 160 | Y146 | 身の回り用品                                  |
| 29 | Y15             | 公的年金給付           | 95  | Y81  | 設備修繕・維持      | 181 | Y147 | たばこ                                     |
| 30 | Y16             | 他の社会保険給付         | 98  | Y82  | 設備材料         | 162 | Y148 | その他の雑費                                  |
| 31 | Y17             | 仕送り金             | 97  | Y83  | 工事その他のサービス   | 163 | Y149 | こづかい(使途不明)                              |
| 32 | Y18             | 特別収入             | 98  | Y84  | 光熱・水道        | 164 | Y150 | 交際費                                     |
| 33 | Y19             | 受贈金              | 99  | Y85  | 電気代          | 165 | Y151 | 食料                                      |
| 34 | Y20             | その他              | 100 | Y86  | ガス代          | 166 | Y152 | 家具・家事用品                                 |
| 35 | Y21             | 実収入以外の収入         | 101 | Y87  | 他の光熱         | 167 | Y153 | 被服及び履物                                  |
| 38 | Y22             | 預貯金引出            | 102 | Y88  | 上下水道料        | 168 | Y154 | 娯楽                                      |
| 37 | Y23             | 保険取金             | 103 | Y89  | 家具・家事用品      | 169 | Y155 | 他の物品サービス                                |
| 38 | Y24             | 個人・企業年金保険取金      | 104 | Y90  | 家庭用耐久財       | 170 | Y156 | 贈与金                                     |
| 39 | Y25             | 他の保険取金           | 105 | Y91  | 家事用耐久財       | 171 | Y157 | 他の交際費                                   |
| 40 | Y26             | 有価証券売却           | 106 | Y92  | 冷暖房用器具       | 172 | Y158 | 仕送り金                                    |
| 41 | Y27             | 株式売却             | 107 | Y93  | 一般家具         | 173 | Y159 | 非消費支出                                   |
| 42 | Y28             | 他の有価証券売却         | 108 | Y94  | 室内装備・装飾品     | 174 | Y160 | 直接税                                     |
| 43 | Y29             | 土地家屋借入金          | 109 | Y95  | 寝具類          | 175 | Y161 | 勤労所得税                                   |
| 44 | Y30             | 他の借入金            | 110 | Y96  | 家事雑貨         | 176 | Y162 | 個人住民税                                   |
| 45 | Y31             | 分割払一括購入借入金       | 111 | Y97  | 家事用消耗品       | 177 | Y163 | 他の税                                     |
| 46 | Y32             | 財産売却             | 112 | Y98  | 家事サービス       | 178 | Y164 | 社会保険料                                   |
| 47 | Y33             | その他              | 113 | Y99  | 被服及び履物       | 179 | Y165 | 公的年金保険料                                 |
| 48 | Y34             | 繰入金              | 114 | Y100 | 和服           | 180 | Y166 | 健康保険料                                   |
| 49 | Y35             | 支出総額             | 115 | Y101 | 洋服           | 181 | Y167 | 介護保険料                                   |
| 50 | Y36             | 実支出              | 116 | Y102 | 男子用洋服        | 182 | Y168 | 他の社会保険料                                 |
| 51 | Y37             | 消費支出             | 117 | Y103 | 婦人用洋服        | 183 | Y169 | 他の非消費支出                                 |
| 52 | Y38             | 食料               | 118 | Y104 | 子供用洋服        | 184 | Y170 | 実支出以外の支出                                |
| 53 | Y39             | 穀類               | 119 | Y105 | シャツ・セーター類    | 185 | Y171 | 預貯金                                     |
| 54 | Y40             | 米                | 120 | Y106 | 男子用シャツ・セーター類 | 186 | Y172 | 保険掛金                                    |
| 55 | Y41             | パン               | 121 | Y107 | 婦人用シャツ・セーター類 | 187 | Y173 | 個人・企業年金保険掛金                             |
| 58 | Y42             | めん類              | 122 | Y108 | 子供用シャツ・セーター類 | 188 | Y174 | 他の保険掛金                                  |
| 57 | Y43             | 他の穀類             | 123 | Y109 | 下着類          | 189 | Y175 | 有価証券購入                                  |
| 58 | Y44             | 魚介類              | 124 | Y110 | 男子用下着類       | 190 | Y176 | 株式購入                                    |
| 59 | Y45             | 生鮮魚介             | 125 | Y111 | 婦人用下着類       | 191 | Y177 | 他の有価証券購入                                |
| 60 | Y46             | 塩干魚介             | 126 | Y112 | 子供用下着類       | 192 | Y178 | 土地家屋借金返済                                |
| 61 | Y47             | 魚肉練製品            | 127 | Y113 | 生地・糸類        | 193 | Y179 | 他の借金返済                                  |
| 62 | Y48             | 他の魚介加工品          | 128 | Y114 | 他の被服         | 194 | Y180 | 分割払一括購入借入金返済                            |
| 63 | Y49             | 肉類               | 129 | Y115 | 履物類          | 195 | Y181 | 財産購入                                    |
| 64 | Y50             | 生鮮肉              | 130 | Y116 | 被服関連サービス     | 198 | Y182 | その他                                     |
| 65 | Y51             | 加工肉              | 131 | Y117 | 保健医療         | 197 | Y183 | 繰越金                                     |
| 66 | Y52             | 乳卵類              | 132 | Y118 | 医薬品          |     |      | [注] Youto001 → Y1, ..., Youto181 → Y181 |

### 付録3 模範解答プログラム

```

1  /* SAScompe2016kitei完成.sas */ options nocenter; %let drive=F;
2  libname giji "&drive:¥全消¥全消擬似マイクロデータ¥zensho_permanent";
3  %include "&drive:¥全消¥全消擬似マイクロデータ¥program¥giji_format.sas";
4
5  **★表1★;
6  data kadai; set giji.zensho;
7    keep S1_Sangyou S1_Sex Y4 Y38 weight allWGT;
8    rename S1_Sangyou =industry S1_Sex=sex Y4=income Y38=food;
9    allWGT=weight;
10 run;
11
12 proc print data=kadai (obs=15); var industry sex income food weight allWGT;
13   format weight F12.6;
14   format allWGT F12.6;
15   title "★表1★① data=kadai (obs=15)"; run;
16
17 proc freq data=kadai;
18   tables industry*sex / missing nocol norow nopercnt;
19   format industry xS1_Sangyou.;
20   format sex xS1_Sex.;
21   title "★表1★② 産業符号×性別 (proc freq) 乗率なし";
22   title2 "表1 a)単純クロス表";
23 run;
24 title2;
25
26 proc freq data=kadai;
27   tables industry*sex / out=table1 missing nocol norow nopercnt;
28   format industry xS1_Sangyou.;
29   format sex xS1_Sex.;
30   weight weight;
31   title "★表1★③ 産業符号×性別 (proc freq) 乗率あり";
32 run;
33
34 proc print data=table1; sum COUNT PERCENT;
35   title "★表1★④ data=table1 産業符号×性別 (proc freq) 乗率あり";
36 run;
37
38 data _NULL_; set kadai end=final;
39   if final then call symputx("allWGT",allWGT); run;
40 %put "&allWGT"; ** 数値変数 allWGT の値をマクロ変数 allWGT に格納する例;
41
42 data newWGT; set kadai end=final; wgt10man=weight/&allWGT*100000;
43   allwgt10man+wgt10man; if final then put allwgt10man; ** wgt10manの合計の確認;
44 run;
45
46 proc print data=newWGT (obs=20); title "★表1★⑤ data=newWGT (obs=20)"; run;
47
48 proc freq data=newWGT;
49   tables industry*sex / out=table1 missing nocol norow nopercnt;
50   format industry xS1_Sangyou.;
51   format sex xS1_Sex.;
52   weight wgt10man;
53   title "★表1★⑥ b) 10万世帯比 産業符号×性別 (小数点あり)";
54 run;
55
56 proc freq data=newWGT;
57   tables industry*sex / out=table1 missing nocol norow nopercnt FORMAT=COMMA6.;
58   format industry xS1_Sangyou.;
59   format sex xS1_Sex.;
60   weight wgt10man;
61   title "★表1★⑦ b) 10万世帯比 産業符号×性別 (小数点なし)";
62 run;
63
64 proc print data=table1; title "★表1★⑧ data=table1"; sum PERCENT; run;

```

```

65
66 proc sgplot data=giji.newWGT;
67   histogram income / weight=weight;
68   title "★2) 経常収支のヒストグラム";
69 run;
70
71 proc sgplot data=giji.newWGT;
72   histogram food / weight=weight;
73   title "★2) 食料費のヒストグラム";
74 run; quit;
75
76 **表2★:
77 /*proc sort data=newWGT; by sex; run;*/
78
79 *経常収入と食料費をまとめて計算;
80 proc univariate data=newWGT noprint;
81   var income food; weight weight;
82   output out=fivePCTL PCTLPTS=20 40 60 80 PCTLPRE=income food;
83   *weightを使うとhistogramを指定してもヒストグラムの出力なし;
84 run;
85
86 proc print data=fivePCTL; title "★表2★① data=fivePCTL: 5分位点: 乗率あり";
87   var income20 income40 income60 income80
88       food20  food40  food60  food80;
89
90 proc transpose data=fivePCTL out=finishTabel2; run;
91
92 proc print data=finishTabel2; title "★表2★② 5分位点 計算完了"; run;
93
94 data TB2SEIRI; set finishTabel2; keep COL1 buni _NAME_;
95   rename COL1=value _NAME_=variable;
96   len=length(_NAME_); buni=compress(substr(_NAME_, len-1, 2) || '%');
97 run;
98
99 proc print data=TB2SEIRI;
100   title "★表2★③ data=TB2SEIRI"; var buni value variable; run;
101
102 data income(rename=value=income) food(rename=value=food);
103   set; if mod(_N_, 2)=0 then output food; *偶数番目は食料費 の分位点;
104       else output income; *奇数番目は経常収入の分位点;
105 run;
106
107 proc print data=income; title "★表2★④ income"; var buni income; run;
108
109 proc print data=food; title "★表2★⑤ food"; var buni food; run;
110
111 data; merge income food; run;
112
113 proc print; title "★表2★⑥ 表2完成"; var buni income food;
114   format income comma7.0;
115   format food comma7.0;
116 run;
117
118 **表3★:
119 data _NULL_; set fivePCTL; call symputx("income20", income20); run;
120 %put "&income20"; *数値変数income20の値をマクロ変数income20に格納する例;
121
122 data fivePCTL; set fivePCTL; *小数点以下を四捨五入する;
123   array temp {8} income20 income40 income60 income80 food20 food40 food60 food80;
124   do i=1 to 8; temp[i]=round(temp[i], 1); end;
125 run;
126
127 proc print; title "★表3★① 小数点以下を四捨五入"; run;
128

```

```

129 options macrogen;
130 data _NULL_; set fivePCTL;
131 call symputx("income20", income20);
132 call symputx("income40", income40);
133 call symputx("income60", income60);
134 call symputx("income80", income80);
135 call symputx("food20", food20);
136 call symputx("food40", food40);
137 call symputx("food60", food60);
138 call symputx("food80", food80);
139 run;
140
141 %put "&income20": *数値変数income20等の値をマクロ変数income20等に格納;
142 %put "&income40";
143 %put "&income60";
144 %put "&income80";
145 %put "&food20";
146 %put "&food40";
147 %put "&food60";
148 %put "&food80";
149
150 %macro format;
151 proc format;
152 value income low - < &income20 ="&income20. 円未満"
153 &income20 - < &income40 ="&income20. 円以上"
154 &income40 - < &income60 ="&income40. 円以上"
155 &income60 - < &income80 ="&income60. 円以上"
156 &income80 < - high ="&income80. 円以上"
157 .="計";
158
159 value food low - < &food20 ="&food20. 円未満"
160 &food20 - < &food40 ="&food20. 円以上"
161 &food40 - < &food60 ="&food40. 円以上"
162 &food60 - < &food80 ="&food60. 円以上"
163 &food80 < - high ="&food80. 円以上"
164 .="計";
165 run;
166 %mend;
167
168 %format;
169
170 proc freq data=newWGT; table income*sex / norow nocol nopercnt FORMAT=COMMA6.;
171 format income income.;
172 format sex XS1_SEX.;
173 weight wgt10man;
174 title "★表3★② 経常収入"; run;
175
176 proc freq data=newWGT; table food*sex / norow nocol nopercnt FORMAT=COMMA6.;
177 format food food.;
178 format sex XS1_SEX.;
179 weight wgt10man;
180 title "★表3★③ 食料費"; run;
181
182 **表4★:
183 data X; set newWGT; foodX=food; X=1; *表4の作成のための工夫; run;
184
185 proc tabulate data=X;
186 class sex income food;
187 var foodX X;
188 table sex*(income ALL) ALL,
189 food*(foodX*SUMWGT*F=comma7.0) X*(SUMWGT*F=comma7.0);
190 weight wgt10man;
191 format income income.;
192 format food food.;
193 format sex XS1_SEX.;
194 title "★表4★① 男女別";
195 run;

```

```

196
197 *★表 5 ★;
198 data table5; set newWGT; foodLOG=log10(food); run;
199
200 proc print data=table5 (obs=15); title "★表 5 ★① data=table5 (obs=15)"; run;
201
202 proc tabulate data=table5 out=result5;
203   class sex income food;
204   var foodLOG wgt10man;
205   tables sex*(income ALL),
206         foodLOG*(SUMWGT*F=6.0 MEAN);
207   weight wgt10man;
208   format income income.;
209   format sex X$1_SEX.;
210   title "★表 5 ★② proc tabulateから表示";
211 run;
212
213 proc print data=result5;
214   var sex income foodLOG_SumWgt foodLOG_Mean;
215   title "★表 5 ★③ proc tabulateから出力のdataset[result5] : 幾何平均なし"; run;
216
217 proc sort data=result5; by sex; run;
218 proc print data=result5;
219   var sex income foodLOG_SumWgt foodLOG_Mean;
220   format foodLOG_SumWgt F6.0;
221   format foodLOG_Mean F7.4;
222   title "★表 5 ★④ proc tabulateから出力のdataset[result5] : 幾何平均なし"; run;
223
224 data final_result5; set result5; geomean=10**foodLOG_Mean; *幾何平均計算;
225 label income="経常収支" foodLOG_SumWgt="世帯数"
226        foodLOG_Mean="常用対数平均" geomean="幾何平均";
227 run;
228
229 proc print data=final_result5 label;
230   var sex income foodLOG_SumWgt foodLOG_Mean geomean;
231   format foodLOG_SumWgt F6.0;
232   format foodLOG_Mean F7.4;
233   format geomean COMMA6.0;
234   title "★表 5 ★⑤ data=final_result5 : 表 5 完成"; run;
235
236 data plot; set final_result5; *折れ線グラフ作成の準備;
237 keep sex income geomean;
238 if income=. then delete; *経常収支の「計」を削除;
239 run;
240
241 proc print data=plot; title "図 5 折れ線グラフ用 data=plot"; run;
242
243 proc sgplot data=plot;
244   title "図 5 折れ線グラフ";
245   *グループ分けのグラフの場合、「/」以降にgroup=で指定;
246   series x=income y=geomean / group=sex;
247   *styleattrsでグループ毎の線のスタイル設定;
248   styleattrs
249     datacontrastcolors =(blue red)
250     datalinepatterns=(solid dot)
251   ;
252   xaxis type=discrete;
253   format sex X$1_SEX.;
254 run;
255 quit;

```

付録 4 SASプログラム実行後の結果Viewer画面

★表1★① data=kadai (obs=15)

| Obs | industry | sex | income    | food     | Weight    | allWGT     |
|-----|----------|-----|-----------|----------|-----------|------------|
| 1   |          | 1   | 211011.15 | 37941.54 | 18.900000 | 18.900000  |
| 2   |          | 1   | 580649.86 | 68062.70 | 15.433330 | 34.333330  |
| 3   |          | 1   | 423207.04 | 67993.23 | 15.433330 | 49.766660  |
| 4   |          | 1   | 299831.27 | 32238.18 | 21.033330 | 70.799990  |
| 5   |          | 1   | 280483.95 | 30854.72 | 8.933330  | 79.733330  |
| 6   |          | 1   | 311914.61 | 35803.69 | 8.933330  | 88.666660  |
| 7   |          | 1   | 295731.14 | 58880.19 | 16.900000 | 105.566660 |
| 8   |          | 1   | 244441.20 | 39041.89 | 8.933330  | 115.499990 |
| 9   |          | 1   | 192940.31 | 43411.93 | 10.986670 | 126.466660 |
| 10  |          | 1   | 249051.14 | 50348.08 | 14.100000 | 140.566660 |
| 11  | 9        | 1   | 255514.31 | 30464.31 | 17.200000 | 157.766660 |
| 12  | 9        | 1   | 637979.83 | 34834.20 | 13.600000 | 171.366660 |
| 13  | 9        | 1   | 211772.81 | 52946.08 | 18.166670 | 189.533330 |
| 14  |          | 1   | 269278.40 | 33950.48 | 15.311110 | 204.844440 |
| 15  |          | 1   | 123950.88 | 37516.85 | 13.866670 | 218.711110 |

★表1★② 産業符号×性別 (proc freq) 乗率あり  
表1 a)単純クロス表

The FREQ Procedure

| Frequency         | Table of industry by sex |         |      |       |
|-------------------|--------------------------|---------|------|-------|
|                   | industry(産業符号)           | sex(性別) |      | Total |
| 男                 |                          | 女       |      |       |
| 不詳                |                          | 9704    | 1591 | 11295 |
| 鉱業                |                          | 3       | 0    | 3     |
| 建設業               |                          | 2134    | 9    | 2143  |
| 製造業               |                          | 5688    | 173  | 5861  |
| 電気・ガス・熱供給・水道業     |                          | 222     | 0    | 222   |
| 情報通信業             |                          | 547     | 0    | 547   |
| 運輸業               |                          | 1235    | 3    | 1238  |
| 卸売・小売業            |                          | 2083    | 201  | 2284  |
| 金融・保険業            |                          | 472     | 39   | 511   |
| 不動産業              |                          | 40      | 0    | 40    |
| 飲食店・宿泊業           |                          | 50      | 32   | 82    |
| 医療・福祉             |                          | 371     | 270  | 641   |
| 教育・学習支援業          |                          | 1059    | 89   | 1148  |
| 複合サービス事業          |                          | 321     | 3    | 324   |
| サービス業(他に分類されないもの) |                          | 1816    | 106  | 1922  |
| 公務(他に分類されないもの)    |                          | 3656    | 130  | 3786  |
| Total             |                          | 25381   | 2648 | 32027 |

★表1★③ data=table1 産業符号×性別 (proc freq) 乗率あり

| Obs | industry          | sex | CGUNT     | PERCENT |
|-----|-------------------|-----|-----------|---------|
| 1   | 不詳                | 男   | 154607.10 | 31.204  |
| 2   | 不詳                | 女   | 23813.67  | 4.806   |
| 3   | 鉱業                | 男   | 33.83     | 0.007   |
| 4   | 建設業               | 男   | 31442.97  | 6.246   |
| 5   | 建設業               | 女   | 153.67    | 0.031   |
| 6   | 製造業               | 男   | 88374.70  | 17.837  |
| 7   | 製造業               | 女   | 2527.57   | 0.510   |
| 8   | 電気・ガス・熱供給・水道業     | 男   | 3455.87   | 0.697   |
| 9   | 情報通信業             | 男   | 10008.03  | 2.020   |
| 10  | 運輸業               | 男   | 19082.93  | 3.852   |
| 11  | 運輸業               | 女   | 30.37     | 0.006   |
| 12  | 卸売・小売業            | 男   | 32916.97  | 6.644   |
| 13  | 卸売・小売業            | 女   | 2920.57   | 0.589   |
| 14  | 金融・保険業            | 男   | 7169.07   | 1.447   |
| 15  | 金融・保険業            | 女   | 515.43    | 0.104   |
| 16  | 不動産業              | 男   | 711.37    | 0.144   |
| 17  | 飲食店・宿泊業           | 男   | 782.27    | 0.158   |
| 18  | 飲食店・宿泊業           | 女   | 531.23    | 0.107   |
| 19  | 医療・福祉             | 男   | 5063.77   | 1.022   |
| 20  | 医療・福祉             | 女   | 4011.10   | 0.810   |
| 21  | 教育・学習支援業          | 男   | 15407.07  | 3.110   |
| 22  | 教育・学習支援業          | 女   | 1334.93   | 0.269   |
| 23  | 複合サービス事業          | 男   | 4881.67   | 0.981   |
| 24  | 複合サービス事業          | 女   | 54.07     | 0.011   |
| 25  | サービス業(他に分類されないもの) | 男   | 28943.27  | 5.842   |
| 26  | サービス業(他に分類されないもの) | 女   | 1613.37   | 0.326   |
| 27  | 公務(他に分類されないもの)    | 男   | 52969.77  | 10.691  |
| 28  | 公務(他に分類されないもの)    | 女   | 2130.17   | 0.430   |
|     |                   |     | 495464.77 | 100.000 |

★表1★④ 産業符号×性別 (proc freq) 乗率あり

The FREQ Procedure

| Frequency         | Table of industry by sex |          |         |          |
|-------------------|--------------------------|----------|---------|----------|
|                   | industry(産業符号)           | sex(性別)  |         | Total    |
| 男                 |                          | 女        |         |          |
| 不詳                |                          | 154607   | 23813.7 | 178421   |
| 鉱業                |                          | 33.83333 | 0       | 33.83333 |
| 建設業               |                          | 31443    | 153.667 | 31596.6  |
| 製造業               |                          | 88374.7  | 2527.57 | 90902.3  |
| 電気・ガス・熱供給・水道業     |                          | 3455.87  | 0       | 3455.87  |
| 情報通信業             |                          | 10006    | 0       | 10006    |
| 運輸業               |                          | 19082.9  | 30.367  | 19113.3  |
| 卸売・小売業            |                          | 32917    | 2920.57 | 35837.5  |
| 金融・保険業            |                          | 7169.07  | 515.433 | 7684.5   |
| 不動産業              |                          | 711.367  | 0       | 711.367  |
| 飲食店・宿泊業           |                          | 782.267  | 531.233 | 1313.5   |
| 医療・福祉             |                          | 5063.77  | 4011.1  | 9074.87  |
| 教育・学習支援業          |                          | 15407.1  | 1334.93 | 16742    |
| 複合サービス事業          |                          | 4881.67  | 54.0667 | 4935.73  |
| サービス業(他に分類されないもの) |                          | 28943.3  | 1613.37 | 30556.8  |
| 公務(他に分類されないもの)    |                          | 52969.8  | 2130.17 | 55099.9  |
| Total             |                          | 495829   | 39638.1 | 499465   |



★表1★⑤ b) 10万世帯比 産業符号×性別 (小数点あり)

The FREQ Procedure

★表1★⑤ data=newWGT (obs=20)

| Obs | sex | Industry | Weight  | Income    | food     | allWGT  | wgt10man | allwgt10man |
|-----|-----|----------|---------|-----------|----------|---------|----------|-------------|
| 1   | 1   |          | 18.9000 | 211011.15 | 37941.54 | 18.900  | 3.81460  | 3.8146      |
| 2   | 1   |          | 15.4333 | 580640.86 | 68062.70 | 34.333  | 3.11492  | 6.9295      |
| 8   | 1   |          | 15.4333 | 423207.84 | 67993.23 | 49.767  | 3.11492  | 10.0444     |
| 4   | 1   |          | 21.0333 | 296881.27 | 32238.18 | 70.800  | 4.24517  | 14.2896     |
| 5   | 1   |          | 8.9333  | 280489.95 | 30854.72 | 79.733  | 1.80302  | 16.0928     |
| 6   | 1   |          | 8.9333  | 311914.61 | 35803.69 | 88.667  | 1.80302  | 17.8957     |
| 7   | 1   |          | 16.9000 | 255731.14 | 58880.19 | 105.567 | 3.41094  | 21.3066     |
| 8   | 1   |          | 9.9333  | 244441.20 | 38041.99 | 115.500 | 2.00485  | 23.3114     |
| 9   | 1   |          | 10.9667 | 192940.31 | 43411.93 | 126.467 | 2.21341  | 25.5249     |
| 10  | 1   |          | 14.1000 | 248051.14 | 50848.08 | 140.567 | 2.84581  | 28.3707     |
| 11  | 1   | 9        | 17.2000 | 255514.31 | 30464.31 | 157.767 | 3.47149  | 31.8422     |
| 12  | 1   | 9        | 13.6000 | 637979.88 | 34834.20 | 171.367 | 2.74490  | 34.5871     |
| 13  | 1   | 9        | 18.1667 | 211772.81 | 52946.06 | 189.533 | 3.66659  | 38.2536     |
| 14  | 1   |          | 15.3111 | 269278.40 | 38950.48 | 204.844 | 3.09025  | 41.3439     |
| 15  | 1   |          | 13.8667 | 123950.88 | 37516.85 | 218.711 | 2.79872  | 44.1426     |
| 16  | 1   |          | 15.3111 | 320849.58 | 50488.01 | 234.022 | 3.09025  | 47.2329     |
| 17  | 1   |          | 14.4000 | 147226.20 | 38034.94 | 248.422 | 2.90636  | 50.1392     |
| 18  | 1   |          | 15.3111 | 427183.07 | 34774.92 | 263.738 | 3.08025  | 53.2295     |
| 19  | 1   |          | 17.0667 | 332853.82 | 27115.22 | 280.800 | 3.44458  | 56.6741     |
| 20  | 1   |          | 17.0667 | 431381.82 | 31465.25 | 297.867 | 3.44458  | 60.1186     |

| Industry(産業符号)    | sex(性別)  |         |         |
|-------------------|----------|---------|---------|
|                   | 男        | 女       | Total   |
| 不詳                | 31204.5  | 4806.33 | 36010.8 |
| 鉱業                | 6.8286   | 0       | 6.8286  |
| 建設業               | 6346.16  | 31.0146 | 6377.17 |
| 製造業               | 17836.7  | 510.141 | 18346.9 |
| 電気・ガス・熱供給・水道業     | 697.5    | 0       | 697.5   |
| 情報通信業             | 2019.52  | 0       | 2019.52 |
| 運輸業               | 3851.52  | 6.12892 | 3857.65 |
| 卸売・小売業            | 6643.65  | 589.46  | 7233.11 |
| 金融・保険業            | 1446.94  | 104.03  | 1550.97 |
| 不動産業              | 143.576  | 0       | 143.576 |
| 飲食店・宿泊業           | 157.885  | 107.219 | 265.105 |
| 医療・福祉             | 1022.02  | 609.583 | 1631.59 |
| 教育・学習支援業          | 3109.62  | 269.431 | 3379.05 |
| 複合サービス事業          | 981.234  | 10.9123 | 992.146 |
| サービス業(他に分類されないもの) | 5841.64  | 325.627 | 6167.27 |
| 公務(他に分類されないもの)    | 10690.93 | 429.933 | 11120.9 |
| Total             | 92000.2  | 7999.79 | 100000  |

★表1★⑥ data=table1

| Obs | Industry          | sex | COUNT    | PERCENT |
|-----|-------------------|-----|----------|---------|
| 1   | 不詳                | 男   | 31204.46 | 31.204  |
| 2   | 不詳                | 女   | 4806.33  | 4.806   |
| 3   | 鉱業                | 男   | 6.83     | 0.007   |
| 4   | 建設業               | 男   | 6346.16  | 6.346   |
| 5   | 建設業               | 女   | 31.01    | 0.031   |
| 6   | 製造業               | 男   | 17836.73 | 17.837  |
| 7   | 製造業               | 女   | 510.14   | 0.510   |
| 8   | 電気・ガス・熱供給・水道業     | 男   | 697.50   | 0.697   |
| 9   | 情報通信業             | 男   | 2019.52  | 2.020   |
| 10  | 運輸業               | 男   | 3851.52  | 3.852   |
| 11  | 運輸業               | 女   | 6.13     | 0.006   |
| 12  | 卸売・小売業            | 男   | 6643.65  | 6.644   |
| 13  | 卸売・小売業            | 女   | 589.46   | 0.589   |
| 14  | 金融・保険業            | 男   | 1446.94  | 1.447   |
| 15  | 金融・保険業            | 女   | 104.03   | 0.104   |
| 16  | 不動産業              | 男   | 143.58   | 0.144   |
| 17  | 飲食店・宿泊業           | 男   | 157.89   | 0.158   |
| 18  | 飲食店・宿泊業           | 女   | 107.22   | 0.107   |
| 19  | 医療・福祉             | 男   | 1022.02  | 1.022   |
| 20  | 医療・福祉             | 女   | 609.56   | 0.610   |
| 21  | 教育・学習支援業          | 男   | 3109.62  | 3.110   |
| 22  | 教育・学習支援業          | 女   | 269.43   | 0.269   |
| 23  | 複合サービス事業          | 男   | 981.23   | 0.981   |
| 24  | 複合サービス事業          | 女   | 10.91    | 0.011   |
| 25  | サービス業(他に分類されないもの) | 男   | 5841.64  | 5.842   |
| 26  | サービス業(他に分類されないもの) | 女   | 325.63   | 0.326   |
| 27  | 公務(他に分類されないもの)    | 男   | 10690.93 | 10.691  |
| 28  | 公務(他に分類されないもの)    | 女   | 429.93   | 0.430   |
|     |                   |     |          | 100.000 |

★表1★⑥ b) 10万世帯比 産業符号×性別 (小数点なし)

The FREQ Procedure

| Industry(産業符号)    | sex(性別) |       |         |
|-------------------|---------|-------|---------|
|                   | 男       | 女     | Total   |
| 不詳                | 31,204  | 4,806 | 36,011  |
| 鉱業                | 7       | 0     | 7       |
| 建設業               | 6,346   | 31    | 6,377   |
| 製造業               | 17,837  | 510   | 18,347  |
| 電気・ガス・熱供給・水道業     | 697     | 0     | 697     |
| 情報通信業             | 2,020   | 0     | 2,020   |
| 運輸業               | 3,852   | 6     | 3,858   |
| 卸売・小売業            | 6,644   | 589   | 7,233   |
| 金融・保険業            | 1,447   | 104   | 1,551   |
| 不動産業              | 144     | 0     | 144     |
| 飲食店・宿泊業           | 158     | 107   | 265     |
| 医療・福祉             | 1,022   | 610   | 1,632   |
| 教育・学習支援業          | 3,110   | 269   | 3,379   |
| 複合サービス事業          | 981     | 11    | 992     |
| サービス業(他に分類されないもの) | 5,842   | 326   | 6,167   |
| 公務(他に分類されないもの)    | 10,691  | 430   | 11,121  |
| Total             | 92,000  | 8,000 | 100,000 |

★表2★① data=fivePCTL:5分位点:乗率あり

| Obs | income20  | income40  | income60  | income80  | food20   | food40   | food60   | food80   |
|-----|-----------|-----------|-----------|-----------|----------|----------|----------|----------|
| 1   | 310531.07 | 404883.13 | 501180.86 | 635875.28 | 49701.02 | 63146.23 | 76055.14 | 93035.80 |

★表2★② 5分位点 計算完了

| Obs | _NAME_   | _LABEL_                        | GOL1      |
|-----|----------|--------------------------------|-----------|
| 1   | income20 | the 20.0000 percentile, income | 310531.07 |
| 2   | food20   | the 20.0000 percentile, food   | 49701.02  |
| 3   | income40 | the 40.0000 percentile, income | 404883.13 |
| 4   | food40   | the 40.0000 percentile, food   | 63146.23  |
| 5   | income60 | the 60.0000 percentile, income | 501180.86 |
| 6   | food60   | the 60.0000 percentile, food   | 76055.14  |
| 7   | income80 | the 80.0000 percentile, income | 635875.28 |
| 8   | food80   | the 80.0000 percentile, food   | 93035.80  |

★表2★③ data=TB2SEIRI

| Obs | bunl | value     | variable |
|-----|------|-----------|----------|
| 1   | 20%  | 310531.07 | income20 |
| 2   | 20%  | 49701.02  | food20   |
| 3   | 40%  | 404883.13 | income40 |
| 4   | 40%  | 63146.23  | food40   |
| 5   | 60%  | 501180.86 | income60 |
| 6   | 60%  | 76055.14  | food60   |
| 7   | 80%  | 635875.28 | income80 |
| 8   | 80%  | 93035.80  | food80   |

★表2★④ income

| Obs | bunl | income    |
|-----|------|-----------|
| 1   | 20%  | 310531.07 |
| 2   | 40%  | 404883.13 |
| 3   | 60%  | 501180.86 |
| 4   | 80%  | 635875.28 |



★表2★⑤ food

| Obs | bunl | food     |
|-----|------|----------|
| 1   | 20%  | 49701.02 |
| 2   | 40%  | 63146.23 |
| 3   | 60%  | 76055.14 |
| 4   | 80%  | 93035.80 |



★表2★⑥ 表2完成

| Obs | bunl | income  | food   |
|-----|------|---------|--------|
| 1   | 20%  | 310,531 | 49,701 |
| 2   | 40%  | 404,883 | 63,146 |
| 3   | 60%  | 501,181 | 76,055 |
| 4   | 80%  | 635,875 | 93,036 |

★表3★① 小数点以下を四捨五入

| Obs | income20 | food20 | income40 | food40 | income60 | food60 | income80 | food80 | i |
|-----|----------|--------|----------|--------|----------|--------|----------|--------|---|
| 1   | 310531   | 49701  | 404883   | 63146  | 501181   | 76055  | 635875   | 93036  | 9 |

★表3★② 経常収入

The FREQ Procedure

| Frequency | Table of income by sex |  |        |       |        |
|-----------|------------------------|--|--------|-------|--------|
|           | sex(性別)                |  |        |       |        |
|           | income(経常収入)           |  | 男      | 女     | Total  |
|           | 310531円未満              |  | 15,494 | 4,506 | 20,000 |
|           | 310531円以上              |  | 18,617 | 1,377 | 19,994 |
|           | 404883円以上              |  | 19,099 | 911   | 20,009 |
|           | 501181円以上              |  | 19,356 | 640   | 19,995 |
|           | 635875円以上              |  | 19,435 | 567   | 20,002 |
|           | Total                  |  | 92,000 | 8,000 | 100000 |

★表3★③ 食料費

The FREQ Procedure

| Frequency | Table of food by sex |  |        |       |        |
|-----------|----------------------|--|--------|-------|--------|
|           | sex(性別)              |  |        |       |        |
|           | food(食料)             |  | 男      | 女     | Total  |
|           | 49701円未満             |  | 16,803 | 3,196 | 19,999 |
|           | 49701円以上             |  | 18,097 | 1,903 | 20,000 |
|           | 63146円以上             |  | 18,734 | 1,268 | 20,001 |
|           | 76055円以上             |  | 19,053 | 948   | 20,001 |
|           | 93036円以上             |  | 19,314 | 685   | 19,999 |
|           | Total                |  | 92,000 | 8,000 | 100000 |

★表4★① 男女別

|     |           | 食料       |          |          |          |          | X      |
|-----|-----------|----------|----------|----------|----------|----------|--------|
|     |           | 49701円未満 | 49701円以上 | 63146円以上 | 76055円以上 | 93036円以上 |        |
|     |           | foodX    | foodX    | foodX    | foodX    | foodX    |        |
|     |           | SumWgt   | SumWgt   | SumWgt   | SumWgt   | SumWgt   | SumWgt |
| 性別  | 經常収入      |          |          |          |          |          |        |
| 男   | 310531円未満 | 5,827    | 3,974    | 2,798    | 1,834    | 1,060    | 15,494 |
|     | 310531円以上 | 5,130    | 4,501    | 3,978    | 2,996    | 2,012    | 18,617 |
|     | 404883円以上 | 2,911    | 4,301    | 4,395    | 4,294    | 3,198    | 19,099 |
|     | 501181円以上 | 1,906    | 3,233    | 4,082    | 4,903    | 5,231    | 19,256 |
|     | 635875円以上 | 1,029    | 2,087    | 3,481    | 5,026    | 7,813    | 19,435 |
|     | All       | 16,803   | 18,097   | 18,734   | 19,053   | 19,314   | 92,000 |
| 女   | 經常収入      |          |          |          |          |          |        |
|     | 310531円未満 | 2,462    | 994      | 565      | 314      | 172      | 4,506  |
|     | 310531円以上 | 371      | 399      | 235      | 211      | 161      | 1,377  |
|     | 404883円以上 | 243      | 225      | 216      | 164      | 63       | 911    |
|     | 501181円以上 | 68       | 134      | 156      | 143      | 140      | 640    |
|     | 635875円以上 | 53       | 152      | 96       | 117      | 149      | 567    |
| All | 3,196     | 1,903    | 1,268    | 948      | 685      | 8,000    |        |
| All | 19,999    | 20,000   | 20,001   | 20,001   | 19,999   | 100,000  |        |

★表5★① data=table5 (obs=15)

| Obs | sex | industry | Weight  | Income    | food     | allWGT  | wgt10man | allwgt10man | foodLOG |
|-----|-----|----------|---------|-----------|----------|---------|----------|-------------|---------|
| 1   | 1   | .        | 18.9000 | 211011.15 | 37941.54 | 18.900  | 3.81460  | 3.8146      | 4.57911 |
| 2   | 1   | .        | 15.4333 | 580649.86 | 68062.70 | 34.333  | 3.11492  | 6.9295      | 4.83291 |
| 3   | 1   | .        | 15.4333 | 423207.94 | 67993.23 | 49.767  | 3.11492  | 10.0444     | 4.83247 |
| 4   | 1   | .        | 21.0333 | 299831.27 | 32238.18 | 70.800  | 4.24517  | 14.2896     | 4.50837 |
| 5   | 1   | .        | 8.9333  | 280483.95 | 30854.72 | 79.733  | 1.80302  | 16.0926     | 4.48932 |
| 6   | 1   | .        | 8.9333  | 311914.61 | 35803.69 | 88.667  | 1.80302  | 17.8957     | 4.55393 |
| 7   | 1   | .        | 16.9000 | 295731.14 | 58880.19 | 105.567 | 3.41094  | 21.3066     | 4.76997 |
| 8   | 1   | .        | 9.9333  | 244441.20 | 39041.99 | 115.500 | 2.00485  | 23.3114     | 4.59153 |
| 9   | 1   | .        | 10.9667 | 192940.31 | 43411.93 | 126.467 | 2.21341  | 25.5249     | 4.63761 |
| 10  | 1   | .        | 14.1000 | 249051.14 | 50348.08 | 140.567 | 2.84581  | 28.3707     | 4.70198 |
| 11  | 1   | 9        | 17.2000 | 255514.31 | 30464.31 | 157.767 | 3.47149  | 31.8422     | 4.48379 |
| 12  | 1   | 9        | 13.6000 | 637979.83 | 34834.20 | 171.367 | 2.74490  | 34.5871     | 4.54201 |
| 13  | 1   | 9        | 18.1667 | 211772.81 | 52946.06 | 189.533 | 3.66659  | 38.2536     | 4.72383 |
| 14  | 1   | .        | 15.3111 | 269278.40 | 33950.48 | 204.844 | 3.09025  | 41.3439     | 4.53085 |
| 15  | 1   | .        | 13.8667 | 123950.88 | 37516.85 | 218.711 | 2.79872  | 44.1426     | 4.57423 |

★表5★③ proc tabulateから出力のdataset「result5」:幾何平均なし

| Obs | sex | income    | foodLOG_SumWgt | foodLOG_Mean |
|-----|-----|-----------|----------------|--------------|
| 1   | 男   | 310531円未満 | 15493.70       | 4.73889      |
| 2   | 男   | 310531円以上 | 18616.73       | 4.78322      |
| 3   | 男   | 404883円以上 | 19099.57       | 4.83680      |
| 4   | 男   | 501181円以上 | 19355.80       | 4.88259      |
| 5   | 男   | 635875円以上 | 19435.42       | 4.93345      |
| 6   | 女   | 310531円未満 | 4505.94        | 4.67130      |
| 7   | 女   | 310531円以上 | 1376.78        | 4.78107      |
| 8   | 女   | 404883円以上 | 910.75         | 4.78132      |
| 9   | 女   | 501181円以上 | 639.70         | 4.86171      |
| 10  | 女   | 635875円以上 | 566.62         | 4.87741      |
| 11  | 男   | 計         | 92000.21       | 4.83952      |
| 12  | 女   | 計         | 7999.79        | 4.73254      |

★表5★② proc tabulateから表示

|    |           | foodLOG |      |
|----|-----------|---------|------|
|    |           | SumWgt  | Mean |
| 性別 | 經常収入      |         |      |
| 男  | 310531円未満 | 15494   | 4.74 |
|    | 310531円以上 | 18617   | 4.78 |
|    | 404883円以上 | 19099   | 4.84 |
|    | 501181円以上 | 19356   | 4.88 |
|    | 635875円以上 | 19435   | 4.93 |
|    | All       | 92000   | 4.84 |
| 女  | 經常収入      |         |      |
|    | 310531円未満 | 4506    | 4.67 |
|    | 310531円以上 | 1377    | 4.78 |
|    | 404883円以上 | 911     | 4.78 |
|    | 501181円以上 | 640     | 4.86 |
|    | 635875円以上 | 567     | 4.88 |
|    | All       | 8000    | 4.73 |

★表5★④ proc tabulateから出力のdataset「result5」:幾何平均

| Obs | sex | income    | foodLOG_SumWgt | foodLOG_Mean |
|-----|-----|-----------|----------------|--------------|
| 1   | 男   | 310531円未満 | 15494          | 4.7389       |
| 2   | 男   | 310531円以上 | 18617          | 4.7832       |
| 3   | 男   | 404883円以上 | 19099          | 4.8368       |
| 4   | 男   | 501181円以上 | 19356          | 4.8826       |
| 5   | 男   | 635875円以上 | 19435          | 4.9335       |
| 6   | 男   | 計         | 92000          | 4.8395       |
| 7   | 女   | 310531円未満 | 4506           | 4.6713       |
| 8   | 女   | 310531円以上 | 1377           | 4.7811       |
| 9   | 女   | 404883円以上 | 911            | 4.7813       |
| 10  | 女   | 501181円以上 | 640            | 4.8617       |
| 11  | 女   | 635875円以上 | 567            | 4.8774       |
| 12  | 女   | 計         | 8000           | 4.7325       |

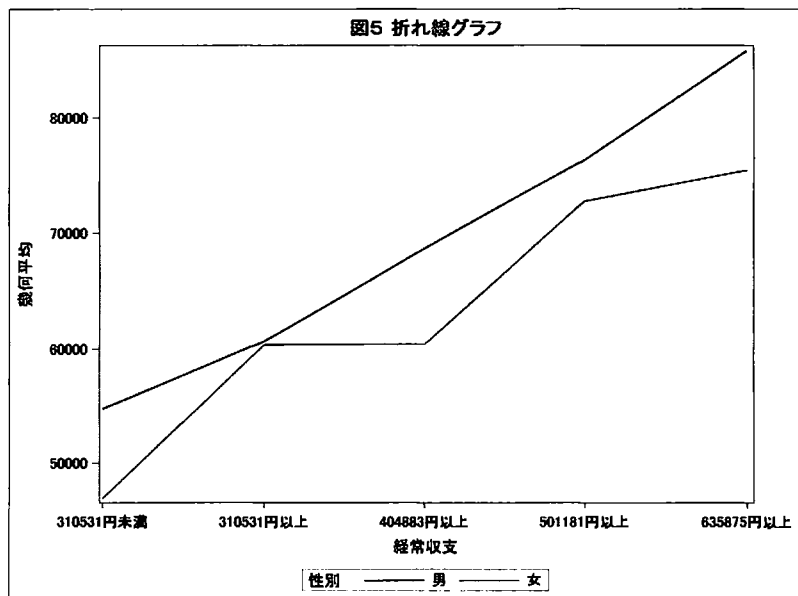
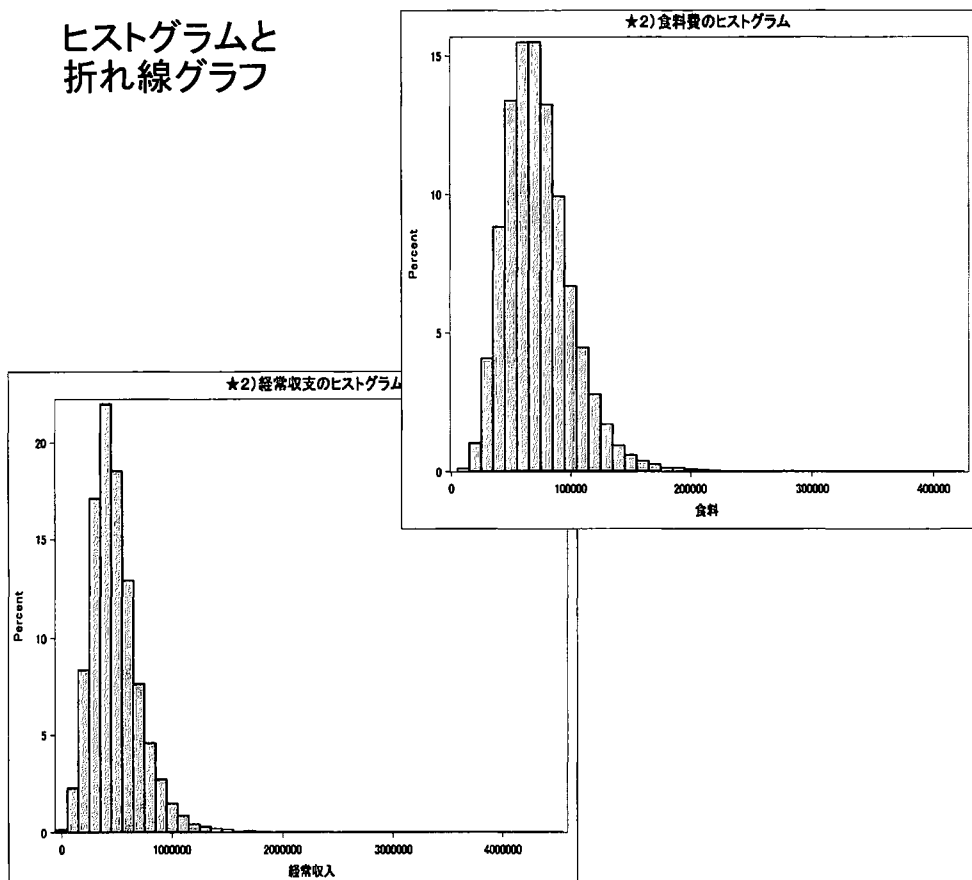
★表5★⑤ data=final\_result5:表5完成

| Obs | 性別 | 經常収支      | 世帯数   | 常用対数平均 | 幾何平均   |
|-----|----|-----------|-------|--------|--------|
| 1   | 男  | 310531円未満 | 15494 | 4.7389 | 54,814 |
| 2   | 男  | 310531円以上 | 18617 | 4.7832 | 60,704 |
| 3   | 男  | 404883円以上 | 19099 | 4.8368 | 68,675 |
| 4   | 男  | 501181円以上 | 19356 | 4.8826 | 76,311 |
| 5   | 男  | 635875円以上 | 19435 | 4.9335 | 85,793 |
| 6   | 男  | 計         | 92000 | 4.8395 | 69,107 |
| 7   | 女  | 310531円未満 | 4506  | 4.6713 | 46,913 |
| 8   | 女  | 310531円以上 | 1377  | 4.7811 | 60,405 |
| 9   | 女  | 404883円以上 | 911   | 4.7813 | 60,439 |
| 10  | 女  | 501181円以上 | 640   | 4.8617 | 72,730 |
| 11  | 女  | 635875円以上 | 567   | 4.8774 | 75,407 |
| 12  | 女  | 計         | 8000  | 4.7325 | 54,018 |

図5 折れ線グラフ用 data=plot

| Obs | sex | income    | geomean  |
|-----|-----|-----------|----------|
| 1   | 男   | 310531円未満 | 54813.86 |
| 2   | 男   | 310531円以上 | 60704.43 |
| 3   | 男   | 404883円以上 | 68674.76 |
| 4   | 男   | 501181円以上 | 76310.70 |
| 5   | 男   | 635875円以上 | 85793.40 |
| 6   | 女   | 310531円未満 | 46913.36 |
| 7   | 女   | 310531円以上 | 60405.14 |
| 8   | 女   | 404883円以上 | 60439.18 |
| 9   | 女   | 501181円以上 | 72729.70 |
| 10  | 女   | 635875円以上 | 75407.31 |

# ヒストグラムと折れ線グラフ



## 付録5 2016年 第4回 SASマイクロデータ分析コンテスト

### 規定課題

#### 1) 産業符号・性別 クロス表

産業符号別・性別に集計用乗率なしの単純クロス集計表、および集計用乗率を使って10万世帯あたり(10万世帯比)のクロス集計表を、表1の様式で作成して下さい。(注:産業符号の「不詳」が先頭に来ても構いません。)

表1

| 産業符号              | a) 単純クロス表 |       |        | b) 10万世帯比でのクロス表 |       |         |
|-------------------|-----------|-------|--------|-----------------|-------|---------|
|                   | 男         | 女     | 合計     | 男               | 女     | 合計      |
| 鉱業                |           |       |        |                 |       |         |
| 建設業               |           |       |        |                 |       |         |
| 製造業               |           |       |        |                 |       |         |
| 電気・ガス・熱供給・水道業     |           |       |        |                 |       |         |
| 情報通信業             |           |       |        |                 |       |         |
| 運輸業               |           |       |        |                 |       |         |
| 卸売・小売業            |           |       |        |                 |       |         |
| 金融・保険業            |           |       |        |                 |       |         |
| 不動産業              |           |       |        |                 |       |         |
| 飲食店・宿泊業           |           |       |        |                 |       |         |
| 医療・福祉             |           |       |        |                 |       |         |
| 教育・学習支援業          |           |       |        |                 |       |         |
| 複合サービス事業          |           |       |        |                 |       |         |
| サービス業(他に分類されないもの) |           |       |        |                 |       |         |
| 公務(他に分類されないもの)    |           |       |        |                 |       |         |
| 不詳                |           |       |        |                 |       |         |
| 合計                | 29,381    | 2,646 | 32,027 | 92,000          | 8,000 | 100,000 |

#### 2) 経常収入および食料費

経常収入および食料費について、集計用乗率を用いて表2のように5分位点(下側20%点, 40%点, 60%点および80%点)A~Fを求めて、表2を完成して下さい。更に、経常収入および食糧費についてそれぞれヒストグラムを描いて下さい。ヒストグラムの縦軸はパーセント(又は10万世帯比)、横軸は経常収入および食糧費の階級区分(単位:円)とします。

| 分位点 | 経常収入    | 食料費    |
|-----|---------|--------|
| 20% | 310,531 | 49,701 |
| 40% | A       | D      |
| 60% | B       | E      |
| 80% | C       | F      |

#### 3) 男女別・経常収入および食料費の5分位分布

男女別に経常収入および食料費について、表2に示す5分位別に、以下の表3に示す10万世帯比のクロス表を作成して下さい。

表3

| 経常収入 |           |        |       | 食料費     |     |          |        |       |         |
|------|-----------|--------|-------|---------|-----|----------|--------|-------|---------|
| 5分位  | 5分位数      | 男      | 女     | 計       | 5分位 | 5分位数     | 男      | 女     | 計       |
| 1    | 310,531未満 |        |       | 20,000  | 1   | 49,701未満 |        |       | 19,999  |
| 2    | 310,531以上 |        |       | 19,993  | 2   | 49,701以上 |        |       | 20,000  |
| 3    | A         |        |       | 20,009  | 3   | D        |        |       | 20,001  |
| 4    | B         |        |       | 19,995  | 4   | E        |        |       | 19,998  |
| 5    | C         |        |       | 20,002  | 5   | F        |        |       | 20,003  |
|      | 計         | 92,000 | 8,000 | 100,000 |     | 計        | 92,000 | 8,000 | 100,000 |

4) 男女別・経常収入・食料費 についての3重クロス表

男女別・経常収入(5分位)別の食料費(5分位)について、表4に示す3重クロス集計表を作成して下さい。

表4

|     | 経常収入<br>5分位 | 食料(5分位) |        |        |        |        | 計       |
|-----|-------------|---------|--------|--------|--------|--------|---------|
|     |             | 1       | 2      | 3      | 4      | 5      |         |
| 男   | 1           |         |        |        |        |        |         |
|     | 2           |         |        |        |        |        |         |
|     | 3           |         |        |        |        |        |         |
|     | 4           |         |        |        |        |        |         |
|     | 5           |         |        |        |        |        |         |
|     | 計           |         |        |        |        |        |         |
| 女   | 1           |         |        |        |        |        |         |
|     | 2           |         |        |        |        |        |         |
|     | 3           |         |        |        |        |        |         |
|     | 4           |         |        |        |        |        |         |
|     | 5           |         |        |        |        |        |         |
|     | 計           |         |        |        |        |        |         |
| 男女計 |             | 19,999  | 20,000 | 20,001 | 19,998 | 20,003 | 100,000 |

5) 男女別・経常収入別に食料費を常用対数変換した平均値の比較

表5に示すように、男女別・経常収入(5分位)別に、食料費の常用対数を計算し平均値を求めて下さい。その平均値について逆対数変換(10<sup>x</sup>)をして幾何平均を計算して下さい。この幾何平均について男女別・経常収入の分位点別に食糧費を比較するために、折れ線グラフを作成して下さい。縦軸を幾何平均(単位:円)、横軸を経常収入(5分位)とします。

表5

|   | 経常収入<br>5分位 | 食料     |        |        |
|---|-------------|--------|--------|--------|
|   |             | 世帯数    | 常用対数平均 | 幾何平均   |
| 男 | 1           |        |        |        |
|   | 2           |        |        |        |
|   | 3           |        |        |        |
|   | 4           |        |        |        |
|   | 5           |        |        |        |
|   | 計           | 92,000 | 4.8395 | 69,103 |
| 女 | 1           |        |        |        |
|   | 2           |        |        |        |
|   | 3           |        |        |        |
|   | 4           |        |        |        |
|   | 5           |        |        |        |
|   | 計           | 8,000  | 4.7325 | 54,013 |

## 統計理論



## 局外トレンドのある回帰モデルの推測

○藤原滉<sup>1</sup>、丸川真友子<sup>1</sup>、富田哲治<sup>1</sup>

(<sup>1</sup>県立広島大学経営情報学部経営情報学科)

経時データに対する回帰モデルにおいて、測定時点が少ない場合は経時変化の記述に直線などの単純な関数形を用いる場合が多い。しかし、測定時点が多い場合は、より複雑な関数形(例えば、多項式関数・スプライン関数・非線形関数)を考える必要がある。その結果、多項式関数ならば次数選択、スプライン関数であれば基底の種類や個数の選択など、解析者に新たな負担が生じる。本発表では、あまり関心はないが無視できない経時変化を局外トレンドとして捉え、局外トレンドの関数形を特定することなく、共変量の効果を推測する方法について説明する。

経時データの例として、少女11人と少年16人に対し、歯科矯正の診断に利用するために頭部X線写真から2つの基準点(脳下垂体と翼上顎裂)の距離[mm]を4時点(8, 10, 12, 14歳)で計測したデータ(Pothoff and Roy: *Biometrika*, 51, 313-326, 1964)を用いる。個体*i*の*t*歳における距離を $y_i(t)$ 、性別を表すダミー変数を $a_i(0=少女, 1=少年)$ とし、次のような回帰モデルを考える。

$$y_i(t) = \beta_0(t) + \beta_1(t)a_i + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma^2)$$

$\beta_j(t)$ は変化係数(Hasti and Tibshirani: *J. Roy. Statist. Soc. Ser. B*, 55(4), 757-796, 1993)と呼ばれる時間とともに変化する回帰係数である。 $\beta_0(t)$ は少女の経時変化、 $\beta_1(t)$ は性差の経時変化を表し、本稿では少女の経時変化 $\beta_0(t)$ を局外トレンドとする。つまり、 $\beta_0(t)$ の関数形を特定せずに $\beta_1(t)$ を推定する。性差の経時変化に直線を用いた推定結果が図1である。共変量の効果の推定値を用いて、事後的に局外トレンドを推定することもでき、その推定結果が図2である。

生存時間解析において、コックスの比例ハザードモデルが広く用いられている理由として、ベースラインハザードの関数形を特定する必要がないことが挙げられ、考慮すべきことが減ることにより解析者への負担が軽減されているからと思われる。本発表の方法においても、同様の利点が期待される。このような方法は、測定時点が多い経時データにおいて有効であると考えられる。本発表では、簡単のため時点間相関は独立として扱ったが、Satoh and Tonda (*Amer. J. Math. Management Sci.*, 35(3), 183-193, 2016) は、成長曲線モデルの枠組みで時点間相関も考慮した推測法を提案している。

図1：共変量効果（性差）の推定結果

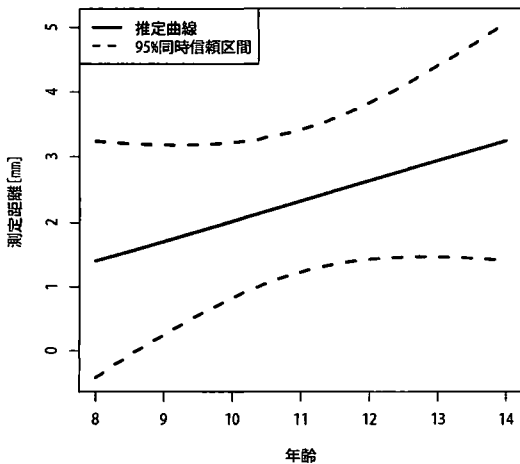
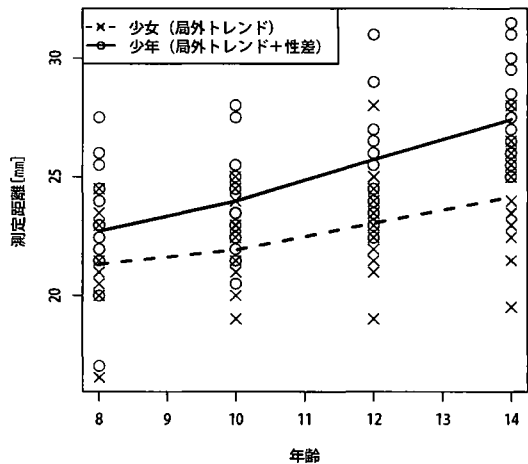


図2：少女と少年の推定曲線



# ベイジアンモデルによる地域人口予測モデルの可能性について

○片桐智志<sup>1</sup>, 山下諭史<sup>1</sup>

(<sup>1</sup>ネイチャーインサイト株式会社)

The possibility of regional population forecasting model by Bayesian model

Satoshi Katagiri and Satoshi Yamashita

Nature Insight CO., LTD.

## 要旨

人口増減を表現したリー・カーター法に基づく階層ベイズモデルをマルコフ連鎖モンテカルロ法によって推定し、岩手県の市町村別人口の予測モデルの作成を試みた。利用できるデータの制約上、全国規模のデータを代用したため完全ではないものの、直線的な人口推計モデルと異なる、確率的モデルによる人口予測の可能性を示すことができた。

キーワード：人口予測 MCMC マルコフ連鎖モンテカルロ法 階層ベイズ リー・カーター (Lee-Carter) 法

## はじめに

[国立社会保障・人口問題研究所, 2013] は、2010年から60年までの50年の間に約3分の1の人口減少が発生するとの推計を示した。これは出生率の低下が長期間続いてきたことが原因のひとつであるため、推計の精度に差こそあれ、(仮に今すぐ出生率の水準がおおきく上昇したとしても)長い人口減少期間が発生するという結果は頑健である。人口の減少が確実となった場合、とりわけ、昼夜人口の差が小さく、住人の生活が限られた範囲に根付いているような地方では、将来の商品の需要を見きわめるため、市町村のような小さな地域の人口を年齢層ごとの分布を含めて適切に予測することが重要となってくる。

市区町村別の推計として、すでに国立社会保障・人口問題研究所の「日本の地域別将来推計人口」が存在するが、これは人口の「推計」という立場をとっており、現在から直近数年分の人口の変化が直線的に続くという仮定のもとでのいわば「参考値」である。本稿では、人口の増減率が動的に変化するモデルを採用することで、より「予測」として適した、市町村別の人口推計モデルの作成を意図している。

## 人口予測の背景

人口が増減する要因は自然増減と社会増減の2つに大別される。自然増減とは死亡と出生による人口の増減で、社会増減とは経済・社会活動のために人間が他地域へ移動することによって発生する増減を意味する。自然増減と社会増減の大きさを適切に予測できれば、そこから人口予測が可能となる。

特に自然増減に関しては、(Lee & Carter, 1992)と(Lee, 1993)においてそれぞれ、死亡率と出生率を予測するモデルが提案され、リー・カーター (Lee-Carter) 法として定着している。国内におけるリー・カーター法の応用研究には、[石井太, 2006]や[南條善治 & 吉永一彦, 2006]ほか多くがある。さらに(Brouhns, et al., 2002)によって、リー・カーター法を発展させポアソン分布を用いた確率モデルが導入され、確率モデルを導入した人口予測の研究がすすんでいる。国内でも[小暮厚之 & 長谷川智弘, 2005]などの研究がある。(Brouhns, et al., 2002)を素朴に発展させ階層ベイズモデルとした(Wiśniowski, et al., 2015)で示された方法に基づいた人口予測を試みる。

## モデル

(Lee & Carter, 1992) にもとづいて、 $t$  年に  $x$  歳である人々の死亡率  $\mu_{x,t}$  を

$$\mu_{x,t} = \exp(\alpha_x + \beta_x \kappa_t)$$

とする。 $\alpha_x$ ,  $\beta_x$  はいずれも年齢ごとの係数で、年次によって変化することがない。一方、 $\kappa_t$  はどの年齢でも同じだが年によって値が変化する。これらの3種類の母数のうち(1)  $\alpha_x$  は時間経過に対して一定な年齢ごとに固有の効果で、(2)  $\beta_x$  は各年齢に固有の死亡率の減少(増加)傾向の大きさであり、(3)  $\kappa_t$  は時間に対して変動するトレンドと解釈できる。この3つに分解し、 $\kappa_t$  に何らかの時系列モデルを仮定することで将来の死亡率の予測を行うのがリー・カーター法となる。さらに、これは死亡率だけでなく出生率に対しても適用できる(Lee, 1993)。リー・カーター法は当初は実際のデータとしては死亡者数だけを利用して特異値分解によってこれらのパラメータを求めていたが、パラメータを人口に対する死亡の発生確率のパラメータとし、人口と死亡者数のデータを用いた確率モデルとして定義しなおした(Brouhns, et al., 2002)によって、特異値よりも直感的に解釈しやすいパラメータとして求めることができるようになった。この式は双線形モデルと呼ばれる形式のため、制約つき最適化計算により求める必要がある。

パラメータの分散を表すため、 $\mu_{x,t}$  の対数は正規分布にしたがうものとする。なお、(Wiśniowski, et al., 2015)では(Renshaw & Haberman, 2006)で導入されたコーホート別効果  $\gamma_x$  をさらに仮定しているが、本稿では一般化のためこのパラメータは使用しない。よって、

$$\ln \mu_{x,t} = \mathcal{N}(\alpha_x + \beta_x \kappa_t, \sigma_\mu^2)$$

とする。つまり、 $\mu_{x,t}$  は正の値に収まるため、発生率として矛盾がない値だけをとる。分散は標準偏差  $\sigma_\mu$  が連続一様分布

$$\sigma_u \sim \mathcal{U}(0, 100)$$

にしたがうようにする。さらに、イベントの発生数をポアソン分布

$$Y_{x,t} \sim \mathcal{Po}(\mu_{x,t} E_{x,t})$$

とする。  $Y_{x,t}$  は実際の発生数、  $E_{x,t}$  はイベントが起こりうる人間の数である。ポアソン分布のパラメータがイベントの発生数の期待値であることから、  $\mu_{x,t}$  は発生率と解釈できる。よって、以降は一般性をもたせるために  $\mu_{x,t}$  を(イベントの)発生率パラメータと呼ぶ。さらに、パラメータ  $\alpha_x$ 、  $\beta_x$ 、  $\kappa_t$  については以下のように事前分布を設定する。

$\alpha_x$  はそれぞれ独立に正規分布に従う。

$$\alpha_x \sim NID(0, 100)$$

$\beta_x$  は相関のある多変量正規分布とする。

$$\beta_{1:X-1} \sim MVN\left(\frac{1}{X}, \tau_\beta \Psi\right),$$

$$\beta_X = 1 - \sum_{i \neq X} \beta_i$$

ただし、共分散行列の係数  $\tau_\beta$  はガンマ分布に従う。

$$\tau_\beta \sim \Gamma(0.001, 0.001)$$

$$\Psi = \begin{bmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 2 \end{bmatrix}$$

$\kappa_t$  はベクトル自己回帰過程とする。ただし、  $\kappa_1$  は固定する。

$$\kappa_t \sim MVN(\phi_0 + \Phi_1 \kappa_{t-1}, \mathbf{T}_\kappa), t = 1, 2, \dots,$$

$$\kappa_0 = \mathbf{1}$$

ただし、  $\phi_0$ 、  $\Phi_1$  の事前分布は要素ごとに独立な正規分布で、分散は逆ウィシャート分布とする。

$$\phi_0 \sim NID(0, 1),$$

$$\Phi_1 \sim NID(0, 1),$$

$$\mathbf{T}_\kappa \sim invWish(2, 2 \times \mathbf{I}_2)$$

とする。また、(Wiśniowski, et al., 2015)で提案されているように  $\kappa_t$  を VAR ではなくドリフト項のあるランダムウォークとした場合についても試し、さらに男女で独立な AR(1)の場合についても推定した。この場合、  $\phi_1$  がドリフト項となり、分散は VAR と同様に逆ウィシャート分布とする。

同様に出生率も求められるが、出生率の場合  $\kappa_t$  は男女別ではなく、女性だけの単変量となり、  $Y_{x,t}$  も男女の出生数の合計となる。また、(Lee, 1993)に則して ARMA(1,1)の場合と AR(1)の場合の2通りで推定した。単変量の場合の分散はいずれも、標準偏差が連続一様分布  $U(0,100)$  となるように事前分布を設定した。これらの  $\kappa_t$  の事前分布の候補のうちどれを採用するかについては、(Spiegelhalter, et al., 2002)によるデビエンス情報基準(DIC)のより小さいものを選ぶようにした。

以上のモデルはいわば階層ベイズモデルの一種であるため、マルコフ連鎖モンテカルロ(MCMC)法でパラメータを求めることができる。この求めたパラメータを用いて、同一出生集団ごとに人口を毎年の人口を計算するコーホート要因法の一つであるロジャース・モデルにより人口を求めることができる。まず、単一地域

のコーホート要因法を考えると、コーホート別出生率とコーホート別人口を掛けて新生児（0歳年齢）人口を求め、さらに前年のコーホート別人口と対応する生残率を掛けることで翌年の予測人口を計算できる。[小池司朗, 2008]によれば、複数の地域ごとにコーホート要因法で人口を計算する場合、地域間の人口の移動の考慮が必要という課題があり、そのなかでもロジャース・モデルはある地域からある地域への転入出率のパラメータを個別に設定したものである。(Wiśniowski, et al., 2015)の方法にならって、以下のようなモデルを仮定する。  $p_{x,t}^k$  を  $t$  年における性別  $k$ ,  $x$  歳の人口とする。同様に、  $b_{x,t}^k$  は出生率、  $s_{x,t}^k$  は生残率、  $i_{x,t}^k$  は転入人口である。

$$\begin{bmatrix} p_{0,t+1}^F \\ p_{1,t+1}^F \\ \vdots \\ p_{X,t+1}^F \\ p_{0,t+1}^M \\ \vdots \\ p_{X,t+1}^M \end{bmatrix} = \begin{bmatrix} 0 & \cdots & b_{x,t}^F & b_{x+1,t}^F & \cdots & 0 \\ s_{0,t}^F & & & & & 0 \\ & s_{1,t}^F & & & & \vdots \\ & & \ddots & & & 0 \\ \mathbf{0} & & & s_{X-1,t}^F & s_{X,t}^F & \\ 0 & \cdots & b_{x,t}^M & b_{x+1,t}^M & \cdots & 0 \\ & & & & & \mathbf{0} \end{bmatrix} \begin{bmatrix} i_{0,t}^F \\ i_{1,t}^F \\ \vdots \\ i_{X,t}^F \\ i_{0,t}^M \\ \vdots \\ i_{X,t}^M \end{bmatrix} + \begin{bmatrix} p_{0,t}^F \\ p_{1,t}^F \\ \vdots \\ p_{X,t}^F \\ p_{0,t}^M \\ \vdots \\ p_{X,t}^M \end{bmatrix}$$

生残率パラメータは以下のように計算する。

$$s_{x,t}^k = \begin{cases} \frac{1 - \mu_{x,t}^{D,k}}{1 + \mu_{x+1,t}^{D,k}} & \text{if } x \leq X \\ \frac{1 - \mu_{x,t}^{D,k}}{1 + \mu_{x,t}^{D,k}} & \text{if } x = X \end{cases}, k = F \text{ or } M$$

ここで  $\mu_{x,t}^{D,k}$  は先のモデルでの死亡率パラメータを意味する。出生率パラメータは、(Wiśniowski, et al., 2015)にならい、男女合計出生率  $b_{x,t}$ ,

$$b_{x,t} = \frac{1}{1 + 0.5} \mu_{0,t}^{D,F} \frac{1}{2} (\mu_{x,t}^B + s_{x,t}^F \mu_{x+1,t}^B)$$

に対し  $1/2.05$  をかけたものを女児出生率に、  $(1 - 1/2.05)$  をかけたものを男児出生率に使っている。

$$b_{x,t}^F = \frac{1}{2.05} b_{x,t},$$

$$b_{x,t}^M = \left(1 - \frac{1}{2.05}\right) b_{x,t}$$

ただし、  $\mu_{x,t}^B$  は出生率パラメータを意味する。便宜上行列を用いて表したが、実際の計算はSQLプロシジャを用いた集計を年ごとに再帰的に実行することで可能である。

## 用いたデータ

本稿の目的は岩手県の市町村別人口の予測であるが、市町村別の死亡・出生人数について年齢で分けた十分なデータを得られないため、死亡数は各年の人口動態調査の「死亡数、性・年齢別」より、人口は国勢調査

の全国結果を補間した人口推計を、出生数は人口動態統計「出生数、性・母の年齢(各歳)、出生順位・嫡出子-嫡出でない別」の1993-2010年間のデータを用いた。岩手県の市町村・性年齢別人口については、「いわての統計情報」(<http://www3.pref.iwate.jp/webdb/view/outside/s14Tokei/top.html>)の岩手県人口移動報告年報の「市町村・男女別人口(各歳・5歳階級)」より入手できるため、ロジャース・モデルによる計算の初期値とした。転入出については全国データの代用ができず、また年齢別の人数を得られない。そこで、人口移動がないと仮定した閉鎖人口のみの場合と、岩手県人口移動報告年報の「市町村別社会動態(転入数・転出数)」から得た過去5年の市町村別データの平均を、転入出数の年齢分布の得られる県全体の値で按分した疑似値で代用した。

## 計算方法と予測結果

MCMC法の計算は、SAS/STAT 14.1のMCMCプロシジャで実施することができるが、パラメータの非常に多い本モデルの計算は処理時間が長大になったため、本稿ではStanというソフトを用いて計算を行った。一方で、SASのMCMCプロシジャそのものを使うことはできなかったものの、診断用の自動マクロはデータセットの体裁をMCMCプロシジャのものに合わせれば使うことができる。SAS/IMLが使えるならば、例えば(Geweke, 1992)によるパラメータ収束の検定は%GEWEKE()マクロで実行でき、(Brooks & Gelman, 1998)による複数系列が同一の点に収束しているかの診断は%GELMAN()マクロで実行できる。そのほかにも多くのマクロがユーザーガイドで紹介されている。ただし、GEWEKEはプロシジャのオプションでも可能である。MCMCプロシジャの出力データセットは以下の表1ようになる。

| 変数名       | 型  | 説明          |
|-----------|----|-------------|
| ITERATION | 数値 | サンプルのインデックス |
| (任意の変数名)  | 数値 | 事後パラメータの値   |
| LOGPRIOR  | 数値 | 事前確率の対数     |
| LOGLIKE   | 数値 | 対数尤度        |
| LOGPOST   | 数値 | 事後確率の対数     |

表 1 SAS の MCMC 診断マクロの入力フォーマット

Stanの出力結果もよく似たフォーマットになり、変数名がSASの命名則に反してないことに気がつけばIMPORTプロシジャで容易に読み込むことができる。Stanでは事後確率がlp\_という名称で自動計算されるため、対数尤度loglikeを計算するようプログラムした結果をcsvで出力したという前提で、プログラム1のようにして診断できる。ただし、パラメータが非常に膨大な場合、1行あたりの読み込みバイト数上限を超える可能性があり注意が必要である。

```
/* csv 読み込み */
proc import file="FOO/BAR.csv"
    out =work. result
```

```

dbms=csv replace ;

run ;
/* 対数事前確率計算・イテレーション追加 */
data work. result ;
  set work. result (rename =(lp__ = logpost)) ;
  logprior = logpost - loglike ;
  iteration + 1 ;
run ;

%GEWEKE(data=work. result) ;

```

プログラム 1 Stan の結果を SAS の MCMC 診断マクロで診断する

DIC を求めたところ、出生率は AR(1) のケースが優れていた。一方で、死亡率のパラメータはドリフトランダムウォークと VAR(1) いずれのケースでも男性に対応するパラメータの事後分布の収束が女性のものと比べて遅かった。そこで、男女で独立な AR(1) を事前分布として設定したものを採用した。以下の図 1 および図 2 は、それぞれ MCMC で推定した死亡率パラメータと出生率パラメータの対数のプロットである。青色が実績期間、灰色が予測期間で、年が新しくなるほど淡色となる。年々、乳幼児の死亡率と高齢者の死亡率が低下していることが顕著である。また、出生率についても、30 代前半までが低下傾向にあるのに対して、それより上の年代は増加傾向にあることがわかる。

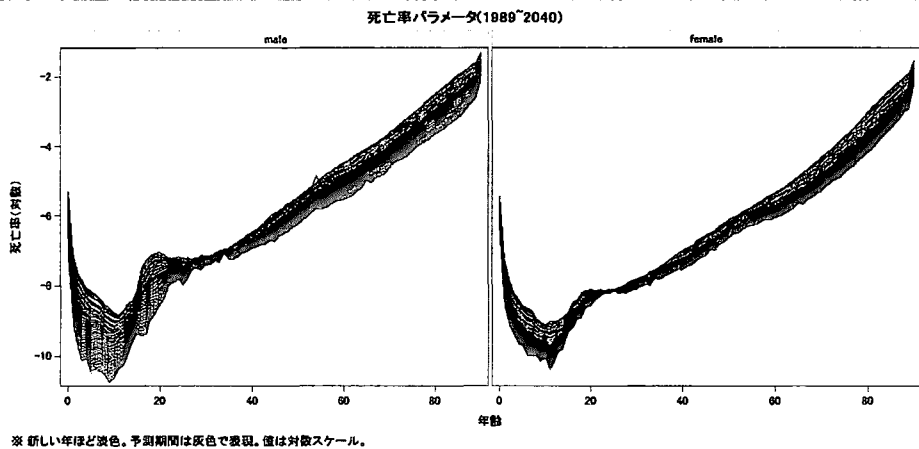


図 1 男女・年齢別死亡率パラメータの推定および予測値 (平均値)

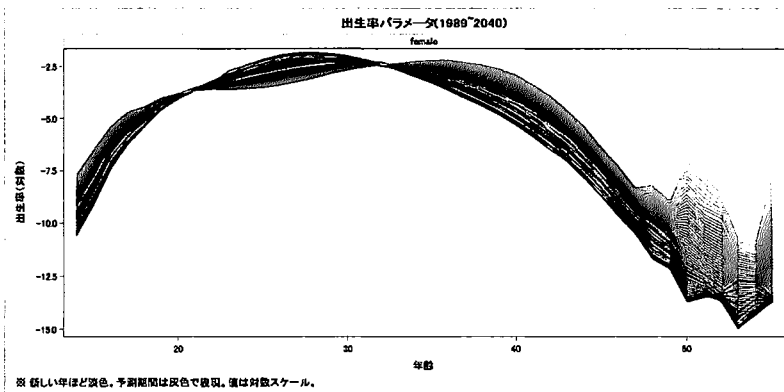


図 2 年齢別出生率パラメータの推定および予測値 (平均値)

このグラフは TEMPLTE プロシジャを用いて作成した。プログラム 2 はその際のマクロの抜粋である。

```
%macro plot_ln_mu_2(ds, year_act_end, entrytitle, entryfootnote);
/* グラデーションになるよう色コードを作成 */
data work. tmp_color_mst(drop = i ) ;
    _HEX_B length = $8 /*16進数カラーコード青 */
    _HEX_G length = $8 /*16進数カラーコード灰 */ ;
do i = 0 to 200 ;
    _HEX_B = cats('CX',PUT(_10, HEX2. ),PUT(_10, HEX2. ) , 'FF') ;
    _HEX_G = cats('CX',PUT(_10, HEX2. ),PUT(_10, HEX2. ) , PUT(_10, HEX2. ) ) ;
output ;
end ;
run;
/*年とカラーコード対応表を作成し、マクロ変数に出力する処理中略 */
/* グラフテンプレートを作成 */
proc template ;
define statgraph sgspaghetti ;
dynamic _AGE _LN_MU _YEAR _SEX ;
dynamic _panelnumber_ ;
begingraph / DataContrastColors=( &contra. ) ;
entrytitle halign=center "&ENTRYTITLE. " ;
entryfootnote halign=left "&ENTRYFOOTNOTE. " ;
layout datalattice columnvar=_SEX ;
/*中略*/
run;
/* 描画 */
proc sgrender data=&ds. template=sgspaghetti ;
```



```

dynamic _AGE="AGE" _LN_MU="'LN_MU'n" _YEAR="YEAR" _SEX="SEX" ;
ODS GRAPHICS / ANTIALIASMAX=4100 ;
run;
%mend plot_ln_mu_2 ;

```

プログラム 2 TEMPLATE プロシジャを利用したグラフ作成プログラム (抜粋)

さらに、以下図 3、図 4 がロジャース・モデルによって予測した岩手県盛岡市の人口を、5 歳刻みで集計しなおして 5 年ごとに人口ピラミッドを作成したものである。これも TEMPLATE プロシジャを利用して作成した。その際に使用したのがプログラム 3 である。計算には MCMC サンプル事後パラメータの平均値を用いた。転入出を考慮しない閉鎖人口が図 3、転入出を考慮した開放人口が図 4 である。図 4 については、擬似的な数値を与えたため、あくまで参考として見られたい。

```

proc template;
  define statgraph sgpyramid;
    dynamic _RANGE_LABEL _MALE2 _RANGE_LABEL2 _FEMALE;
    begingraph;
      entrytitle halign=center "&t. 年";
      entryfootnote halign=left "&ENTRYFOOTNOTE.";
      layout lattice / rowdatarange=data columndatarange=data columns=2 rowgutter=10
columngutter=10 columnweights=(0.5 0.5);
        layout overlay / xaxisopts=( reverse=true) yaxisopts=( display=(LINE)
discreteopts=( tickvaluefitpolicy=none));
          barchart category=_RANGE_LABEL response=_MALE2 / name='bar(h)'
datatransparency=0.5 orient=horizontal groupdisplay=Cluster clusterwidth=1.0;
          endlayout;
        layout overlay / yaxisopts=( display=(TICKVALUES LINE)
discreteopts=( tickvaluefitpolicy=none));
          barchart category=_RANGE_LABEL2 response=_FEMALE / name='bar(h)2'
datatransparency=0.5 orient=horizontal groupdisplay=Cluster clusterwidth=1.0
fillattrs=GraphData2;
          endlayout;
        endlayout;
      endgraph;
    end;
  run;

```

プログラム 3 人口ピラミッドのフォーマット

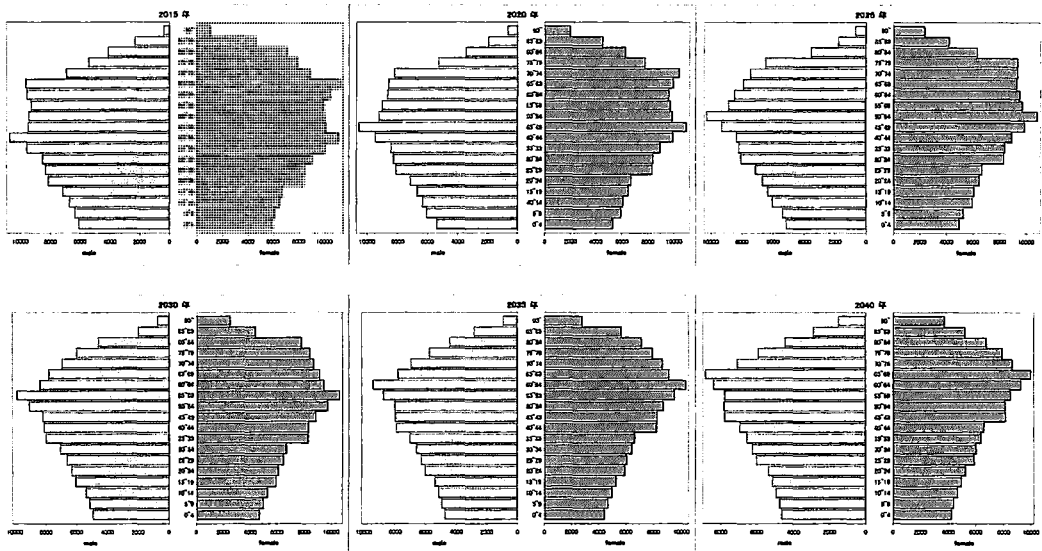


図 3 2015 年から 2040 年までの岩手県盛岡市予測人口 (閉鎖人口)

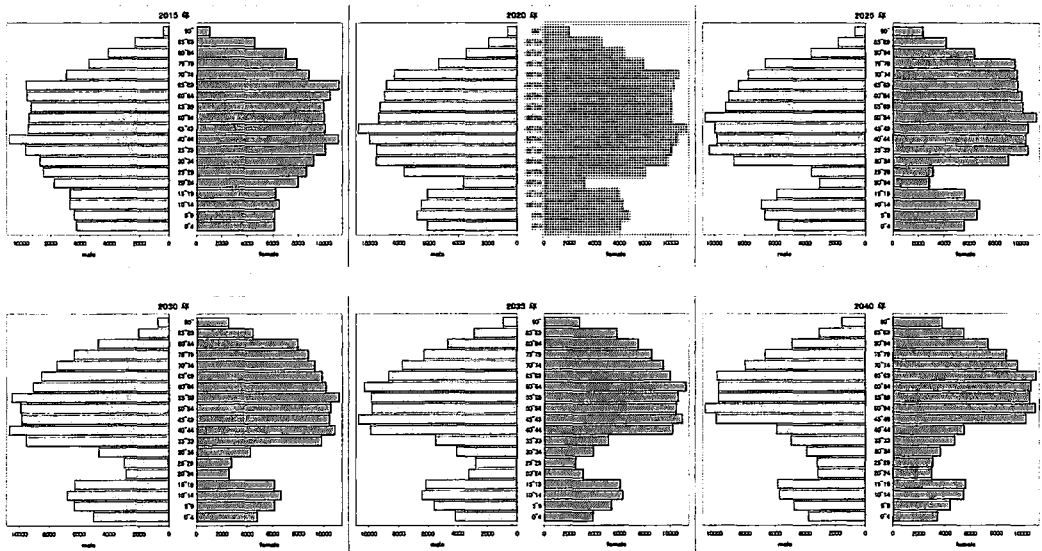


図 4 2015 年から 2040 年までの岩手県盛岡市予測人口 (開放人口)

## 結び

Lee-Carter モデルを拡張した階層ベイズモデルによる人口予測を試みるにあたって、一般公開されている全国の国勢調査データから推定した人口の自然増減パラメータを代用することで、市町村別の人口予測を行った。本来ならば市町村のデータを使用すべきところなので、この結果は仮のものであり、いわばデモンストレーションという形式であるが確率的モデルによる人口予測の可能性を示すことができた。

また、本稿の試みでは時間その他の制約から、SAS MCMC プロシジャを活用することができなかったが、

推定結果の視覚化や、推定パラメータと実際の人口から予測人口を計算する場面では SAS の豊富な機能を活用して実行できることが示された。

## 謝辞

本稿の執筆にあたり、SAS テクニカルサポートご担当各位に MCMC プロシジャの使用法についての相談にご対応いただいたことに感謝の意を表す。ご援助いただきながらプロシジャの活用に至らなかったのはひとえに著者の力不足によるものである。また、通常業務で多忙のなか、校正やプログラム作成で支援いただいた弊社先輩・同僚各位に対してもこの場を借りてお礼を申し上げる。

## 参考文献

- Brooks, S. P. & Gelman, A., 1998. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4), pp. 434-455.
- Brouhns, N., Denuit, M. & Vermunt, J. K., 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3), pp. 373-393.
- Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, Volume 4, pp. 169-193.
- Lee, R. D., 1993. Modeling and forecasting the time series of US fertility: age distribution, range, and ultimate level. *International journal of forecasting*, 9(2), pp. 187-202.
- Lee, R. D. & Carter, L. R., 1992. Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association*, 87(419), p. 659-671.
- Renshaw, A. E. & Haberman, S., 2006. A Cohort-based Extension to the Lee-Carter Model for Mortality Reduction Factor. *Insurance: Mathematics and Economics*, Volume 58, pp. 556-570.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A., 2002. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B*, 64(5), pp. 583-639.
- Wiśniowski, A. et al., 2015. Bayesian population forecasting: extending the Lee-Carter method. *Demography*.
- 国立社会保障・人口問題研究所, 2013. 日本の地域別将来推計人口 (平成 25 年 3 月推計) . [オンライン]  
Available at: <http://www.ipss.go.jp/pp-shicyoson/j/shicyoson13/t-page.asp>
- 小池司朗, 2008. 地域別将来人口推計における人口移動モデルの比較研究. *人口問題研究*, 64[3], pp. 87-111.
- 小暮厚之 長谷川智弘, 2005. 将来生命表の統計モデリング: Lee-Carter 法とその拡張 - ヒューマンセキュリティへの基盤研究. 総合政策学ワーキングペーパーシリーズ, Issue 71.
- 石井太, 2006. 我が国の平均寿命の動向と死亡率推計モデルの検討. *人口問題研究*, 62[3], pp. 21-30.
- 南條善治 吉永一彦, 2006. Lee-Carter 法による日本の将来生命表の推計: 寿命レベル別生命表の利用. 日本大学人口研究所報告シリーズ, Issue 1.

Let's データ分析

# 貧困・中流・富裕世帯における教育費の実態調査

参加カテゴリ：B)SAS または JMP の使用歴 3 年未満

四方克成(個人参加)

Survey of educational expenses in the classification by household income

Katsunari Shiho

## 要約

教育に対する家計負担が高まる現在<sup>1</sup>、親の経済力が子供の進路を左右する<sup>2</sup>。深刻な貧困問題を引き起こす教育格差<sup>3</sup>の実態の一つでも多く明らかにすべく疑似マイクロデータを用いて分析を行った。

全体の3.1%が教育費を支出する貧困世帯であった。貧困、中流、富裕の3区分で見ると、富裕世帯における教育費の支出割合は食料費よりも高く、世帯主年齢40～59歳の富裕世帯では平均支出額が中流世帯の2倍近かった。また、教育が贅沢品に該当する世帯にて、教育費を節約した場合、貧困よりも富裕世帯の家計収支に与える影響の方が大きかった。

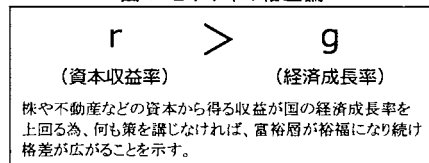
## 1. はじめに

### <貧困の裏にある格差の構造>

世界中で大ベストセラーとなった『21世紀の資本』の著者であるトマ・ピケティ<sup>4</sup>。昨年、彼の来日が発表された頃<sup>5</sup>から、再び新聞や雑誌、テレビなどで「格差」という言葉が取り上げられるようになった。

十年前にこの言葉が取り上げられたのは、派遣社員と正規社員の間における労働市場の二分化がきっかけであった<sup>6</sup>。しかし、今回の場合は、日本の国民経済は根本的に貧富の差が広がる構造になっているというものである(図17)。この不等式をモチーフに、貧困家庭で育った人が非正規労働として働く様子、自分の子供にも十分な教育を受けさせてあげられない生活など、さまざまな実態が報道されている。

図1 ピケティの格差論



### <家計を圧迫する教育費>

教育に対する家計負担は明らかに高まっている。消費者物価指数の変化を調べると、教育の物価指数が他と比べ大幅に上昇していることがわかる。1980年と2015年の比較で、総合が77.2から103.6となっているのに対し、教育は49.4から102.3となっている(表1)。

表1 1980年と2015年における消費者物価指数の違い(2010年=100)

|       | 総合    | 食料    | 住居   | 光熱・水道 | 家具・家事用品 | 被服及び服物 | 保健医療 | 交通・通信 | 教育    | 娯楽   |
|-------|-------|-------|------|-------|---------|--------|------|-------|-------|------|
| 1980年 | 77.2  | 75.8  | 64.0 | 92.1  | 136.5   | 72.9   | 68.1 | 92.0  | 49.4  | 88.2 |
| 2015年 | 103.6 | 106.6 | 99.1 | 116.2 | 94.5    | 104.5  | 99.9 | 103.6 | 102.3 | 98.9 |

出所：総務省統計局「平成22年基準消費者物価指数」参考文献[6]

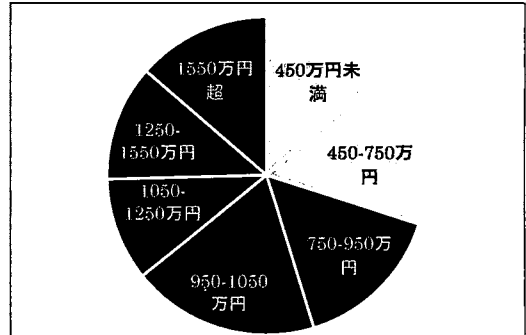
<親の年収に左右される子供の進路>

教育の値段が高い日本の生活において、子供の学力と未来を支えるものは、やはり親の経済力しかないのだろうか。

東大生の親の年収は750万円以上が約70%を占めている(グラフ1)。

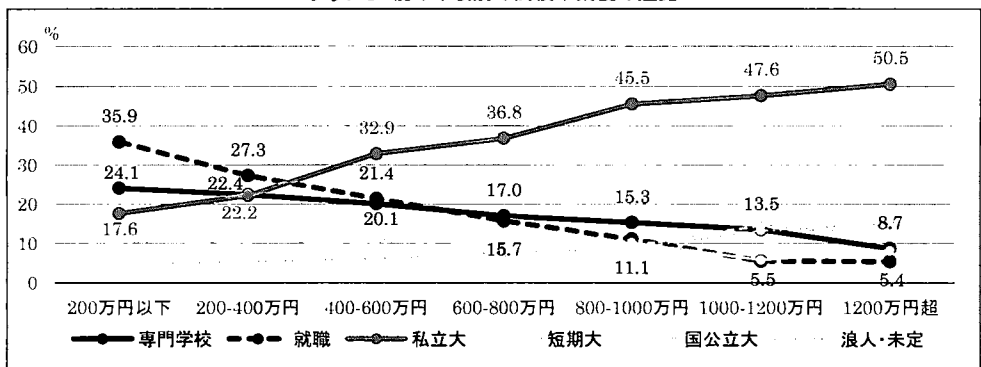
そして、親の年収が高くなるにしたがって、子供の専門学校と就職への進路率は下がり、私立大への進路率が大きく上がっている(グラフ2)。

グラフ1 東大生の親の年収



出所: 東京大学「学生生活実態調査(2014年)」参考文献[7]

グラフ2 親の年収別の高校卒業後の進路



出所: 東京大学(2009)「高校生の進路と親の年収の関連について」参考文献[5]

2. 分析テーマ

<疑似マイクロデータを使って教育費の実態を調査する>

- ① 貧富の差によって、教育費の支出にどのくらいの差があるか。
  - ② 総務省の統計調査により、二人以上総世帯・勤労者世帯にとって、教育費は贅沢品に該当する<sup>8</sup>ことが明らかになっているが、貧富の差による違いはあるか。
  - ③ 教育費に対する過剰な支出をやめて節約した場合、家計に与える影響はどのくらいか。
- 以上、3つの小テーマに沿って分析を行う。

<分析の目的>

限られた条件による分析結果となる為、すべての世帯に対する汎用的な説明として、用いられることは難しいと思料する。しかし、より多くの人の問題意識を高めるきっかけとなって、さらなる分析調査につながり、日本における教育格差の問題が早期に解決されることを願う。

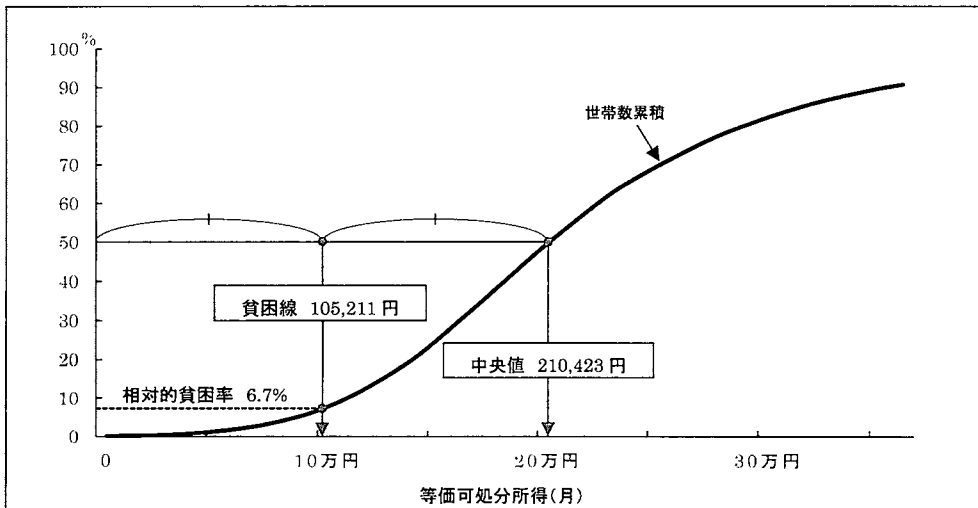
<本論の進行順序>

- ・疑似マイクロデータに含まれる貧困率の基礎分析
- ・データの分類化について
- ・消費構造における教育費の分析 ……小テーマ①に対応
- ・教育費の必需品/贅沢品の分析 ……小テーマ② //
- ・教育費を節約した場合の家計に与える影響の分析 ……小テーマ③ //

3. 疑似マイクロデータに含まれる貧困率の基礎分析

貧困や格差の議論で用いられる指標の一つ、等価方式による相対的貧困率を使って、疑似マイクロデータに含まれる貧困世帯の割合を分析した。結果、等価可処分所得の中央値は 21 万 423 円、貧困線は 10 万 5,211 円、貧困線に満たない相対的貧困率は 6.7%であった(グラフ3 表2)。

グラフ3 等価可処分所得に対する世帯数累積パーセント(集計用乗率使用)



- [用語説明 9]
- 1)相対的貧困率…ある国や地域における相対的貧困者の割合。
  - 2)等価方式による相対的貧困率…等価可処分所得から算出した貧困線未満の人々の割合。
  - 3)貧困線…国民の等価可処分所得の中央値の半分の額。
  - 4)等価可処分所得…可処分所得を等価計算で求めた額。
  - 5)等価計算…値を世帯人員の平方根で除す計算方法。

表2 等価可処分所得の範囲別世帯数と割合(集計用乗率使用)

|                | 等価可処分所得(x)       |                       |                 |
|----------------|------------------|-----------------------|-----------------|
|                | x < 105,211      | 105,211 ≤ x < 210,423 | 210,423 ≤ x     |
| 世帯数 (合計に占める割合) | 33,306 (6.7%)    | 214,425 (43.3%)       | 247,733 (50.0%) |
| 小計 (合計に占める割合)  | 247,732 (50.0%)  |                       | 247,733 (50.0%) |
| 合計             | 495,465 (100.0%) |                       |                 |

#### 4. データの分類化について

##### < 貧困・中流・富裕の区分け >

貧富の差で分析する為、等価可処分所得の中央値を基準に分類条件を設定した（表3）。

表3 貧困・中流・富裕グループの分類条件

|    | 分類条件(等価可処分所得)      | 等価可処分所得の範囲              |
|----|--------------------|-------------------------|
| 貧困 | 中央値×0.5 未満         | 105,211 円未満             |
| 中流 | 中央値×0.5 以上、×1.5 未満 | 105,211 円以上、315,634 円未満 |
| 富裕 | 中央値×1.5 以上         | 315,634 円以上             |

##### < 教育ありと教育なし世帯の設定 >

教育を受けさせている世帯に絞って分析する為、教育費の支出>0の世帯を対象とした。

##### < データ分類結果 >

格差別、教育費の支出あり・なし別に世帯数をクロス集計した結果、教育費の支出ありの貧困世帯は、全体の3.1%であった（表4）。

表4 格差別 教育区分別 世帯数(集計用乗率使用)

|    | 分析対象<br>教育費の支出あり |       | 分析対象外<br>教育費の支出なし |       | 合計      |        |
|----|------------------|-------|-------------------|-------|---------|--------|
|    | 世帯数              | 割合    | 世帯数               | 割合    | 世帯数     | 割合     |
| 貧困 | 15,419           | 3.1%  | 17,887            | 3.6%  | 33,306  | 6.7%   |
| 中流 | 205,899          | 41.6% | 174,604           | 35.2% | 380,504 | 76.8%  |
| 富裕 | 27,749           | 5.6%  | 53,907            | 10.9% | 81,655  | 16.5%  |
| 合計 | 249,066          | 50.3% | 246,398           | 49.7% | 495,465 | 100.0% |

#### 5. 消費構造における教育費の分析

##### < 富裕世帯における教育費は食料費を超える >

教育費の支出あり世帯を対象に貧困世帯、中流世帯、富裕世帯のそれぞれにおける消費構造を把握する為、消費支出(全体)に対する各支出項目の割合を比較した(表5)。

食料や住居、光熱・水道などの生活必需品項目については、富裕世帯よりも貧困世帯の方が割合は高かった。

しかし、教育費については教養娯楽や交際費など生活に不可欠ではない項目とともに貧困世帯から富裕世帯に向かって割合が高くなっている。

注目すべき点として、富裕世帯における教育費の割合は、食料費の割合よりも高く、食料16.9%に対して教育17.7%であった。

表5 消費支出に対する各支出項目の割合  
(集計用乗率使用)

|         | 貧困    | 中流    | 富裕    |
|---------|-------|-------|-------|
| 食料      | 26.4% | 23.5% | 16.9% |
| 住居      | 6.4%  | 4.6%  | 3.7%  |
| 光熱・水道   | 7.8%  | 6.2%  | 4.2%  |
| 家具・家事用品 | 2.9%  | 2.7%  | 2.0%  |
| 被服及び履物  | 4.0%  | 4.2%  | 4.0%  |
| 保健医療    | 3.8%  | 3.4%  | 2.6%  |
| 交通・通信   | 14.1% | 14.0% | 15.0% |
| 教育      | 11.0% | 11.8% | 17.7% |
| 教養娯楽    | 8.8%  | 9.8%  | 8.8%  |
| 諸雑費     | 4.9%  | 5.5%  | 5.6%  |
| こづかい    | 5.0%  | 7.7%  | 8.8%  |
| 交際費     | 3.7%  | 4.7%  | 5.3%  |
| 仕送り金    | 0.9%  | 1.9%  | 5.3%  |

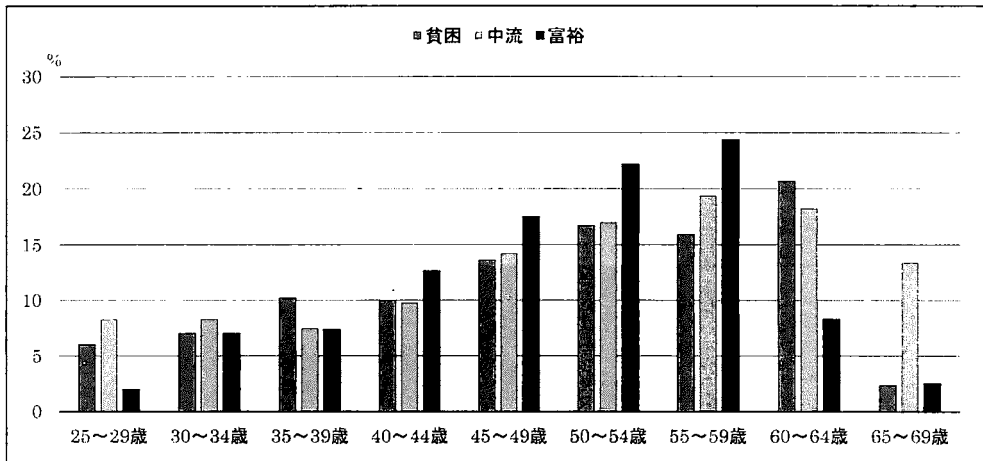


＜富裕世帯における教育費の支出割合は年齢階級による高低差が大きい＞

教育費の支出割合の違いについて、世帯主年齢に対する傾向を見る為、年齢5歳階級別にヒストグラムを使って比較した(グラフ4)。

全体的に世帯主年齢が高い方が、教育費の支出割合が高いことがわかる。富裕世帯においては、貧困・中流世帯に比べ、年齢階級による高低差が大きい。支出割合が高めの50～59歳の範囲では20%以上となっていた。

グラフ4 世帯主年齢5歳階級別 教育費の支出割合(集計用乗率使用)

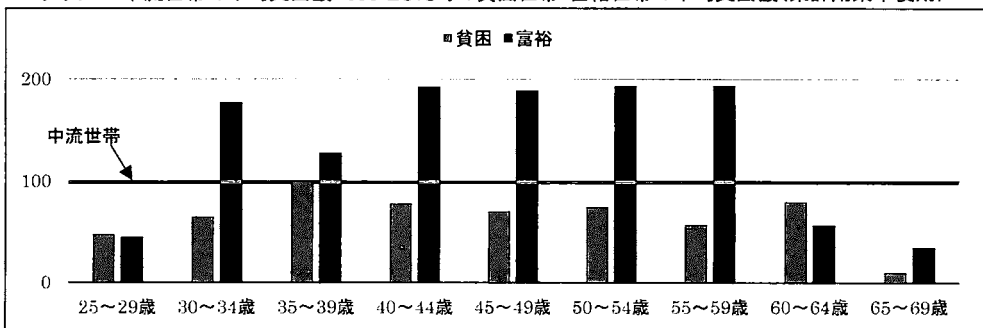


＜富裕世帯における教育費の支出額は40～59歳で中流世帯の2倍近い＞

教育費の支出額の差を把握する為、中流世帯の平均支出額を100として、貧困・富裕世帯の平均支出額をヒストグラムにして比較した(グラフ5)。

富裕世帯では、40～59歳の年齢階級で中流世帯の2倍に近い支出額となっていた。また、富裕世帯の支出額は年齢階級による高低差が大きいこともわかった。

グラフ5 中流世帯の平均支出額=100とした時の貧困世帯・富裕世帯の平均支出額(集計用乗率使用)



## 6. 教育費の必需品/贅沢品の分析

＜40～59歳の年齢階級では貧富の差に関わらず教育は贅沢品に該当＞

格差別、年齢階級別に教育費の性質が必需品か贅沢品かを判別する為に支出弾力性を計算した。計算にあたり、消費支出に対する教育費の回帰式を求める段階で、係数（限界性向）が有意ではない格差・年齢階級の組み合わせは除外した（ $p < 0.05$ ）。

計算結果を分析すると、40～59歳の年齢階級においては、貧富の差に関係なく、教育費は贅沢品に該当することがわかった（表6）。

表6 消費支出に対する教育費の支出弾力性(集計用乗率使用)

|        | 贅沢品に該当 |      | 必需品に該当 |
|--------|--------|------|--------|
|        | 貧困     | 中流   | 富裕     |
| 25-29歳 | 1.03   | 0.77 |        |
| 30-34歳 | 0.91   | 0.46 |        |
| 35-39歳 | 1.65   | 1.06 | 0.74   |
| 40-44歳 | 1.21   | 1.62 | 1.96   |
| 45-49歳 | 1.51   | 2.29 | 1.56   |
| 50-54歳 | 1.85   | 2.14 | 1.18   |
| 55-59歳 | 1.55   | 1.48 | 1.59   |
| 60-64歳 |        | 0.98 | 1.94   |
| 65-69歳 |        | 2.36 |        |

[参考<sup>10)</sup>

支出弾力性とは、消費支出総額が1%変化する時に各支出項目が何%変化するかを示した指標である。支出弾力性が1.00未満の支出項目は基礎的支出(必需品)に分類され、食料、家賃、光熱費、保健医療サービスなどが該当する。1.00以上の支出項目は選択的支出(贅沢品)に分類され、教育費、教養娯楽用耐久財、月謝などが該当する。

[支出弾力性の計算]

- 1) 格差別、年齢階級別に消費支出に対する教育費の回帰式を求める。

$$\text{教育費}(Y) = \text{限界性向}(a) \cdot \text{消費支出}(X) + \text{切片}(b)$$

- 2) 限界性向(a)について、有意差が認められない(0.05超)場合、その格差、年齢階級の組み合わせは除外する。
- 3) 格差別、年齢階級別の限界性向(a)、平均消費支出( $\bar{X}$ )、平均教育費( $\bar{Y}$ )を使って、支出弾力性を求める。

$$\text{支出弾力性}(\eta) = \frac{0.01 \times \text{限界性向}(a) \times \text{平均消費支出}(\bar{X})}{\text{平均教育費}(\bar{Y})} \times 100$$

## 7. 教育費を節約した場合の家計に与える影響の分析

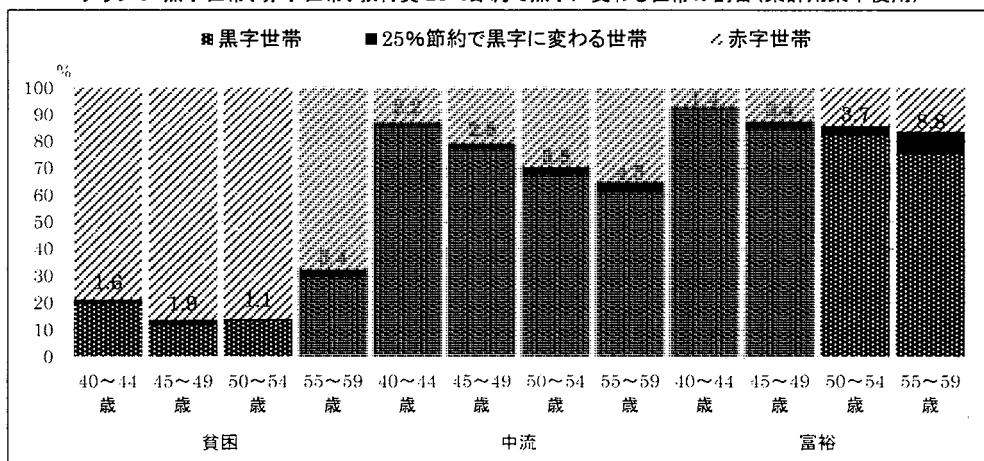
<節約の影響は貧困世帯よりも富裕世帯の方が大きい>

貧困・中流・富裕世帯において、教育費が贅沢品に該当した世帯主年齢 40～59 歳の世帯を対象に、教育費を節約した場合の家計に与える影響を分析した。

家計収支が赤字（実収入-実支出<0）の世帯が、教育費を節約した場合に黒字（実収入-実支出+教育費×節約率>0）に変わる割合を調べる方法で行った（グラフ6）。節約率は25%で設定。

結果、貧困世帯は赤字世帯の割合が大きく、節約によって黒字に変わる世帯は1.1～3.4%の間にとどまり、節約の影響は小さかった。富裕世帯は赤字世帯の割合は小さく、節約によって黒字に変わる世帯は55～59歳にて8.8%となり、最も大きく節約の影響があらわれた。

グラフ6 黒字世帯、赤字世帯、教育費25%節約で黒字に変わる世帯の割合（集計用乗率使用）



## 8. 結論 : 分析結果まとめ

<消費構造における教育費の分析（小テーマ①）>

- ・ 富裕世帯における教育費の支出割合は食料費の割合よりも高い。
- ・ 富裕世帯における教育費の支出割合は年齢階級による高低差が大きい。50～59歳の範囲では支出割合は20%を超えている。
- ・ 富裕世帯における教育費の支出額は、40～59歳の範囲で中流世帯の2倍近い。

<教育費の必需品/贅沢品の分析（小テーマ②）>

- ・ 40～59歳の範囲では、貧富の差に関係なく、教育費は贅沢品に該当する。

<教育費を節約した場合の家計に与える影響の分析（小テーマ③）>

- ・ 節約の影響は貧困世帯よりも富裕世帯の方が大きい。

## 9. 考察

### <メッセージ>

分析目的に対して得られた結論は二つある。一つ目は40～59歳の範囲では、貧富の差に関係なく教育は贅沢品に該当するなか、貧富の差によって支出に大きな格差があるということ。富裕世帯では贅沢品であっても、食料費よりも多くの教育費を支出しており、貧困世帯にとつての贅沢品とは性質が異なる。したがって、性質をあらわす新たな指標の発明も求められる。

二つ目は貧困世帯において、教育費の節約が家計にほとんど効果がないこと。貧困世帯は家計が既に赤字の状態でも、学歴が左右する現実に向け、教育費を支払う決断に踏み切っているかもしれない。教育費によってさらなる貧困化が進んでいる可能性がある。

この二点が論文を読んで頂いた方へ伝えたいメッセージである。

### <分析結果の適用範囲>

当分析結果は、以下の点を考慮して使用すべきである。

- ・ 等価可処分所得の中央値を基準に貧困世帯、中流世帯、富裕世帯に分類している。  
(収入階級、可処分所得による分析とは異なる可能性がある)
- ・ 疑似マイクロデータを用いており、単身世帯、無業世帯は含まれていない。
- ・ 教育費は、子供に対してだけではなく、自身や他の親族に対する支出の可能性もある。

## 10. 本文中の注について

- 1… 1. はじめに<家計を圧迫する教育費>の要約。
- 2… 1. はじめに<親の年収に左右される子供の進路>の要約。
- 3… 参考文献[1]を読んだ感想要約。家庭事情や親が病気などの理由で経済水準が低く、まともな教育を受けられずに育った子供たちが、親になり現在も苦しい生活を続けている。貧困の連鎖が断ち切れない社会問題がルポ形式で書かれている。
- 4… 参考文献[11]の要約。
- 5… 参考文献[9]、来日期間は2015年1月29日(木)～2月1日(日)。
- 6… 参考文献[4]、経済協力開発機構(OECD)が日本に対する報告書の中で指摘したことが話題になった。日本における非正規労働が1995年から2005年にかけて19%から30%に拡大し、そして、非正規労働の賃金が正規労働の40%しかないことが「経済格差」を引き起こすことになる。
- 7… 文章は参考文献[2][3]を要約、図は参考文献[2]41ページ引用。
- 8… 参考文献[8]のデータに基づいた根拠。
- 9… 参考文献[10]の要約。
- 10… 参考文献[12]の引用。

## 11. 参考文献

### <図書・雑誌>

- [1] 保坂渉(2015)『子どもの貧困連鎖』新潮社
- [2] 中川雅之(2015)「2000万人の貧困」『日経ビジネス』日経 BP 社 2015年3月23日発行
- [3] 藤田康範(2015)『コミックでわかるピケティ入門』中経出版

### <オンライン>

- [4] 経済協力開発機構(OECD)『Economic Surveys: Japan 2006』(2016年6月17日)  
[http://www.keepeek.com/Digital-Asset-Management/oced/economics/oced-economic-surveys-japan-2006\\_eco\\_surveys-jpn-2006-en#page196](http://www.keepeek.com/Digital-Asset-Management/oced/economics/oced-economic-surveys-japan-2006_eco_surveys-jpn-2006-en#page196)
- [5] 東京大学(2009/7/31)『高校生の進路と親の年収の関連について』  
<http://ump.p.u-tokyo.ac.jp/crump/resource/crump090731.pdf>
- [6] 総務省統計局『平成22年基準消費者物価指数』  
<http://www.e-stat.go.jp/SG1/estat/List.do?bid=000001033700&cycode=0>
- [7] 東京大学『学生生活実態調査(2014年)』  
<http://www.u-tokyo.ac.jp/content/400037152.pdf>
- [8] 総務省統計局『家計調査 収支項目分類及びその内容例示(平成27年(2015年)1月改定)』  
支払 消費支出(交通・通信, 教育, 教養娯楽, その他の消費支出)(エクセル:266KB)  
<http://www.stat.go.jp/data/kakei/kou27/reiji27.htm>
- [9] みずず書房(最終更新2015/2/2)『21世紀の資本のトマ・ピケティ 来日』(2016年6月17日)  
<http://www.msiz.co.jp/topics/07876/>
- [10] 総務省統計局(2015/12/18)『相対的貧困率等に関する調査分析結果』(2016年6月17日)  
<http://www.stat.go.jp/data/zensho/2009/pdf/hinkonritsu.pdf>
- [11] Wikipedia(最終更新2016/6/13)『21世紀の資本』(2016年6月17日)  
<https://ja.wikipedia.org/wiki/21%E4%B8%96%E7%B4%80%E3%81%AE%E8%B3%87%E6%9C%AC>
- [12] 総務省統計局『家計調査 支出弾力性の計算方法について』(2016年6月17日)  
<http://www.stat.go.jp/data/kakei/kou17/dan17.htm>

付録 規定課題

1) 産業符号・性別 クロス表

表 1

| 産業符号              | a) 単純クロス表 |       |        | b) 10万人世帯比でのクロス表 |       |         |
|-------------------|-----------|-------|--------|------------------|-------|---------|
|                   | 男         | 女     | 合計     | 男                | 女     | 合計      |
| 鉱業                | 3         | 0     | 3      | 7                | 0     | 7       |
| 建設業               | 2,134     | 9     | 2,143  | 6,346            | 31    | 6,377   |
| 製造業               | 5,668     | 173   | 5,841  | 17,837           | 510   | 18,347  |
| 電気・ガス・熱供給・水道業     | 222       | 0     | 222    | 697              | 0     | 697     |
| 情報通信業             | 547       | 0     | 547    | 2,020            | 0     | 2,020   |
| 運輸業               | 1,235     | 3     | 1,238  | 3,852            | 6     | 3,858   |
| 卸売・小売業            | 2,083     | 201   | 2,284  | 6,644            | 589   | 7,233   |
| 金融・保険業            | 472       | 39    | 511    | 1,447            | 104   | 1,551   |
| 不動産業              | 40        | 0     | 40     | 144              | 0     | 144     |
| 飲食店・宿泊業           | 50        | 32    | 82     | 158              | 107   | 265     |
| 医療・福祉             | 371       | 270   | 641    | 1,022            | 810   | 1,832   |
| 教育・学習支援業          | 1,059     | 89    | 1,148  | 3,110            | 269   | 3,379   |
| 複合サービス事業          | 321       | 3     | 324    | 981              | 11    | 992     |
| サービス業(他に分類されないもの) | 1,816     | 106   | 1,922  | 5,842            | 326   | 6,167   |
| 公務(他に分類されないもの)    | 3,656     | 130   | 3,786  | 10,691           | 430   | 11,121  |
| 不詳                | 9,704     | 1,591 | 11,295 | 31,204           | 4,806 | 36,011  |
| 合計                | 29,381    | 2,646 | 32,027 | 92,000           | 8,000 | 100,000 |

2) 経常収入および食料費

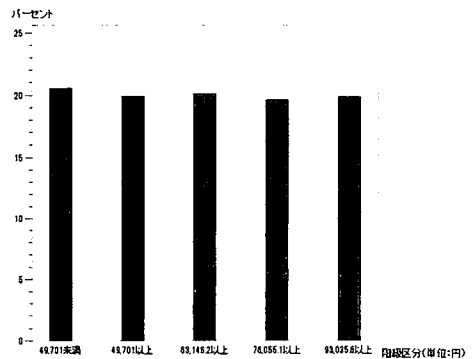
表 2

| 分位点 | 経常収入    | 食料費    |
|-----|---------|--------|
| 20% | 310,531 | 49,701 |
| 40% | 404,883 | 63,146 |
| 60% | 501,181 | 76,055 |
| 80% | 635,875 | 93,036 |

経常収入 ヒストグラム



食料費 ヒストグラム



3) 男女別・経常収入および食料費の5分位分布

表 3

経常収入

| 5分位 | 5分位数      | 男      | 女     | 計       |
|-----|-----------|--------|-------|---------|
| 1   | 310,531未満 | 15,494 | 4,506 | 20,000  |
| 2   | 310,531以上 | 18,617 | 1,377 | 19,993  |
| 3   | 404,883以上 | 19,099 | 911   | 20,009  |
| 4   | 501,181以上 | 19,356 | 640   | 19,995  |
| 5   | 635,875以上 | 19,435 | 567   | 20,002  |
|     | 計         | 92,000 | 8,000 | 100,000 |

食料費

| 5分位 | 5分位数       | 男      | 女     | 計       |
|-----|------------|--------|-------|---------|
| 1   | 49,701未満   | 16,803 | 3,196 | 19,999  |
| 2   | 49,701以上   | 18,097 | 1,903 | 20,000  |
| 3   | 63,146.2以上 | 18,734 | 1,268 | 20,001  |
| 4   | 76,055.1以上 | 19,049 | 948   | 19,998  |
| 5   | 93,035.8以上 | 19,317 | 685   | 20,003  |
|     | 計          | 92,000 | 8,000 | 100,000 |

4) 男女別・経常収入・食料費についての3重クロス表

表 4

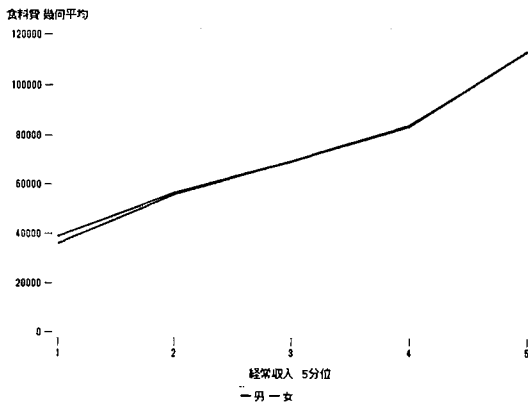
| 経常収入5分位 |   | 食料(5分位) |        |        |        |        | 計       |
|---------|---|---------|--------|--------|--------|--------|---------|
|         |   | 1       | 2      | 3      | 4      | 5      |         |
| 男       | 1 | 5,827   | 3,974  | 2,798  | 1,831  | 1,064  | 15,494  |
|         | 2 | 5,130   | 4,501  | 3,978  | 2,996  | 2,012  | 18,617  |
|         | 3 | 2,911   | 4,301  | 4,395  | 4,294  | 3,198  | 19,099  |
|         | 4 | 1,906   | 3,233  | 4,082  | 4,903  | 5,231  | 19,356  |
|         | 5 | 1,029   | 2,087  | 3,481  | 5,026  | 7,813  | 19,435  |
|         | 計 | 16,803  | 18,097 | 18,734 | 19,049 | 19,317 | 92,000  |
| 女       | 1 | 2,462   | 994    | 565    | 314    | 172    | 4,506   |
|         | 2 | 371     | 399    | 235    | 211    | 161    | 1,377   |
|         | 3 | 243     | 225    | 216    | 164    | 63     | 911     |
|         | 4 | 68      | 134    | 156    | 143    | 140    | 640     |
|         | 5 | 53      | 152    | 96     | 117    | 149    | 567     |
|         | 計 | 3,196   | 1,903  | 1,268  | 948    | 685    | 8,000   |
| 男女計     |   | 19,999  | 20,000 | 20,001 | 19,998 | 20,003 | 100,000 |

5) 男女別・経常収入別に食料費を常用対数変換した平均値の比較

表 5

| 経常収入5分位 | 世帯数 | 食料     |        |         |
|---------|-----|--------|--------|---------|
|         |     | 常用対数平均 | 幾何平均   |         |
| 男       | 1   | 16,803 | 4.5902 | 38,922  |
|         | 2   | 18,097 | 4.7527 | 56,585  |
|         | 3   | 18,734 | 4.8410 | 69,343  |
|         | 4   | 19,049 | 4.9232 | 83,792  |
|         | 5   | 19,317 | 5.0537 | 113,162 |
|         | 計   | 92,000 | 4.8395 | 69,103  |
| 女       | 1   | 3,196  | 4.5557 | 35,950  |
|         | 2   | 1,903  | 4.7479 | 55,963  |
|         | 3   | 1,268  | 4.8404 | 69,247  |
|         | 4   | 948    | 4.9207 | 83,311  |
|         | 5   | 685    | 5.0546 | 113,397 |
|         | 計   | 8,000  | 4.7325 | 54,013  |

食料費 幾何平均 (男女別・経常収支の分位点別)



## 付録 規定課題の SAS プログラム

### 1) 産業符号・性別 クロス表

```

/*出力形式の準備*/
PROC FORMAT LIB=SASUSER;
  VALUE sex 1 = "男" 2 = "女";
  VALUE sangyo 1 = "農業" 2 = "林業" 3 = "漁業" 4 = "鉱業" 5 = "建設業" 6 = "製造業"
  7 = "電気・ガス・熱供給・水道業" 8 = "情報通信業" 9 = "運輸業" 10 = "卸売・小売業"
  11 = "金融・保険業" 12 = "不動産業" 13 = "飲食店・宿泊業" 14 = "医療・福祉"
  15 = "教育・学習支援業" 16 = "複合サービス事業" 17 = "サービス業(他に分類されないもの)"
  18 = "公務(他に分類されないもの)" 19 = "その他(非就業を含む)" 99 = "不詳";
RUN;

/*表作成用データの準備*/
PROC SQL;
  CREATE TABLE SASUSER.K1data AS
  SELECT S1_Sex FORMAT=sex. AS Sex, 1 AS Setai, Weight*10**5/495465 AS JmnWeight,
  (CASE S1_Sangyou WHEN "VV" THEN 99 ELSE INPUT(S1_Sangyou,best12.)END) FORMAT=sangyo. AS Sangyou
  FROM SASUSER.GIJDATA;

/*ラベルの準備*/
%let label01 = 合計; %let label02 = a) 単純クロス表; %let label03 = b) 10 万人世帯比でのクロス表; %let label04 = 産業符号;

/*表の作成(表 1)*/ TITLE "表 1";
PROC TABULATE DATA= SASUSER.K1data;
  VAR Setai;
  CLASS Sex Sangyou;
  TABLE Sangyou=(LABEL="" ALL=(LABEL="&label01"));

  Setai=(LABEL="&label02")*F=COMMA6.*(
  Sex=(LABEL=""*N=(LABEL="" ALL=(LABEL="&label01")*N=(LABEL="")))
  Setai=(LABEL="&label03")*F=COMMA7.*(
  Sex=(LABEL=""*SumWgt=(LABEL="" ALL=(LABEL="&label01")*SumWgt=(LABEL="")))
  / BOX=(LABEL="&label04" STYLE={JUST=LEFT VJUST=BOTTOM}) MISSTEXT="0"; WEIGHT JmnWeight;
RUN; QUIT;

```

### 2) 経常収入および食料費

```

/*表作成用データの準備*/
PROC SQL;
  CREATE TABLE SASUSER.K2data AS
  SELECT Youto004 LABEL="経常収入" AS Income, Youto038 LABEL="食料費" AS Food, Weight AS Weight
  FROM SASUSER.GIJDATA;

/*ラベルの準備*/
%let label05 = 20%; %let label06 = 40%; %let label07 = 60%; %let label08 = 80%; %let label09 = 分位点;

/*表の作成(表 2)*/ TITLE "表 2";
PROC TABULATE DATA=SASUSER.K2data;
  VAR Income Food;
  TABLE P20={LABEL="&label05" STYLE={JUST=RIGHT}} P40={LABEL="&label06" STYLE={JUST=RIGHT}}
  P60={LABEL="&label07" STYLE={JUST=RIGHT}} P80={LABEL="&label08" STYLE={JUST=RIGHT}};

  Income*F=COMMA7. Food*F=COMMA6.
  / BOX=(LABEL="&label09" STYLE={JUST=RIGHT});
  WEIGHT Weight;
RUN;

/*グラフ出力形式の準備*/
PROC FORMAT LIB=SASUSER;
  VALUE income 1 = "310,531 未満" 2 = "310,531 以上" 3 = "404,883 以上"
  4 = "501,181 以上" 5 = "635,875 以上";
  VALUE food 1 = "49,701 未満" 2 = "49,701 以上" 3 = "63,146.2 以上"
  4 = "76,055.1 以上" 5 = "93,035.8 以上";
RUN;

```



## 2) 経常収入および食料費【続き】

```

/*グラフ作成用データの準備*/
PROC SQL;
CREATE TABLE SASUSER.K2dataGprt AS
SELECT (CASE WHEN 310531 > Youto004 THEN 1 WHEN 404883 > Youto004 >= 310531 THEN 2
        WHEN 501181 > Youto004 >= 404883 THEN 3 WHEN 635875 > Youto004 >= 501181 THEN 4
        WHEN Youto004 >= 635875 THEN 5 END) FORMAT=income. AS IncomeQuint,
(CASE WHEN 49701.0 > Youto038 THEN 1 WHEN 63146.2 > Youto038 >= 49701 THEN 2
        WHEN 76055.1 > Youto038 >= 63146.2 THEN 3 WHEN 93035.8 > Youto038 >= 76055.1 THEN 4
        WHEN Youto038 >= 93035.8 THEN 5 END) FORMAT=food. AS FoodQuint,
Weight*10**5 / 495464 AS JmnSetai
FROM SASUSER.GIJDATA;

/*グラフ表示形式の準備*/
Axis1 LABEL=(“パーセント”); Axis2 LABEL=NONE VALUE=NONE; Axis3 LABEL=(“階級区分(単位:円)”);

/*グラフの作成(経常収入 ヒストグラム)*/ TITLE “経常収入 ヒストグラム”;
PROC GCHART DATA=SASUSER.K2dataGprt;
VBAR JmnSetai
/
GROUP=IncomeQuint LEVELS=1 TYPE=PCT MISSING
RAXIS=Axis1 MAXIS=Axis2 GAXIS=Axis3;
RUN;

/*グラフの作成(食料費 ヒストグラム)*/ TITLE “食料費 ヒストグラム”;
PROC GCHART DATA=SASUSER.K2dataGprt;
VBAR JmnSetai
/
GROUP=FoodQuint LEVELS=1 TYPE=PCT MISSING
RAXIS=Axis1 MAXIS=Axis2 GAXIS=Axis3;
RUN; QUIT;

```

## 3) 男女別・経常収入および食料費の5分位分布

```

/*出力形式の準備*/
PROC FORMAT LIB=SASUSER;
VALUE sex 1 = “男” 2 = “女”;
VALUE income 1 = “310,531 未満” 2 = “310,531 以上” 3 = “404,883 以上”
4 = “501,181 以上” 5 = “635,875 以上”;
VALUE food 1 = “49,701 未満” 2 = “49,701 以上” 3 = “63,146.2 以上”
4 = “76,055.1 以上” 5 = “93,035.8 以上”;
RUN;

/*表作成用データの準備*/
PROC SQL;
CREATE TABLE SASUSER.K3K4data AS
SELECT S1_Sex FORMAT=sex. AS Sex,
1 AS Setai,
(CASE WHEN 310531 > Youto004 THEN 1 WHEN 404883 > Youto004 >= 310531 THEN 2
        WHEN 501181 > Youto004 >= 404883 THEN 3 WHEN 635875 > Youto004 >= 501181 THEN 4
        WHEN Youto004 >= 635875 THEN 5 END) FORMAT=income. AS IncomeQuint,
(CASE WHEN 310531 > Youto004 THEN 1 WHEN 404883 > Youto004 >= 310531 THEN 2
        WHEN 501181 > Youto004 >= 404883 THEN 3 WHEN 635875 > Youto004 >= 501181 THEN 4
        WHEN Youto004 >= 635875 THEN 5 END) AS IncomeQuintNo,
(CASE WHEN 49701.0 > Youto038 THEN 1 WHEN 63146.2 > Youto038 >= 49701 THEN 2
        WHEN 76055.1 > Youto038 >= 63146.2 THEN 3 WHEN 93035.8 > Youto038 >= 76055.1 THEN 4
        WHEN Youto038 >= 93035.8 THEN 5 END) FORMAT=food. AS FoodQuint,
(CASE WHEN 49701.0 > Youto038 THEN 1 WHEN 63146.2 > Youto038 >= 49701 THEN 2
        WHEN 76055.1 > Youto038 >= 63146.2 THEN 3 WHEN 93035.8 > Youto038 >= 76055.1 THEN 4
        WHEN Youto038 >= 93035.8 THEN 5 END) AS FoodQuintNo,
Weight*10**5/495465 AS JmnWeight
FROM SASUSER.GIJDATA;

/*ラベルの準備*/
%let label10=計; %let label11=5 分位 5 分位数;

```

### 3) 男女別・経常収入および食料費の5分位分布【続き】

```

/*表の作成(表 3 経常収入)*/ TITLE1 "表 3"; TITLE2 "経常収入";
PROC TABULATE DATA=SASUSER.K3K4data;
  VAR   Setai;
  CLASS Sex IncomeQuint  IncomeQuintNo;
  TABLE IncomeQuintNo={LABEL="" STYLE(CLASSLEV)={JUST=RIGHT}}*
         IncomeQuint={LABEL="" STYLE(CLASSLEV)={JUST=RIGHT}} ALL={LABEL=""&label10" STYLE={JUST=RIGHT}},

         Setai={LABEL=""*(Sex={LABEL=""}*SumWgt={LABEL=""}*F=COMMA6.
         ALL={LABEL=""&label10"}*SumWgt={LABEL=""}*F=COMMA7.)
         / BOX={LABEL=""&label11"};
  WEIGHT JmnWeight;

RUN;

/*表の作成(表 3 食料費)*/ TITLE2 "食料費";
PROC TABULATE DATA=SASUSER.K3K4data;
  VAR   Setai;
  CLASS Sex FoodQuint  FoodQuintNo;
  TABLE FoodQuintNo={LABEL="" STYLE(CLASSLEV)={JUST=RIGHT}}*
         FoodQuint={LABEL="" STYLE(CLASSLEV)={JUST=RIGHT}} ALL={LABEL=""&label10" STYLE={JUST=RIGHT}},

         Setai={LABEL=""*(Sex={LABEL=""}*SumWgt={LABEL=""}*F=COMMA6.
         ALL={LABEL=""&label10"}*SumWgt={LABEL=""}*F=COMMA7.)
         / BOX={LABEL=""&label11"};
  WEIGHT JmnWeight;

RUN; QUIT;

```

### 4) 男女別・経常収入・食料費についての3重クロス表

```

/*表作成用データは規定課題3と共通*/

/*ラベルの準備*/
%let label12 = 計; %let label13 = 男女計; %let label14 = 食料(5分位); %let label15 = 経常収入 5分位;

/*表の作成(表 4)*/ TITLE "表 4";
PROC TABULATE DATA=SASUSER.K3K4data;
  VAR   Setai;
  CLASS Sex IncomeQuintNo  FoodQuintNo;
  TABLE Sex={LABEL="" STYLE(CLASSLEV)={JUST=RIGHT}}*(STYLE={JUST=RIGHT}}*(
         IncomeQuintNo={LABEL="" STYLE(CLASSLEV)={JUST=RIGHT}}*SumWgt={LABEL=""}*F=COMMA6.
         ALL={LABEL=""&label12" STYLE={JUST=RIGHT}}*SumWgt={LABEL=""}*F=COMMA6.
         )
         ALL={LABEL=""&label13" STYLE={JUST=RIGHT}}*SumWgt={LABEL=""}*F=COMMA7.,

         Setai={LABEL=""}*FoodQuintNO={LABEL=""&label14"}
         Setai={LABEL=""&label12" STYLE={VJUST=BOTTOM}}
         / BOX={LABEL=""&label15" STYLE={VJUST=BOTTOM}};
  WEIGHT JmnWeight;

RUN; QUIT;

```

### 5) 男女別・経常収入別に食料費を常用対数変換した平均値の比較

```

/*出力形式の準備*/
PROC FORMAT LIB=SASUSER;
  VALUE sex 1 = "男" 2 = "女";
  VALUE foodNo 1 = "1" 2 = "2" 3 = "3" 4 = "4" 5 = "5" 6 = "計";
RUN;

```

## 5) 男女別・経常収入別に食料費を常用対数変換した平均値の比較【続き】

```

/*表とグラフ作成用データの準備*/
PROC SQL;
  CREATE TABLE SASUSER.K5data AS
  SELECT S1_Sex FORMAT=sex, AS Sex, 1*Weight*10**5/495465 AS JmnSetai, Weight AS Weight,
  (CASE WHEN 49701.0 > Youto038 THEN 1 WHEN 63146.2 > Youto038 >= 49701 THEN 2
  WHEN 76055.1 > Youto038 >= 63146.2 THEN 3 WHEN 93035.8 > Youto038 >= 76055.1 THEN 4
  WHEN Youto038 >= 93035.8 THEN 5 END) AS FoodQuintNo,
  LOG10(Youto038) AS FoodLG
  FROM SASUSER.GJIDATA;

PROC SQL;
  CREATE TABLE SASUSER.K5data2 AS
  SELECT Sex, FoodQuintNo FORMAT=FoodNo., JmnSetai, FoodLGmean, ROUND(10**FoodLGmean,1) AS FoodGeoM
  FROM (SELECT Sex, FoodQuintNo, ROUND(SUM(JmnSetai),1) AS JmnSetai,
  ROUND(SUM(FoodLG*Weight)/SUM(Weight),0.0001) AS FoodLGmean
  FROM SASUSER.K5data
  GROUP BY Sex, FoodQuintNo OUTER UNION CORR
  SELECT Sex, 6 AS FoodQuintNo, ROUND(SUM(JmnSetai),1) AS JmnSetai,
  ROUND(SUM(FoodLG*Weight)/SUM(Weight),0.0001) AS FoodLGmean
  FROM SASUSER.K5data
  GROUP BY Sex)
  ORDER BY Sex, FoodQuintNo;

/*ラベルの準備*/
%let label16 = 世帯数; %let label17 = 食料; %let label18 = 常用対数平均;
%let label19 = 幾何平均; %let label20 = 経常収入 5 分位;

/*表の作成(表 5)*/TITLE "表 5";
PROC TABULATE
  DATA =SASUSER.K5data2;
  VAR JmnSetai FoodLGmean FoodGeoM;
  CLASS Sex FoodQuintNo;
  TABLE Sex={LABEL="" STYLE(CLASSLEV)={JUST=CENTER}}*
  FoodQuintNo={LABEL="" STYLE(CLASSLEV)={JUST=CENTER}},

  JmnSetai={LABEL=""*F=COMMA6.*Sum={LABEL=""&label16"}
  FoodLGmean={LABEL=""&label17"}*F=COMMA6.4*Sum={LABEL=""&label18"}
  FoodGeoM={LABEL=""*F=COMMA6.*Sum={LABEL=""&label19"}
  / BOX={LABEL=""&label20" STYLE={VJUST=BOTTOM}};

RUN;

/*グラフ作成用データの準備*/
PROC SQL;
  CREATE TABLE SASUSER.K5data3 AS
  SELECT Sex, FoodQuintNo FORMAT=FoodNo., JmnSetai, FoodLGmean, ROUND(10**FoodLGmean,1) AS FoodGeoM
  FROM (SELECT Sex, FoodQuintNo, ROUND(SUM(JmnSetai),1) AS JmnSetai,
  ROUND(SUM(FoodLG*Weight)/SUM(Weight),0.0001) AS FoodLGmean
  FROM SASUSER.K5data
  GROUP BY Sex, FoodQuintNo);

/*グラフ表示形式の準備*/
SYMBOL1 INTERPOL=JOIN HEIGHT=10pt VALUE=NONE LINE=1 WIDTH=2 CV=_STYLE_;
SYMBOL2 INTERPOL=JOIN HEIGHT=10pt VALUE=NONE LINE=1 WIDTH=2 CV =_STYLE_;
Legend1 LABEL="";
Axis4 STYLE=1 WIDTH=1 MINOR=NONE LABEL=("&label17 食料費 幾何平均");
Axis5 STYLE=1 WIDTH=1 MINOR=NONE LABEL=("&label19 経常収入 5分位");

/*グラフの作成*/TITLE "食料費 幾何平均 (男女別・経常収入の分位点別)";
PROC GPLOT
  DATA = SASUSER.K5data3;
  PLOT FoodGeoM * FoodQuintNo=Sex
  /
  VAXIS=Axis4 HAXIS=Axis5 VZERO LEGEND=Legend1;
RUN; QUIT;

```

# 消費税率の引上げ及び軽減税率制度の導入は 低所得者にどのような影響を与えるか

参加カテゴリ：A) パワーユーザ（使用経験年数は問わない）

チーム名：「チーム ベイ」

松本 光右 株式会社 浜銀総合研究所 情報戦略コンサルティング部

三ツ井 誠 株式会社 浜銀総合研究所 情報戦略コンサルティング部

安田 浩平 株式会社 浜銀総合研究所 情報戦略コンサルティング部

## 要旨

本論文では、平成 31 年に実施が予定されている「消費税率の引上げ及び軽減税率制度の導入」について、主に低所得者への影響分析を行った。本分析により、低所得者層の負担緩和という導入目的において、軽減税率を飲食料品に適用することは妥当性があることが示された。また、軽減税率の適用範囲を拡大する場合は、交通・通信費を対象にすることが有効であることが示された。一方で、消費税率の引上げの影響により、全世帯の 3% 弱の家計収支が赤字になる可能性があり、その影響度は世帯のライフステージ等によって異なる。より精緻な分析を行うためには、世帯を消費行動ごとにグループ化することが有効だろう。

キーワード：消費税率引上げ、軽減税率制度、平成 16 年全国消費実態調査

## 1. 背景と目的

平成 31 年に実施が予定されている消費税率 10% への引上げに伴い、軽減税率制度が導入される公算が高い。軽減税率制度は低所得者への配慮の観点から導入されるものであり、消費税の対象となるもののうち、「酒類・外食を除く飲食料品」と「定期購読契約が締結された週 2 回以上発行される新聞」の消費税率を 8%（標準税率は 10%）に据え置くものである。

本論文の目的は、「消費税率の引上げ及び軽減税率制度の導入」が主に低所得者の家計にどのような影響を与えるかを調べることである。分析は以下の 2 つの視点を中心とする。

- I. 減税対象の妥当性：軽減税率制度は低所得者を優遇する制度になっているか
- II. 生活水準の保護：消費税率引上げにより、家計の赤字化を招く恐れがないか

「消費税率の引上げ及び軽減税率制度の導入」にあたっては、「制度の税収への影響度」や「軽減税率制度への対応による社会全体のコスト増加」など様々な論点があるが、本分析では「低所得者の家計への影響」に主眼を置くものとする。

## 2. 分析方針

### 2-1. 検証方法

「Ⅰ. 減税対象の妥当性」の検証方法はいくつかの方法が考えられるが、本論文では「実支出全体に占める当該項目の支出額の割合が、低所得者ほど大きくなっているか」と読みかえることとする。

「Ⅱ. 生活水準の保護」の検証では、「消費税率5%下では黒字であるが、消費税率10%になると赤字になる世帯」について分析を行う。本論文を執筆している時点では消費税率は8%であるが、使用データが「平成16年全国消費実態調査」に基づくデータであるため、消費税率は5%→10%に変化するものとする。また、平成16年と現在では、各種税率や所得の水準、物価、消費行動等に変化が生じているが、本分析ではこれらの点については考慮しないものとする。

### 2-2. 消費税率引上げによって増加する支出額の算出方法

本章では、消費税率が5%から10%に引上げられることによって増加する支出額の算出方法について整理する。

全国消費実態調査は「実支出」と「実支出以外の支出」に分けられる。「実支出以外の支出」は、預貯金や保険掛金、繰越金など、貯蓄性の支出で構成比されているため、本論文では「実支出」のみを検証の対象とする。

「実支出」は、

- A. 消費税の対象であり、軽減税率の対象となるもの
- B. 消費税の対象であり、軽減税率の対象とならないもの
- C. 消費税の対象とならないもの

の3種類に分けられる。「実支出」の各項目の消費税率について、図表1のように整理を行い、これに従い消費税率引上げの支出額を算出した。

#### 消費税率引上げによって増加する支出額の算出方法

- ① 消費税の対象となるもの（図表1のA、B）

$$\text{支出額（増税）} = \text{支出額} \times \text{税率} \div 1.05$$

- ② 消費税の対象とならないもの（図表1のC）

$$\text{支出額（増税）} = \text{支出額}$$

軽減税率の対象が「酒類・外食を除く飲食料品」であることを考えると、「038\_食料」の支出額から「075\_酒類」の支出額と「077\_一般外食」の支出額を控除したものに軽減税率をかけるべきであるが、これらが全体に与える影響度は小さいため、今回の分析では「038\_食料」に一律で軽減税率を適用するものとする。

図表 1 税率一覧

| 項目             | 消費税5% | 消費税10%+軽減税率 |
|----------------|-------|-------------|
| 038 食料         | 5%    | 8%          |
| 079 住居         | 0%    | 0%          |
| 084 光熱・水道      | 5%    | 10%         |
| 089 家具・家事用品    | 5%    | 10%         |
| 099 被服及び履物     | 5%    | 10%         |
| 119 医薬品        | 5%    | 10%         |
| 119 健康維持用機取品   | 5%    | 10%         |
| 120 保健医療用品・器具  | 5%    | 10%         |
| 121 保健医療サービス   | 0%    | 0%          |
| 122 交通・通信      | 5%    | 10%         |
| 130 授業料等       | 0%    | 0%          |
| 131 教科書・学習参考教材 | 5%    | 10%         |
| 132 補習教育       | 5%    | 10%         |
| 133 教養娯楽       | 5%    | 10%         |
| 142 その他の消費支出   | 5%    | 10%         |
| 160 直接税        | 0%    | 0%          |
| 164 社会保険料      | 0%    | 0%          |
| 189 他非消費支出     | 0%    | 0%          |

- A. 消費税の対象であり、軽減税率の対象となるもの  
 B. 消費税の対象であり、軽減税率の対象とならないもの  
 C. 消費税の対象とならないもの

### 3. 減税対象の妥当性に関する検証

本章では「減税対象の妥当性」に関する検証を行う。冒頭でも述べたが、実支出全体に占める当該項目の支出額の割合が、低所得者ほど大きくなっている場合、減税対象に妥当性があると判断する。また、減税対象が実支出全体に占める構成比にも注目する。

図表 2 は実支出全体に占める各項目の支出の割合を示したものであり、この表の結果について以下のように整理できる。

|                                                                                 |
|---------------------------------------------------------------------------------|
| A) 年間収入が小さいほど割合が大きいもの<br>「038_食料」「079_住居」「084_光熱・水道」「089_家具・家事用品」「122_交通・通信」    |
| B) 年間収入が大きいほど割合が大きいもの<br>「130_授業料等」「133_教養娯楽」「142_その他の消費支出」「160_直接税」「164_社会保険料」 |

減税対象が妥当かは、対象が A) の項目に含まれているかを確認すればよく、「038\_食料」はその条件を満たしている。A) に含まれるそれぞれの項目について、実支出全体に占める当該項目の支出の構成比を確認すると、

「038\_食料」 > 「122\_交通・通信」 > 「084\_光熱・水道」 > 「089\_家具・家事用品」

※ 「079\_住居」は消費税の対象外であるため除外している

であるため、「038\_食料」を軽減税率の対象とすることは、実支出全体へのインパクトが大きいことがわかる。

以上のことから、飲食料品に軽減税率を適用することは妥当であると考えられる。一方で、「038\_食料」に軽減税率 8% を適用することで、低所得者の最低限度の生活水準が保護されるかは別途検証する必要がある。次章では、消費税率引上げによって増加する支出額を元に家計収支の分析を行い、消費税率の引上げによって低所得者の家計状況がどう変化するかについて分析を行うものとする。

図表2、年間収入別：支出内訳（消費税5%）

|                | 年間収入    |             |             |              |          | 経構成比    |             |             |              |          |
|----------------|---------|-------------|-------------|--------------|----------|---------|-------------|-------------|--------------|----------|
|                | 400万円未満 | 400万円～600万円 | 600万円～800万円 | 800万円～1000万円 | 1000万円以上 | 400万円未満 | 400万円～600万円 | 600万円～800万円 | 800万円～1000万円 | 1000万円以上 |
| 038 食料         | 51,730  | 62,830      | 73,531      | 82,730       | 92,946   | 20.99%  | 19.58%      | 18.78%      | 17.49%       | 15.64%   |
| 079 住居         | 24,513  | 22,120      | 15,776      | 14,152       | 12,267   | 9.95%   | 6.89%       | 4.03%       | 2.99%        | 2.06%    |
| 084 光熱・水道      | 15,388  | 17,284      | 19,188      | 21,002       | 23,355   | 6.24%   | 5.39%       | 4.90%       | 4.44%        | 3.93%    |
| 089 家具・家事用品    | 6,507   | 7,812       | 9,290       | 10,604       | 11,783   | 2.64%   | 2.43%       | 2.37%       | 2.24%        | 1.98%    |
| 099 被服及び履物     | 8,183   | 10,815      | 13,750      | 17,326       | 20,969   | 3.32%   | 3.37%       | 3.51%       | 3.66%        | 3.53%    |
| 118 医薬品        | 1,312   | 1,421       | 1,680       | 1,799        | 2,043    | 0.53%   | 0.44%       | 0.43%       | 0.38%        | 0.34%    |
| 119 健康保持用摂取品   | 613     | 735         | 889         | 1,102        | 1,327    | 0.25%   | 0.23%       | 0.23%       | 0.23%        | 0.22%    |
| 120 保健医療用品・器具  | 1,789   | 2,046       | 2,233       | 2,336        | 2,467    | 0.73%   | 0.64%       | 0.57%       | 0.49%        | 0.42%    |
| 121 保健医療サービス   | 5,342   | 6,272       | 6,567       | 6,815        | 7,892    | 2.17%   | 1.95%       | 1.68%       | 1.44%        | 1.33%    |
| 122 交通・通信      | 33,050  | 40,760      | 47,175      | 55,353       | 63,671   | 3.41%   | 2.70%       | 2.05%       | 1.70%        | 1.07%    |
| 130 授業料等       | 7,832   | 11,852      | 17,574      | 23,007       | 31,782   | 3.18%   | 3.69%       | 4.43%       | 4.86%        | 5.35%    |
| 131 教科書・学習参考教材 | 118     | 218         | 295         | 439          | 493      | 0.05%   | 0.07%       | 0.08%       | 0.09%        | 0.08%    |
| 132 補習教育       | 1,258   | 2,022       | 3,991       | 5,151        | 6,346    | 0.51%   | 0.63%       | 1.02%       | 1.09%        | 1.07%    |
| 133 教養娯楽       | 17,593  | 25,132      | 32,029      | 37,248       | 44,251   | 7.14%   | 7.83%       | 8.18%       | 7.87%        | 7.45%    |
| 142 その他の消費支出   | 40,250  | 58,068      | 76,079      | 98,962       | 140,946  | 16.33%  | 18.10%      | 19.43%      | 20.92%       | 23.71%   |
| 160 直接税        | 7,995   | 15,855      | 26,162      | 38,674       | 58,773   | 3.24%   | 4.94%       | 6.68%       | 8.17%        | 9.89%    |
| 164 社会保険料      | 22,844  | 35,499      | 45,212      | 56,230       | 72,775   | 9.27%   | 11.06%      | 11.55%      | 11.89%       | 12.24%   |
| 169 他の非消費支出    | 96      | 119         | 122         | 162          | 278      | 0.04%   | 0.04%       | 0.03%       | 0.03%        | 0.05%    |

※図表2の対象は全世帯

#### 4. 生活水準の保護の観点

本章では、消費税率の引上げによって、主に低所得者が家計収支で黒字を維持できるかについて、検証を行う。検証にあたり、以下のように赤字家計と黒字家計を定義する。

|                  |                  |
|------------------|------------------|
| 赤字家計：実収入－実支出 < 0 | 黒字家計：実収入－実支出 ≥ 0 |
|------------------|------------------|

##### 4-1. 低所得者の定義

低所得者の明確な定義は存在しないが、本論文では「年間収入」が400万円未満の層を低所得者と定義する。図表3からわかるように、低所得者は「年間収入」が下位10%程度の世帯である。なお、本分析では地域による所得の水準差は考慮しないものとする。

図表3 年収収入の分布

| 年間収入              | 世帯数(10万世帯比) | 累積構成比   |
|-------------------|-------------|---------|
| 0円 以上 100万円未満     | 180         | 0.18%   |
| 100万円 以上 200万円未満  | 1,124       | 1.30%   |
| 200万円 以上 300万円未満  | 3,483       | 4.79%   |
| 300万円 以上 400万円未満  | 7,812       | 12.70%  |
| 400万円 以上 500万円未満  | 12,500      | 25.20%  |
| 500万円 以上 600万円未満  | 14,221      | 39.42%  |
| 600万円 以上 700万円未満  | 13,790      | 53.21%  |
| 700万円 以上 800万円未満  | 12,073      | 65.28%  |
| 800万円 以上 900万円未満  | 9,695       | 74.98%  |
| 900万円 以上 1000万円未満 | 7,534       | 82.51%  |
| 1000万円 以上         | 17,488      | 100.00% |

※図表3の対象は全世帯

##### 4-2. 消費税率引上げにより赤字家計となる世帯数

図表4は、消費税率5%と消費税率10%状況下における年間収入別の赤字家計と黒字家計の分布である。消費税率引上げによって黒字家計から赤字家計になる世帯数は、2,882世帯（10万世帯比）であり、そのうち低所得者の世帯数は392世帯（10万世帯比）である。

消費税率引上げによって赤字家計になる世帯の構成比を示した (A)-(B) の値は、年収 200～300 万円が一番大きく、次いで年収 300～400 万円の値が大きい。従って、消費税率引上げは低所得者への影響が相対的に大きいといえる。

図表 4 消費税率別：赤字家計と黒字家計の集計

| 税率                | 世帯数(10万世帯比) |        |             |        | 構成比    |        |             |        | 赤字家計率  |             |         |
|-------------------|-------------|--------|-------------|--------|--------|--------|-------------|--------|--------|-------------|---------|
|                   | 消費税5%       |        | 消費税10%+軽減税率 |        | 消費税5%  |        | 消費税10%+軽減税率 |        | (A)    | (B)         | (A)-(B) |
|                   | 赤字家計        | 黒字家計   | 赤字家計        | 黒字家計   | 赤字家計   | 黒字家計   | 赤字家計        | 黒字家計   | 消費税5%  | 消費税10%+軽減税率 |         |
| 収支                |             |        |             |        |        |        |             |        |        |             |         |
| 0円 以上 100万円未満     | 137         | 44     | 139         | 42     | 0.14%  | 0.04%  | 0.14%       | 0.04%  | 75.88% | 76.82%      | 1.04%   |
| 100万円 以上 200万円未満  | 707         | 416    | 721         | 403    | 0.71%  | 0.42%  | 0.72%       | 0.40%  | 62.64% | 64.11%      | 1.21%   |
| 200万円 以上 300万円未満  | 1,578       | 1,905  | 1,599       | 1,784  | 1.58%  | 1.91%  | 1.70%       | 1.78%  | 45.30% | 48.79%      | 3.49%   |
| 300万円 以上 400万円未満  | 2,624       | 5,288  | 2,880       | 5,032  | 2.62%  | 5.29%  | 2.88%       | 5.03%  | 33.16% | 36.40%      | 3.24%   |
| 400万円 以上 500万円未満  | 3,431       | 9,088  | 3,714       | 8,726  | 3.43%  | 9.07%  | 3.77%       | 8.73%  | 27.45% | 30.19%      | 2.74%   |
| 500万円 以上 600万円未満  | 3,459       | 10,762 | 3,848       | 10,373 | 3.46%  | 10.76% | 3.85%       | 10.37% | 24.32% | 27.06%      | 2.74%   |
| 600万円 以上 700万円未満  | 2,945       | 10,845 | 3,334       | 10,456 | 2.95%  | 10.85% | 3.33%       | 10.46% | 21.36% | 24.18%      | 2.82%   |
| 700万円 以上 800万円未満  | 2,496       | 9,578  | 2,842       | 9,231  | 2.50%  | 9.58%  | 2.84%       | 9.23%  | 20.67% | 23.54%      | 2.87%   |
| 800万円 以上 900万円未満  | 2,172       | 7,523  | 2,480       | 7,215  | 2.17%  | 7.52%  | 2.48%       | 7.21%  | 22.40% | 25.58%      | 3.18%   |
| 900万円 以上 1000万円未満 | 1,530       | 6,004  | 1,730       | 5,804  | 1.53%  | 6.00%  | 1.73%       | 5.80%  | 20.31% | 22.96%      | 2.66%   |
| 1000万円 以上         | 3,206       | 14,282 | 3,719       | 13,769 | 3.21%  | 14.28% | 3.72%       | 13.77% | 18.33% | 21.27%      | 2.93%   |
| 合計                | 24,284      | 75,718 | 27,166      | 72,834 | 24.28% | 75.72% | 27.17%      | 72.83% | 24.28% | 27.17%      | 2.88%   |

※図表 4 の対象は全世帯

#### 4-3. 増税後の低所得者の家計状況

図表 5 は低所得者（年間収入 400 万円未満）の世帯を、赤字家計と黒字家計に分類し、10 万世帯あたりの世帯数を算出したものである。本章の図表 5、7 は結果の一部のみを掲載し、結果の全体は 7 章の付録に記載している。

黒字家計と赤字家計を比較すると、赤字家計は「世帯人員：2 人」と「年齢：60～64 歳」の構成比が高いことがわかる。ここから推測される人物像は「年金を満額受給する前の世帯が、貯蓄等を切り崩して生活している像」である。これらの世帯は計画的な赤字家計とみなすことができるため、低所得者の家計状況を分析する際には、除くことが望ましい。次節では「消費税 5% 時は黒字家計であった世帯」のみを対象とすることで、計画的な赤字家計を排除した分析を行う。

図表 5 低所得者の家計状況（消費税 10%+軽減税率）（一部抜粋）

| 消費税率10%+軽減税率 | 世帯数(10万世帯比) |       | 構成比    |        | 消費税率10%+軽減税率 | 世帯数(10万世帯比) |      | 構成比    |      |
|--------------|-------------|-------|--------|--------|--------------|-------------|------|--------|------|
|              | 赤字家計        | 黒字家計  | 赤字家計   | 黒字家計   |              | 赤字家計        | 黒字家計 | 赤字家計   | 黒字家計 |
| ■世帯人員        |             |       |        |        |              |             |      |        |      |
| 2人           | 2,289       | 2,572 | 42.08% | 25.43% |              |             |      |        |      |
| 3人           | 1,743       | 2,218 | 32.04% | 30.54% |              |             |      |        |      |
| 4人           | 1,018       | 1,798 | 18.71% | 24.77% |              |             |      |        |      |
| 5人           | 311         | 528   | 5.72%  | 7.28%  |              |             |      |        |      |
| 6人           | 61          | 83    | 1.12%  | 1.15%  |              |             |      |        |      |
| 7人           | 18          | 52    | 0.32%  | 0.72%  |              |             |      |        |      |
| 8人           | 0           | 6     | 0.00%  | 0.08%  |              |             |      |        |      |
| 9人           | 0           | 2     | 0.00%  | 0.03%  |              |             |      |        |      |
| ■年齢          |             |       |        |        |              |             |      |        |      |
| 01.24歳未満     |             | 75    |        | 1.38%  |              | 95          |      | 1.31%  |      |
| 02.25～29歳    |             | 314   |        | 5.77%  |              | 738         |      | 10.17% |      |
| 03.30～34歳    |             | 538   |        | 9.89%  |              | 1,047       |      | 14.42% |      |
| 04.35～39歳    |             | 340   |        | 6.26%  |              | 786         |      | 10.82% |      |
| 05.40～44歳    |             | 350   |        | 6.44%  |              | 443         |      | 6.10%  |      |
| 06.45～49歳    |             | 238   |        | 4.38%  |              | 327         |      | 4.51%  |      |
| 07.50～54歳    |             | 381   |        | 7.00%  |              | 353         |      | 4.86%  |      |
| 08.55～59歳    |             | 335   |        | 6.16%  |              | 489         |      | 6.73%  |      |
| 09.60～64歳    |             | 556   |        | 10.21% |              | 368         |      | 5.07%  |      |
| 10.65～69歳    |             | 131   |        | 2.41%  |              | 161         |      | 2.22%  |      |
| 11.70～74歳    |             | 14    |        | 0.26%  |              | 43          |      | 0.59%  |      |
| 12.75歳以上     |             | 3     |        | 0.06%  |              | 7           |      | 0.10%  |      |
| 99.不詳        |             | 2,163 |        | 36.78% |              | 2,403       |      | 33.09% |      |

※図表 5 の対象は「年間収入 400 万円未満」の世帯



#### 4-4. 増税により家計が赤字になる世帯の特徴

図表6は「消費税率5%時は黒字家計であった世帯」を対象に、「消費税率10%時に赤字家計となる世帯」と「消費税率10%時に黒字家計となる世帯」について、実支出額の比較を行ったものである。図表6の縦構成比に着目すると、ほとんどの項目で赤字家計と黒字家計に差がみられないが、「160\_直接税」は赤字家計の方が支出額、構成比ともに低いという特徴がある。一般に、年間収入が同一の場合、扶養家族が多いほど直接税が少なくなるため、ここから赤字家計世帯は扶養家族が多いことがわかる。また、赤字家計世帯の「084\_光熱・水道」の支出額が、黒字家計世帯よりも大きい。これは、赤字家計世帯の方が黒字家計世帯よりも世帯人員が多いことを暗示している（図表8より、光熱・水道費と世帯人員の関係は明白である）。

図表7は黒字家計と赤字家計について、属性の分布を比較したものである。図表6で得られた示唆と同様に、赤字家計は黒字家計よりも世帯人員が多い。また、住居の建て方に着目すると、一戸建の構成比が高いことがわかる。

以上を踏まえると、消費税率引上げによって、黒字家計から赤字家計になる世帯には、子供のいる核家族世帯が一定数含まれているのではないかと考えられる。

図表6 消費税5%時は黒字家計であった世帯の平均支出

| 消費税率10%+軽減税率   | 平均支出額   |         | 縦構成比    |         |
|----------------|---------|---------|---------|---------|
|                | 赤字家計    | 黒字家計    | 赤字家計    | 黒字家計    |
| 038 食料         | 59,478  | 50,185  | 22.69%  | 22.02%  |
| 079 住居         | 21,053  | 22,461  | 8.03%   | 9.86%   |
| 084 光熱・水道      | 17,630  | 15,889  | 6.73%   | 6.88%   |
| 089 家具・家事用品    | 10,043  | 5,553   | 3.83%   | 2.44%   |
| 099 被服及び履物     | 8,988   | 7,070   | 3.43%   | 3.10%   |
| 118 医薬品        | 1,278   | 1,204   | 0.49%   | 0.53%   |
| 119 健康保持用採取品   | 807     | 448     | 0.31%   | 0.20%   |
| 120 保健医療用品・器具  | 1,800   | 1,764   | 0.69%   | 0.77%   |
| 121 保健医療サービス   | 5,227   | 4,291   | 1.99%   | 1.88%   |
| 122 交通・通信      | 35,716  | 28,311  | 13.63%  | 12.42%  |
| 130 授業料等       | 5,748   | 5,611   | 2.19%   | 2.46%   |
| 131 教科書・学習参考教材 | 74      | 88      | 0.03%   | 0.04%   |
| 132 補習教育       | 1,053   | 872     | 0.40%   | 0.38%   |
| 133 教養娯楽       | 19,540  | 15,792  | 7.46%   | 6.93%   |
| 142 その他の消費支出   | 46,147  | 35,977  | 17.61%  | 15.79%  |
| 160 直接税        | 6,484   | 8,420   | 2.47%   | 3.69%   |
| 164 社会保険料      | 21,010  | 24,051  | 8.02%   | 10.55%  |
| 169 他の非消費支出    | 26      | 99      | 0.01%   | 0.04%   |
| 実支出計           | 262,098 | 227,888 | 100.00% | 100.00% |

※図表6の対象は「年間収入400万円未満」かつ「消費税率5%時は黒字家計」の世帯

図表7 消費税5%時は黒字家計であった世帯の平均支出（一部抜粋）

| 消費税率10%+軽減税率 | 世帯数(10万世帯比) |       | 構成比   |       |
|--------------|-------------|-------|-------|-------|
|              | 赤字家計        | 黒字家計  | 赤字家計  | 黒字家計  |
| 世帯人員         |             |       |       |       |
| 2人           | 126         | 2,572 | 2.16% | 5.43% |
| 3人           | 125         | 2,218 | 1.82% | 5.04% |
| 4人           | 104         | 1,798 | 2.46% | 4.77% |
| 5人           | 29          | 528   | 7.46% | 7.28% |
| 6人           | 5           | 83    | 1.31% | 1.15% |
| 7人           | 3           | 52    | 0.79% | 0.72% |
| 8人           | 0           | 6     | 0.00% | 0.08% |
| 9人           | 0           | 2     | 0.00% | 0.03% |

| 消費税率10%+軽減税率    | 世帯数(10万世帯比) |       | 構成比    |        |
|-----------------|-------------|-------|--------|--------|
|                 | 赤字家計        | 黒字家計  | 赤字家計   | 黒字家計   |
| 居住の建て方          |             |       |        |        |
| 01 一戸建          | 190         | 3,161 | 23%    | 5.53%  |
| 02 長屋棟          | 6           | 84    | 1.54%  | 0.88%  |
| 03 共同住宅(1-2階建)  | 17          | 661   | 4.23%  | 9.10%  |
| 04 共同住宅(3-5階建)  | 85          | 1,272 | 12.60% | 17.52% |
| 05 共同住宅(6-10階建) | 2           | 120   | 0.52%  | 1.65%  |
| 06 共同住宅(11階建以上) | 0           | 50    | 0.00%  | 0.68%  |
| 99 不詳           | 94          | 1,934 | 23.87% | 6.63%  |

※図表7の対象は「年間収入400万円未満」かつ「消費税率5%時は黒字家計」の世帯

図表8 世帯人員と光熱・水道費の関係

|      |      | 世帯数(10万世帯比) |             |              |              |              |          |
|------|------|-------------|-------------|--------------|--------------|--------------|----------|
|      |      | 084 光熱・水道費  |             |              |              |              |          |
|      |      | 5000円未満     | 5000～10000円 | 10000～15000円 | 15000～20000円 | 20000～25000円 | 25000円以上 |
| 世帯人員 | 2人   | 120         | 3,268       | 10,001       | 6,468        | 2,287        | 1,235    |
|      | 3人   | 112         | 1,397       | 7,613        | 9,203        | 5,160        | 3,158    |
|      | 4人   | 65          | 866         | 6,142        | 10,083       | 7,876        | 6,423    |
|      | 5人以上 | 25          | 303         | 1,548        | 4,110        | 4,840        | 7,696    |

|      |      | 世帯数(10万世帯比) 構成比 |             |              |              |              |          |
|------|------|-----------------|-------------|--------------|--------------|--------------|----------|
|      |      | 084 光熱・水道費      |             |              |              |              |          |
|      |      | 5000円未満         | 5000～10000円 | 10000～15000円 | 15000～20000円 | 20000～25000円 | 25000円以上 |
| 世帯人員 | 2人   | 0.51%           | 13.98%      | 42.78%       | 27.67%       | 9.78%        | 5.28%    |
|      | 3人   | 0.42%           | 5.24%       | 23.58%       | 34.54%       | 19.37%       | 11.85%   |
|      | 4人   | 0.21%           | 2.75%       | 19.53%       | 32.06%       | 25.04%       | 20.42%   |
|      | 5人以上 | 0.14%           | 1.64%       | 8.36%        | 22.19%       | 26.13%       | 41.55%   |

※図表8の対象は全世帯

## 5. 結論

### 5-1. まとめ

冒頭で述べた2つの視点について、結論をまとめる。

#### I. 減税対象の妥当性：軽減税率制度は低所得者を優遇する制度になっているか

年間収入が低いほど、実支出全体に占める食料費の割合が高く、また、食料費は実支出全体の2割を占める影響度の大きい項目であることから、飲食料品を減税対象にすることは妥当性があると考えられる。軽減税率を導入する場合、導入しない場合と比べて、実支出全体が0.35%程度軽減されるだろう。

※計算式： $20\% \times (1 - 1.08 \div 1.10) \approx 0.35\%$

#### II. 生活水準の保護：消費税率引上げにより、家計の赤字化を招く恐れはないか

消費税率を引き上げた場合、赤字家計世帯が増加する可能性があり、その可能性は年間収入が200～400万円の世帯が相対的に高い。また、同じ年間収入でもライフステージ等によって実支出額が異なるため、世帯人員が多い場合などに赤字家計になりやすい。

以上より、軽減税率の対象には妥当性があるが、その効果は実支出の0.35%と限定的である。そのため、軽減税率を適用したとしても、赤字家計となる世帯が一定数発生する可能性があるといえる。

### 5-2. 分析から得られる示唆

消費税率の引上げによって、特定世帯の家計状況が過度に厳しくなった場合の対策として、以下の2つ手段が考えられる。

1つ目は、実収入を上げる施策を打つことである。低所得者層の勤労所得を上げることや、給付金の付与等が考えられる。

2つ目は、直接税・間接税問わず、特定世帯を救済する税制を整備し、実支出を減らすことだ。もし、軽減税率制度で実現する場合は、交通・通信費に軽減税率を拡大することが考えられる。ただし、この方法によって得られる効果は限定的であることに留意が必要だ。

また、今回の分析では、世帯のライフステージ等の違いにより、同じ年間収入でも消費税率引上げが与える影響の大きさが違うことがわかった。より精緻な分析を行うためには、世帯を消費行動ごとにグループ化することが有効だろう。

## 6. 参考文献

- ・総務省統計局「平成16年全国消費実態調査」 <http://www.stat.go.jp/data/zensho/2004/>
- ・国税庁「消費税の軽減税率制度について」 <https://www.nta.go.jp/zeimokubetsu/shohi/keigenzeiritsu/>

# 7. 付録：図表

図表5 低所得者の家計状況(消費税10%+軽減税率)

| 消費税10%+軽減税率            | 世帯数(10万世帯単位) |       | 世帯数    |        |
|------------------------|--------------|-------|--------|--------|
|                        | 黒字家計         | 黒字家計  | 黒字家計   | 黒字家計   |
| 世帯収入                   |              |       |        |        |
| 250万円未満                | 1,524        | 1,044 | 18.02% | 14.38% |
| 300万円未満                | 1,035        | 1,184 | 19.03% | 18.31% |
| 350万円未満                | 1,255        | 2,046 | 23.07% | 28.13% |
| 400万円未満                | 1,625        | 2,988 | 29.88% | 37.12% |
| 無職者人員                  |              |       |        |        |
| 1人                     | 2,289        | 2,572 | 42.08% | 35.43% |
| 2人                     | 1,743        | 2,218 | 32.04% | 30.54% |
| 3人                     | 1,018        | 1,798 | 18.71% | 24.77% |
| 4人                     | 311          | 528   | 5.72%  | 7.28%  |
| 5人                     | 61           | 83    | 1.12%  | 1.15%  |
| 6人                     | 18           | 52    | 0.32%  | 0.72%  |
| 7人                     | 0            | 8     | 0.00%  | 0.08%  |
| 8人                     | 0            | 2     | 0.00%  | 0.03%  |
| 無職者人員                  |              |       |        |        |
| 1人                     | 3,910        | 4,540 | 71.80% | 62.33% |
| 2人                     | 1,154        | 2,198 | 21.22% | 30.28% |
| 3人                     | 152          | 225   | 2.79%  | 3.11%  |
| 4人                     | 25           | 18    | 0.46%  | 0.25%  |
| 5人                     | 197          | 279   | 3.63%  | 3.83%  |
| 世帯の消費                  |              |       |        |        |
| 01 衣類                  | 1,859        | 2,308 | 4.18%  | 11.79% |
| 02 防犯衣類                | 304          | 593   | 5.60%  | 8.17%  |
| 03 靴類・履物・アクセサリー類       | 2,025        | 2,915 | 37.24% | 40.16% |
| 04 その他(ブロック売り、レンジ売りなど) | 11           | 5     | 0.21%  | 0.07%  |
| 05 合計                  | 1,239        | 1,439 | 22.78% | 19.82% |
| 居住費の増減                 |              |       |        |        |
| 01 一戸建て                | 2,420        | 3,161 | 44.50% | 46.53% |
| 02 共同住宅                | 60           | 64    | 1.05%  | 0.88%  |
| 03 共同住宅(1-2階建)         | 404          | 661   | 7.43%  | 9.10%  |
| 04 共同住宅(3-5階建)         | 654          | 1,272 | 16.45% | 17.52% |
| 05 共同住宅(6-10階建)        | 86           | 120   | 1.59%  | 1.65%  |
| 06 共同住宅(11階建以上)        | 35           | 50    | 0.65%  | 0.68%  |
| 07 合計                  | 1,539        | 1,934 | 29.29% | 26.83% |
| 居住費の負担割合               |              |       |        |        |
| 01 世帯収入(世帯員数)          | 2,207        | 2,856 | 40.58% | 34.4%  |
| 02 世帯収入(世帯員数)          | 252          | 335   | 4.63%  | 4.81%  |
| 03 世帯収入(世帯員数)          | 1,454        | 1,959 | 26.73% | 26.98% |
| 04 世帯収入(世帯員数)          | 690          | 1,226 | 12.68% | 16.89% |
| 05 世帯収入(世帯員数)          | 137          | 141   | 2.51%  | 1.84%  |
| 06 世帯収入(世帯員数)          | 68           | 106   | 1.24%  | 1.46%  |
| 07 合計                  | 18           | 5     | 0.29%  | 0.07%  |
| 08 合計                  | 617          | 632   | 11.34% | 8.70%  |
| 居住費                    |              |       |        |        |
| 01 居住                  | 3,434        | 5,839 | 60.44% | 66.42% |
| 02 合計                  | 2,005        | 1,421 | 36.86% | 19.57% |
| 年齢別                    |              |       |        |        |
| 01 24歳未満               | 75           | 95    | 1.38%  | 1.31%  |
| 02 25-29歳              | 314          | 738   | 5.77%  | 10.17% |
| 03 30-34歳              | 538          | 1,047 | 9.89%  | 14.42% |
| 04 35-39歳              | 340          | 788   | 6.26%  | 10.82% |
| 05 40-44歳              | 350          | 443   | 6.44%  | 6.10%  |
| 06 45-49歳              | 238          | 327   | 4.38%  | 4.51%  |
| 07 50-54歳              | 381          | 353   | 7.00%  | 4.86%  |
| 08 55-59歳              | 335          | 489   | 6.16%  | 6.73%  |
| 09 60-64歳              | 558          | 368   | 10.21% | 5.07%  |
| 10 65-69歳              | 131          | 161   | 2.41%  | 2.22%  |
| 11 70-74歳              | 14           | 43    | 0.26%  | 0.59%  |
| 12 75歳以上               | 3            | 7     | 0.08%  | 0.10%  |
| 03 合計                  | 2,163        | 2,403 | 39.78% | 33.09% |
| 居住費・貯蓄費の別              |              |       |        |        |
| 01 貯蓄                  | 3,833        | 6,179 | 70.47% | 68.17% |
| 02 貯蓄(うち、パート)          | 1,438        | 978   | 28.43% | 13.47% |
| 03 合計                  | 189          | 104   | 3.10%  | 1.43%  |
| 居住費区分                  |              |       |        |        |
| 01 居住                  | 5,122        | 6,953 | 94.17% | 95.47% |
| 02 合計                  | 278          | 302   | 5.11%  | 4.18%  |
| 03 合計                  | 39           | 5     | 0.72%  | 0.08%  |
| 居住費負担                  |              |       |        |        |
| 01 1-4人                | 383          | 509   | 6.87%  | 8.97%  |
| 02 5-9人                | 1,620        | 2,197 | 29.78% | 30.26% |
| 03 10-49人              | 1,642        | 2,470 | 30.19% | 34.02% |
| 04 500-999人            | 163          | 188   | 2.99%  | 2.59%  |
| 05 1000人以上             | 371          | 369   | 6.82%  | 5.08%  |
| 06 居住費負担割合             | 278          | 302   | 5.11%  | 4.18%  |
| 07 合計                  | 1,603        | 1,228 | 18.45% | 16.91% |
| 居住費負担                  |              |       |        |        |
| 04 居住                  | 1            | 0     | 0.03%  | 0.00%  |
| 05 居住費                 | 410          | 775   | 7.53%  | 10.68% |
| 06 居住費                 | 616          | 1,015 | 11.33% | 13.98% |
| 07 電気・ガス・水道費・水増費       | 9            | 5     | 0.17%  | 0.07%  |
| 08 借入金返済               | 10           | 27    | 0.19%  | 0.31%  |
| 09 借入金                 | 206          | 320   | 3.79%  | 4.54%  |
| 10 貯蓄・小遣金              | 424          | 600   | 2.84%  | 7.03%  |
| 11 生活費・貯蓄              | 45           | 32    | 0.80%  | 0.43%  |
| 12 学費・教育               | 21           | 3     | 0.39%  | 0.05%  |
| 13 貯蓄・貯蓄               | 80           | 41    | 1.11%  | 0.57%  |
| 14 借入金・借入金             | 135          | 138   | 2.46%  | 1.80%  |
| 15 借入金・学費・学費           | 32           | 19    | 0.59%  | 0.27%  |
| 16 借入金・借入金             | 8            | 37    | 0.14%  | 0.51%  |
| 17 一戸建て・借入金(借入金以外の)    | 418          | 510   | 2.85%  | 7.03%  |
| 18 借入金(借入金以外の)         | 125          | 118   | 1.89%  | 1.63%  |
| 07 合計                  | 2,895        | 3,609 | 53.05% | 49.77% |
| 居住費負担                  |              |       |        |        |
| 01 借入金負担割合             | 3,141        | 4,683 | 57.77% | 64.80% |
| 02 借入金負担割合             | 140          | 146   | 2.58%  | 2.02%  |
| 03 借入金負担               | 1,281        | 1,808 | 23.56% | 22.16% |
| 04 借入金負担1              | 28           | 51    | 0.51%  | 0.71%  |
| 05 借入金負担2              | 103          | 118   | 1.89%  | 1.63%  |
| 06 合計                  | 746          | 653   | 13.71% | 8.99%  |

図表7 消費税5%時は黒字家計であった世帯の平均支出

| 消費税10%+軽減税率            | 世帯数(10万世帯単位) |       | 世帯数    |        |
|------------------------|--------------|-------|--------|--------|
|                        | 黒字家計         | 黒字家計  | 黒字家計   | 黒字家計   |
| 世帯収入                   |              |       |        |        |
| 250万円未満                | 55           | 1,044 | 14.03% | 14.38% |
| 300万円未満                | 82           | 1,184 | 20.84% | 18.31% |
| 350万円未満                | 93           | 2,046 | 23.75% | 28.13% |
| 400万円未満                | 163          | 2,988 | 41.36% | 37.12% |
| 無職者人員                  |              |       |        |        |
| 1人                     | 128          | 2,572 | 27.16% | 35.43% |
| 2人                     | 125          | 2,218 | 31.82% | 30.54% |
| 3人                     | 104          | 1,798 | 26.48% | 24.77% |
| 4人                     | 29           | 528   | 7.46%  | 7.28%  |
| 5人                     | 5            | 83    | 1.31%  | 1.15%  |
| 6人                     | 3            | 52    | 0.79%  | 0.72%  |
| 7人                     | 0            | 8     | 0.00%  | 0.08%  |
| 8人                     | 0            | 2     | 0.00%  | 0.03%  |
| 無職者人員                  |              |       |        |        |
| 1人                     | 251          | 4,540 | 63.90% | 62.33% |
| 2人                     | 118          | 2,198 | 29.53% | 30.28% |
| 3人                     | 3            | 225   | 1.27%  | 3.11%  |
| 4人                     | 3            | 18    | 0.21%  | 0.25%  |
| 5人                     | 18           | 278   | 4.49%  | 3.83%  |
| 世帯の消費                  |              |       |        |        |
| 01 衣類                  | 138          | 2,308 | 4.52%  | 11.79% |
| 02 防犯衣類                | 14           | 593   | 3.80%  | 8.17%  |
| 03 靴類・履物・アクセサリー類       | 145          | 2,915 | 26.79% | 40.16% |
| 04 その他(ブロック売り、レンジ売りなど) | 0            | 5     | 0.00%  | 0.07%  |
| 05 合計                  | 99           | 1,439 | 25.10% | 19.82% |
| 居住費の増減                 |              |       |        |        |
| 01 一戸建て                | 190          | 3,161 | 49.23% | 46.53% |
| 02 共同住宅                | 6            | 64    | 1.54%  | 0.88%  |
| 03 共同住宅(1-2階建)         | 17           | 661   | 4.23%  | 9.10%  |
| 04 共同住宅(3-5階建)         | 85           | 1,272 | 21.60% | 17.52% |
| 05 共同住宅(6-10階建)        | 2            | 120   | 0.52%  | 1.65%  |
| 06 共同住宅(11階建以上)        | 0            | 50    | 0.00%  | 0.68%  |
| 07 合計                  | 94           | 1,934 | 38.77% | 26.83% |
| 居住費の負担割合               |              |       |        |        |
| 01 世帯収入(世帯員数)          | 155          | 2,856 | 40.40% | 34.4%  |
| 02 世帯収入(世帯員数)          | 38           | 335   | 9.73%  | 4.81%  |
| 03 世帯収入(世帯員数)          | 91           | 1,959 | 23.18% | 26.98% |
| 04 世帯収入(世帯員数)          | 50           | 1,226 | 12.83% | 16.89% |
| 05 世帯収入(世帯員数)          | 19           | 141   | 4.83%  | 1.84%  |
| 06 世帯収入(世帯員数)          | 5            | 106   | 1.32%  | 1.46%  |
| 07 合計                  | 0            | 5     | 0.00%  | 0.07%  |
| 08 合計                  | 34           | 632   | 8.71%  | 8.70%  |
| 居住費                    |              |       |        |        |
| 01 居住                  | 284          | 5,839 | 59.44% | 66.42% |
| 02 合計                  | 109          | 1,421 | 27.68% | 19.57% |
| 年齢別                    |              |       |        |        |
| 01 24歳未満               | 12           | 95    | 3.18%  | 1.31%  |
| 02 25-29歳              | 15           | 738   | 3.81%  | 10.17% |
| 03 30-34歳              | 37           | 1,047 | 9.32%  | 14.42% |
| 04 35-39歳              | 30           | 788   | 7.55%  | 10.82% |
| 05 40-44歳              | 29           | 443   | 7.15%  | 6.10%  |
| 06 45-49歳              | 6            | 327   | 1.86%  | 4.51%  |
| 07 50-54歳              | 31           | 353   | 7.78%  | 4.86%  |
| 08 55-59歳              | 36           | 489   | 9.06%  | 6.73%  |
| 09 60-64歳              | 30           | 368   | 7.54%  | 5.07%  |
| 10 65-69歳              | 6            | 161   | 1.82%  | 2.22%  |
| 11 70-74歳              | 4            | 43    | 1.08%  | 0.59%  |
| 12 75歳以上               | 0            | 7     | 0.00%  | 0.10%  |
| 03 合計                  | 159          | 2,403 | 39.43% | 33.09% |
| 居住費・貯蓄費の別              |              |       |        |        |
| 01 貯蓄                  | 317          | 6,179 | 69.72% | 68.17% |
| 02 貯蓄(うち、パート)          | 62           | 978   | 15.78% | 13.47% |
| 03 合計                  | 14           | 104   | 3.49%  | 1.43%  |
| 居住費区分                  |              |       |        |        |
| 01 居住                  | 372          | 6,953 | 96.60% | 95.47% |
| 02 合計                  | 18           | 302   | 4.54%  | 4.18%  |
| 03 合計                  | 3            | 5     | 0.81%  | 0.08%  |
| 居住費負担                  |              |       |        |        |
| 01 1-4人                | 16           | 509   | 4.15%  | 6.97%  |
| 02 5-9人                | 99           | 2,197 | 25.25% | 30.26% |
| 03 10-49人              | 124          | 2,470 | 31.54% | 34.02% |
| 04 500-999人            | 15           | 188   | 3.87%  | 2.59%  |
| 05 1000人以上             | 31           | 369   | 7.90%  | 5.08%  |
| 06 居住費負担割合             | 18           | 302   | 4.54%  | 4.18%  |
| 07 合計                  | 89           | 1,228 | 27.74% | 16.91% |
| 居住費負担                  |              |       |        |        |
| 04 居住                  | 0            | 0     | 0.00%  | 0.00%  |
| 05 居住費                 | 38           | 775   | 9.26%  | 10.68% |
| 06 居住費                 | 52           | 1,015 | 13.32% | 13.98% |
| 07 電気・ガス・水道費・水増費       | 0            | 5     | 0.00%  | 0.07%  |
| 08 借入金返済               | 0            | 27    | 0.00%  | 0.31%  |
| 09 借入金                 | 22           | 320   | 5.48%  | 4.54%  |
| 10 貯蓄・小遣金              | 33           | 600   | 6.39%  | 7.03%  |
| 11 生活費・貯蓄              | 0            | 32    | 0.00%  | 0.43%  |
| 12 学費・教育               | 0            | 3     | 0.00%  | 0.05%  |
| 13 貯蓄・貯蓄               | 9            | 41    | 2.18%  | 0.57%  |
| 14 借入金・借入金             | 5            | 138   | 1.23%  | 1.80%  |
| 15 借入金・学費・学費           | 0            | 19    | 0.00%  | 0.27%  |
| 16 借入金・借入金             | 0            | 37    | 0.00%  | 0.51%  |
| 17 一戸建て・借入金(借入金以外の)    | 10           | 510   | 2.53%  | 7.03%  |
| 18 借入金(借入金以外の)         | 2            | 118   | 2.24%  | 1.63%  |
| 07 合計                  | 215          | 3,609 | 46.67% | 49.77% |
| 居住費負担                  |              |       |        |        |
| 01 借入金負担割合             | 261          | 4,683 | 60.00% | 64.80% |
| 02 借入金負担割合             | 10           | 146   | 2.57%  | 2.02%  |
| 03 借入金負担               | 68           | 1,609 | 27.28% | 22.16% |
| 04 借入金負担1              | 3            | 51    | 0.81%  | 0.71%  |
| 05 借入金負担2              | 2            | 118   | 2.24%  | 1.63%  |
| 06 合計                  | 42           | 653   | 10.79% | 8.99%  |

規定課題

課題1) 産業符号・性別 クロス表

産業符号別・性別に集計用乗率なしの単純クロス集計表，および集計用乗率を使って 10 万世帯あたり（10 万世帯比）のクロス集計表を，表 1 の様式で作成して下さい。（注：産業符号の「不詳」が先頭に来て構いません。）

解答1)

| 産業符号              | a)単純クロス表 |       |        | b)10万世帯比でのクロス表 |       |         |
|-------------------|----------|-------|--------|----------------|-------|---------|
|                   | 男        | 女     | 合計     | 男              | 女     | 合計      |
| 鉱業                | 3        | 0     | 3      | 7              | 0     | 7       |
| 建設業               | 2,134    | 9     | 2,143  | 6,346          | 31    | 6,377   |
| 製造業               | 5,668    | 173   | 5,841  | 17,837         | 510   | 18,347  |
| 電気・ガス・熱供給・水道業     | 222      | 0     | 222    | 697            | 0     | 697     |
| 情報通信業             | 547      | 0     | 547    | 2,020          | 0     | 2,020   |
| 運輸業               | 1,235    | 3     | 1,238  | 3,852          | 6     | 3,858   |
| 卸売・小売業            | 2,083    | 201   | 2,284  | 6,644          | 589   | 7,233   |
| 金融・保険業            | 472      | 39    | 511    | 1,447          | 104   | 1,551   |
| 不動産業              | 40       | 0     | 40     | 144            | 0     | 144     |
| 飲食店・宿泊業           | 50       | 32    | 82     | 158            | 107   | 265     |
| 医療・福祉             | 371      | 270   | 641    | 1,022          | 810   | 1,832   |
| 教育・学習支援業          | 1,059    | 89    | 1,148  | 3,110          | 269   | 3,379   |
| 複合サービス業           | 321      | 3     | 324    | 981            | 11    | 992     |
| サービス業(他に分類されないもの) | 1,816    | 106   | 1,922  | 5,842          | 326   | 6,167   |
| 公務(他に分類されないもの)    | 3,656    | 130   | 3,786  | 10,691         | 430   | 11,121  |
| 不詳                | 9,704    | 1,591 | 11,295 | 31,204         | 4,806 | 36,011  |
| 合計                | 29,381   | 2,646 | 32,027 | 92,000         | 8,000 | 100,000 |

freq プロシジャを用いて集計した。

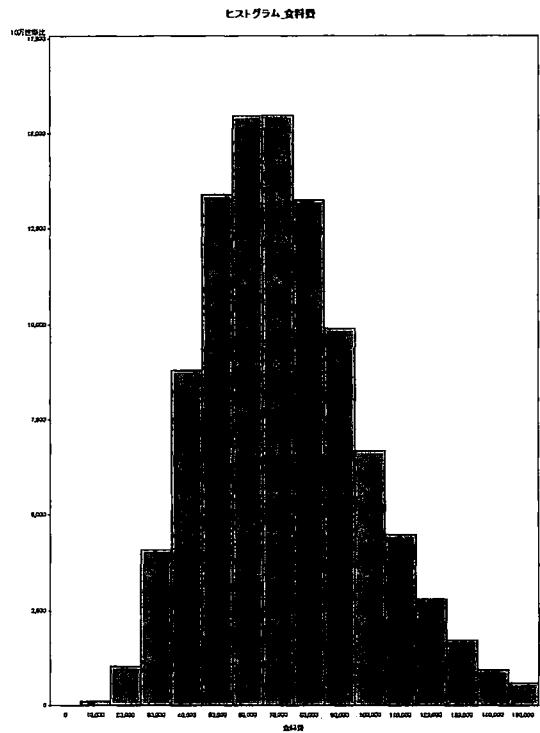
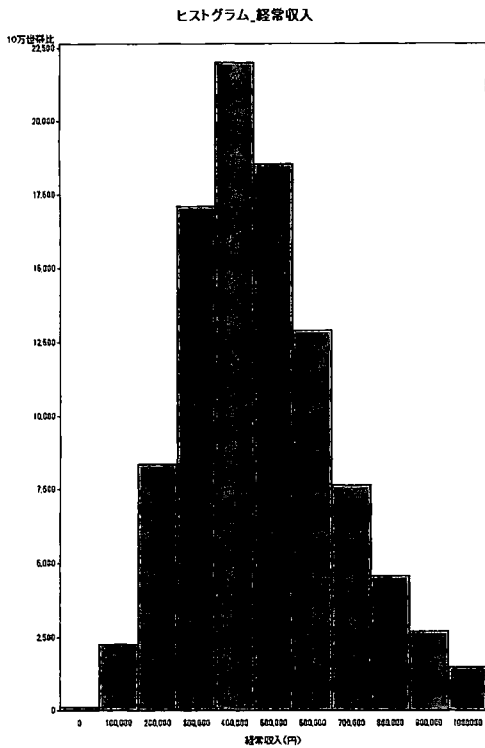
課題2) 経常収入および食料費

経常収入および食料費について，集計用乗率を用いて表 2 のように 5 分位点（下側 20%点，40%点，60%点および 80%点） A～F を求めて，表 2 を完成して下さい。更に，経常収入および食糧費についてそれぞれヒストグラムを描いて下さい。ヒストグラムの縦軸はパーセント（又は 10 万世帯比），横軸は経常収入および食糧費の階級区分（単位：円）とします。

解答2)

| 分位点 | 経常収入    | 食料費    |
|-----|---------|--------|
| 20% | 310,531 | 49,701 |
| 40% | 404,883 | 63,146 |
| 60% | 501,181 | 76,055 |
| 80% | 635,875 | 93,036 |

means プロシジャを用いて分位点の算出を行った。



GCHART プロシジャを用いてヒストグラムを作成した。

課題3) 男女別・経常収入および食料費の5分位分布

男女別に経常収入および食料費について、表2に示す5分位別に、以下の表3に示す10万世帯比のクロス表を作成して下さい。

解答3)

表3

| 経常収入 |           | 男      | 女     | 計       |
|------|-----------|--------|-------|---------|
| 5分位  | 5分位数      |        |       |         |
| 1    | 310,531未満 | 15,494 | 4,506 | 20,000  |
| 2    | 310,531以上 | 18,617 | 1,377 | 19,993  |
| 3    | 404,883以上 | 19,099 | 911   | 20,009  |
| 4    | 501,181以上 | 19,356 | 640   | 19,995  |
| 5    | 635,875以上 | 19,435 | 567   | 20,002  |
| 計    |           | 92,000 | 8,000 | 100,000 |

| 食料費 |          | 男      | 女     | 計       |
|-----|----------|--------|-------|---------|
| 5分位 | 5分位数     |        |       |         |
| 1   | 49,701未満 | 16,803 | 3,196 | 19,999  |
| 2   | 49,701以上 | 18,097 | 1,903 | 20,000  |
| 3   | 63,146以上 | 18,734 | 1,268 | 20,001  |
| 4   | 76,055以上 | 19,049 | 948   | 19,998  |
| 5   | 93,036以上 | 19,317 | 685   | 20,003  |
| 計   |          | 92,000 | 8,000 | 100,000 |

tabulate プロシジャを用いて集計した。

課題4) 男女別・経常収入・食料費 についての3重クロス表

男女別・経常収入(5分位)別の食料費(5分位)について、表4に示す3重クロス集計表を作成して下さい。

解答4)

表4

|     | 経常収入<br>5分位 | 食料(5分位) |        |        |        |        | 計       |
|-----|-------------|---------|--------|--------|--------|--------|---------|
|     |             | 1       | 2      | 3      | 4      | 5      |         |
| 男   | 1           | 5,827   | 3,974  | 2,798  | 1,831  | 1,064  | 15,494  |
|     | 2           | 5,130   | 4,501  | 3,978  | 2,996  | 2,012  | 18,617  |
|     | 3           | 2,911   | 4,301  | 4,395  | 4,294  | 3,198  | 19,099  |
|     | 4           | 1,906   | 3,233  | 4,082  | 4,903  | 5,231  | 19,356  |
|     | 5           | 1,029   | 2,087  | 3,481  | 5,026  | 7,813  | 19,435  |
|     | 計           | 16,803  | 18,097 | 18,734 | 19,049 | 19,317 | 92,000  |
| 女   | 1           | 2,462   | 994    | 565    | 314    | 172    | 4,506   |
|     | 2           | 371     | 399    | 235    | 211    | 161    | 1,377   |
|     | 3           | 243     | 225    | 216    | 164    | 63     | 911     |
|     | 4           | 68      | 134    | 156    | 143    | 140    | 640     |
|     | 5           | 53      | 152    | 96     | 117    | 149    | 567     |
|     | 計           | 3,196   | 1,903  | 1,268  | 948    | 685    | 8,000   |
| 男女計 |             | 19,999  | 20,000 | 20,001 | 19,998 | 20,003 | 100,000 |

tabulate プロシジャを用いて集計した。

課題5) 男女別・経常収入別に食料費を常用対数変換した平均値の比較

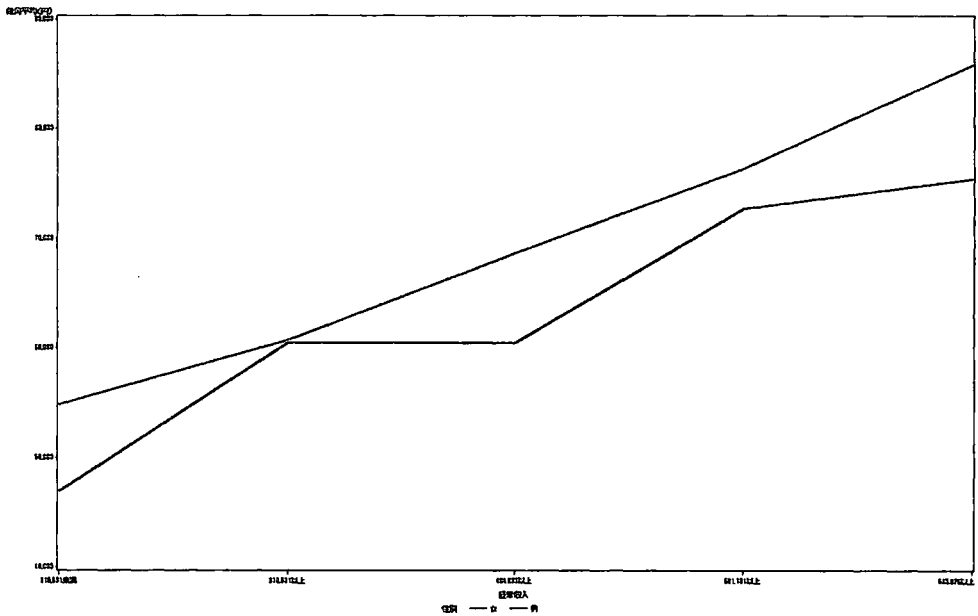
表5に示すように、男女別・経常収入(5分位)別に、食料費の常用対数を計算し平均値を求めて下さい。その平均値について逆対数変換(10<sup>x</sup>)をして幾何平均を計算して下さい。この幾何平均について男女別・経常収入の分位点別に食糧費を比較するために、折れ線グラフを作成して下さい。縦軸を幾何平均(単位:円)、横軸を経常収入(5分位)とします。

解答 5)

表 5

|   | 経常収入 |        | 食料     |        |
|---|------|--------|--------|--------|
|   | 5分位  | 世帯数    | 常用対数平均 | 幾何平均   |
| 男 | 1    | 15,494 | 4.7389 | 54,815 |
|   | 2    | 18,617 | 4.7832 | 60,702 |
|   | 3    | 19,099 | 4.8368 | 68,675 |
|   | 4    | 19,356 | 4.8826 | 76,313 |
|   | 5    | 19,435 | 4.9335 | 85,803 |
|   | 計    | 92,000 | 4.8395 | 69,103 |
| 女 | 1    | 4,506  | 4.6713 | 46,914 |
|   | 2    | 1,377  | 4.7811 | 60,409 |
|   | 3    | 911    | 4.7813 | 60,437 |
|   | 4    | 640    | 4.8617 | 72,728 |
|   | 5    | 567    | 4.8774 | 75,405 |
|   | 計    | 8,000  | 4.7325 | 54,013 |

tabulate プロシジャを用いて集計した。



gplot プロシジャを用いて集計した。



# シングルマザー世帯の家計の実態

## －片働き世帯と共働き世帯との比較から－

参加カテゴリー：C)学部学生・大学院修士課程（社会人院生は除く）

チーム名：中国からの三人娘

賈 慧玲（カ ケイレイ）、鄧 時韻（トウ ジイン）、李 慧蛟（リ ケイジョウ）、  
兵庫県立大学大学院応用情報科学研究科

JIA HUILING, DENG SHIYUN, LI HUIJIAO,  
Graduate School of Applied Informatics, University of Hyogo

### 要約

留学生と同じように厳しい生活が強いられると思われるシングルマザー世帯に焦点を当てて、片働きと共働きの夫婦の揃った普通の子育て世帯と比較をしながら、シングルマザー世帯の家計の実態をとくに食費の観点から、内食と中食と外食の比率に注目して分析を行いました。

キーワード：シングルマザー、エンゲル係数、内食、中食、外食、就業形態

## 1 はじめに

私たちは中国（中華人民共和国）からの留学生で、物価水準の違いから学費や生活費を両親からの仕送りだけでは賅えず、勉学のほかにアルバイトをして、経済的に厳しい生活を送っています。以前から日本人は中国人と比較して裕福な生活を送っていると感じていましたが、新聞やテレビなどの報道を見ると、シングルマザーの家計が私たちと同様に経済的に厳しいだけでなく、子育ても大変だということが分かってきました。シングルマザーが増えた一番の原因として挙げられるのが離婚の増加です。厚生労働省[1]によれば、離婚をした場合、約8割の子供は母親が引き取ると言われていました。今回、2004年の全国消費実態調査の教育用疑似マイクロデータを使って分析ができるチャンスがあると分かった時に、日本のシングルマザー世帯は、一体どんな生活を送っているのかを家計の面から分析してみることに興味を湧きました。

提供された全国消費実態調査の教育用疑似マイクロデータを用いて、JMPの「一変量の分布」を適用して最初に行ったのが性別の構成の確認です。世帯主の性別から見ると、女性は8.26%で10%にも満たないので、調査対象となっている2人以上世帯の中でシングルマザー世帯である可能性が高い世帯は、割合が低いことが分かりました。総務省[4]によれば、2010年の国勢調査で一般世帯数は51,842千世帯であり、そのうち、1

人世帯は 16,785 千世帯なので、2 人以上世帯は 35,057 千世帯となります。この 2 人以上世帯のうち、夫婦と子供から成る世帯は 14,440 千世帯であり、シングルマザー世帯が含まれているひとり親と子供から成る世帯は 4,523 千世帯となっています。また、総務省統計研修所[3]によれば、日本全国のシングルマザー世帯は 108 万世帯ほどであるとされており、近年は増加傾向にあります。

なお、シングルマザーの定義は、総務省統計研修所[3]によれば、「以下の 2 種類のいずれかに該当する世帯の母である。①「母子世帯」の母、「母子世帯」とは、未婚、死別又は離別の母と、その未婚の 20 歳未満の子のみから成る一般世帯をいう。②「上記①の母子と他の世帯員からなる一般世帯」の母、「上記①の母子と他の世帯員からなる一般世帯」とは、例えば、上記①の母子とその祖父母が同居している一般世帯などが該当する。」とされており、実際の分析では、「子と同居で配偶者のいない女性（15～49 歳）」として世帯数を推計し、2010 年のシングルマザーは 1,082 千人、そのうち「母子世帯」の母が 756 千人、「他の世帯員がいる世帯」の母が 326 千人としています。

私たちは、以上のことを調べた上で、シングルマザー世帯の家計の現状、そして夫婦と子供から成る普通の世帯との比較で、食生活を「内食」と「中食」と「外食」に分けてその支出額や構成がどうなっているのかを大学院の講義で使用法を学んだ JMP を使用して調べてみることにしました。

## 2 分析の方法

分析にあたっては、まず最初に、シングルマザー世帯と比較の対象となる夫婦と子供からなる世帯を区別する必要があります。総務省統計研修所[3]では、上述のように「子と同居で配偶者のいない女性（15～49 歳）」と定義して分析をしていますが、私たちが使用する教育用疑似マイクロデータでは、家族の構成員の年齢や続き柄に関する変数は含まれていません。

そこで、2 人以上世帯のデータが提供されているので、性別が女性であれば「他の世帯員がいる世帯」の可能性はありますが、細かいことを気にしては分析ができなくなってしまうので、「母子世帯」でシングルマザー世帯であるとみなすことにしました。ただし、母親の年齢が高齢であれば、私たちが分析の対象としたい「死別あるいは離婚で父親がおらず、母親が働いて高校生以下の子供を育てている」シングルマザー世帯とは言えないので、年齢で対象を絞り込むことにしました。また、有業人員が 2 人以上であれば、母親以外に世帯内で働いている子供あるいは祖父母などがあることになるので、有業人員が 2 人以上の世帯もシングルマザー世帯からは除外することにしました。

さらに、有業人員が一人でも、年金を受給している両親と同居している可能性があるため、このような世帯を排除するために、公的年金給付がゼロの世帯を抽出することにしました。

シングルマザーの年齢については、18歳の高校生の子供がいると考えると、高齢出産も進んでいるので、35歳で子供を産んだとして、53歳のシングルマザーがいる可能性があります。そのため、総務省統計研修所[3]では15～49歳の女性に限定していますが、私たちは54歳までに対象を広げて、年齢が54歳以下で性別が女性の世帯をシングルマザーの世帯と考え、世帯人員が2人であればシングルマザーと子供が1人、世帯人員が3人であればシングルマザーと子供が2人、世帯人員が4人以上であればシングルマザーと子供が3人以上の世帯と考えることにしました。

一方、シングルマザー世帯と比較するために、性別が男性の世帯を夫婦の世帯であるとみなして、公的年金給付がゼロの世帯に限定し、有業人員が1人であれば片働き世帯、有業人員が2人であれば共働き世帯と考えて、世帯人員が3人であれば有業人員に応じて片働き子供1人世帯か共働き子供1人世帯、世帯人員が4人であれば有業人員に応じて片働き子供2人世帯か共働き子供2人世帯、世帯人員が5人であれば有業人員に応じて片働き子供3人以上世帯か共働き子供3人以上世帯というように考えて、世帯を分類することにしました。その結果、私たちが分析の対象としたい子育て世帯は、シングルマザーか夫婦片働きか夫婦共働きかという世帯の勤労形態と、子供が1人か2人か3人以上かという子供の人数で、9つのパターンに分類することになりました。

以上の分類のために使用したのは、図1に示す計算式で、度数に10万世帯比用集計乗率を入れて「一変量の分布」を適用した結果、図2に示すように、子育て世帯を9つの類型に分類することができました。図2からは、本研究での対象となるシングルマザー世帯は882世帯でかなり少ないということが分かりました。

| 水準             | 度数    | 割合      |
|----------------|-------|---------|
| シングルマザー-子供1人   | 328   | 0.00761 |
| シングルマザー-子供2人   | 419   | 0.00972 |
| シングルマザー-子供3人以上 | 135   | 0.00313 |
| 片働き子供1人        | 6355  | 0.14747 |
| 片働き子供2人        | 9884  | 0.22936 |
| 片働き子供3人以上      | 5197  | 0.12060 |
| 共働き子供1人        | 3661  | 0.08495 |
| 共働き子供2人        | 7724  | 0.17923 |
| 共働き子供3人以上      | 6758  | 0.15682 |
| その他            | 2633  | 0.06111 |
| 合計             | 43094 | 1.00000 |

図2 子育て世帯の分類結果

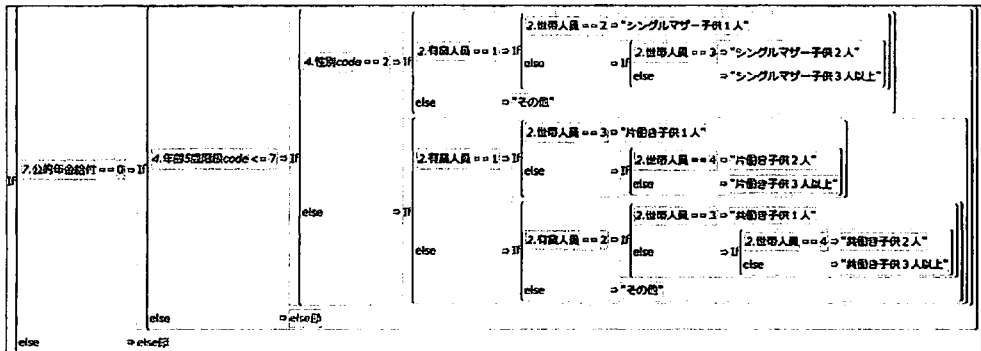


図1 世帯分類のための JMP の計算式



表 1 子供の人数別の経常収入、消費支出、食料費の世帯類型間の平均の差の検定結果

|          | 経常収入   |         | 消費支出  |         | 食料     |         |
|----------|--------|---------|-------|---------|--------|---------|
|          | F値     | p値      | F値    | p値      | F値     | p値      |
| 子供1人世帯   | 183.48 | <0.0001 | 52.25 | <0.0001 | 78.8   | <0.0001 |
| 子供2人世帯   | 295.89 | <0.0001 | 51.59 | <0.0001 | 114.75 | <0.0001 |
| 子供3人以上世帯 | 107.28 | <0.0001 | 13.7  | <0.0001 | 0.41   | 0.66    |

表 2-1 食料の割合(エンゲル係数)

|        | 食料費/経常収入 |        |        |
|--------|----------|--------|--------|
|        | シングルマザー  | 片働き    | 共働き    |
| 子供1人   | 17.37%   | 14.05% | 12.92% |
| 子供2人   | 22.53%   | 15.80% | 14.44% |
| 子供3人以上 | 22.76%   | 15.85% | 13.67% |

表 2-2 消費支出の割合

|        | 消費支出/経常収入 |        |        |
|--------|-----------|--------|--------|
|        | シングルマザー   | 片働き    | 共働き    |
| 子供1人   | 90.26%    | 67.69% | 64.58% |
| 子供2人   | 90.88%    | 67.75% | 63.42% |
| 子供3人以上 | 81.20%    | 68.61% | 63.48% |

帯) から 76 千円 (子供 2 人世帯) の範囲で、世帯類型間の違いは認められません。

ただし、子供の人数別の世帯類型による分散分析の結果をまとめたものが表 1 ですが、子供 3 人以上世帯の食料費を除いては、いずれも F 検定で 1% 有意水準で有意となっており、子供の人数別にみると、子供 1 人世帯、子供 2 人世帯、そして子供 3 人以上世帯のいずれにおいても、シングルマザー世帯、片働き世帯、共働き世帯の世帯類型間で経常収入や消費支出や食料費の平均に差があるという結果が得られています。

一方、消費支出と食料がそれぞれ経常収入に占める割合を計算してまとめたものを表 2-1 と表 2-2 に示しています。これらの表から、シングルマザー世帯は経常収入に占める消費支出の割合が 80% を超えており、特に子供 1 人世帯と子供 2 人世帯では 90% を超えていて、貯蓄はほとんどできず、子供の進学費用の準備もできない状況が示されています。子供の人数に関係なく消費支出の経常収入に占める割合が 60% 台の片働き世帯および共働き世帯と比較して、シングルマザー世帯の生活がいかに厳しいかが窺える結果となっていると言えます。

### 3-2 世帯類型間の就業状況の比較

厚生労働省[1]の平成 23 年度全国母子世帯等調査結果報告によれば、日本のシングルマザーの 2011 年の就業率は 80.6% で、2006 年の 84.5% から 3.9 ポイントの悪化となっています。また、就業形態では、「正規の職員・従業員」が 39.4% であるのに対して「パート・アルバイト等」が 47.4% と最も多くなっており、2006 年の調査と比較して、「パート・アルバイト等」(前回調査では「臨時・パート」)の割合が 3.8 ポイント増加し、「正規の職員・従業員」(前回調査では「常用雇用者」)が 3.1 ポイント減少している状況であるとされています。

さらに、調査されたシングルマザーの中で、母子世帯になる前に不就労であった人のうち、69.1% は就労していますが、25.4% は依然として不就労のままであり、就労して

いる 69.1%のシングルマザーのうち、正規の職員・従業員は 31.1%で、パート・アルバイト等が 57.4%となっており、厳しい就業環境であることが示されています。

2004 年の全国消費実態調査に基づく教育用疑似マイクロデータでこのシングルマザーの厳しい就業環境を確認するために、就業・非就業の別と世帯類型との間で「二変量の関係」を適用してみた結果が図 4 です。図 4 から、教育用疑似マイクロデータの構成世帯はすべて有業者が少なくとも 1 人はいる世帯ですが、片働き世帯と共働き世帯ではほぼ 100%が正規での就業であるのに対して、シングルマザー世帯は非正規のパートとしての就業の比率が高くなっており、就業状態が安定しておらず、その結果として、前項で考察したように経常収入が低い結果につながっていると言えます。

| 行%               | 就業     | 就業 (パート) | 不詳    |
|------------------|--------|----------|-------|
| シングルマザー-子供 1 人   | 66.42  | 33.58    | 0.00  |
| シングルマザー-子供 2 人   | 39.42  | 56.34    | 4.24  |
| シングルマザー-子供 3 人以上 | 30.99  | 40.08    | 28.93 |
| 片働き子供 1 人        | 99.90  | 0.10     | 0.00  |
| 片働き子供 2 人        | 100.00 | 0.00     | 0.00  |
| 片働き子供 3 人以上      | 100.00 | 0.00     | 0.00  |
| 共働き子供 1 人        | 100.00 | 0.00     | 0.00  |
| 共働き子供 2 人        | 99.87  | 0.00     | 0.13  |
| 共働き子供 3 人以上      | 100.00 | 0.00     | 0.00  |
| その他              | 93.87  | 5.38     | 0.74  |

図 4 世帯類型別の就業・非就業の別

### 3-3 世帯類型別の内食と中食と外食の比率の比較

これまでの分析から、シングルマザー世帯、片働き世帯、共働き世帯の世帯類型間で経常収入は大きく異なり、消費支出にも差があるにも関わらず、食料はそれほど大きくは異ならないという結果が得られました。しかし、家事に専業で従事できる主婦のいる片働き世帯と比較すると、家事に時間が多くは割けないシングルマザー世帯と共働き世帯は、家庭内で調理をする内食の比率が低く、外食や中食の比率が高くなることが予想されます。また、シングルマザー世帯は共働き世帯と比較すると経常収入が大幅に低いからです、外食よりは中食の比率が高くなるのではないかと考えられます。

そこで、内食と中食と外食の 3 つの食事形態の支出の割合を、シングルマザー世帯、片働き世帯、共働き世帯の世帯類型別に計算して、その結果をまとめてみました。結果は表 3 に示す通りです。

表 3 から、内食に関しては、共働き世帯は家事に従事できる時間が少なく、調理に割ける時間も短くなるので、他の 2 つの世帯類型と比べて比率が低くなっています。一方、

表3 世帯類型別の3つの食事形態の比率

|    | シングルマザー | 片働き   | 共働き   |
|----|---------|-------|-------|
| 内食 | 61.09   | 61.88 | 58.59 |
| 中食 | 15.73   | 13.05 | 13.73 |
| 外食 | 23.18   | 25.06 | 27.68 |

片働き世帯では、専業主婦のいる可能性が高いので、内食の比率が高くなっています。シングルマザー世帯の場合は、仕事もしながら家事も一人ですることになるので、経済的には一番安い内食で対応したくても、なかなか調理の時間がとれないかもしれません。

経済的な要因を考えると、外食は高くつくので、シングルマザー世帯が外食をする比率は低くなっています。その結果として、中食の比率が一番高いのは、シングルマザー世帯となっています。

以上の分析は、単に比率を比較しただけですが、統計的に3つの食事形態の比率に世帯類型間で違いがあるかどうかを確認するために、「二変量の関係」を適用して、10万世帯比集計用乗数を用いて、一元配置の分散分析を行ってみました。結果は、図5、図6、図7に示す通りです。

これらの分散分析の結果から、いずれも1%有意水準で高度に有意な結果となっており、食料にかかる支出の内訳は、シングルマザー世帯では中食に、片働き世帯では内食

| 分散分析     |       |           |         |          |            |
|----------|-------|-----------|---------|----------|------------|
| 要因       | 自由度   | 平方和       | 平均平方    | F値       | p値(Prob>F) |
| 列228     | 2     | 106993.3  | 53496.7 | 569.1920 | <.0001*    |
| 誤差       | 40458 | 3802499.1 | 94.0    |          |            |
| 全体(修正済み) | 40460 | 3909492.4 |         |          |            |

| 各水準の平均  |         |         |         |        |        |
|---------|---------|---------|---------|--------|--------|
| 水準      | 数       | 平均      | 標準誤差    | 下側95%  | 上側95%  |
| シングルマザー | 881.845 | 61.0898 | 0.32647 | 60.450 | 61.730 |
| 共働き     | 18142.6 | 58.5913 | 0.07198 | 58.450 | 58.732 |
| 片働き     | 21436.3 | 61.8842 | 0.06622 | 61.754 | 62.014 |

図5 内食の比率の世帯類型間の差の検定

| 分散分析     |       |           |         |          |            |
|----------|-------|-----------|---------|----------|------------|
| 要因       | 自由度   | 平方和       | 平均平方    | F値       | p値(Prob>F) |
| 列228     | 2     | 9318.12   | 4659.06 | 219.1370 | <.0001*    |
| 誤差       | 40458 | 860169.28 | 21.26   |          |            |
| 全体(修正済み) | 40460 | 869487.41 |         |          |            |

| 各水準の平均  |         |         |         |        |        |
|---------|---------|---------|---------|--------|--------|
| 水準      | 数       | 平均      | 標準誤差    | 下側95%  | 上側95%  |
| シングルマザー | 881.845 | 15.7340 | 0.15527 | 15.430 | 16.038 |
| 共働き     | 18142.6 | 13.7269 | 0.03423 | 13.660 | 13.794 |
| 片働き     | 21436.3 | 13.0529 | 0.03149 | 12.991 | 13.115 |

図 6 中食の比率の世帯類型間の差の検定

| 分散分析     |       |           |         |          |            |
|----------|-------|-----------|---------|----------|------------|
| 要因       | 自由度   | 平方和       | 平均平方    | F値       | p値(Prob>F) |
| 列228     | 2     | 75620.3   | 37810.1 | 406.6477 | <.0001*    |
| 誤差       | 40458 | 3761760.0 | 93.0    |          |            |
| 全体(修正済み) | 40460 | 3837380.3 |         |          |            |

| 各水準の平均  |         |         |         |        |        |
|---------|---------|---------|---------|--------|--------|
| 水準      | 数       | 平均      | 標準誤差    | 下側95%  | 上側95%  |
| シングルマザー | 881.845 | 23.1761 | 0.32471 | 22.540 | 23.813 |
| 共働き     | 18142.6 | 27.6819 | 0.07159 | 27.542 | 27.822 |
| 片働き     | 21436.3 | 25.0629 | 0.06586 | 24.934 | 25.192 |

図 7 外食の比率の世帯類型間の差の検定

に、共働き世帯では外食にかかる比率が他の世帯類型と比較して相対的に高くなっているという私たちの仮説が支持されたと言えます。

また、より詳しく分析するために、子供の人数別にこれまでの分析を行った結果を図8から図10にまとめてみました。

これらの図からは、子供1人世帯と子供2人世帯とは同じような食事形態の比率が見られますが、子供3人以上世帯では、少し異なった比率となっていることが分かります。特に、内食の比率がシングルマザー世帯で片働き世帯よりも高くなっており、中食の比率は世帯類型の中ではもっとも低くなっています。使用しているデータは2004年の全国消費実態調査に基づくもので、第1子と第2子には月額5千円、第3子以降は月額1万円という当時の児童手当の影響があるのかもしれませんが、他には、家事の手伝いができる年齢の高い子供がいるといった家族構成の影響があるのかもしれないと考えられます。

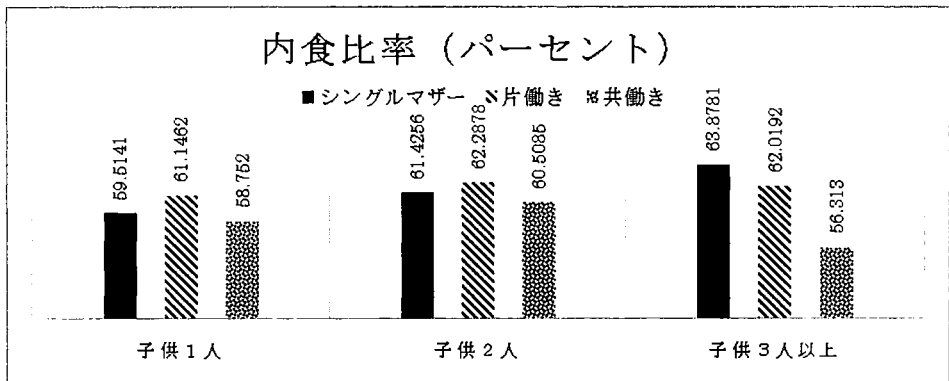


図 8 内食の比率の相違



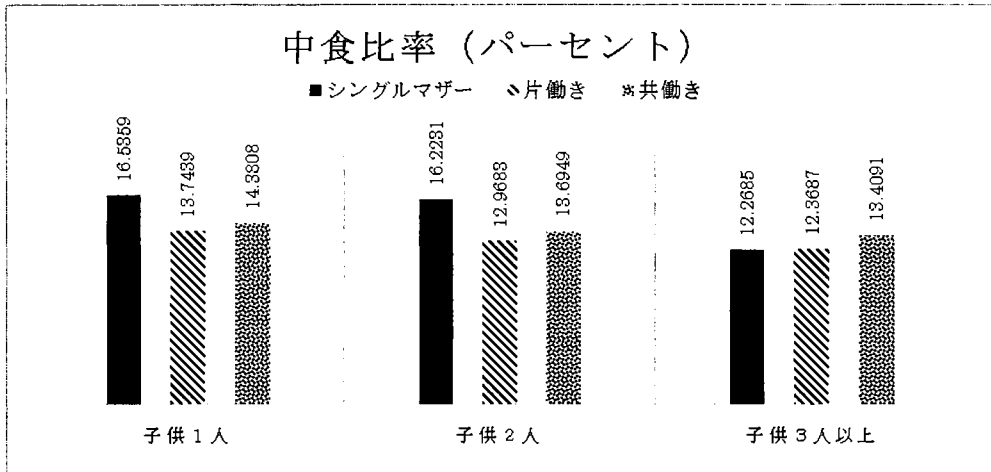


図 9 中食の比率の相違

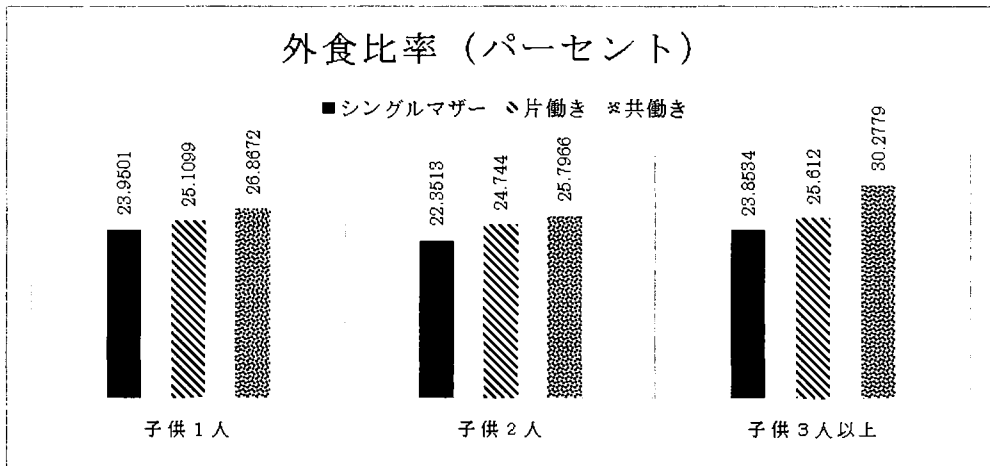


図 10 外食の比率の相違

## 4 おわりに

私たちの自由課題の研究では、シングルマザー世帯の家計の実態を明らかにすることに焦点を当てて、食料を「内食」と「中食」と「外食」という三つの食事形態に分けて、シングルマザー世帯と片働きおよび共働きの普通の子育て世帯との支出構造を比較してみました。さらに、消費支出や経常収入の状況や、就業状態についても分析を行いました。

内閣府男女共同参画局[2]によると、育児や介護などで仕事を中断しやすい女性、もしくはシングルマザーは、低収入で不安定な雇用につきやすい就業構造があると言われ

ています。そして、そのような低収入で不安定な就業形態の積み重ねから年金水準は低くなり、高齢期においても経済的基盤が弱くなることが指摘されています。私たち留学生も、本国からの仕送りだけでは学費や生活費を賄うことは厳しく、パート・アルバイトなどの就業形態で働きながら勉強を続ける必要があり、時間的余裕はほとんどありません。その結果、シングルマザー世帯と同じような食事形態となっているように思います。

今、貧困の連鎖が社会的問題となっていますが、今回の研究でシングルマザー世帯は普通の子育て世帯と比較して厳しい家計のもとで生活していることが明らかになりました。私たちは、あともう少しの期間を頑張って勉強を続けて社会人となれば、厳しい生活から離れることができると思いますが、シングルマザー世帯は大変だと思います。今後は、児童手当や奨学金の充実などで、シングルマザーが少しでも余裕をもって子育てができればよいと感じました。

#### 参考文献

- [1]厚生労働省, 平成 23 年度全国母子世帯等調査結果報告 ([http://www.mhlw.go.jp/seisakunitsuite/bunya/kodomo/kodomo\\_kosodate/boshi-katei/boshi-setai\\_h23/](http://www.mhlw.go.jp/seisakunitsuite/bunya/kodomo/kodomo_kosodate/boshi-katei/boshi-setai_h23/)), 2016 年 6 月 16 日最終閲覧.
- [2]内閣府男女共同参画局, 男女共同参画白書 (概要版) 平成 22 年版 ([http://www.gender.go.jp/whitepaper/h22/gaiyou/html/honpen/b1\\_s05.html](http://www.gender.go.jp/whitepaper/h22/gaiyou/html/honpen/b1_s05.html)), 2016 年 6 月 16 日最終閲覧.
- [3]総務省統計研修所, 「シングル・マザーの最近の状況 (2010 年)」, 2012 (<http://www.stat.go.jp/training/2kenkyu/pdf/zuhyou/single4.pdf>), 2016 年 6 月 16 日最終閲覧.
- [4]総務省, 「平成 22 年国勢調査人口等基本集計結果 結果の概要」 (<http://www.stat.go.jp/data/kokusei/2010/kihon1/pdf/gaiyou1.pdf>), 2016 年 6 月 16 日最終閲覧.

## 規定課題

参加カテゴリー：C)学部学生・大学院修士課程（社会人院生は除く）

チーム名：中国からの三人娘

賈 慧玲（カ ケイレイ）、鄧 時韻（トウ ジイン）、李 慧蛟（リ ケイジョウ）、  
兵庫県立大学大学院応用情報科学研究科

JIA HUILING, DENG SHIYUN, LI HUIJIAO,  
Graduate School of Applied Informatics, University of Hyogo

### 問題 1

「産業符号」という列の名前を右クリックして、「列プロパティ」の中の「値の順序」をクリックする。ここで、レポートに表示する順番に産業符号の水準を並べ替える。また、「性別」の方も同じやり方で順序を変更する。

そして、「二変量の関係」を適用する。説明変数に「産業符号」を、目的変数に「性別」を設定して、「OK」をクリックすると、以下の単純クロス表が得られる。

4.性別

| 度数                | 男     | 女    |       |
|-------------------|-------|------|-------|
| 鉱業                | 3     | 0    | 3     |
| 建設業               | 2134  | 9    | 2143  |
| 製造業               | 5668  | 173  | 5841  |
| 電気・ガス・熱供給・水道業     | 222   | 0    | 222   |
| 情報通信業             | 547   | 0    | 547   |
| 運輸業               | 1235  | 3    | 1238  |
| 卸売・小売業            | 2083  | 201  | 2284  |
| 金融・保険業            | 472   | 39   | 511   |
| 不動産業              | 40    | 0    | 40    |
| 飲食店・宿泊業           | 50    | 32   | 82    |
| 医療・福祉             | 371   | 270  | 641   |
| 教育・学習支援業          | 1059  | 89   | 1148  |
| 複合サービス事業          | 321   | 3    | 324   |
| サービス業（他に分類されないもの） | 1816  | 106  | 1922  |
| 公務（他に分類されないもの）    | 3656  | 130  | 3786  |
| 不詳                | 9704  | 1591 | 11295 |
|                   | 29381 | 2646 | 32027 |

次に、新しい列として「10万世帯比集計用乗率」を作る。計算式を使って、

「集計用乗率」 \* 100000/495465

を入力すれば、10万世帯比集計用乗率が計算される。

もう一回、最初と同様の「二変量の関係」を適用するが、今回は「10万世帯比集計用乗

率」を「重み」のところに入れることで、以下の10万世帯比のクロス表が得られる。

#### 4.性別

| 度数                | 男     | 女    |        |
|-------------------|-------|------|--------|
| 鉱業                | 7     | 0    | 7      |
| 建設業               | 6346  | 31   | 6377   |
| 製造業               | 17837 | 510  | 18347  |
| 電気・ガス・熱供給・水道業     | 697   | 0    | 697    |
| 情報通信業             | 2020  | 0    | 2020   |
| 運輸業               | 3852  | 6    | 3858   |
| 卸売・小売業            | 6644  | 589  | 7233   |
| 金融・保険業            | 1447  | 104  | 1551   |
| 不動産業              | 144   | 0    | 144    |
| 飲食店・宿泊業           | 158   | 107  | 265    |
| 医療・福祉             | 1022  | 810  | 1832   |
| 教育・学習支援業          | 3110  | 269  | 3379   |
| 複合サービス事業          | 981   | 11   | 992    |
| サービス業（他に分類されないもの） | 5842  | 326  | 6167   |
| 公務（他に分類されないもの）    | 10691 | 430  | 11121  |
| 不詳                | 31204 | 4806 | 36011  |
|                   | 92000 | 8000 | 100000 |

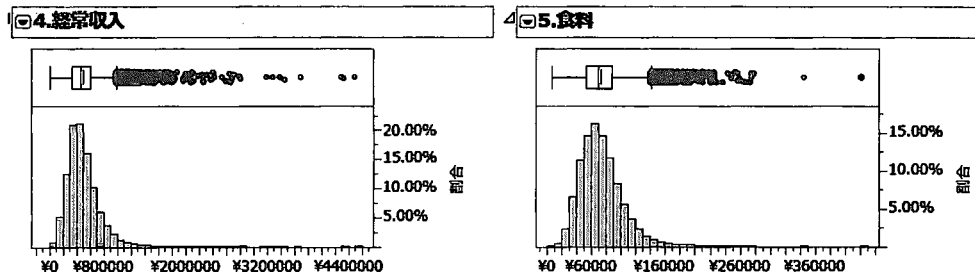
#### 問題 2

「一変量の分布」を適用する。Y列に求める「経常収入」と「食料」を入れて、度数に「集計用乗率」を入れる。「表示オプション」の「分位点の間隔の設定」で0.2を設定すれば、5分位点が以下のように示される。

| 4.経常収入 |             | 5.食料 |             |
|--------|-------------|------|-------------|
| 分位点    |             | 分位点  |             |
| 100%   | 最大値 4501874 | 100% | 最大値 422939  |
| 80%    | 635875      | 80%  | 93035.8     |
| 60%    | 501181      | 60%  | 76055.1     |
| 40%    | 404883      | 40%  | 63146.2     |
| 20%    | 310531      | 20%  | 49701       |
| 0%     | 最小値 0       | 0%   | 最小値 5976.84 |

また、「ヒストグラムオプション」で、選択されている「縦に表示」を取り消し、「割合軸」をチェックして、「Y軸の設定」の「形式」に「パーセント」を選び、「X軸の設定」の

「形式」に「通貨」を選ぶことで、課題のヒストグラムが得られる。



### 問題 3

新しい列「10万世帯比経常収入」と「10万世帯比食料費」を作る。計算式の IF 関数を使って、以下のように5分位階級をそれぞれ作成し、新しい列に計算結果を入れる。

図 5

The left screenshot shows the formula editor for '10万世帯経常収入' with the following IF statements:

- $4. \text{経常収入} < 310531 \Rightarrow 1$
- $310531 \leq 4. \text{経常収入} < 404883 \Rightarrow 2$
- If  $404883 \leq 4. \text{経常収入} < 501181 \Rightarrow 3$
- $501181 \leq 4. \text{経常収入} < 635875 \Rightarrow 4$
- $635875 \leq 4. \text{経常収入} < 4501875 \Rightarrow 5$

The right screenshot shows the formula editor for '10万世帯食料費' with the following IF statements:

- $5. \text{食料} < 49701 \Rightarrow 1$
- $49701 \leq 5. \text{食料} < 63146.2 \Rightarrow 2$
- If  $63146.2 \leq 5. \text{食料} < 76055.1 \Rightarrow 3$
- $76055.1 \leq 5. \text{食料} < 93035.8 \Rightarrow 4$
- $5. \text{食料} < 422940 \Rightarrow 5$

「二変量の関係」を分析する。説明変数に「10万世帯比経常収入」と「10万世帯比食料費」を、目的変数に「性別」を設定して、「10万世帯比集計用乗率」を「重み」のところにいれると、求めたいクロス表が得られる。

④ 10万世帯経常収入と4.性別の分割表に対する分析

重み: 10万世帯集計用乗率

▷モザイク図

④分割表

|           |   | 4.性別  |      |        |  |
|-----------|---|-------|------|--------|--|
|           |   | 度数    | 男    | 女      |  |
| 10万世帯経常収入 | 1 | 15494 | 4506 | 20000  |  |
|           | 2 | 18617 | 1377 | 19993  |  |
|           | 3 | 19099 | 911  | 20009  |  |
|           | 4 | 19356 | 640  | 19995  |  |
|           | 5 | 19435 | 567  | 20002  |  |
|           | 計 | 92000 | 8000 | 100000 |  |

④ 10万世帯食料費と4.性別の分割表に対する分析

重み: 10万世帯集計用乗率

▷モザイク図

④分割表

|          |   | 4.性別  |      |        |  |
|----------|---|-------|------|--------|--|
|          |   | 度数    | 男    | 女      |  |
| 10万世帯食料費 | 1 | 16803 | 3196 | 19999  |  |
|          | 2 | 18097 | 1903 | 20000  |  |
|          | 3 | 18734 | 1268 | 20001  |  |
|          | 4 | 19049 | 948  | 19998  |  |
|          | 5 | 19317 | 685  | 20003  |  |
|          | 計 | 92000 | 8000 | 100000 |  |

以上の結果に基づき、エクセルで作表して課題のクロス表が得られる。

表 1

経常収入

食料費

| 5分位 | 5分位数       | 男      | 女     | 計       | 5分位 | 5分位数        | 男      | 女     | 計       |
|-----|------------|--------|-------|---------|-----|-------------|--------|-------|---------|
| 1   | 310,531 未満 | 15494  | 4506  | 20,000  | 1   | 49,701 未満   | 16803  | 3196  | 19,999  |
| 2   | 310,531 以上 | 18617  | 1377  | 19,993  | 2   | 49,702 以上   | 18097  | 1903  | 20,000  |
| 3   | 404,883 以上 | 19099  | 911   | 20,009  | 3   | 63,146.2 以上 | 18734  | 1268  | 20,001  |
| 4   | 501,181 以上 | 19356  | 640   | 19,995  | 4   | 76,055.1 以上 | 19049  | 948   | 19,998  |
| 5   | 635,875 以上 | 19435  | 567   | 20,002  | 5   | 93,035.8 以上 | 19317  | 685   | 20,003  |
|     | 計          | 92,000 | 8,000 | 100,000 |     | 計           | 92,000 | 8,000 | 100,000 |

問題 4

「二変数の関係」を適用する。説明変数に「10万世帯比経常収入」を、目的変数に「10万世帯比食料費」を設定して、「10万世帯比集計用乗率」を「重み」に設定して、「性別」を「By」に設定することで、以下の結果が得られる。

④ 10万世帯経常収入と10万世帯食料費の分割表に対する分析

4.性別=女

重み: 10万世帯集計用乗率

▷モザイク図

④分割表

|           |   | 10万世帯食料費 |      |      |     |     |      |
|-----------|---|----------|------|------|-----|-----|------|
|           |   | 度数       | 1    | 2    | 3   | 4   | 5    |
| 10万世帯経常収入 | 1 | 2462     | 994  | 565  | 314 | 172 | 4506 |
|           | 2 | 371      | 399  | 235  | 211 | 161 | 1377 |
|           | 3 | 243      | 225  | 216  | 164 | 63  | 911  |
|           | 4 | 68       | 134  | 156  | 143 | 140 | 640  |
|           | 5 | 53       | 152  | 96   | 117 | 149 | 567  |
|           | 計 | 3196     | 1903 | 1268 | 948 | 685 | 8000 |

④ 10万世帯経常収入と10万世帯食料費の分割表に対する分析

4.性別=男

重み: 10万世帯集計用乗率

▷モザイク図

④分割表

|           |   | 10万世帯食料費 |       |       |       |       |       |
|-----------|---|----------|-------|-------|-------|-------|-------|
|           |   | 度数       | 1     | 2     | 3     | 4     | 5     |
| 10万世帯経常収入 | 1 | 5827     | 3974  | 2798  | 1831  | 1064  | 15494 |
|           | 2 | 5130     | 4501  | 3978  | 2996  | 2012  | 18617 |
|           | 3 | 2911     | 4301  | 4395  | 4294  | 3198  | 19099 |
|           | 4 | 1906     | 3233  | 4082  | 4903  | 5231  | 19356 |
|           | 5 | 1029     | 2087  | 3481  | 5026  | 7813  | 19435 |
|           | 計 | 16803    | 18097 | 18734 | 19049 | 19317 | 92000 |

以上の結果を用いてエクセルで作表することで、課題の3重クロス表が得られる。

表 2

|     | 經常収入<br>5 分位 | 食料(5 分位) |        |        |        |        | 計       |
|-----|--------------|----------|--------|--------|--------|--------|---------|
|     |              | 1        | 2      | 3      | 4      | 5      |         |
| 男   | 1            | 5827     | 3974   | 2798   | 1831   | 1064   | 15494   |
|     | 2            | 5130     | 4501   | 3978   | 2996   | 2012   | 18617   |
|     | 3            | 2911     | 4301   | 4395   | 4294   | 3198   | 19099   |
|     | 4            | 1906     | 3233   | 4082   | 4903   | 5231   | 19356   |
|     | 5            | 1029     | 2087   | 3481   | 5026   | 7813   | 19435   |
|     | 計            | 16803    | 18097  | 18734  | 19049  | 19317  | 92000   |
| 女   | 1            | 2462     | 994    | 565    | 314    | 172    | 4506    |
|     | 2            | 371      | 399    | 235    | 211    | 161    | 1377    |
|     | 3            | 243      | 225    | 216    | 164    | 63     | 911     |
|     | 4            | 68       | 134    | 156    | 143    | 140    | 640     |
|     | 5            | 53       | 152    | 96     | 117    | 149    | 567     |
|     | 計            | 3196     | 1903   | 1268   | 948    | 685    | 8000    |
| 男女計 |              | 19,999   | 20,000 | 20,001 | 19,998 | 20,003 | 100,000 |

問題 5

新しい「常用対数食料費」という名前の列を作る。計算式で超越関数の  $\text{Log}_{10}$  を使って、 $\text{Log}_{10}[\text{食料費}]$  を入力して、食料の常用対数を求める。

続いて、まず、「二変量の関係」を適用する。説明変数に「10 万世帯比經常収入」を、目的変数に「10 万世帯比食料費」を設定して、「10 万世帯比集計用乗率」を「重み」のところにに入れて、「性別」を「By」に入れて分析を実行すると、男性と女性のそれぞれで以下のような分割表が得られる。この分割表から世帯数が分かる。

| 男性       |       | 女性       |       |       |       |       |
|----------|-------|----------|-------|-------|-------|-------|
| 分割表      |       | 分割表      |       |       |       |       |
| 10万世帯食料費 |       | 10万世帯食料費 |       |       |       |       |
| 度数       | 1     | 2        | 3     | 4     | 5     |       |
| 1        | 5827  | 3974     | 2798  | 1831  | 1064  | 15494 |
| 2        | 5130  | 4501     | 3978  | 2996  | 2012  | 18617 |
| 3        | 2911  | 4301     | 4395  | 4294  | 3198  | 19099 |
| 4        | 1906  | 3233     | 4082  | 4903  | 5231  | 19356 |
| 5        | 1029  | 2087     | 3481  | 5026  | 7813  | 19435 |
|          | 16803 | 18097    | 18734 | 19049 | 19317 | 92000 |

| 度数 | 1    | 2    | 3    | 4   | 5   |      |
|----|------|------|------|-----|-----|------|
| 1  | 2462 | 994  | 565  | 314 | 172 | 4506 |
| 2  | 371  | 399  | 235  | 211 | 161 | 1377 |
| 3  | 243  | 225  | 216  | 164 | 63  | 911  |
| 4  | 68   | 134  | 156  | 143 | 140 | 640  |
| 5  | 53   | 152  | 96   | 117 | 149 | 567  |
|    | 3196 | 1903 | 1268 | 948 | 685 | 8000 |

次に、もう一度、「二変量の関係」を適用する。ただし、今回は目的変数を「10 万世帯比食料費」に代えて「常用対数食料費」に変更する。これで、多少の誤差がありますが、5 分

位別に、食料費の常用対数の平均値が以下のように得られます。

| 男性     |         |         |         |        |        | 女性     |         |         |         |        |        |
|--------|---------|---------|---------|--------|--------|--------|---------|---------|---------|--------|--------|
| 各水準の平均 |         |         |         |        |        | 各水準の平均 |         |         |         |        |        |
| 水準     | 数       | 平均      | 標準誤差    | 下側95%  | 上側95%  | 水準     | 数       | 平均      | 標準誤差    | 下側95%  | 上側95%  |
| 1      | 15493.7 | 4.73889 | 0.00121 | 4.7365 | 4.7413 | 1      | 4505.93 | 4.67130 | 0.00252 | 4.6664 | 4.6762 |
| 2      | 18616.7 | 4.78322 | 0.00110 | 4.7811 | 4.7854 | 2      | 1376.78 | 4.78107 | 0.00456 | 4.7721 | 4.7900 |
| 3      | 19098.6 | 4.83680 | 0.00109 | 4.8347 | 4.8389 | 3      | 910.748 | 4.78132 | 0.00560 | 4.7703 | 4.7923 |
| 4      | 19355.8 | 4.88259 | 0.00108 | 4.8805 | 4.8847 | 4      | 639.698 | 4.86171 | 0.00668 | 4.8486 | 4.8748 |
| 5      | 19435.4 | 4.93345 | 0.00108 | 4.9313 | 4.9356 | 5      | 566.622 | 4.87741 | 0.00710 | 4.8635 | 4.8913 |

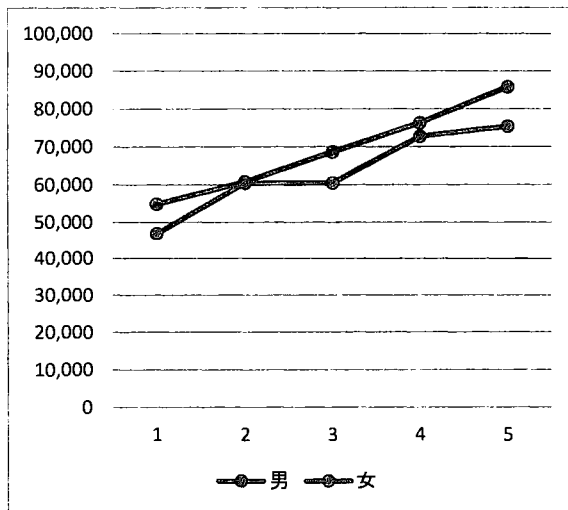
この男女別で5分位階層別の常用対数食料費平均を用いて

$10^{\wedge}$  (常用対数食料費平均)

を計算することで、求めたい幾何平均が次のように得られます。

|   | 経常収入<br>5分位 | 食料     |        |        |
|---|-------------|--------|--------|--------|
|   |             | 世帯数    | 常用対数平均 | 幾何平均   |
| 男 | 1           | 15,494 | 4.7389 | 54,814 |
|   | 2           | 18,617 | 4.7832 | 60,704 |
|   | 3           | 19,099 | 4.8368 | 68,675 |
|   | 4           | 19,356 | 4.8826 | 76,312 |
|   | 5           | 19,435 | 4.9335 | 85,793 |
|   | 計           | 92,000 | 4.8395 | 69,103 |
| 女 | 1           | 4,506  | 4.6713 | 46,914 |
|   | 2           | 1,377  | 4.7811 | 60,405 |
|   | 3           | 911    | 4.7813 | 60,439 |
|   | 4           | 640    | 4.8617 | 72,729 |
|   | 5           | 567    | 4.8774 | 75,407 |
|   | 計           | 8,000  | 4.7325 | 54,013 |

求められているグラフは、以下のようになります。





# 世代別及び経常収入別に見た官民の平均貯蓄率の差異に関する考察

田中健一郎

兵庫県立大学大学院 応用情報科学研究科  
参加カテゴリーB

## 要約

官公庁に努める人（官公）と民間企業に勤める人（民間）で、貯蓄率にどのような違いが出るのかを考察するために、教育用疑似マイクロデータを用いて、統計解析ソフト JMP により分析を行った。官公と民間とでは、明らかに官公の方が、平均貯蓄率が高い結果となったが、年齢 5 歳階級別および経常収入の 5 つの階層別に改めて検定を行うと、平均貯蓄率の差は、一概にどちらが高いと言えない結果となった。また官公と民間の共通点として、ともに平均貯蓄率のピークは 35 歳～39 歳にあり、40 歳以降は減少し、定年前の 55 歳～59 歳でようやく回復することがわかった。婚姻していると考えられる 2 人以上世帯では、子供の教育費がかからない貯められるときに貯めておくことが肝要ということがわかった。

キーワード：教育用疑似マイクロデータ、平均貯蓄率、公務員、官公、民間

## 1. はじめに

公務員は、労働三権（団結権、団体交渉権、団体行動権）の制約がある一方で、身分の保証があり、所得は安定している。そういった国や自治体を就職先として選んだ公務員は非公務員に比べリスク回避的であると推測できる。

リスク回避的であるとした場合、それは貯蓄行動に表れ、質素儉約に努め、貯蓄率が高くなるのではないかと考えられる。しかし逆に、安定した所得が得られるのであれば、将来の不安は少なくなるので、貯蓄は少なくともよいと考え、貯蓄率が低くなるとも考えられる。

官民で貯蓄率に差があるかどうかが本研究のテーマであるが、このテーマを選んだ理由は、筆者自身が公務員であること、そして、同じ公務員で、同じ年代、同じ収入階層の勤労者がどれくらいの額を貯蓄に回しているのか、官民で貯蓄額に違いがでるのかどうかということに興味を持ったことによる。

以下では官民の貯蓄率の差異に関する考察を行う。

## 2. 平均貯蓄率の求め方

総務省統計局[1]の全国消費実態調査の定義によると、平均貯蓄率とは、可処分所得に対する貯蓄純増の割合のことであり、計算式で表すと下記ようになる。

$$\text{平均貯蓄率} = \left( (\text{預貯金} - \text{預貯金引出}) + (\text{保険料} - \text{保険金}) \right) / \text{可処分所得} \times 100$$

可処分所得とは実収入から税金、社会保険料などの非消費支出を差し引いた額で、いわゆる手取り収入のことを指す。

何をもって官と民を分けるかであるが、教育用疑似マイクロデータの変数として利用可能な世帯主の企業区分を用いる。以降ではこの区分名に従って官を官公、民を民間と記すこととする。

官公・民間の 2 群間の平均の差を検定したいので、まず企業区分不詳は分析の対象から外すこととした。そして、今回は世帯ごとの貯蓄率の差を調べたいことから、群ごとに総預貯金や総可処分所得などを求めてから貯蓄率の平均を算出するのではなく、世帯ごとの貯蓄率を求めてから、それを群ごとに平均したものを、群ごとの平均貯蓄率とすることとした。

その際、計算式の性質上、絶対値が 100% を大きく超えることがあり、そのような特異な例が全体に与える影響が極めて大きくなることから絶対値の大きい上下 0.5% を外れ値として分析の対象から除外することとした。具体的には、図 1 に示すように、JMP の「一変量の分布」を適用して得られた 99.5% 点と 0.5% 点の値から、分析の

| 分位点    |               |
|--------|---------------|
| 100.0% | 最大値 68221.8%  |
| 99.5%  | 80.4%         |
| 97.5%  | 58.8%         |
| 90.0%  | 43.5%         |
| 75.0%  | 四分位点 29.6%    |
| 50.0%  | 中央値 13.1%     |
| 25.0%  | 四分位点 -7.9%    |
| 10.0%  | -38.3%        |
| 2.5%   | -102.9%       |
| 0.5%   | -253.3%       |
| 0.0%   | 最小値 -44321.1% |

図 1 貯蓄率の分位点

対象とする貯蓄率の上限を 80.4%、下限を-253.3%として、この上限と下限から外れる世帯を除外した。

### 3. 貯蓄率の分布

分析の手始めとして、官公・民営合わせた全体（不詳除く）の貯蓄率の分布を表示する。なお、分析にあたっては、規定課題で用いた 10 万世帯比用乗率を用いて、10 万世帯当たりで分析を行うこととした。

不詳と外れ値を除いた 98,815 世帯を対象として、一変量の分布で列に貯蓄率、10 万世帯比用乗率を度数に設定し、ヒストグラムにしたものが図 2 である。

図 2 から、貯蓄率は平均値付近に世帯が収束する正規分布のように見えるが、JMP の一変量の分布から連続分布のあてはめ → 正規 → 適合度 → Kolmogorov-Smirnov-Lilliefors の検定を行ってみたところ p 値は 0.01 となり、この分布は正規分布であるという帰無仮説は棄却され、正規分布ではないことがわかる。

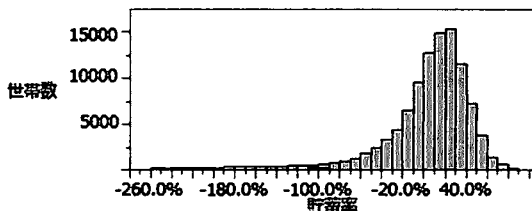


図 2 貯蓄率別世帯数のヒストグラム

### 4. 官公と民営の平均貯蓄率の差の検定

本研究の目的は、官公と民間との間で貯蓄率に違いがあるかどうか、あるとすればどのような違いであるかを確認することである。そこで、まず、単純に、官公庁に勤める人と、民営企業に勤める人の貯蓄率の平均に統計的な有意差があるかどうかを調べることにした。

JMP の「2 変量の関係」で、貯蓄率を目的変数に、企業区分を説明変数に、10 万世帯比用の集計用乗率を度数に入れて分析を実行し、結果画面のメニューから、「平均/ANOVA/プーリングした t 検定」を選択することで、図 3 の結果を得た。

図 3 より、平均値は官公が 9.81%で、民営が 5.36%であり、分散分析の F 値は 220.78 で p 値は <0.0001 であることから、1%有意水準で官公と民間の貯蓄率には有意差があることが分かった。

しかし、貯蓄率は年齢や収入に応じて変わると思われる。したがって、次に、年齢や収入が同じ場合に、官公と民間の 2 つの集団の貯蓄率に違いが出るかどうかの検定を行う。

| 分散分析     |       |           |         |          |             |
|----------|-------|-----------|---------|----------|-------------|
| 要因       | 自由度   | 平方和       | 平均平方    | F 値      | p 値(Prob>F) |
| 4.企業区分   | 1     | 31.140    | 31.1397 | 220.7765 | <.0001*     |
| 誤差       | 98813 | 13937.199 | 0.1410  |          |             |
| 全体(修正済み) | 98814 | 13968.339 |         |          |             |

| 各水準の平均 |         |          |         |         |         |
|--------|---------|----------|---------|---------|---------|
| 水準     | 数       | 平均       | 標準誤差    | 下側95%   | 上側95%   |
| 官公     | 19525.2 | 0.098148 | 0.00269 | 0.09288 | 0.10342 |
| 民営     | 79289.7 | 0.053565 | 0.00133 | 0.05095 | 0.05618 |

図 3 官公と民営の平均貯蓄率

### 5. 5 分位別経常収入および年齢 5 歳階級別の官・民の分布

JMP で「グラフの作成」を選択し、X に規定課題で作成した 5 分位別経常収入、重ね合わせに企業区分、度数に規定課題で作成した 10 万世帯用乗率を設定して作成したグラフが図 4 である。

一方、JMP で同じく「グラフの作成」を選択し、X に年齢 5 歳階級別、重ね合わせに企業区分、度数に 10 万世帯用乗率を設定して作成したグラフが図 5 である。

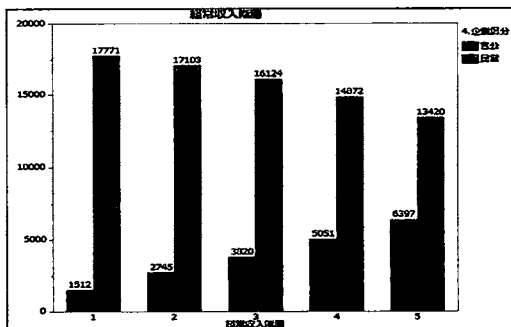


図 4 官公・民間の経常収入 5 階層別世帯数

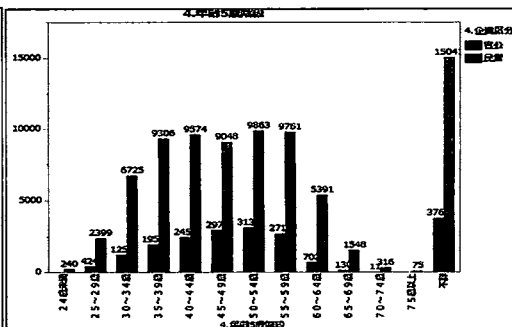


図 5 官公・民間の年齢 5 歳階級別世帯数

図4と図5から、経常収入5階層別と年齢5歳階級別では、それぞれの分布に特徴があることがわかる。集計用乗率で補正されているとは言え、全国消費実態調査の回答者には偏りがあるので、グラフに示されたデータの特徴が官民の年齢や経常収入の分布を一概に正確に表しているとはいいがたいが、グラフからは、公務員は経常収入が多い世帯の割合が高く、若者が少なく、50歳～54歳が最も多くなり、60歳を超えると急激に減少しているという特徴が挙げられる。

### 6. 経常収入5分位階層別の平均貯蓄率の差の検定

JMPで「一変量の分布」を選択し、Y列に経常収入と食料、度数に10万世帯用集計用乗率をセットし、5分位点を求め、第1の5分位階級として310,531円未満、第2の5分位階級として310,531円以上404,883円未満、第3の5分位階級として404,883円以上501,181円未満、第4の5分位階級として501,181円以上635,875円未満、第5の5分位階級として635,875円以上の5つの階層に世帯を分ける。

続いて、JMPで「2変量の関係」を選択し、貯蓄率を目的変数に、経常収入の階層を説明変数に、10万世帯用の集計用乗率を度数に入れて一元配置プロット図を表示させる。その上で、メニューから平均/ANOVA/プーリングしたt検定を選択して得られた経常収入5分位階層別の平均貯蓄率が図6である。

図6より、経常収入の5分位階層が大きい（経常収入が多い）ほど平均貯蓄率が高くなっている。そして、一元配置分析のF検定により、F値は3617.66でp値は<0.0001であり、1%有意水準で経常収入は平均貯蓄率に有意差をもたらしていることが判明した。

さらに Tukey-Kramer 法による多重比較を行うと、図7に示す結果が得られ、図7より、すべての経常収入5分位階層間で、有意差があることが認められる。差の順位レポートでの水準間（5分位階層間）での平均貯蓄率の差を見れば、経常収入の多寡が貯蓄率に与える影響はかなり大きいことがわかる。

| 分散分析     |       |           |         |          |            |
|----------|-------|-----------|---------|----------|------------|
| 要因       | 自由度   | 平方和       | 平均平方    | F値       | p値(Prob>F) |
| 経常収入階層   | 4     | 1784.336  | 446.084 | 3617.657 | <.0001*    |
| 誤差       | 98810 | 12184.003 | 0.123   |          |            |
| 全体(修正済み) | 98814 | 13968.339 |         |          |            |

| 各水準の平均 |         |          |         |         |         |
|--------|---------|----------|---------|---------|---------|
| 水準     | 数       | 平均       | 標準偏差    | 下側95%   | 上側95%   |
| 1      | 19282.9 | -0.19082 | 0.00253 | -0.1958 | -0.1859 |
| 2      | 19848.3 | 0.04380  | 0.00249 | 0.0389  | 0.0487  |
| 3      | 19944   | 0.10812  | 0.00249 | 0.1032  | 0.1130  |
| 4      | 19922.5 | 0.14613  | 0.00249 | 0.1413  | 0.1510  |
| 5      | 19817.3 | 0.19712  | 0.00249 | 0.1922  | 0.2020  |

図6 経常収入5分位階層別の平均貯蓄率

| 差の順位レポート |    |           |           |           |           |         |
|----------|----|-----------|-----------|-----------|-----------|---------|
| 水準       | 水準 | 差         | 差の標準偏差    | 下側信頼限界    | 上側信頼限界    | p値      |
| 5        | 1  | 0.3879406 | 0.0035520 | 0.3782513 | 0.3976299 | <.0001* |
| 4        | 1  | 0.3369524 | 0.0035474 | 0.3272757 | 0.3466291 | <.0001* |
| 3        | 1  | 0.2989395 | 0.0035465 | 0.2892654 | 0.3086136 | <.0001* |
| 2        | 1  | 0.2346207 | 0.0035507 | 0.2249351 | 0.2443062 | <.0001* |
| 5        | 2  | 0.1533200 | 0.0035263 | 0.1437009 | 0.1629391 | <.0001* |
| 4        | 2  | 0.1023318 | 0.0035216 | 0.0927254 | 0.1119381 | <.0001* |
| 5        | 3  | 0.0890011 | 0.0035221 | 0.0793936 | 0.0986087 | <.0001* |
| 3        | 2  | 0.0643188 | 0.0035207 | 0.0547151 | 0.0739226 | <.0001* |
| 5        | 4  | 0.0509882 | 0.0035230 | 0.0413781 | 0.0605983 | <.0001* |
| 4        | 3  | 0.0380129 | 0.0035174 | 0.0284181 | 0.0476078 | <.0001* |

図7 経常収入5分位階層別の平均貯蓄率の Tukey-Kramer 法による多重比較の結果

7. 年齢5歳階級別の平均貯蓄率の差の検定

經常収入の多寡が貯蓄率に影響を及ぼすことが明らかとなったが、それでは年齢の違いは貯蓄率の違いにどのような影響を及ぼすのであろうか。

JMPで「2変数の関係」を選択し、貯蓄率を目的変数に、5歳階級別年齢を説明変数に、10万世帯用集計用乗率を度数に入れて得られた一元配置プロット図において、メニューから平均/ANOVA/プーリングしたt検定を選択して得られた分散分析表と各水準の平均の要約表が図8である。

図8から、年齢5歳階級別の貯蓄率の平均は、35～39歳の階級区分までは階級区分が上がるほど高くなるが、40～44歳と45～49歳の階級区分で段階的に大きく下がり、50歳～54歳の階級区分で少し回復し、55～59歳の階級区分で大きく上昇、そして60歳以降はまた下降するという波を描くことが示されている。

| 分散分析     |       |           |         |          |            |
|----------|-------|-----------|---------|----------|------------|
| 要因       | 自由度   | 平方和       | 平均平方    | F値       | p値(Prob>F) |
| 4.年齢5歳階級 | 12    | 190.598   | 15.8831 | 113.9001 | <.0001*    |
| 誤差       | 98802 | 13777.741 | 0.1394  |          |            |
| 全体(修正済み) | 98814 | 13968.339 |         |          |            |

| 各水準の平均 |         |          |         |         |         |
|--------|---------|----------|---------|---------|---------|
| 水準     | 数       | 平均       | 標準誤差    | 下側95%   | 上側95%   |
| 24歳未満  | 240.343 | 0.02135  | 0.02409 | -0.0259 | 0.0686  |
| 25～29歳 | 2823.34 | 0.09249  | 0.00703 | 0.0787  | 0.1063  |
| 30～34歳 | 7977.71 | 0.09407  | 0.00418 | 0.0859  | 0.1023  |
| 35～39歳 | 11259.1 | 0.12598  | 0.00352 | 0.1191  | 0.1329  |
| 40～44歳 | 12033.2 | 0.08948  | 0.00340 | 0.0828  | 0.0961  |
| 45～49歳 | 12022.3 | 0.04540  | 0.00341 | 0.0387  | 0.0521  |
| 50～54歳 | 13001.5 | 0.04695  | 0.00327 | 0.0405  | 0.0534  |
| 55～59歳 | 12475.1 | 0.09482  | 0.00334 | 0.0883  | 0.1014  |
| 60～64歳 | 6093.79 | -0.05099 | 0.00478 | -0.0604 | -0.0416 |
| 65～69歳 | 1678.85 | -0.00577 | 0.00911 | -0.0236 | 0.0121  |
| 70～74歳 | 327.302 | 0.10819  | 0.02064 | 0.0677  | 0.1486  |
| 75歳以上  | 75.007  | 0.14189  | 0.04312 | 0.0574  | 0.2264  |
| 不詳     | 18807.4 | 0.03121  | 0.00272 | 0.0259  | 0.0365  |

図8 年齢5歳階級別の平均貯蓄率の一元配置分析

また、図9の分散分析表から、F値は113.90でp値は0.0001を下回っており、1%有意水準で5歳階級別年齢は貯蓄率に有意な影響を及ぼしていることが判明した。

さらに年齢の個々の5歳階級間での平均貯蓄率の差を検定するためにTukey-Kramer法による多重比較を行ってみると、図9に示すような結果を得た。図9からは、30~34歳と35~39歳、35~39歳と40~44歳、40~44歳と45~49歳、50~54歳と55~59歳、55~59歳と60~64歳、60~64歳と65~69歳、65~69歳と70~74歳の年齢階級の間で、すなわち、30歳から74歳までの間では、45~49歳と50~54歳の間を除くすべての連続する年連階級間で1%有意水準で平均貯蓄率に有意差があることが分かる。

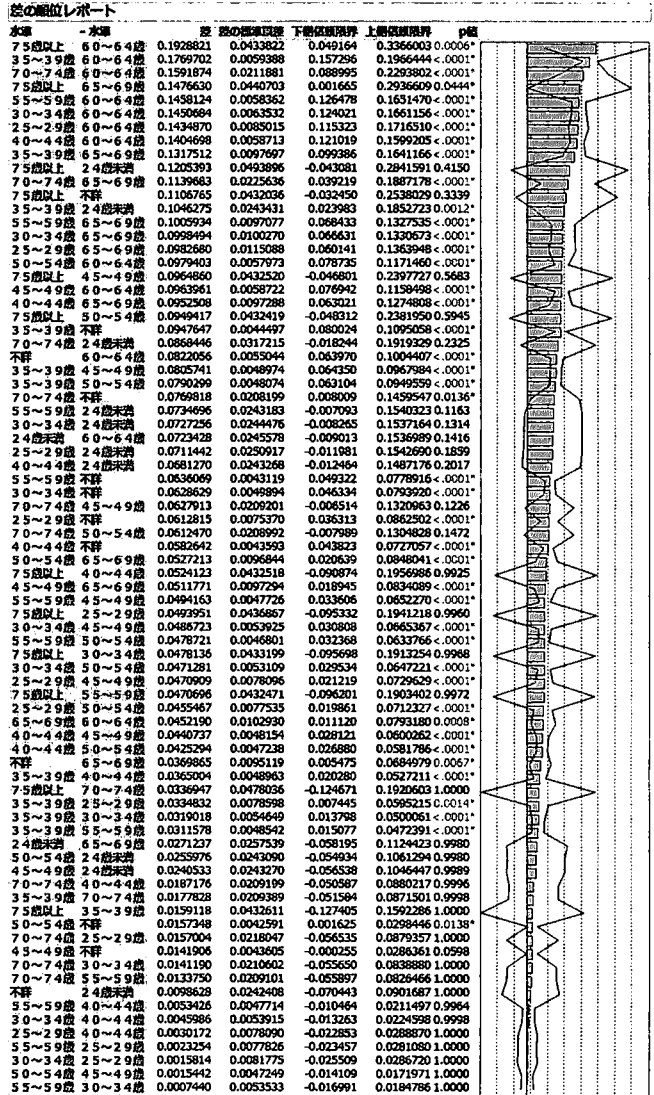


図9 年齢5歳階級別の平均貯蓄率のTukey-Kramer法による多重比較の結果

### 8. 経常収入5分位階層別に見た官民の平均貯蓄率の違い

第6節での分析から、経常収入が多いほど、貯蓄率が高くなる傾向があることがわかった。それでは、経常収入が同じ5分位階層だとしたとき、官民で貯蓄率に違いが出るのだろうか。JMPで経常収入の各5分位階層ごとにサンプルを抽出し、「2変量の関係」で平均貯蓄率の差の検定を行った結果をExcelを使用して加工し、グラフ化した結果を示したものが、以下に示す表1と図10である。なお、表1の欄外のアスタリスクは、\*\*\*は1%水準で有意、\*\*は5%水準で有意、\*は10%水準で有意を示す（以下も同様）。また、セルにグレーの色かけをしている方が、平均貯蓄率が高いことを示す。

表1と図10から、経常収入が同じ5分位階層であるとき、民営のほうが、官公よりも貯蓄率がいずれの階層でも高くなっていることが明らかとなった。第4節では、経常収入でグループ分けすることなく分析したが、そこでは官公の方が、平均貯蓄率で民営を大きく上回っていた。それに関わらず、経常収入の5分位階級でグループ分けして官公と民間との間の平均貯蓄率を比較した結果では、サンプル数が民営では経常収入5分位階級が高くなるにつれて減少していくのに対して、官公では経常収入5分位階級が高くなるにつれてサンプル数が大きく増えていくことを反映して、平均貯蓄率の大小が完全に逆転した結果が得られているのである。

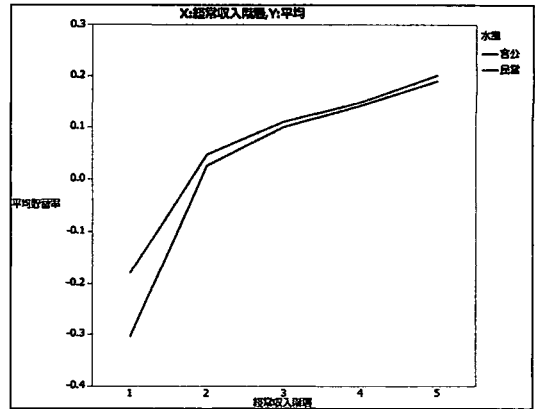


図10 経常収入5分位階級別の官民の平均貯蓄率

表1 経常収入5分位階級別に見た官民の平均貯蓄率

| 経常収入階層 | N(貯蓄率, 官公) | N(貯蓄率, 民営) | 平均(貯蓄率, 官公) | 平均(貯蓄率, 民営) |
|--------|------------|------------|-------------|-------------|
| 1      | 1,512      | 17,771     | -30.5%      | -18.1% ***  |
| 2      | 2,745      | 17,103     | 2.5%        | 4.7% ***    |
| 3      | 3,820      | 16,124     | 10.0%       | 11.0% **    |
| 4      | 5,051      | 14,872     | 14.2%       | 14.8%       |
| 5      | 6,397      | 13,420     | 18.9%       | 20.1% ***   |
| 計      | 19,525     | 79,290     | 9.8%        | 5.4% ***    |

### 9. 年齢5歳階級別に見た官民の平均貯蓄率の違い

経常収入5分位階級別に見た官民の平均貯蓄率を見た場合には、階級別では民間の平均貯蓄率の方が官公の平均貯蓄率よりも高いが、全体では官公の平均貯蓄率の方が民間の平均貯蓄率よりも高いという、合色の誤謬を想起させる興味深い結果が得られたが、年齢5歳階級別に見た官民の平均貯蓄率の違いを見た場合はどうなるのであろうか。

前節での分析と同様に、JMPのサンプルの抽出機能と「2変量の関係」を適用した1元配置分析による結果に基づいて作表と作図をした結果を表2と図11に示している。

表2および図11からは、25～29歳階級から55～59歳階級までは官公の方が平均貯蓄率が高く、60歳を超えると民営の方の平均貯蓄率が高くなる。今回使用している教育用疑似マイクロデータは2004年の調査データに基づくものであり、現在は再雇用や定年延長などで65歳までの雇用が一般的となっているが、調査当時では公務員は60歳で定年を迎えることが一般的であったため、定年を迎える年齢までは、ずっと官公の方が民間よりも平均貯蓄率が高いという結果になったものと考えられる。

年齢5歳階級別で官公と民会の共通の特徴としては、どちらも35歳～39歳で平均貯蓄率が最も高くなり、以降は低下し、定年前の55歳～59歳で再度高くなるということが挙げられる。40歳以降で平均貯蓄率が低くなる理由として、介護保険料の納付が40歳から始まることや、住宅を取得する年齢となって住宅ローン返済が始まることに加え、子供が中学や高校に進学して教育費が家計の負担になっていることなどが考えられる。

教育費や住宅ローンの影響については、「教育」や「土地家屋借金返済」などの変数と間の関係を調べれば明らかとなると考えられるが、これについては今後の課題としたい。

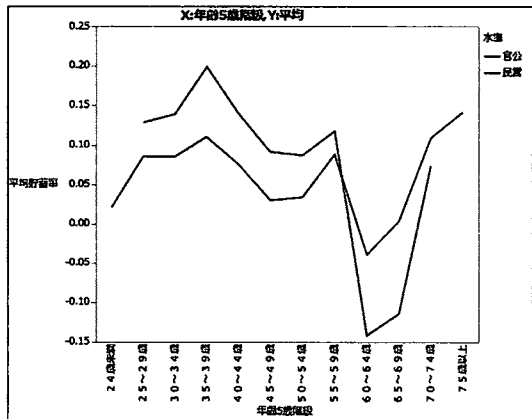


図11 年齢5歳階級別に見た官民の平均貯蓄率

表2 年齢5歳階級別に見た官民の平均貯蓄率

| 4.年齢5歳階級 | N(貯蓄率, 官公) | N(貯蓄率, 民営) | 平均(貯蓄率, 官公) | 平均(貯蓄率, 民営) |
|----------|------------|------------|-------------|-------------|
| 24歳未満    | 0          | 240        |             | 2.1%        |
| 25～29歳   | 424        | 2,399      | 12.9%       | 8.6% ***    |
| 30～34歳   | 1,253      | 6,725      | 13.9%       | 8.6% ***    |
| 35～39歳   | 1,953      | 9,306      | 19.9%       | 11.1% ***   |
| 40～44歳   | 2,459      | 9,574      | 14.0%       | 7.6% ***    |
| 45～49歳   | 2,974      | 9,048      | 9.2%        | 3.0% ***    |
| 50～54歳   | 3,138      | 9,863      | 8.7%        | 3.4% ***    |
| 55～59歳   | 2,715      | 9,761      | 11.8%       | 8.8% ***    |
| 60～64歳   | 702        | 5,391      | -14.2%      | -3.9% ***   |
| 65～69歳   | 130        | 1,548      | -11.4%      | 0.3% ***    |
| 70～74歳   | 11         | 316        | 7.4%        | 10.9% ***   |
| 75歳以上    | 0          | 75         |             | 14.2%       |
| 不詳       | 3,766      | 15,041     | 5.3%        | 2.6% ***    |
| すべて      | 19,525     | 79,290     | 9.8%        | 5.4% ***    |

10. 年齢5歳階級別および経常収入5分位階層別に見た官民の平均貯蓄率の違い

最後に年齢5歳階級別と経常収入5分位階層別とを組み合わせると見た官民の平均貯蓄率の違いについて分析する。

かなり煩雑な作業となったが、これまでの分析と同様に JMP のサンプルの抽出機能と「2変数の関係」を適用して実施した1元配置分析の結果から作表および作図した結果を表3と図12および図13に示している。

表3より、比較可能な年齢5歳階級と経常収入5分位階層の組み合わせ46通りのうち、21通りで官公が民間の貯蓄率を、一方で25通りでは民営が官公の貯蓄率を上回っている。

年齢で特徴的なことは、50～54歳、55歳～59歳、60～64歳の3つの階級は、いずれの経常収入5分位階層でも民営の方が官公よりも貯蓄率が高い結果となっていることである。これは、退職金の額の官民の相違によるものと考えられるが、教育用疑似マイクロデータでは、退職金を補足することができなかったため、この仮説を検証することはできなかった。

上述の50歳から64歳までの年齢階級以外の5歳階級では、経常収入5分位階層によって大小関係が異なっている。

表3 年齢5歳階級別および経常収入5分位階級別に見た官民の平均貯蓄率

| 4.年齢5歳階級 | 経常収入階層 | N(官公)  | N(民営)  | 平均(貯蓄率, 官公) | 平均(貯蓄率, 民営) |     |
|----------|--------|--------|--------|-------------|-------------|-----|
| 24歳未満    | 1      | 0      | 45     |             | -12.9%      |     |
|          | 2      | 0      | 18     |             | 31.5%       |     |
|          | 3      | 0      | 3      |             | 11.5%       |     |
|          | 4      | 0      | 1      |             | 63.7%       |     |
| 25~29歳   | 1      | 42     | 336    | -9.8%       | 0.9%        | *** |
|          | 2      | 43     | 249    | 15.5%       | 10.9%       |     |
|          | 3      | 20     | 104    | 26.2%       | 17.6%       | *** |
|          | 4      | 17     | 49     | 29.1%       | 21.9%       | *** |
|          | 5      | 6      | 19     | 17.5%       | 18.3%       |     |
| 30~34歳   | 1      | 77     | 601    | -0.7%       | -7.4%       | **  |
|          | 2      | 151    | 753    | 9.3%        | 11.5%       | *   |
|          | 3      | 99     | 429    | 16.6%       | 17.4%       |     |
|          | 4      | 84     | 210    | 24.1%       | 20.3%       | *   |
| 35~39歳   | 1      | 23     | 111    | 38.2%       | 16.4%       | *   |
|          | 2      | 31     | 510    | -15.4%      | -11.2%      |     |
|          | 3      | 137    | 762    | 14.8%       | 8.2%        | *** |
|          | 4      | 200    | 788    | 18.9%       | 15.0%       | *** |
|          | 5      | 186    | 564    | 23.4%       | 21.6%       |     |
| 40~44歳   | 1      | 117    | 238    | 32.4%       | 26.5%       | *** |
|          | 2      | 41     | 443    | -59.9%      | -21.6%      | *** |
|          | 3      | 88     | 616    | 3.9%        | 3.4%        |     |
|          | 4      | 214    | 777    | 12.6%       | 10.7%       | *   |
|          | 5      | 246    | 735    | 16.6%       | 14.1%       | *** |
| 45~49歳   | 1      | 246    | 464    | 28.0%       | 23.1%       | *** |
|          | 2      | 32     | 360    | -18.1%      | -27.4%      | *   |
|          | 3      | 70     | 457    | -0.3%       | -6.5%       | **  |
|          | 4      | 194    | 647    | 2.1%        | 6.7%        | *** |
|          | 5      | 284    | 767    | 11.6%       | 10.0%       |     |
| 50~54歳   | 1      | 429    | 682    | 14.8%       | 14.0%       |     |
|          | 2      | 25     | 428    | -30.5%      | -27.3%      |     |
|          | 3      | 62     | 472    | -5.1%       | -3.7%       |     |
|          | 4      | 125    | 585    | 4.7%        | 5.1%        |     |
|          | 5      | 285    | 734    | 5.0%        | 8.3%        | **  |
| 55~59歳   | 1      | 572    | 949    | 14.6%       | 15.5%       |     |
|          | 2      | 38     | 510    | -27.0%      | -24.8%      |     |
|          | 3      | 77     | 520    | -7.6%       | 2.3%        | **  |
|          | 4      | 130    | 579    | 0.4%        | 9.5%        | *** |
|          | 5      | 241    | 669    | 15.7%       | 15.7%       |     |
| 60~64歳   | 1      | 470    | 917    | 19.1%       | 24.4%       | *** |
|          | 2      | 65     | 634    | -52.4%      | -28.2%      | *** |
|          | 3      | 55     | 364    | -15.7%      | 0.5%        | *** |
|          | 4      | 39     | 268    | 5.2%        | 8.7%        |     |
|          | 5      | 45     | 240    | 3.1%        | 14.1%       | *** |
| 65~69歳   | 1      | 39     | 197    | 12.5%       | 25.2%       | *** |
|          | 2      | 22     | 193    | -48.6%      | -20.1%      | *** |
|          | 3      | 15     | 119    | 8.4%        | 6.9%        |     |
|          | 4      | 5      | 74     | 31.7%       | 11.3%       |     |
|          | 5      | 5      | 70     | 17.8%       | 19.6%       |     |
| 70~74歳   | 1      | 4      | 45     | 19.2%       | 22.6%       |     |
|          | 2      | 4      | 33     | 32.5%       | -10.1%      | **  |
|          | 3      | 1      | 29     | -64.9%      | 20.2%       | *** |
|          | 4      | 0      | 21     |             | 17.0%       |     |
|          | 5      | 0      | 11     |             | 21.6%       |     |
| 75歳以上    | 1      | 0      | 13     |             | 24.4%       |     |
|          | 2      | 0      | 4      |             | -2.5%       |     |
|          | 3      | 0      | 3      |             | 4.5%        |     |
|          | 4      | 0      | 4      |             | 16.6%       |     |
|          | 5      | 0      | 2      |             | 47.2%       |     |
| 不詳       | 1      | 0      | 2      |             | 53.2%       |     |
|          | 2      | 159    | 1681   | -38.4%      | -17.0%      | *** |
|          | 3      | 219    | 1026   | -2.1%       | 6.0%        | *** |
|          | 4      | 264    | 783    | 10.6%       | 12.1%       |     |
|          | 5      | 292    | 610    | 14.0%       | 18.0%       | *** |
|          | 計      | 19,525 | 79,290 | 9.8%        | 5.4%        | *** |



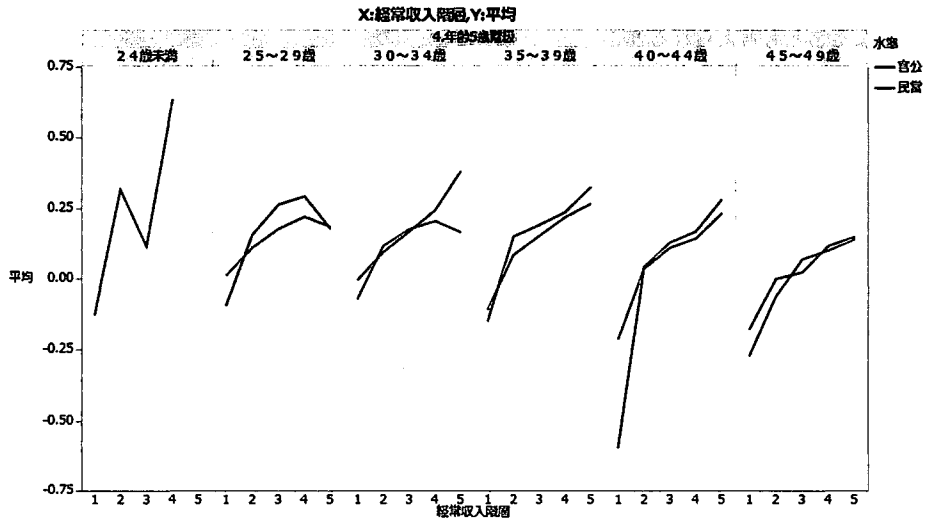


図 12 年齢 5 歳階級別および経常収入 5 分位階級別に見た官民の平均貯蓄率 (2 の 1)

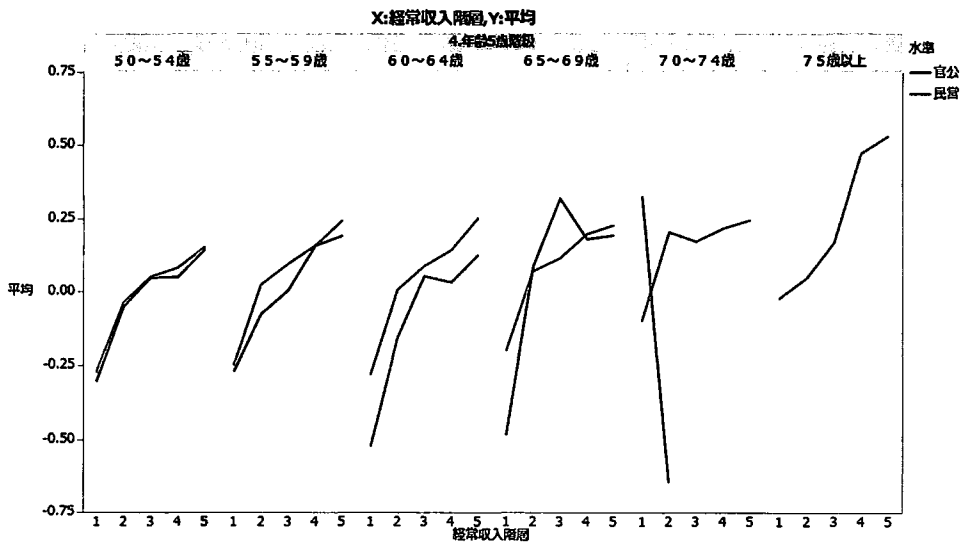


図 13 年齢 5 歳階級別および経常収入 5 分位階級別に見た官民の平均貯蓄率 (2 の 2)

### 11. まとめ

本研究では、世帯ごとの収入や預貯金、年齢 5 歳階級、企業区分などの変数を使って、官民の平均貯蓄率の違いについて分析を行った。

全体で見ると、官の方が民よりも貯蓄率が高いが、同じ経常収入の 5 分位階層内での比較では、民の方が官をわずかに上回るという、合成の誤謬を想起させる興味深い結果を得た。

しかし、年齢 5 歳階級を考慮に入れた分析では、経常収入の 5 分位階層によって結果が異なり、官民の貯蓄率の違いにおける法則性を見つけることはできなかった。

本研究での分析結果をどう役立てるのかについては、お金を貯めるのはいつかということを考えるのに役立つと考えられる。年齢 5 歳階級別に見た貯蓄率のピークは官も民も 35~39 歳にきている。40 歳からは介護保険料の納付が必要になり、住宅ローンに加え子供の教育費も大きな負担になってくる。したがって、39 歳までの期間にお金を貯め逃すと、子供の独立までの期間が苦しくなる。晩婚化が進んでいる近年では、子供が独立するのは貯蓄率が上向く 50 歳~54 歳や老後に向けての貯蓄率が高まる 55~

59 歳ではなく、60 歳以降になってからという世帯も多く、このような世帯では、老後のための貯蓄が十分ではなくなることも考えられる。

それぞれの家族構成や、働き方にもよるが「備えあれば憂いなし」ということわざ通り、貯め時を逃さないことが、今後のゆとりある生活を送るには最重要であると感じた。

#### 参考文献

[1]総務省統計局, 「平成 26 年全国消費実態調査 用語の解説 15. 計算式」  
(<http://www.stat.go.jp/data/zensho/2014/kaisetsu.htm>) (2016. 6. 12 参照)

規 定 課 題

田中健一郎

兵庫県立大学大学院 応用情報科学研究科 応用情報科学専攻  
参加カテゴリーB

1)

単純クロス表の作成

JMP の表の作成で、列を性別、行を産業符号に設定し、集計統計量の追加にチェックマークを入れて実行した。

10 万世帯比でのクロス表の作成

元々の集計用乗率をそのまま用いると合計で 495464.76574 世帯となるので、10 万世帯比にするため集計用乗率に 100,000 / 495,464.76574 を乗じた 10 万世帯用集計用乗率を列として新たに追加した。

JMP の表の作成で、列を性別、行を産業符号に設定し、重みに 10 万世帯用集計用乗率を設定し、集計統計量の追加にチェックマークを入れて実行した。

a)単純クロス表

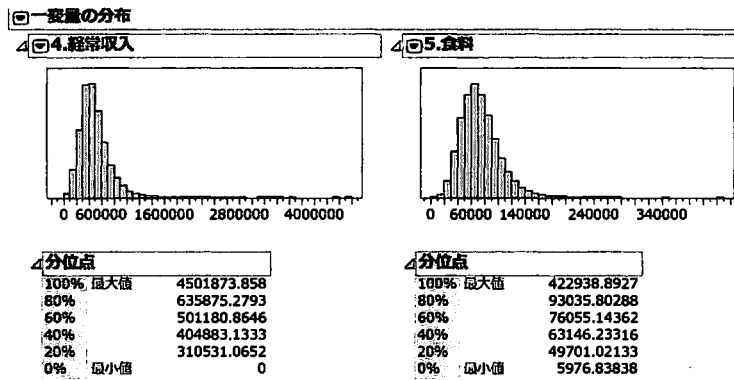
| 4.産業符号            | 4.性別  |      |       |
|-------------------|-------|------|-------|
|                   | 男     | 女    | すべて   |
| 鉱業                | 3     | 0    | 3     |
| 建設業               | 2134  | 9    | 2143  |
| 製造業               | 5668  | 173  | 5841  |
| 電気・ガス・熱供給・水道業     | 222   | 0    | 222   |
| 情報通信業             | 547   | 0    | 547   |
| 運輸業               | 1235  | 3    | 1238  |
| 卸売・小売業            | 2083  | 201  | 2284  |
| 金融・保険業            | 472   | 39   | 511   |
| 不動産業              | 40    | 0    | 40    |
| 飲食店・宿泊業           | 50    | 32   | 82    |
| 医療・福祉             | 371   | 270  | 641   |
| 教育・学習支援業          | 1059  | 89   | 1148  |
| 複合サービス事業          | 321   | 3    | 324   |
| サービス業（他に分類されないもの） | 1816  | 106  | 1922  |
| 公務（他に分類されないもの）    | 3656  | 130  | 3786  |
| 不詳                | 9704  | 1591 | 11295 |
| すべて               | 29381 | 2646 | 32027 |

b)10 万世帯比でのクロス表

| 4.産業符号            | 4.性別  |      |        |
|-------------------|-------|------|--------|
|                   | 男     | 女    | すべて    |
| 鉱業                | 7     | 0    | 7      |
| 建設業               | 6346  | 31   | 6377   |
| 製造業               | 17837 | 510  | 18347  |
| 電気・ガス・熱供給・水道業     | 697   | 0    | 697    |
| 情報通信業             | 2020  | 0    | 2020   |
| 運輸業               | 3852  | 6    | 3858   |
| 卸売・小売業            | 6644  | 589  | 7233   |
| 金融・保険業            | 1447  | 104  | 1551   |
| 不動産業              | 144   | 0    | 144    |
| 飲食店・宿泊業           | 158   | 107  | 265    |
| 医療・福祉             | 1022  | 810  | 1832   |
| 教育・学習支援業          | 3110  | 269  | 3379   |
| 複合サービス事業          | 981   | 11   | 992    |
| サービス業（他に分類されないもの） | 5842  | 326  | 6167   |
| 公務（他に分類されないもの）    | 10691 | 430  | 11121  |
| 不詳                | 31204 | 4806 | 36011  |
| すべて               | 92000 | 8000 | 100000 |

2)

JMP の分析で一変量の分布を選択。Y 列に経常収入と食料、度数に 10 万世帯用集計用条率をセットして分析、分位点の間隔を 0.2 に変更してテーブルを出力し、ヒストグラムは表示されたものをコピー



3)

新規に経常収入階層、食料階層の列を追加する。2) で求めた分位点を各行に反映させるため、計算式にそれぞれ IF 文を設定する。

経常収入の場合

```
If(0 <= :Name("4.経常収入") < 310531, 1, 310531 <= :Name("4.経常収入") < 404883, 2, 404883 <= :Name("4.経常収入") < 501181, 3, 501181 <= :Name("4.経常収入") < 635875, 4, 635875 <= :Name("4.経常収入"), 5)
```

食料の場合

```
If(0 <= :Name("5.食料") < 49701, 1, 49701 <= :Name("5.食料") < 63146, 2, 63146 <= :Name("5.食料") < 76055, 3, 76055 <= :Name("5.食料") < 93035, 4, 93035 <= :Name("5.食料"), 5)
```

JMP の表の作成で、列を性別及び食料階層、行を経常収入階層に設定し、重みに 10 万世帯用集計用乗率を設定し、集計統計量の追加にチェックマークを入れて実行した。

| 経常収入階層 | 4.性別  |      |        |
|--------|-------|------|--------|
|        | 男     | 女    | すべて    |
| 1      | 15494 | 4506 | 20000  |
| 2      | 18617 | 1377 | 19994  |
| 3      | 19099 | 911  | 20009  |
| 4      | 19356 | 640  | 19995  |
| 5      | 19435 | 567  | 20002  |
| すべて    | 92000 | 8000 | 100000 |

| 食料階層 | 4.性別  |      |        |
|------|-------|------|--------|
|      | 男     | 女    | すべて    |
| 1    | 16803 | 3196 | 19999  |
| 2    | 18097 | 1903 | 20000  |
| 3    | 18734 | 1268 | 20001  |
| 4    | 19049 | 948  | 19998  |
| 5    | 19317 | 685  | 20003  |
| すべて  | 92000 | 8000 | 100000 |

4)

JMP の二変量の関係で、Y に食料階層 X に経常収入階層、BY に性別、度数に 10 万世帯用集計用乗率をセットして分析

☑ 経常収入階層と食料階層の分割表に対する分析 4.性別=男

度数: 10万世帯用集計用条率

4 ☑ 分割表

|        |   | 食料階層  |       |       |       |       |       |
|--------|---|-------|-------|-------|-------|-------|-------|
|        |   | 1     | 2     | 3     | 4     | 5     | 合計    |
| 経常収入階層 | 1 | 5827  | 3974  | 2798  | 1831  | 1064  | 15494 |
|        | 2 | 5130  | 4501  | 3978  | 2996  | 2012  | 18617 |
|        | 3 | 2911  | 4301  | 4395  | 4294  | 3198  | 19099 |
|        | 4 | 1906  | 3233  | 4082  | 4903  | 5231  | 19356 |
|        | 5 | 1029  | 2087  | 3481  | 5026  | 7813  | 19435 |
| 合計     |   | 16803 | 18097 | 18734 | 19049 | 19317 | 92000 |

☑ 経常収入階層と食料階層の分割表に対する分析 4.性別=女

度数: 10万世帯用集計用条率

4 ☑ 分割表

|        |   | 食料階層 |      |      |     |     |      |
|--------|---|------|------|------|-----|-----|------|
|        |   | 1    | 2    | 3    | 4   | 5   | 合計   |
| 経常収入階層 | 1 | 2462 | 994  | 565  | 314 | 172 | 4506 |
|        | 2 | 371  | 399  | 235  | 211 | 161 | 1377 |
|        | 3 | 243  | 225  | 216  | 164 | 63  | 911  |
|        | 4 | 68   | 134  | 156  | 143 | 140 | 640  |
|        | 5 | 53   | 152  | 96   | 117 | 149 | 567  |
| 合計     |   | 3196 | 1903 | 1268 | 948 | 685 | 8000 |

5)

食料の常用対数列を新たに作成し、JMP の二変量の関係で Y に常用対数 (食料)、X に経常収入階層、BY に性別、度数に 10 万世帯用集計用条率をセットして分析。

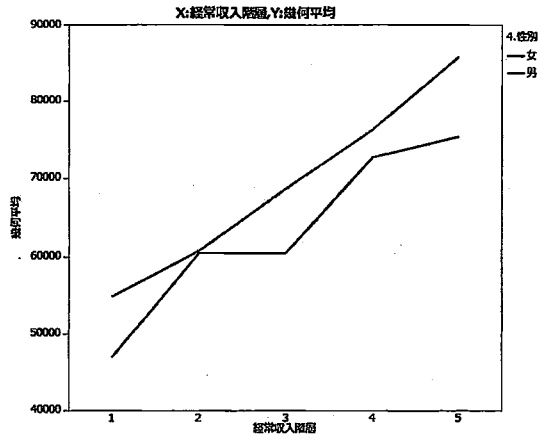
メニューから平均 ANOVA を選択し、各水準の平均のところまで右クリックし、連結したテーブルの作成を選ぶ。

合計行がないので、新規で追加し、平均 ANOVA の結果表から転記。

常用対数平均を小数点以下第 5 位で四捨五入し、幾何平均列を新規作成。計算式を  $10^{\wedge}$  常用対数平均と入力。

|    | 4.性別 | 経常収入階層 | 世帯数   | 常用対数平均 | 幾何平均  |
|----|------|--------|-------|--------|-------|
| 1  | 男    | 1      | 15494 | 4.7389 | 54815 |
| 2  | 男    | 2      | 18617 | 4.7832 | 60702 |
| 3  | 男    | 3      | 19099 | 4.8368 | 68675 |
| 4  | 男    | 4      | 19356 | 4.8826 | 76313 |
| 5  | 男    | 5      | 19435 | 4.9335 | 85803 |
| 6  | 男    | 合計     | 92000 | 4.8395 | 69103 |
| 7  | 女    | 1      | 4506  | 4.6713 | 46914 |
| 8  | 女    | 2      | 1377  | 4.7811 | 60409 |
| 9  | 女    | 3      | 911   | 4.7813 | 60437 |
| 10 | 女    | 4      | 640   | 4.8617 | 72728 |
| 11 | 女    | 5      | 567   | 4.8774 | 75405 |
| 12 | 女    | 合計     | 8000  | 4.7325 | 54013 |

合計行は除外したあと、JMP のグラフの作成で、Y に幾何平均、X に経常収入階層、重ね合わせに性別を設定する



# 一戸建て・共同住居住まい世帯における傾向の比較

カテゴリーB

田中 基史

株式会社コネクティボ システム開発グループ

## 1. はじめに

居住に関して昨今、生活者のライフスタイルの多様化により、一戸建てまたは共同住宅のどちらを選択すべきなのか非常に悩ましい問題である。各々どちらも利点・欠点が存在するため、一概にどちらが良いということは難しい。

そのため本稿では擬似マイクロデータを用いて、一戸建て・共同住宅住まい世帯に関して様々な観点から客観的にどのような傾向があるのか比較を行った。また、年間収入・平均消費性向・平均貯蓄率の指標を用いてどのような傾向があるか比較を行った。

## 2. 方法

擬似マイクロデータを用いて、変数「Tatekata」から一戸建てを「1」、共同住宅を「2」として分別する。その際に、本稿では対象外の「長屋建」・「その他」・「不詳」データは除外する。分類する2グループに関して比較を行う。

1 つ目の比較を行うにあたり「年齢 5 歳階級」・「世帯人数」・「有業人員」・「職業」・「産業」・「住居の所有関係」はそれぞれ変数「S1\_Age」・「SetaiJinin」・「ShuugyouJinin」・「S1\_Shokugyou」・「S1\_Sangyou」・「Shoyuu」を使用する。尚、各々の「不詳」データは除くものとする。

2 つ目の比較を行うにあたり「年間収入」・「平均消費性向」・「平均貯蓄率」の指標を用いる。「年間収入」は変数「Youto001」を使用する。また、「平均消費性向」・「平均貯蓄率」は式(1)の「可処分所得」を算出し、(1)式を用いて式(2)(3)となる[1][2]。

$$\text{可処分所得 [円]} = \text{実収入} - \text{非消費支出} \quad (1)$$

実収入:Youto003, 非消費支出:Youto159

$$\text{平均消費性向 [\%]} = \frac{\text{消費支出}}{\text{可処分所得}} \times 100 \quad (2)$$

消費支出:Youto0037

$$\text{平均貯蓄率 [\%]} = \frac{(\text{預貯金} - \text{預貯金引出}) + (\text{保険料} - \text{保険金})}{\text{可処分所得}} \times 100 \quad (3)$$

預貯金:Youto171, 預貯金引出:Youto022, 保険料:Youto172, 保険金:Youto023

### 3. 結果

#### 3.1. 一戸建てまたは共同住宅住まい世帯数の比較

一戸建てまたは共同住宅住まい世帯のグループで、「年齢5歳階級」・「世帯人数」・「有業人員」・「職業」・「産業」・「住居の所有関係」を比較した。比較結果を表1(a)-(f)、図1(a)-(f)に示す。

表1. 一戸建てまたは共同住宅住まい世帯数

(a) 年齢5歳階級

|      |   | 年齢5歳階級 |        |        |        |        |        |        |        |        |        |        |
|------|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|      |   | 24歳未満  | 25～29歳 | 30～34歳 | 35～39歳 | 40～44歳 | 45～49歳 | 50～54歳 | 55～59歳 | 60～64歳 | 65～69歳 | 70～74歳 |
| 住居区分 |   |        |        |        |        |        |        |        |        |        |        |        |
| 一戸建て | N | 0      | 87     | 680    | 1464   | 2040   | 2197   | 2643   | 2748   | 1014   | 229    | 29     |
| 共同住宅 | N | 9      | 189    | 415    | 367    | 240    | 94     | 23     | 25     | 18     | 0      | 0      |

(b) 世帯人員

|      |   | 世帯人員 |      |      |      |     |     |
|------|---|------|------|------|------|-----|-----|
|      |   | 2人   | 3人   | 4人   | 5人   | 6人  | 7人  |
| 住居区分 |   |      |      |      |      |     |     |
| 一戸建て | N | 2594 | 3280 | 4620 | 2076 | 440 | 111 |
| 共同住宅 | N | 277  | 471  | 583  | 46   | 3   | 0   |

(c) 有業人員

|      |   | 有業人員 |      |      |     |    |
|------|---|------|------|------|-----|----|
|      |   | 1人   | 2人   | 3人   | 4人  | 5人 |
| 住居区分 |   |      |      |      |     |    |
| 一戸建て | N | 4970 | 6463 | 1416 | 279 | 3  |
| 共同住宅 | N | 1021 | 349  | 10   | 0   | 0  |

(d) 職業

|      |   | 職業    |            |      |       |       |
|------|---|-------|------------|------|-------|-------|
|      |   | 常用労働者 | 臨時及び日々雇労働者 | 民間職員 | 官公職員1 | 官公職員2 |
| 住居区分 |   |       |            |      |       |       |
| 一戸建て | N | 4219  | 6          | 5203 | 329   | 3374  |
| 共同住宅 | N | 358   | 0          | 660  | 136   | 226   |

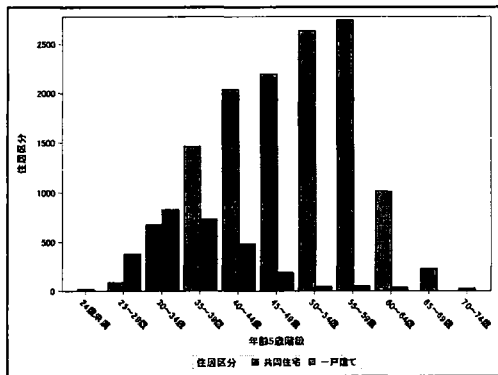
(e) 産業

|      |   | 産業 |      |      |               |       |     |        |        |      |        |       |              |          |                   |                |
|------|---|----|------|------|---------------|-------|-----|--------|--------|------|--------|-------|--------------|----------|-------------------|----------------|
|      |   | 農業 | 建設業  | 製造業  | 電気・ガス・熱供給・水道業 | 情報通信業 | 運輸業 | 卸売・小売業 | 金融・保険業 | 不動産業 | 飲食・宿泊業 | 医療・福祉 | 教育・学問・文化・娯楽業 | 複合サービス産業 | サービス業(他に分類されないもの) | 公務(他に分類されないもの) |
| 住居区分 |   |    |      |      |               |       |     |        |        |      |        |       |              |          |                   |                |
| 一戸建て | N | 3  | 1373 | 3860 | 139           | 270   | 780 | 1257   | 285    | 17   | 29     | 351   | 812          | 194      | 1023              | 2728           |
| 共同住宅 | N | 0  | 101  | 465  | 14            | 85    | 85  | 148    | 48     | 0    | 0      | 10    | 42           | 0        | 98                | 223            |

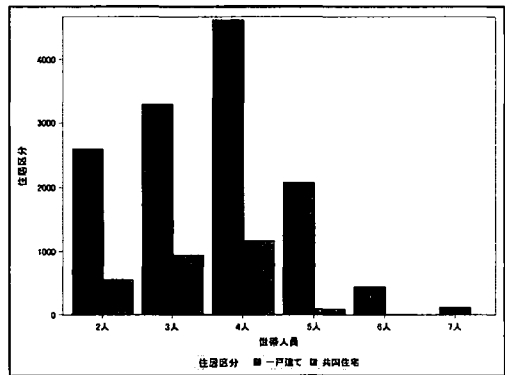
(f) 住居の所有関係

|      |   | 住居の所有関係    |            |              |            |                |                 |
|------|---|------------|------------|--------------|------------|----------------|-----------------|
|      |   | 持ち家(世帯員名義) | 持ち家(その他名義) | 民営賃貸住宅(設備等別) | 県市区町村営賃貸住宅 | 都市再生機構・公社等賃貸住宅 | 社宅・公務員住宅(借上げ含む) |
| 住居区分 |   |            |            |              |            |                |                 |
| 一戸建て | N | 12879      | 107        | 129          | 0          | 0              | 16              |
| 共同住宅 | N | 273        | 0          | 686          | 130        | 13             | 378             |

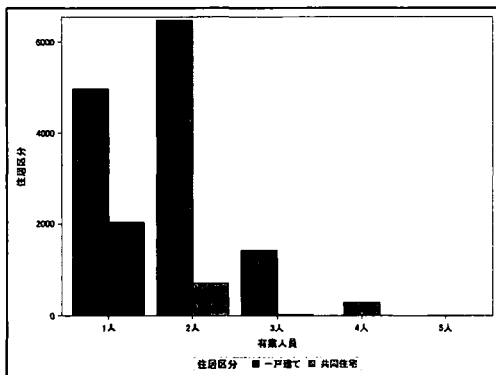




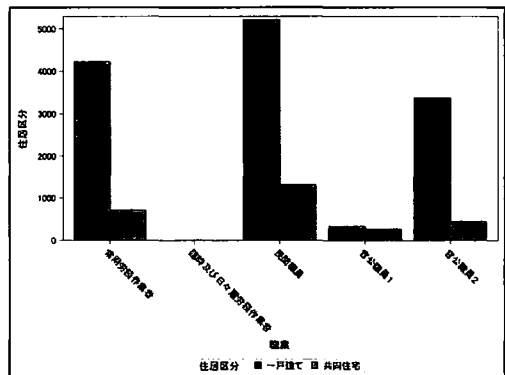
(a) 年齢5歳階級



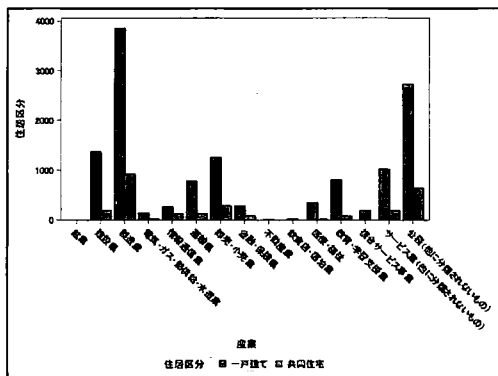
(b) 世帯人員



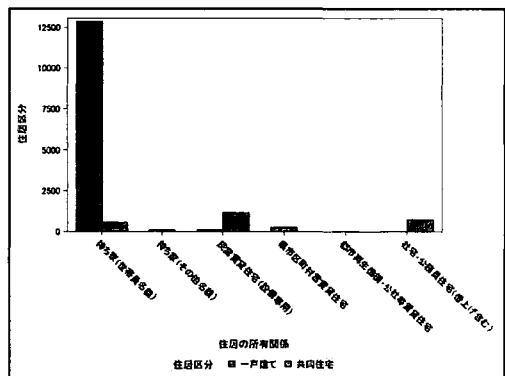
(c) 有業人員



(d) 職業



(e) 産業



(f) 住居の所有関係

図1. 一戸建て・共同住宅住まい世帯数

表1, 図1より, 「年齢5歳階級」の比較では, 一戸建て住まい世帯で『55~59歳』が, 共同住宅住まい世帯では『30~34歳』がピークとなり, 65歳以上の世帯数は0となった. 「世帯人員」の比較では一戸建て・共同住宅住まい世帯共に『4人』がピークとなったが, 「有業人員」では, 一戸建て世帯は『2人』, 共同住宅住まい世帯は『1人』がピークとなった. 「職業」の比較では一戸建て・共同

住宅住まい世帯共に『民間職員』が最も多くなった。「産業」の比較でも一戸建・共同住宅世帯共に『製造業』が最も多くなった。「住居の所有関係」の比較では、一戸建て住まい世帯は『持ち家(世帯員名義)』が最も多かったのに対し、共同住居住まい世帯は『民営賃貸住宅(設備専用)』が最も多くなる結果となった。

### 3.2. 年間収入および貯蓄・消費の比較

一戸建て・共同住宅住まい世帯グループにおける年齢5歳階級・世帯人員別「年間収入」・「平均消費性向」・「平均貯蓄率」の算術平均値を用いて比較を行う。比較結果を表2, 3に示す。

表2. 年齢5歳階級別「年間収入」・「平均消費性向」・「平均貯蓄率」・「教育」の比較

|          |           |           | 年齢5歳階級  |         |         |         |         |         |         |         |         |         |         |         |
|----------|-----------|-----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|          |           |           | 24歳未満   | 25～29歳  | 30～34歳  | 35～39歳  | 40～44歳  | 45～49歳  | 50～54歳  | 55～59歳  | 60～64歳  | 65～69歳  | 70～74歳  |         |
| 住居区分     |           |           |         |         |         |         |         |         |         |         |         |         |         |         |
|          | 一戸建て      | 年間収入【千円】  | Mean    | 0       | 4786.55 | 5516.47 | 6592.73 | 7712.55 | 8567.71 | 9276.23 | 8080.41 | 6579.21 | 6431.25 | 6648.75 |
|          |           | 平均消費性向【%】 | Mean    | 0       | 74.71   | 73.02   | 67.75   | 74.48   | 82.35   | 85.84   | 59.85   | 100.57  | 34.52   | 81.80   |
| 平均貯蓄率【%】 |           | Mean      | 0       | 6.57    | 10.55   | 15.17   | 9.68    | 6.20    | -7.27   | 33.42   | -8.31   | 7.58    | 3.79    |         |
| 共同住宅     | 年間収入【千円】  | Mean      | 3265.45 | 4630.44 | 5438.48 | 6592.60 | 7626.65 | 8563.03 | 5119.09 | 7644.58 | 5855.69 | 0       | 0       |         |
|          | 平均消費性向【%】 | Mean      | 116.05  | 87.90   | 85.66   | 76.41   | 85.04   | 89.32   | 80.88   | 77.35   | 111.19  | 0       | 0       |         |
|          | 平均貯蓄率【%】  | Mean      | -28.64  | 8.49    | 13.63   | 16.23   | 7.65    | -8.63   | 12.39   | 8.10    | -2.92   | 0       | 0       |         |

表3. 世帯人員別「年間収入」・「平均消費性向」・「平均貯蓄率」・「教育」の比較

|          |           |           | 世帯人員    |         |         |         |         |         |         |
|----------|-----------|-----------|---------|---------|---------|---------|---------|---------|---------|
|          |           |           | 2人      | 3人      | 4人      | 5人      | 6人      | 7人      |         |
| 住居区分     |           |           |         |         |         |         |         |         |         |
|          | 一戸建て      | 年間収入【千円】  | Mean    | 7261.46 | 8034.75 | 8229.29 | 8808.31 | 9310.27 | 9431.58 |
|          |           | 平均消費性向【%】 | Mean    | 61.17   | 76.70   | 88.08   | 80.37   | 81.35   | 77.72   |
| 平均貯蓄率【%】 |           | Mean      | 30.84   | 10.59   | -1.28   | 7.61    | 6.35    | 15.59   |         |
| 共同住宅     | 年間収入【千円】  | Mean      | 6291.30 | 5737.43 | 6604.10 | 6809.47 | 7379.33 | 0       |         |
|          | 平均消費性向【%】 | Mean      | 80.62   | 84.84   | 84.60   | 84.83   | 112.69  | 0       |         |
|          | 平均貯蓄率【%】  | Mean      | 18.29   | 13.93   | 4.13    | 9.59    | -18.27  | 0       |         |

表2, 3より、「年間収入」の比較では、年齢別では一戸建て住まい世帯は『50～54歳』が最大となり、共同住宅住まい世帯は『45～49歳』が最大となった。また、世帯人員別では共に世帯人員が増加するごとに『年間収入』も増加する結果となった。「平均消費性向」の比較では、共同住宅住まい世帯の方が高くなることがわかった。「平均貯蓄率」の比較では、全体的に一戸建て住まい世帯の方が高くなることがわかった。また、世帯人員別では共に『2人』の場合が最大値となった。

## 4. まとめ

本稿では、擬似マイクロデータを用いて様々な観点から一戸建て・共同住宅住まい世帯の比較を行った。

『3.1 一戸建てまたは共同住宅住まい世帯数の比較』では、一戸建て住まい世帯は年齢が上がるごとに世帯数が増加しているのに対し、共同住宅住まい世帯では『30～34 歳』がピークとなり以降は減少する傾向になった。これは年齢が上がるにつれ共同住宅から一戸建てへと住み替えが行われた等が考えられる。また、「世帯人数」では共に『4 人』が最大となったが、「有業人員」では一戸建て住まい世帯ではピークが『2 人』となっているのに対し、共同住宅住まい世帯ではピークが『1 人』となり、『2 人』以降は減少する傾向になった。本データからは「有業人員」が増えるにつれ共同住宅が選択されなくなる傾向になった。「職業」・「産業」では共に最大が『民間職員』『製造業』となり特にこれといった傾向が見られなかった。「住居の所有関係」では、一戸建て住まい世帯は圧倒的に持ち家が多いが、共同住宅住まい世帯では賃貸の方が多くなった。このことから、持ち家とする場合、一戸建てを選択する傾向が強いことがわかった。

『3.2 年間収入および貯蓄・消費の比較』では、一戸建て住まい世帯は『50～54 歳』で「年間収入」が最大になるのに対し、共同住宅住まい世帯では『45～49 歳』で最大となった。また、年齢別の各々の最大値を比較した場合、713.3[千円]の差があった。消費性向の割合に関して、一戸建て住まい世帯では、年齢では『60～64 歳』が世帯人数では『4 人』が最大となった。共同住宅住まい世帯では、年齢では『24 歳未満』が世帯人数では『6 人』が最大となった。消費性向に伴い貯蓄率は各々の年齢・世帯人数で最小値となった。また、消費性向の割合では、世帯人員の一戸建て住まい世帯の最小値は 61.17 [%]だが、共同住宅住まい世帯の最小値は 80.62 [%]となった。貯蓄率の割合では、世帯人員の共同住宅住まい世帯の最大値は 18.29 [%]であるのに対し、一戸建て住まい世帯の最大値は 30.84 [%]と約 2 倍の結果となった。このことから、一戸建て住まい世帯では年齢別または世帯人員別で「年間収入」・「平均貯蓄率」が高く、「平均消費性向」が低くなる傾向があることがわかった。

以上の結果より、世帯主が 30 歳未満の場合は共同住宅を選択する傾向があり、世帯主の年齢が高齢になる、世帯の有業人員・世帯人員が増加するにつれて「年間年収」が増加することから、経済的な余裕が増すことにより一戸建てを選択する傾向がある結果となった。

## 参考文献

[1] 統計局ホームページ/ 家計

<http://www.stat.go.jp/data/getujidb/2i.htm>

[2] 統計局ホームページ/平成 26 年全国消費実態調査 用語の解説

<http://www.stat.go.jp/data/zensho/2014/kaisetsu.htm>

付録 規定課題

表 1

| 産業符号              | 男     | 女    | 合計    | 男     | 女    | 合計     |
|-------------------|-------|------|-------|-------|------|--------|
| 鉱業                | 3     | 0    | 3     | 7     | 0    | 7      |
| 建設業               | 2134  | 9    | 2143  | 6346  | 31   | 6377   |
| 製造業               | 5668  | 173  | 5841  | 17837 | 510  | 18347  |
| 電気・ガス・熱供給・水道業     | 222   | 0    | 222   | 697   | 0    | 697    |
| 情報通信業             | 547   | 0    | 547   | 2020  | 0    | 2020   |
| 運輸業               | 1235  | 3    | 1238  | 3852  | 6    | 3858   |
| 卸売・小売業            | 2083  | 201  | 2284  | 6644  | 589  | 7233   |
| 金融・保険業            | 472   | 39   | 511   | 1447  | 104  | 1551   |
| 不動産業              | 40    | 0    | 40    | 144   | 0    | 144    |
| 飲食店・宿泊業           | 50    | 32   | 82    | 158   | 107  | 265    |
| 医療・福祉             | 371   | 270  | 641   | 1022  | 810  | 1832   |
| 教育・学習支援業          | 1059  | 89   | 1148  | 3110  | 269  | 3379   |
| 複合サービス事業          | 321   | 3    | 324   | 981   | 11   | 992    |
| サービス業(他に分類されないもの) | 1816  | 106  | 1922  | 5842  | 326  | 6167   |
| 公務(他に分類されないもの)    | 3656  | 130  | 3786  | 10691 | 430  | 11121  |
| 不詳                | 9704  | 1591 | 11295 | 31204 | 4806 | 36011  |
| 合計                | 29381 | 2646 | 32027 | 92000 | 8000 | 100000 |

表 2

| 分位点 | 経常収入   | 食料    |
|-----|--------|-------|
| 20% | 310531 | 49701 |
| 40% | 404883 | 63146 |
| 60% | 501181 | 76055 |
| 80% | 635875 | 93036 |

経常収入(右)と食料費(左)のそれぞれのヒストグラム (単位: 円)

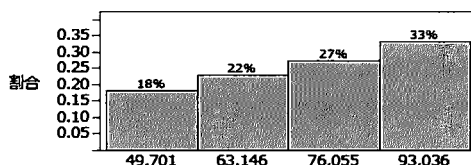
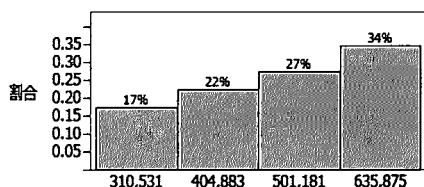


表 3 経常収入

| 5分位 5分位数    | 男     | 女    | 計      |
|-------------|-------|------|--------|
| 1 310531 未満 | 15494 | 4506 | 20000  |
| 2 310531 以上 | 18617 | 1377 | 19994  |
| 3 404883 以上 | 19099 | 911  | 20009  |
| 4 501181 以上 | 19356 | 640  | 19995  |
| 5 635875 以上 | 19435 | 567  | 20002  |
| 計           | 92000 | 8000 | 100000 |

表 3 食料費

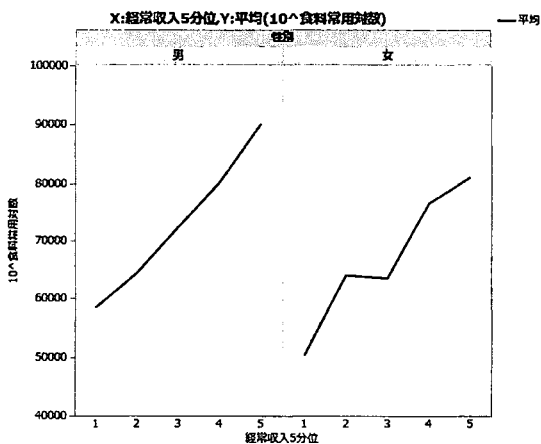
| 5 分位 5 分位数 | 男     | 女    | 計      |
|------------|-------|------|--------|
| 1 49701 未満 | 16803 | 3196 | 19999  |
| 2 49701 以上 | 18097 | 1903 | 20000  |
| 3 63146 以上 | 18734 | 1268 | 20001  |
| 4 76055 以上 | 19053 | 948  | 20001  |
| 5 93036 以上 | 19314 | 685  | 19999  |
| 計          | 92000 | 8000 | 100000 |

表 4

|     |   | 食料(5 分位)     |       |       |       |       |       |        |
|-----|---|--------------|-------|-------|-------|-------|-------|--------|
|     |   | 經常收入<br>5 分位 | 1     | 2     | 3     | 4     | 5     | 計      |
| 男   | 1 |              | 5827  | 3974  | 2798  | 1834  | 1060  | 15494  |
|     | 2 |              | 5130  | 4501  | 3978  | 2996  | 2012  | 18617  |
|     | 3 |              | 2911  | 4301  | 4395  | 4294  | 3198  | 19099  |
|     | 4 |              | 1906  | 3233  | 4082  | 4903  | 5231  | 19356  |
|     | 5 |              | 1029  | 2087  | 3481  | 5026  | 7813  | 19435  |
|     | 計 |              | 16803 | 18097 | 18734 | 19053 | 19314 | 92000  |
| 女   | 1 |              | 2462  | 994   | 565   | 314   | 172   | 4506   |
|     | 2 |              | 371   | 399   | 235   | 211   | 161   | 1377   |
|     | 3 |              | 243   | 225   | 216   | 164   | 63    | 911    |
|     | 4 |              | 68    | 134   | 156   | 143   | 140   | 640    |
|     | 5 |              | 53    | 152   | 96    | 117   | 149   | 567    |
|     | 計 |              | 3196  | 1903  | 1268  | 948   | 685   | 8000   |
| 男女計 |   |              | 19999 | 20000 | 20001 | 19998 | 20003 | 100000 |

表 5

|    |           | 食料    |        |       |
|----|-----------|-------|--------|-------|
| 性別 | 經常收入 5 分位 | 世帯数   | 常用対数平均 | 幾何平均  |
| 男  | 1         | 15494 | 4.7389 | 54814 |
|    | 2         | 18617 | 4.7832 | 60704 |
|    | 3         | 19099 | 4.8368 | 68675 |
|    | 4         | 19356 | 4.8826 | 76311 |
|    | 5         | 19435 | 4.9335 | 85793 |
|    | 計         | 92000 | 4.8395 | 69103 |
| 女  | 1         | 4506  | 4.6713 | 46913 |
|    | 2         | 1377  | 4.7811 | 60405 |
|    | 3         | 911   | 4.7813 | 60439 |
|    | 4         | 640   | 4.8617 | 72730 |
|    | 5         | 567   | 4.8774 | 75407 |
|    | 計         | 8000  | 4.7325 | 54013 |



# 収入と医療費の関係から見える「医療格差」に関して

参加カテゴリー：C)

村木 翔太

加唐 誠剛東

九州大学薬学府薬物動態学分野

## 1. はじめに

わが国では、1947年に日本国憲法が施行され、第25条、いわゆる「生存権」が規定されて以降、国民皆保険制度を基盤として、雇用保険、社会福祉、介護保険など様々な社会保障制度が整備されてきた[1]。例えば、医療保険制度は、すべての国民に医療サービスを提供するための制度であり、国民は一定割合の自己負担金で医療を受けることができる。こういった制度もあり、日本の医療制度は、基本的に公平であるとされているが、それでも「医療格差」は解消し切れていないとの報告もあり[2]、様々な要因による格差があることが知られている。

本研究では、擬似マイクロデータを用いて、収入と医療費の関係に着目し、「医療格差」の実態を調査することを目的とした。

## 2. 方法

### 年間収入と保健医療費の分布

年間収入と保健医療費について分布の特徴を把握するため、集計用乗率を用いて5分位点（20%、40%、60%および80%点）を求め、またそれぞれのヒストグラムを作成した。

### 年間収入別の保健医療費の関係

年間収入と保健医療費の関係を検討するため、保健医療費とその内訳項目（医薬品費、健康保持用摂取品費、保健医療用品・器具費、保健医療サービス費）について、年間収入の5分位別に平均を算出した。

### 年間収入別の医療費の負担

保健医療費が家計に与える負担と、収入の関係を検討するため、保健医療費とその内訳項目が消費支出に占める割合を求め、年間収入の5分位別に平均を算出した。

平成16年全国消費実態調査の個票データに基づいて、独立行政法人統計センターが作成した教育用擬似マイクロデータを使用した。また、解析ソフトはSAS 9.3を使用した。

## 3. 結果

### 年間収入と保健医療費の分布

年間収入と保健医療費について、集計用乗率を使った10万世帯比での5分位点と、ヒストグラムをそれぞれ表1、図1に示した。

表 1

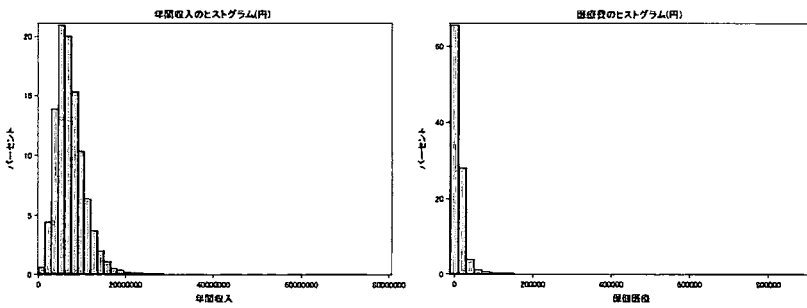
収入(円)

| 分位点 | 年間収入      | 収入総額      | 実収入     | 経常収入    |
|-----|-----------|-----------|---------|---------|
| 20% | 4,618,753 | 618,870   | 317,220 | 310,531 |
| 40% | 6,039,453 | 788,330   | 412,716 | 404,883 |
| 60% | 7,544,301 | 962,933   | 509,299 | 501,181 |
| 80% | 9,636,868 | 1,210,818 | 646,585 | 635,875 |

医療費(円)

| 分位点 | 保健医療   | 医薬品   | 健康保持用摂取品 | 保健医療用品・器具 | 保健医療サービス |
|-----|--------|-------|----------|-----------|----------|
| 20% | 2,774  | 339   | 0        | 234       | 1,183    |
| 40% | 5,442  | 739   | 56       | 651       | 2,563    |
| 60% | 8,730  | 1,305 | 324      | 1,421     | 4,538    |
| 80% | 14,887 | 2,358 | 1,048    | 3,051     | 8,162    |

図 1



年間収入別の保健医療費の関係

年間収入別の保健医療費とその内訳項目をまとめた(表 2)。表側は年間収入の 5 分位を、表体は保健医療費の金額と、その内訳項目の金額(医薬品費, 健康保持用摂取品費, 保健医療用品・器具費, 保健医療サービス費)それぞれの合計値を示している。

表 2 より, 年間収入が多い(少ない)世帯ほど, 保健医療費と内訳項目の費用は多い(少ない)傾向が見られた。

表 2

年間収入 vs. 医療費(円)

|              | 保健医療        | 医薬品        | 健康保持用摂取品   | 保健医療用品・器具  | 保健医療サービス    |
|--------------|-------------|------------|------------|------------|-------------|
| 年間収入         |             |            |            |            |             |
| 4,819,000円未満 | 62,037,877  | 8,539,343  | 4,803,978  | 12,282,932 | 36,931,729  |
| 4,819,000円以上 | 68,161,148  | 9,505,402  | 4,515,180  | 13,020,039 | 39,120,527  |
| 6,039,000円以上 | 71,383,228  | 10,634,496 | 5,615,419  | 13,925,549 | 41,207,761  |
| 7,544,000円以上 | 78,277,558  | 11,286,088 | 6,945,731  | 14,695,616 | 43,350,122  |
| 9,837,000円以上 | 88,702,220  | 12,939,011 | 8,276,133  | 15,744,401 | 49,740,674  |
| 合計           | 382,561,628 | 52,904,340 | 29,658,441 | 69,648,236 | 210,350,808 |

## 年間収入別の医療費の負担

年間収入と、医療費が消費支出に占める割合の関係を表 3 に示した。表側は年間収入の 5 分位を、表体は保健医療費の金額と、その内訳項目の金額（医薬品費、健康保持用摂取品費、保健医療用品・器具費、保健医療サービス費）を消費支出で除した数値の平均値を示している。

表 3 より、年間収入が多い（少ない）世帯ほど、医療費が消費支出に占める割合は低い（高い）傾向が見られた。保健医療費の各内訳項目に着目すると、健康保持用摂取品に限っては年間収入による割合の変化が見られなかったが（2.6～2.9%）、その他の項目では保健医療費と同様の傾向が見られた。

表 3

年間収入 vs. 消費支出における 医療費の割合 (%)

|              | 保健医療 | 医薬品  | 健康保持用摂取品 | 保健医療用品・器具 | 保健医療サービス |
|--------------|------|------|----------|-----------|----------|
| 年間収入         |      |      |          |           |          |
| 4,819,000円未満 | 4.05 | 0.81 | 0.27     | 0.83      | 2.35     |
| 4,819,000円以上 | 3.75 | 0.58 | 0.26     | 0.78      | 2.16     |
| 6,039,000円以上 | 3.60 | 0.55 | 0.28     | 0.70      | 2.07     |
| 7,544,000円以上 | 3.37 | 0.51 | 0.31     | 0.65      | 1.80     |
| 9,837,000円以上 | 3.09 | 0.48 | 0.29     | 0.57      | 1.78     |
| 合計           | 3.58 | 0.54 | 0.28     | 0.70      | 2.05     |

## 4. 考察・まとめ

まず、年間収入と医療費の関係を金額として比較してみると、年間収入が多い世帯と比べて年間収入が少ない世帯では医療費も少ないという結果が得られた。しかし、年間収入と医療費が消費支出に占める割合として比較したところ、年間収入が多い世帯と比べて年間収入が少ない世帯では医療費が消費支出に占める医療費の割合は高いという結果が得られた。これは、医療費が単純に収入に依存して出費されているわけではないことを示唆している。そのため、収入の低い世帯では、全体の消費支出に占める医療費の割合が大きくなり、高収入世帯に比べその負担が相対的に大きくなっていることが考えられる。

さらに、保健医療費が消費支出に占める割合を細分化した項目ごとに見てみると、医薬品、保健医療サービス、保険医療用品・器具の 3 項目と、健康保持用摂取品の 1 項目で、年間収入との関係に違いが見られた。前者の 3



項目では、保健医療で見られた傾向と同様に、収入が低い世帯での負担が相対的に大きくなっているが、健康保持用摂取品に関しては、収入による消費支出に占める割合にほとんど違いが見られなかった。前者の3項目は、疾病治療の目的で出費される項目である一方で、健康保持用摂取品は栄養成分の補給など健康増進のために用いる食品であり、疾病予防の目的で出費される項目であると予想される。そのため、収入が多く家計に余裕のある世帯で、健康保持用摂取品をより購入する傾向にあるのではないかと考えられる。

本邦では、医薬品、保健医療サービス、保険医療用品・器具は、ほとんどが医療保険の対象となっている。このように全国民が健康な生活を送ることに必要な医療サービスは医療保険の対象になっており、低所得者でも必要な医療を受けられることが理想だと考える。しかし、医療保険の自己負担割合は、高額医療制度等の一部の例を除き、原則収入によらず一定である。本研究でも、低所得者と高所得者の家計における医療費負担は同じであるとは言える結果ではなかった。通常の医療費の自己負担割合においても収入による変動を導入することで、収入による医療費負担の相対的格差を是正することができるかもしれない。

## 5. 参考

[1] 厚生労働省 「平成 24 年版厚生労働白書 - 社会保障を考える -」

< <http://www.mhlw.go.jp/wp/hakusyo/kousei/12/> >

[2] 池上直己 「医療における格差—構造的特性と政策的対応」

*医療経済研究* 18(1), 5-21, (2006)

## 付録 規定課題

以下に規定課題の表及び図を示す。ソフトウェアはSAS 9.4を使用した。

### 1) 産業符号・性別 クロス表

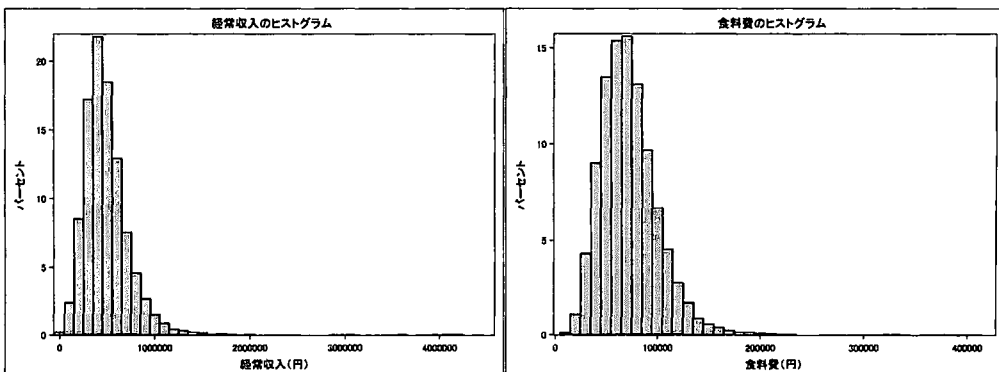
表1

| 産業符号              | a) 単純クロス表 |       |        | b) 10万世帯比でのクロス表 |       |         |
|-------------------|-----------|-------|--------|-----------------|-------|---------|
|                   | 男         | 女     | 合計     | 男               | 女     | 合計      |
| 不詳                | 9,704     | 1,591 | 11,295 | 31,204          | 4,808 | 36,011  |
| 鉱業                | 3         | 0     | 3      | 7               | 0     | 7       |
| 建設業               | 2,134     | 9     | 2,143  | 6,346           | 31    | 6,377   |
| 製造業               | 5,868     | 173   | 5,841  | 17,837          | 510   | 18,347  |
| 電気・ガス・熱供給・水道業     | 222       | 0     | 222    | 697             | 0     | 697     |
| 情報通信業             | 547       | 0     | 547    | 2,020           | 0     | 2,020   |
| 運輸業               | 1,235     | 3     | 1,238  | 3,852           | 6     | 3,858   |
| 卸売・小売業            | 2,083     | 201   | 2,284  | 6,844           | 589   | 7,233   |
| 金融・保険業            | 472       | 39    | 511    | 1,447           | 104   | 1,551   |
| 不動産業              | 40        | 0     | 40     | 144             | 0     | 144     |
| 飲食店・宿泊業           | 50        | 32    | 82     | 158             | 107   | 265     |
| 医療・福祉             | 371       | 270   | 641    | 1,022           | 810   | 1,832   |
| 教育・学習支援業          | 1,059     | 89    | 1,148  | 3,110           | 269   | 3,379   |
| 複合サービス事業          | 321       | 3     | 324    | 981             | 11    | 992     |
| サービス業(他に分類されないもの) | 1,816     | 106   | 1,922  | 5,842           | 326   | 6,167   |
| 公務(他に分類されないもの)    | 3,656     | 130   | 3,786  | 10,891          | 430   | 11,121  |
| 合計                | 29,381    | 2,646 | 32,027 | 92,000          | 8,000 | 100,000 |

### 2) 経常収入および食料費

表2

| 分位点 | 経常収入    | 食料費    |
|-----|---------|--------|
| 20% | 310,531 | 49,701 |
| 40% | 404,883 | 63,146 |
| 60% | 501,181 | 76,055 |
| 80% | 635,875 | 93,036 |



3) 男女別・経常収入および食料費の5分位分布

表3

表3

| 経常収入 |           |        |       | 食料費     |     |          |        |       |         |
|------|-----------|--------|-------|---------|-----|----------|--------|-------|---------|
| 5分位  | 5分位数      | 男      | 女     | 計       | 5分位 | 5分位数     | 男      | 女     | 計       |
| 1    | 310,531未満 | 15,494 | 4,506 | 20,000  | 1   | 49,701未満 | 16,803 | 3,198 | 19,999  |
| 2    | 310,531以上 | 18,617 | 1,377 | 19,993  | 2   | 49,701以上 | 18,097 | 1,903 | 20,000  |
| 3    | 404,883以上 | 19,099 | 911   | 20,009  | 3   | 63,148以上 | 18,734 | 1,288 | 20,001  |
| 4    | 501,181以上 | 19,356 | 840   | 19,995  | 4   | 76,055以上 | 19,053 | 948   | 20,001  |
| 5    | 635,875以上 | 19,435 | 567   | 20,002  | 5   | 93,036以上 | 19,314 | 685   | 19,999  |
| 計    |           | 92,000 | 8,000 | 100,000 | 計   |          | 92,000 | 8,000 | 100,000 |

4) 男女別・経常収入・食料費についての3重クロス表

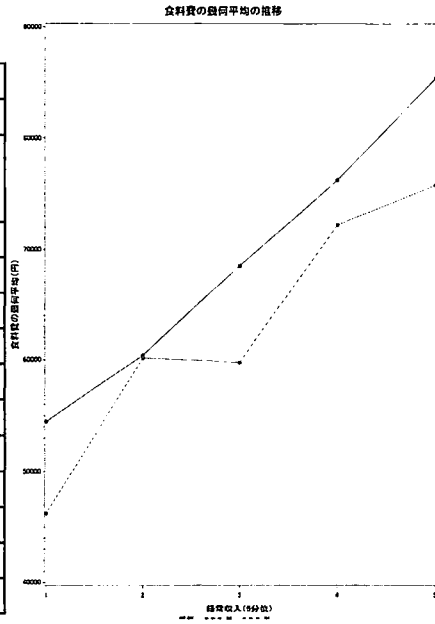
表4

| 性別  | 経常収入5分位 | 食料(5分位) |        |        |        |         | 計      |
|-----|---------|---------|--------|--------|--------|---------|--------|
|     |         | 1       | 2      | 3      | 4      | 5       |        |
| 男   | 1       | 5,827   | 3,974  | 2,798  | 1,834  | 1,060   | 15,494 |
|     | 2       | 5,130   | 4,501  | 3,978  | 2,998  | 2,012   | 18,617 |
|     | 3       | 2,911   | 4,301  | 4,395  | 4,294  | 3,198   | 19,099 |
|     | 4       | 1,906   | 3,233  | 4,082  | 4,903  | 5,231   | 19,356 |
|     | 5       | 1,029   | 2,087  | 3,481  | 5,028  | 7,813   | 19,435 |
|     | 計       | 16,803  | 18,097 | 18,734 | 19,053 | 19,314  | 92,000 |
| 女   | 1       | 2,482   | 994    | 565    | 314    | 172     | 4,506  |
|     | 2       | 371     | 399    | 235    | 211    | 161     | 1,377  |
|     | 3       | 243     | 225    | 216    | 164    | 63      | 911    |
|     | 4       | 88      | 134    | 156    | 143    | 140     | 640    |
|     | 5       | 53      | 152    | 98     | 117    | 149     | 567    |
|     | 計       | 3,198   | 1,903  | 1,268  | 948    | 685     | 8,000  |
| 男女計 | 19,999  | 20,000  | 20,001 | 20,001 | 19,999 | 100,000 |        |

5) 男女別・経常収入別に食料費を常用対数変換した平均値の比較

表5

|   |             | 食料費    |        |        |
|---|-------------|--------|--------|--------|
|   |             | 世帯数    | 常用対数平均 | 幾何平均   |
| 男 | 経営収入<br>5分位 |        |        |        |
|   | 1           | 15,494 | 4.7363 | 54,485 |
|   | 2           | 18,817 | 4.7811 | 60,402 |
|   | 3           | 19,099 | 4.8357 | 68,502 |
|   | 4           | 19,356 | 4.8822 | 76,244 |
|   | 5           | 19,435 | 4.9313 | 85,371 |
|   | 計           | 92,000 | 4.8376 | 68,805 |
| 女 | 1           | 4,508  | 4.6651 | 48,245 |
|   | 2           | 1,377  | 4.7792 | 60,144 |
|   | 3           | 911    | 4.7783 | 59,745 |
|   | 4           | 640    | 4.8587 | 72,226 |
|   | 5           | 567    | 4.8796 | 75,791 |
|   | 計           | 8,000  | 4.7252 | 53,117 |



## 付録 規定課題の SAS プログラム

```
/*-----*/
/* プログラム①
/*-----*/
/* ① read_giji_zensho(all_at_once).sas */ options nocenter macrogen;
*以下の5つのマクロ変数を指定してください。;
%let path=c:\¥SAScontest; *「全消擬似マイクロデータ・プロジェクト」用 path 名;
%let ds_folder=zensho_permanent; *パーマネント SAS データセット用フォルダ名;
%let csvFolder=DATA; *教育用擬似マイクロデータの CSV ファイルが保存されるフォルダ名;
%let dsname=zensho; *マイクロデータから作成されるパーマネント SAS データセット名;
%let SurveyName=全国消費実態調査：教育用擬似マイクロデータ; *調査名:結果の表示時に使用される;
libname zensho "&path¥&ds_folder";
*全国消費実態調査:擬似マイクロデータの全ての CSV ファイル名の指定(ワイルドカード使用);
filename FL "&path¥&csvFolder¥*.csv";
data zensho.&dsname;
    length X1-X197 $15; *全ての変数を一旦文字型変数として取り込む;
    infile FL dsd LRECL=32767;
    input X1-X197;
run;
proc print data=zensho.&dsname (obs=10); title "①-1 &SurveyName (先頭 10 件)"; run;
data zensho.&dsname; set; drop i X1-X197; *文字型変数を数値型変数に変換;
    array X {197}; *文字型変数;
    do i=1 to 13; if index(X[i],"V") then X[i]="." ; *数値型の欠損値(不詳);
        if X[i]=" " then X[i]="999"; *数値型のコード(非..);
    end;
    array F {197} 8; *数値変数;
    do i=1 to 197; F[i]=X[i]; end; *文字型変数を数値型変数に変換;
run;
proc print data=zensho.&dsname (obs=10); title "①-2 &SurveyName (先頭 10 件)"; run;
```

```

/*-----*/
/* プログラム②
/*-----*/
/* ② giji_zensho_codetable_import.sas */ *符号表を SAS データセットに変換;
%let Codetable=GII_2004zensho_layout_code.csv; *符号表のファイル名;;
data zensho.code_table;
length X1-X11 $100;
infile "&path%&CodeTable" dsd firstobs=9;
input X1-X11;
run;
data zensho.code_table; *項目名に対応する変数名に rename する;
  set;
  rename X1=lineNO
         X2=Vlabel
         X3=strata
         X4=position
         X5=bytes
         X6=xhaichi
         X7=Vtype
         X8=xshubetsu
         X9=variable
         X10=Vcode
         X11=VcodeContents;
run;
proc print data=zensho.code_table; title "② &SurveyName.符号表"; run;

```

```

/*-----*/
/* プログラム③
/*-----*/
/* ③ label_rename.sas */ *教育用擬似マイクロデータの変数ラベル作成と変数の rename;
filename out1 "&path%label.txt"; *変数ラベル定義用テキスト;
filename out2 "&path%rename.txt"; *変数 F1~F197 の rename 用テキスト;
* 全ての変数に変数ラベルを付すための LABEL 文の中味を外部ファイルに出力;
data label_define; keep var_label variable Vlabel;
  file out1; *LABEL 文の中味を外部ファイルに出力;
  length var_label $ 100;
  set zensho.code_table; if variable=" " then delete;
  * 「Y」で始まる変数名を後で全て「Y」で始まる変数名(★数値型★)に置換するための
  rename 文用変数名の作成;
  variable=tranwrd(variable,"Youto00","Y");
  variable=tranwrd(variable,"Youto0","Y");
  variable=tranwrd(variable,"Youto","Y");
  var_label=compress(variable) || '=' || compress(Vlabel || '');
  put var_label; *最終的に作成された全ての変数に付与するラベル文の作成;
run;
proc print; title "③-1 label"; var var_label variable Vlabel; run;
data _null_;
  file out2;
  *変数 F1~F197 を「符号表」の変数名に変える rename 文の中味をテキストファイルに出力;
  set label_define;
  statement=compress("F" || _N_ || "=" || variable);
  put statement;
run;
data zensho.&dsname; *変数 F1~F197 を rename をする;
  set zensho.&dsname;
  rename
  %include "&path%rename.txt";
run;
proc print data=zensho.&dsname(obs=5); title "③-2 変数 F1~F197 の rename: (先頭 5 件)"; run;
data zensho.&dsname; *全ての変数にラベルを付ける; set; label %include "&path%label.txt"; run;
*マイクロデータのパーマネントデータセット完成;
proc print data=zensho.&dsname(obs=5);
  title "③-4 &SurveyName. パーマネント SAS データセット(先頭 5 件:変数名表示)"; run;
proc print data=zensho.&dsname(obs=5) label;
  title "③-5 &SurveyName. パーマネント SAS データセット(先頭 5 件:変数ラベル表示)"; run;

```

```

/*-----*/
/* プログラム④
/*-----*/
/* ④ create_format.sas */ /*マイクロデータ用の proc format 作成;
filename out3 "&path%program%giji_micro_format.sas";
data format; keep formatName Vcode VcodeContents;
    set zensho.code_table; /*符号表のデータセットを使用;
    retain sw formatName;
    if Vlabel="," then sw=0;
    if variable NE " " AND Vtype=" " /*文字型変数*/
        then do; sw=1; formatName=variable; output; return; end;
    if sw=1 then output;
run;
proc print; title "④-1 format 対象変数(1)"; var formatName Vcode VcodeContents; run;
*★===== dataset 'format'に対して、「V」と「△」の処理を追加する=====★;
data format; set format;
    if index(Vcode,"V") > 0 then Vcode=".";
    if (formatName="S1_KigyokuKibo" OR formatName="S1_KigyokuKibun") AND Vcode="△"
        then Vcode="999";
    else Vcode=kcompress(Vcode,"△");
run;
proc print; title "④-2 V converted to period"; run;
*★===== dataset 'format'に対する「V」と「△」の追加処理終了=====★;
proc sort data=format out=format1; by formatName; run;
data format1; set format1; by formatName;
    if first.formatName then statement2='value ' || compress(formatName || 'X');
    statement=compress(Vcode || '=' || VcodeContents || '');
run;
proc print;
    title "④-3 format 対象変数(2)";
    var formatName Vcode VcodeContents statement statement2;
run;
data _null_; file out3;
    set format1 end=final; by formatName;
    if _n_=1 then put "proc format; *&SurveyName;";
    if first.formatName then put statement2;
    if last.formatName then put "    " statement ";" /;
        else put "    " statement;
    if final then put "run;";
run;

```



/\*-----\*/

/\* プログラム⑤

/\*-----\*/

proc format; \*全国消費実態調査：教育用擬似マイクロデータ；

value KouzouX

1="木造"  
2="防火木造"  
3="鉄骨・鉄筋コンクリート造"  
4="その他（ブロック造り、レンガ造りなど）"  
.="不詳";

value S1\_AgeX

1="24歳未満"  
2="25～29歳"  
3="30～34歳"  
4="35～39歳"  
5="40～44歳"  
6="45～49歳"  
7="50～54歳"  
8="55～59歳"  
9="60～64歳"  
10="65～69歳"  
11="70～74歳"  
12="75歳以上"  
.="不詳";

value S1\_KigyouKiboX

1="1～4人"  
2="5～29人"  
3="30～499人"  
4="500～999人"  
5="1000人以上"  
999="非就業又は官公"  
.="不詳";

value S1\_KigyouKubunX

1="民営"  
2="自営"  
3="官公"  
999="非就業"  
.="不詳";

value S1\_SangyouX

1="農業"  
2="林業"  
3="漁業"

4="鉱業"  
5="建設業"  
6="製造業"  
7="電気・ガス・熱供給・水道業"  
8="情報通信業"  
9="運輸業"  
10="卸売・小売業"  
11="金融・保険業"  
12="不動産業"  
13="飲食店・宿泊業"  
14="医療・福祉"  
15="教育・学習支援業"  
16="複合サービス事業"  
17="サービス業（他に分類されないもの）"  
18="公務（他に分類されないもの）"  
19="その他（非就業を含む）"  
.= "不詳";

value S1\_SexX

1="男"  
2="女";

value S1\_ShokugyouX

1="常用労務作業者"  
2="臨時及び日々雇労務作業者"  
3="民間職員"  
4="官公職員1"  
5="官公職員2"  
6="商人及び職人"  
7="個人経営者"  
8="農林漁業従事者"  
9="法人経営者"  
10="自由業者"  
11="その他"  
12="無職"  
.= "不詳";

value S1\_ShugyouX

1="就業"  
2="うちパート"  
3="非就業"  
4="うち仕事を探している"  
.= "不詳";

value SetaiKubunX

1="勤労"

2="勤労以外"

3="無職";

value ShoyuuX

1="持ち家（世帯員名義）"

2="持ち家（その他名義）"

3="民営賃貸住宅（設備専用）"

4="民営賃貸住宅（設備共用）"

5="県市区町村営賃貸住宅"

6="都市再生機構・公社等賃貸住宅"

7="社宅・公務員住宅（借上げ含む）"

8="借間"

9="寮・寄宿舎"

.="不詳";

value TatakataX

1="一戸建"

2="長屋建"

3="共同住宅（1・2階建）"

4="共同住宅（3～5階建）"

5="共同住宅（6～10階建）"

6="共同住宅（11階建以上）"

7="その他"

.="不詳";

run;

```

/*.....*/
/* プログラム⑥
/*.....*/

%let program_path=C:\SAScontest\program;
*① ミクロデータ本体を SAS データセットに取り込む;
%include "&program_path\read_giji_zensho(all_at_once).sas";
*② 符号表を SAS データセットに変換;
%include "&program_path\giji_zensho_codetable_import.sas";
*③ 教育用擬似ミクロデータの変数ラベル作成と変数の rename;
%include "&program_path\label_rename.sas";
*④ proc format を作成する;
%include "&program_path\create_format.sas";
*⑤ format の実行;
%include "&program_path\giji_micro_format.sas";

/* 規定課題 1 */
data zensho.&dsname;
set zensho.&dsname;
Weight2 = Weight/495465*100000;
run;

ODS PDF FILE="c:\SAScontest\規定課題 1.pdf";
proc tabulate data=zensho.&dsname MISSING format=comma8.;
title "表 1";
class S1_Sangyou S1_Sex;
var Weight2;
tables S1_Sangyou=" ALL="合計", N="a) 単純クロス表"(S1_Sex=" ALL="合計) Weight2="b) 10 万世帯比
でのクロス表"(S1_Sex="sum=" ALL="合計"*sum=") / MISSTEXT="0" BOX="産業符号";
format S1_Sangyou S1_SangyouX. S1_Sex S1_SexX.;
run;
ODS PDF CLOSE;

/* 規定課題 2 */
ODS PDF FILE="c:\SAScontest\規定課題 2.pdf";
proc tabulate data=zensho.&dsname MISSING format=comma15.;
title "表 2";
var Y4 Y38;
weight Weight;
tables (P20="20%" P40="40%" P60="60%" P80="80%"), (Y4="経常収入" Y38="食料費") / MISSTEXT="0"
BOX="分位点";
run;

```

```

proc format;
value clKeijoushunyū low<310531 ="1"
      310531-<404883 ="2"
      404883-<501181 ="3"
      501181-<635875 ="4"
      635875-high ="5";

run;

proc format;
value clShokuryouhi low<49701 ="1"
      49701-<63146 ="2"
      63146-<76055 ="3"
      76055-<93036 ="4"
      93036-high ="5";

run;

proc format;
value clKeijoushunyū_lb low<310531 =" 1          310,531 未満"
      310531-<404883 =" 2          310,531 以上"
      404883-<501181 =" 3          404,883 以上"
      501181-<635875 =" 4          501,181 以上"
      635875-high =" 5          635,875 以上";

run;

proc format;
value clShokuryouhi_lb low<49701 =" 1          49,701 未満"
      49701-<63146 =" 2          49,701 以上"
      63146-<76055 =" 3          63,146 以上"
      76055-<93036 =" 4          76,055 以上"
      93036-high =" 5          93,036 以上";

run;

proc sgplot data=zensho.&dsname;
title "經常収入のヒストグラム";
histogram Y4;
axis label = "經常収入 (円) ";
run;

proc sgplot data=zensho.&dsname;
title "食料費のヒストグラム";
histogram Y38;
axis label = "食料費 (円) ";

```

```

run;
ODS PDF CLOSE;

/* 規定課題 3*/
ODS PDF FILE="c:\¥SAScontest¥規定課題 3.pdf";
proc tabulate data=zensho.&dsname MISSING format=comma7.;
title "表 3";
class Y4 S1_Sex;
classlev Y4 / style=[cellwidth=4 cm];
var weight2;
tables (Y4="5 分位          5 分位数" ALL="計")*weight2=""*sum="", S1_Sex="" ALL="計" / RTS=16
MISSTEXT="0" BOX="經常收入" ROW=FLOAT;
format Y4 clKeijoushunyub_lb. S1_Sex S1_SexX.;
run;

proc tabulate data=zensho.&dsname MISSING format=comma7.;
title "表 3";
class Y38 S1_Sex;
classlev Y38 / style=[cellwidth=4 cm];
var weight2;
tables (Y38="5 分位          5 分位数" ALL="計")*weight2=""*sum="", S1_Sex="" ALL="計" / MISSTEXT="0"
BOX="食料費" ROW=FLOAT;
format Y38 clShokuryouhib_lb. S1_Sex S1_SexX.;
run;
ODS PDF CLOSE;

/* 規定課題 4*/
ODS PDF FILE="c:\¥SAScontest¥規定課題 4.pdf";
proc tabulate data=zensho.&dsname MISSING format=comma7.;
title "表 4";
class Y4 Y38 S1_Sex;
var weight2;
tables ((S1_Sex="性別" * (Y4="經常收入 5 分位" ALL="計")) ALL="男女計") * weight2=""*sum="", Y38="食料
(5 分位)" ALL="計" / MISSTEXT="0" ROW=FLOAT;
format Y4 clKeijoushunyub. Y38 clShokuryouhib. S1_Sex S1_SexX.;
run;
ODS PDF CLOSE;

/* 規定課題 5*/
data zensho.&dsname;
set zensho.&dsname;
LOG_Y38 = log10(Y38);

```

```

run;

proc tabulate data=zensho.&dsname format=comma7. out=task5(drop=_type_ _page_ _table_);
title "表 5";
class Y4 Y38 S1_Sex;
var weight2 LOG_Y38;
tables (S1_Sex * (Y4 ALL)), ALL*weight2 LOG_Y38*mean*f=comma7.4 / MISSTEXT="0";
format Y4 clKeijoushuny. S1_Sex S1_SexX.;
run;

data task5;
set task5;
GM = 10**(LOG_Y38_Mean);
input Y4_ch $;
cards;
1
2
3
4
5
1
2
3
4
5
計
計
;
run;

ODS PDF FILE="c:\SAScontest\規定課題 5.pdf";
proc tabulate data=task5 MISSING format=comma7. order=data;
title "表 5";
class Y4_ch S1_Sex;
var Weight2_Sum LOG_Y38_Mean GM;
tables (S1_Sex=" " * Y4_ch="經常收入 5 分位"), ALL=""*Weight2_Sum="世帯数"*sum=" " ALL="食料費
"*LOG_Y38_Mean="常用対数平均"*sum=""*f=comma7.4 ALL=""*GM="幾何平均"*sum=" " / MISSTEXT="0";
format S1_Sex S1_SexX.;
run;

SYMBOL1 I=join V=dot C=blue;
SYMBOL2 I=join V=dot C=red;

```

```
axis3 label=(h=1.5 "經常収入 (5 分位) ");
axis4 label=(h=1.5 ANGLE=90 ROTATE=0 "食料費の幾何平均 (円) ");
proc gplot data=task5 (obs=10);
title "食料費の幾何平均の推移";
plot GM * Y4_ch=S1_Sex / haxis=AXIS3 vaxis=AXIS4;
run;
quit;
ODS PDF CLOSE;
```

#### 【参考プログラム】

擬似マイクロデータの整理にあたり，統計情報研究開発センターの周防節雄氏のプログラム (<http://mighty.gk.u-hyogo.ac.jp/confidential/suoh20141121.html>) を，プログラム①～⑤として用いた。



## 業種で見る中小企業の収入 ～大企業との比較～

チーム名 : FEG\_Freshers  
チームリーダー : 野田 斉  
チームメンバー : 山内健太、立花潤  
(株式会社金融エンジニアリング・グループ)

### 要旨

中小企業勤務者のうち、大企業勤務者と比較して高収入かつ時間的余裕を持つ者の割合を業種別に集計した。結果として、就職・転職活動における業界選びの参考となる知見が得られた。

キーワード：企業規模、世帯実収入、中央値、野菜・海藻、健康保持用摂取品、旅行

### 1. はじめに

就職・転職時の企業選びにおいて、企業規模は無視できない項目である。一般的には「大企業」であるほど高収入である、福利厚生が厚い、安定しているといったイメージが持たれている。「大企業病（大手病）」といった言葉も就職活動時にはよく聞かれ、世間の大企業への選好度合いは大きい。しかし、日本国内の実に99%以上は「中小企業」であり、これらの企業に対して過度の偏見を持つことは好ましくない。そこで、本稿では中小企業の中でも大企業と比べて遜色のない収入を得ていて、かつ時間的な余裕のある生活を送っている人の割合（QOL<sup>1</sup>が高い人の割合）を業界別に調査する。一口に中小企業といっても、どの業界であれば適切な収入と労働時間で生活できる確率が高いかを知ることは、就職・転職活動においても参考になると考えられる。

### 2. 収入の現状

まずは世間で持たれているイメージの通り、中小企業勤務者の収入は大企業勤務者のそれと比較して本当に低いのか確認する<sup>2</sup>。図表 2-1 は世帯主の年齢階級ごとに世帯月収の中央値をグラフ化したものである。（本稿では収入としてデータ中の「Youto003（各世帯の実収入）」を使用している<sup>3</sup>。）この図を見る限り、全ての年齢階級で世帯主が中小企業勤めの場合の世帯月収は大企業勤めの場合を下回っている。

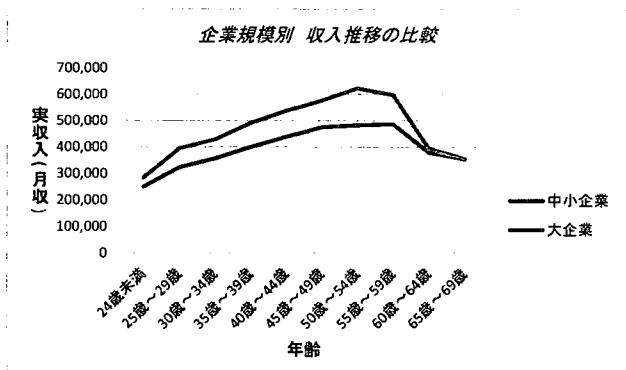
<sup>1</sup> QOLとは Quality of Life の略語であり、個人の人生の質、生活の質を指す。人間らしい生き方ができているか、人生に幸福を感じているかという観点の言葉であり、単純に収入や財産などの生活水準を指す言葉ではない。従って本稿では、収入に加えて労働時間を考慮した場合に QOL という言葉を用いる。

<sup>2</sup> 本稿では社員数が 5～499 人の場合を中小企業、1000 人以上の場合を大企業としている。大企業に 500～999 人規模を含めない理由は、一般に超大企業（大手企業）と認識される企業は 1000 人を超える規模だという想定のため、中小企業が収入や時間面でそれらの企業に優るかどうかを調査するためである。また 1～4 人規模の企業は零細企業として、中小企業には含めていない。

<sup>3</sup> 本来であれば世帯主個人の収入で分析するべきだが、データから正確に世帯主個人の収入を把握できないことから、世帯収入を使用している。

次に次章「3.産業別の高収入確率」では、中小企業勤めであっても大企業に優る収入を得ている者が、どの業界に多く属しているのかを確認する。

図表 2-1



### 3. 産業別の高収入確率

まず中小企業でありながらも「大企業と比べて高収入」という状態の定義として、大企業の実収入の中央値を超えれば「高収入」とする。中央値は大企業の年齢階級ごとに算出<sup>4</sup>し、中小企業と大企業で同一の年齢階級同士を比較する。そして各階級で大企業を上回った人の割合を業種ごとに求める。また、対象とする年齢階級は”20歳未満”から”60歳~64歳”までとしている。この方法によって、各業界で「大企業と比べて高収入」になることのできる確率を推定する<sup>5</sup>。

図表 3-1 は集計の結果である。

図表 3-1 業界別 高収入比率

| 順位 | 業界        | 高年収比率(%) | 高年収者数(人) |
|----|-----------|----------|----------|
| 1  | 電気ガス熱供給水道 | 67.2     | 117      |
| 2  | 鉱業        | 59.9     | 20       |
| 3  | 教育学習支援    | 57.4     | 952      |
| 4  | 金融保険      | 43.3     | 550      |
| 5  | 不動産       | 42.7     | 154      |
| 6  | 情報通信      | 42.3     | 1,796    |
| 7  | 複合サービス事業  | 38.6     | 509      |
| 8  | 医療福祉      | 35.7     | 1,851    |
| 9  | サービス業     | 31.7     | 6,305    |
| 10 | 製造        | 29.2     | 12,920   |
| 11 | 卸売り小売     | 25.8     | 6,040    |
| 12 | 建設        | 25.6     | 6,075    |
| 13 | 飲食宿泊      | 24.9     | 224      |
| 14 | 運輸        | 23.6     | 2,550    |
| -  | 不詳        | 27.3     | 21,265   |
| -  | 業界全体      | 28.5     | 61,326   |

<sup>4</sup> 大企業の実収入の中央値はそれぞれ (24歳未満=284679.51) (25歳~29歳=395026.07) (30歳~34歳=428615.90) (35歳~39歳=490612.44) (40歳~44歳=536533.47) (45歳~49歳=577023.80) (50歳~54歳=622720.55) (55歳~59歳=596912.52) (60歳~64歳=393320.21) (不詳=472963.68) である。単位は円/月。

<sup>5</sup> 確率を推定するとしたが、ここではつまり、「大企業と比べて高収入」だとされる人数の割合を産業別に算出するということである。

この表は企業規模が中小であっても、大企業の実収入の中央値を超えている人（高収入の人）の比率を示している。高収入比率が50%を超えている業界は、全体の傾向として大企業よりも中小企業のほうが高収入ということになるが、これは直感に反する結果である。この結果を信じるのであれば、これらの業界（電気・ガス・熱供給・水道業、鉱業<sup>6</sup>、教育学習支援）への就職を狙ってみるのも良いだろう。

高年収比率が40%台の業界（金融保険業、不動産業、情報通信業）は傾向として大企業よりも低収入であるが、この集計結果では上位に位置していることから、他業界との比較において高収入になる確率が高いと言える。それに対して高収入比率が20%台の業界（製造業、卸売り小売業、建設業、飲食宿泊業、運輸業）に属した場合は、集計結果によると高収入になる確率が低いと言える。

#### 4. 「働きすぎ」の懸念

収入が大きければ良いという訳でもなく、得られる収入に対して業務負担が過多であれば、余裕のある生活を送れているとは言い難い。特に、高収入であるほど長時間労働をしている可能性も考えられる。そこで新たに労働時間の指標を導入し、中小企業の高収入者が大企業に比べて長時間労働をしていないか検証する。しかし労働時間を示す変数はデータに含まれていないため、労働時間指標を新たに作成する必要がある。本稿では「野菜指標」と「旅行指標」の2つを作成した<sup>7</sup>。

まず「野菜指標」については式4-1によって求める。

$$\text{式 4-1} \quad \text{野菜指標} = \frac{\text{健康保持用摂取品への支出額}}{\text{野菜・海藻への支出額}}$$

野菜への支出が少なく（分母が小さい）、健康保持用摂取品への支出が多い（分子が大き）い場合にこの指標の数値は大きくなる。したがってこの数値が大きいくほど、野菜の出費に対して健康保持用摂取品の支出が大きいくということである。野菜指標の値が大きいく状態を本稿では「野菜を食べる時間がないので、栄養を摂る手段を健康保持用摂取品に置き換えている」と解釈する（つまりこの指標は「小さい」ほど好ましい<sup>8</sup>）。さらにこれを「野菜を調理する時間がないために仕方なくサプリメントを服用している」と解釈することで、「野菜指標」を「時間に余裕があるか」の指標としている。比較のために大企業内で野菜指標の平均を計算し（平均は0.11）、これを基準とする。そして中小企業に勤める個人の野菜指標値が、この平均値より下回れば野菜指標による基準を達成したとする。

次に「旅行指標」は式4-2によって計算する。

$$\text{式 4-2} \quad \text{旅行指標} = \text{宿泊費} + \text{パック旅行費}$$

<sup>6</sup> 鉱業についてはサンプル数が少なく、高齢層に偏っていたので信頼できる結果ではない。

<sup>7</sup> 分析したデータは全国消費実態調査（の擬似マイクロデータ）であるため、データのほとんどが収入と支出に関する項目である。そこから時間に関する指標を作成する方法はいくつも考えられるが、本稿では比較的豊富な“食料”と“教養娯楽”に関するデータを活用することを考えた。

<sup>8</sup> しかし健康保持用摂取品への支出額が0の場合は、この指標の値も0になる。この状態は「十分に野菜を摂れているのでサプリメントは必要ない」のか、「健康意識が低く、十分に野菜を摂れていないにも関わらずサプリメントも服用していない」のか区別がつかない。よってこの指標が基準を下回ることには大きな意味は無いが、少なくとも基準を上回ってしまうことは好ましくない。

これは単純に時間がある人ほど旅行に行くことと仮定することで時間指標に成り得ている。こちらにも比較のため大企業の平均を求めている（5739.20円）。旅行指標の値が大企業平均よりも上回れば、旅行に行く時間的余裕があると言えるので基準達成とする。

これらの時間指標を使って、中小企業の高収入者の中で「大企業と比べて時間に余裕がある<sup>9</sup>人」の比率を業種ごとに算出する。図表 4-1 は業界別の高収入比率（図表 3-1）をグラフ化したもの、図表 4-2<sup>10</sup>、図表 4-3 はそれぞれ野菜指標と旅行指標による高収入者の時間指標達成率のグラフである。また図表 4-4 は各比率における順位の変動を業界毎にまとめたものであり、順位の変動の際に参照願いたい。

まず野菜指標を導入した場合、図表 4-1,4-2 を比較すると、順位の変動は各所に起きているが、特に「不動産業」の変動が顕著である。順位が最下位に落ちている上、指標達成率自体も約 35%程であり、他業界が 60%を越えている中で一際低い数値である。この結果から、不動産業は高収入になることと引き換えに長時間労働をする必要があることが推察される。同様に「教育学習支援業」も大きく順位を落としている事がわかる。また図表 4-2 を見ると、不動産業界を除いて時間指標達成率はおおむね 70%を超えている。これは一般に高収入だからといって長時間労働をしている（働きすぎている）という懸念はなく、むしろ高収入である人ほど労働時間が短い傾向を示唆している。この理由として、例えば高収入な人ほど優秀であり、効率的に業務を遂行した結果、労働時間を短く抑えることができているといったものが考えられる。

次に旅行指標を導入した場合、野菜指標と同様に「不動産業」と「教育学習支援業」は大きく順位を落としている（図表 4-1,4-3）。特に不動産業は指標達成率が約 50%と、他業界（おおむね 60%以上）と比較して低い数値である。したがって収入と引き換えに労働時間が増加していることが推察される。また旅行指標を導入した場合に顕著な結果として、「金融保険業」が順位を下げていることが挙げられる。「電気・ガス・熱供給・水道業」も順位を落としているが、図表 3-1 で当該業の高収入者数を見ると非常に小さい数値である。そのためサンプル数が少ないことで変動が大きく出たと考える方が自然である。

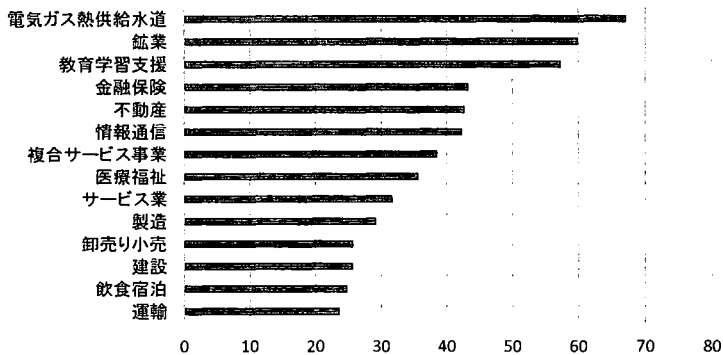
以上の結果をまとめると、不動産業は高収入となるにあたって「働きすぎ」の印象は拭えない。教育学習支援業、金融保険業にも同様のことが示唆される。そしてほとんどの業界では高収入者の時間指標達成率が高いため、高収入であるのは長時間働いているからということではないと言える。

<sup>9</sup> 大企業と比べて長時間労働をしていない、余暇時間を確保できるという意味。

<sup>10</sup> サンプル数の問題上、「鉱業」は図表 4-2,4-3,4-4 から抜け落ちている。

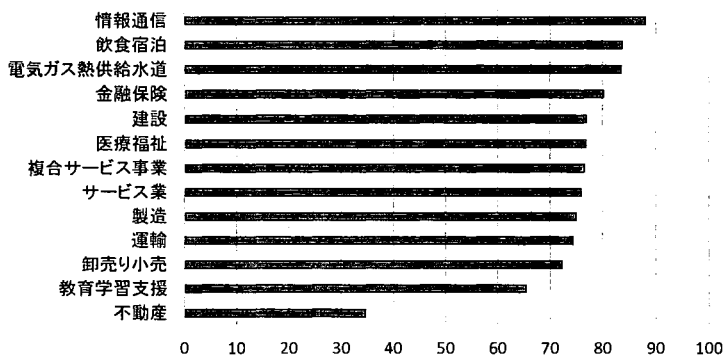
図表 4-1

業界別 高収入比率



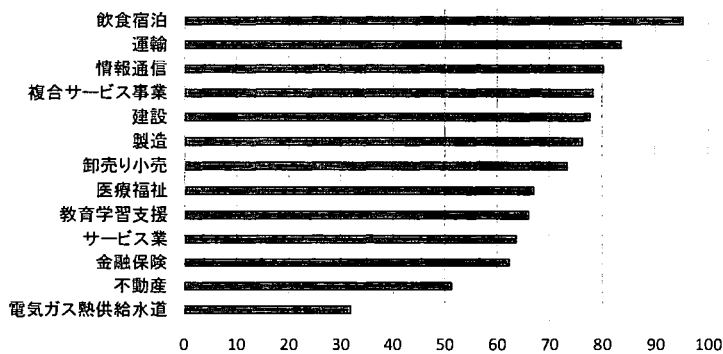
図表 4-2

業界別 高収入者の時間指標達成率(野菜指標)



図表 4-3

業界別 高収入者の時間指標達成率(旅行指標)



図表 4-4 各比率での順位

| 業界       | 高収入比率 | 野菜指標達成率 | 旅行指標達成率 |
|----------|-------|---------|---------|
| 電気ガス水道熱  | 1     | 3       | 13      |
| 教育学習支援   | 2     | 12      | 9       |
| 金融保険     | 3     | 4       | 11      |
| 不動産      | 4     | 13      | 12      |
| 情報通信     | 5     | 1       | 3       |
| 複合サービス事業 | 6     | 7       | 4       |
| 医療福祉     | 7     | 6       | 8       |
| サービス業    | 8     | 8       | 10      |
| 製造       | 9     | 9       | 6       |
| 卸売り小売    | 10    | 11      | 7       |
| 建設       | 11    | 5       | 5       |
| 飲食宿泊     | 12    | 2       | 1       |
| 運輸       | 13    | 10      | 2       |

### 5. 高収入かつ時間に余裕がある確率

最後に、総合的に見て QOL の高い生活を送れる可能性が高い業種を調査する。つまり高収入になる確率と時間に余裕がある確率の同時確率(高収入かつ時間に余裕がある人の、全体数に対する比率)を算出する。ここでは時間指標において、野菜指標と旅行指標の両方を達成した場合に、「時間に余裕がある」とする。

結果を図表 5-1 に示した。大企業と比べて高収入であることに加えて2つの時間指標を導入したため、図表 3-1 の結果とはかなり違っている。前章で見たとおり、時間指標を達成する確率が低い不動産業は下位にランクしている。図表 3-1 で高収入になる確率が低かった業種(運輸業や卸売小売業)も下位となっている。それに対して、どの指標でも成績の良かった「情報通信業」が 29.04%で首位となった。10人のうち3人は、同業種の大企業に勤めている平均的な人たちより、ゆとりのある生活を送ることができると言える。

図表 5-1 業界別 高 QOL 比率

| 順位 | 業界        | 高QOL比率(%) | 該当者数(人) |
|----|-----------|-----------|---------|
| 1  | 情報通信      | 29.0      | 1,233   |
| 2  | 教育学習支援    | 26.3      | 437     |
| 3  | 複合サービス事業  | 22.5      | 297     |
| 4  | 金融保険      | 22.1      | 281     |
| 5  | 飲食宿泊      | 19.7      | 177     |
| 6  | 医療福祉      | 19.7      | 1,022   |
| 7  | 製造        | 17.0      | 7,519   |
| 8  | 建設        | 16.3      | 3,870   |
| 9  | サービス業     | 16.1      | 3,209   |
| 10 | 不動産       | 14.9      | 54      |
| 11 | 運輸        | 14.8      | 1,598   |
| 12 | 卸売小売      | 13.8      | 3,242   |
| 13 | 電気ガス熱供給水道 | 10.5      | 18      |
| -  | 不詳        | 15.8      | 12,298  |
| -  | 業界全体      | 16.4      | 35,253  |

## 6. おわりに

本稿では、初めに中小企業を収入という観点で大企業と比較した。次に時間の指標として野菜指標と旅行指標を導入することで、中小企業の高収入者が長時間労働をしていないかを調べた。最後に高収入かつ時間に余裕のある中小企業労働者の割合を求めた。その結果として、情報通信業の労働者のうち約 30 パーセントが大企業と比較して余裕のある生活を送れているという結論に至った。加えて不動産業は高収入である反面、長時間労働であるという結論も得られた。

また本分析では野菜指標と旅行指標を作成したが、使用する変数を恣意的に定めた事で分析の幅を狭めてしまったのではないかという懸念がある。クラスター分析など、大量の変数をモデルに組み込むことが出来る分析手法を用いれば、収入に対して特徴のある変数を抽出できたと考えられる。この点は本分析の今後の課題としたい。

以上

付録 規定課題

1) 産業符号・性別クロス表

表1

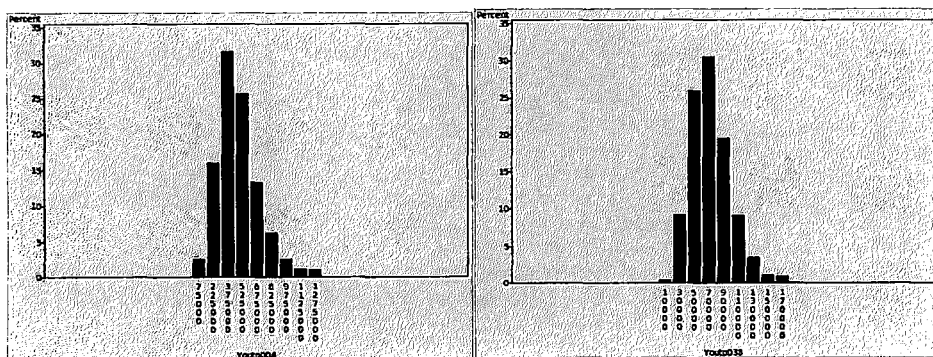
| 産業符号              | a) 単純クロス表 |       |        | b) 10万世帯比でのクロス表 |       |         |
|-------------------|-----------|-------|--------|-----------------|-------|---------|
|                   | 男         | 女     | 合計     | 男               | 女     | 合計      |
| 鉱業                | 3         | .     | 3      | 7               | .     | 7       |
| 建設業               | 2,134     | 9     | 2,143  | 6,346           | 31    | 6,377   |
| 製造業               | 5,668     | 173   | 5,841  | 17,837          | 510   | 18,347  |
| 電気・ガス・熱供給・水道業     | 222       | .     | 222    | 698             | .     | 698     |
| 情報通信業             | 547       | .     | 547    | 2,020           | .     | 2,020   |
| 運輸業               | 1,235     | 3     | 1,238  | 3,852           | 6     | 3,858   |
| 卸売・小売業            | 2,083     | 201   | 2,284  | 6,644           | 590   | 7,233   |
| 金融・保険業            | 472       | 39    | 511    | 1,447           | 104   | 1,551   |
| 不動産業              | 40        | .     | 40     | 144             | .     | 144     |
| 飲食店・宿泊業           | 50        | 32    | 82     | 158             | 107   | 265     |
| 医療・福祉             | 371       | 270   | 641    | 1,022           | 810   | 1,832   |
| 教育・学習支援業          | 1,059     | 89    | 1,148  | 3,110           | 269   | 3,379   |
| 複合サービス事業          | 321       | 3     | 324    | 981             | 11    | 992     |
| サービス業(他に分類されないもの) | 1,816     | 106   | 1,922  | 5,842           | 326   | 6,167   |
| 公務(他に分類されないもの)    | 3,656     | 130   | 3,786  | 10,691          | 430   | 11,121  |
| 不詳                | 9,704     | 1,591 | 11,295 | 31,204          | 4,806 | 36,011  |
| 合計                | 29,381    | 2,646 | 32,027 | 92,000          | 8,000 | 100,000 |

2) 経常収入および食料費

表2

| 分位点 | 経常収入    | 食料費    |
|-----|---------|--------|
| 20% | 310,531 | 49,701 |
| 40% | 404,883 | 63,146 |
| 60% | 501,181 | 76,055 |
| 80% | 635,875 | 93,036 |

左：経常収入のヒストグラム 右：食料費のヒストグラム





1) 男女別・経常収入および食料費の5分位分布

表3

| 経常収入 |           |        |       |         | 食料費 |          |        |       |         |
|------|-----------|--------|-------|---------|-----|----------|--------|-------|---------|
| 5分位  | 5分位数      | 男      | 女     | 計       | 5分位 | 5分位数     | 男      | 女     | 計       |
| 1    | 310,531未満 | 15,494 | 4,506 | 20,000  | 1   | 49,701未満 | 16,803 | 3,196 | 19,999  |
| 2    | 310,531以上 | 18,617 | 1,377 | 19,994  | 2   | 49,701以上 | 18,097 | 1,903 | 20,000  |
| 3    | 404,883以上 | 19,095 | 911   | 20,006  | 3   | 63,146以上 | 18,734 | 1,268 | 20,001  |
| 4    | 501,181以上 | 19,359 | 640   | 19,999  | 4   | 76,055以上 | 19,049 | 948   | 19,998  |
| 5    | 635,875以上 | 19,435 | 567   | 20,002  | 5   | 93,036以上 | 19,317 | 685   | 20,003  |
| 計    |           | 92,000 | 8,000 | 100,000 | 計   |          | 92,000 | 8,000 | 100,000 |

2) 男女別・経常収入・食料費についての3重クロス表

表4

|     | 経常収入<br>5分位 | 食料5分位  |        |        |        |        | 計       |
|-----|-------------|--------|--------|--------|--------|--------|---------|
|     |             | 1      | 2      | 3      | 4      | 5      |         |
| 男   | 1           | 5,827  | 3,974  | 2,798  | 1,831  | 1,064  | 15,494  |
|     | 2           | 5,130  | 4,501  | 3,978  | 2,996  | 2,012  | 18,617  |
|     | 3           | 2,911  | 4,301  | 4,395  | 4,294  | 3,198  | 19,099  |
|     | 4           | 1,906  | 3,233  | 4,082  | 4,903  | 5,231  | 19,356  |
|     | 5           | 1,029  | 2,087  | 3,481  | 5,026  | 7,813  | 19,435  |
|     | 計           | 16,803 | 18,097 | 18,734 | 19,049 | 19,317 | 92,000  |
| 女   | 1           | 2,462  | 994    | 565    | 314    | 172    | 4,506   |
|     | 2           | 371    | 399    | 235    | 211    | 161    | 1,377   |
|     | 3           | 243    | 225    | 216    | 164    | 63     | 911     |
|     | 4           | 68     | 134    | 156    | 143    | 140    | 640     |
|     | 5           | 53     | 152    | 96     | 117    | 149    | 567     |
|     | 計           | 3,196  | 1,903  | 1,268  | 948    | 685    | 8,000   |
| 男女計 |             | 19,999 | 20,000 | 20,001 | 19,998 | 20,003 | 100,000 |

3) 男女別・経常収入別に食料費を常用対数変換した平均値の比較

表5

|   | 経常収入<br>5分位 | 食料     |        |        |
|---|-------------|--------|--------|--------|
|   |             | 世帯数    | 常用対数平均 | 幾何平均   |
| 男 | 1           | 15,494 | 4.7389 | 54,815 |
|   | 2           | 18,617 | 4.7832 | 60,702 |
|   | 3           | 19,099 | 4.8368 | 68,675 |
|   | 4           | 19,356 | 4.8826 | 76,313 |
|   | 5           | 19,435 | 4.9335 | 85,803 |
|   | 計           | 92,000 | 4.8395 | 69,103 |
| 女 | 1           | 4,506  | 4.6713 | 46,914 |
|   | 2           | 1,377  | 4.7811 | 60,409 |
|   | 3           | 911    | 4.7813 | 60,437 |
|   | 4           | 640    | 4.8617 | 72,728 |
|   | 5           | 567    | 4.8774 | 75,405 |
|   | 計           | 8,000  | 4.7325 | 54,013 |

付録 規定課題の SAS プログラム

```
/*proc import で読み込んだ上で*/
```

```
/* データを縦に結合 */
```

```
data data.d_all ;
```

```
    set data.d1_5000 data.d5001_10000 data.d10001_15000 data.d15001_20000  
data.d20001_25000 data.d25001_30000 data.d30001_32027;  
run ;
```

```
/*10万人比のウェイトを作る*/
```

```
proc means data=data.D_ALL sum ;
```

```
    var weight ;
```

```
    output out=weightsum (drop=_:);
```

```
run;
```

```
data weightsum2(rename=(weight=weight_sum));
```

```
    set weightsum ;
```

```
run;
```

```
data data.D_ALL_w ;
```

```
    retain weight_sum2 ;
```

```
    merge data.D_ALL
```

```
        weightsum2 ;
```

```
    if weight_sum = . then weight_sum=weight_sum2 ;
```

```
    else weight_sum2 = weight_sum ;
```

```
    new_weight = weight / weight_sum *100000 ;
```

```
    drop weight_sum
```

```
        weight_sum2 ;
```

```
run ;
```

```
/*規定課題 1*/
```

```
data data.kiteil ;
```

```
    set data.D_ALL_w ;
```

```
    if S1_Sangyou="VV" then S1_Sangyou="99";
```

```
    S1_Sangyou2 = input(S1_Sangyou, 8.) ;
```

```
    keep S1_Sex S1_Sangyou2 new_weight weight ;
```

```
run ;
```

```
proc format ;
```

```

value Sex
1="男"
2="女"
;
value Sangyou
1="農業"
2="林業"
3="漁業"
4="鉱業"
5="建設業"
6="製造業"
7="電気・ガス・熱供給・水道業"
8="情報通信業"
9="運輸業"
10="卸売・小売業"
11="金融・保険業"
12="不動産業"
13="飲食店・宿泊業"
14="医療・福祉"
15="教育・学習支援業"
16="複合サービス事業"
17="サービス業（他に分類されないもの）"
18="公務（他に分類されないもの）"
19="その他（非就業を含む）"
99="不詳"
;
run ;

proc tabulate
data = data.kiteil
missing
format = comma7.
;
class S1_Sex S1_Sangyou2 ;
table S1_Sangyou2="産業符号" ALL="合計", S1_Sex="性別" ALL="合計";
format S1_Sex Sex. S1_Sangyou2 Sangyou. ;
run ;

```

```

proc tabulate
  data = data.kitei1
  missing
  format = comma7.
  ;
  class S1_Sex S1_Sangyou2 ;
  var new_weight ;
  table (S1_Sangyou2="産業符号" ALL="合計")*new_weight , S1_Sex="性別" ALL="
合計" ;
  format S1_Sex Sex. S1_Sangyou2 Sangyou. ;
run ;

```

/\*規定課題 2\*/

```

data data.kitei2 ;
  set data.D_ALL_w ;
  keep Youto004 Youto038 weight ;
run ;

```

```

proc univariate
  data = data.kitei2 ;
  var Youto004 Youto038 ;
  weight weight ;
  output out = gobunni
  pctlpts = 20 40 60 80
  pctlpre = p004_ p038_
  ;
run ;

```

```

data data.hyou2 ;
  x = 0.2 ;
  y = 310531.0652 ;
  z = 49701.02133 ;
  output ;
  x = 0.4 ;
  y = 404883.1333 ;
  z = 63146.23316 ;
  output ;
  x = 0.6 ;

```

```

y = 501180.8646 ;
z = 76055.14362 ;
output ;
x = 0.8 ;
y = 635875.2793 ;
z = 93035.80288 ;
output ;
label x = "分位点"
      y = "經常收入"
      z = "食料費"
;
format x percent8. y comma7. z comma6. ;
run ;

title "表 2" ;
proc print data = data.hyou2 noobs label ;
run ;
title ;

proc univariate
  data = data.kitei2 ;
  var Youto004 Youto038 ;
run ;

proc gchart data = data.kitei2 ;
  vbar Youto004 /
  type = pct
  midpoints = 75000 225000 375000 525000 675000 825000 975000 1125000
1275000 ;
run ;

proc gchart data = data.kitei2 ;
  vbar Youto038 /
  type = pct
  midpoints = 10000 30000 50000 70000 90000 110000 130000 150000 170000 ;
run ;

```

```

/*規定課題 3*/
data data.kitei3 ;
    set data.D_ALL_w ;
    keep Youto004 Youto038 new_weight S1_Sex ;
run ;

data data.kitei3s ;
    set data.kitei3 ;
        if Youto004 >= 635875.2793 then bunnrui = 5 ;
        else if Youto004 >= 501180.8646 then bunnrui = 4 ;
        else if Youto004 >= 404883.1333 then bunnrui = 3 ;
        else if Youto004 >= 310531.0652 then bunnrui = 2 ;
        else
            bunnrui = 1 ;

        if Youto038 >= 93035.80288 then bunnrui2 = 5 ;
        else if Youto038 >= 76055.14362 then bunnrui2 = 4 ;
        else if Youto038 >= 63146.23316 then bunnrui2 = 3 ;
        else if Youto038 >= 49701.02133 then bunnrui2 = 2 ;
        else
            bunnrui2 = 1 ;
run ;

proc tabulate
    data = data.kitei3s
    format = comma7. ;
    class bunnrui S1_Sex ;
    var new_weight ;
    table (bunnrui = "5 分位" ALL = "計")*new_weight , S1_Sex ALL = "計" ;
    format S1_Sex Sex. ;
run ;

proc tabulate
    data = data.kitei3s
    format = comma7. ;
    class bunnrui2 S1_Sex ;
    var new_weight ;
    table (bunnrui2 = "5 分位" ALL = "計")*new_weight , S1_Sex ALL = "計" ;
    format S1_Sex Sex. ;
run ;

```

```

/*規定課題 4*/
proc tabulate
    data = data.kitei3s
    format = comma7. ;
    class bunnrui S1_Sex bunnrui2;
    var new_weight ;
    table (S1_Sex*(bunnrui = "經常收入 5 分位" ALL = "計") ALL = "男女計
")*new_weight , bunnrui2 = "食料 5 分位" ALL = "計";
    format S1_Sex Sex.;
run ;

/*規定課題 5*/
data data.kitei5 ;
    set data.kitei3s ;
    logYouto038 = log10(Youto038) ;
run ;

proc tabulate data = data.kitei5 out = kika ;
    class S1_Sex bunnrui ;
    var Youto038 logYouto038 ;
    weight new_weight ;
    table S1_Sex*(bunnrui = "經常收入 5 分位" ALL = "計") , Youto038*(sumwgt = "世
帶數")*f=comma7. logYouto038*(mean = "常用對數平均")*f=6.4 ;
    format S1_Sex sex. ;
run ;

data kikaheikinn ;
    set kika ;
    kikaheikinn = 10**logYouto038_mean ;
run ;

```

マイナス金利下におけるローン事業の発展性と  
疑似マイクロデータに見るローン利用者の行動特性  
(参加カテゴリー：B)

内田晃秀  
筑波銀行 営業企画部

Akihide Uchida  
Marketing Planning Division, Tsukubabank, Ltd.

## 要旨

バブル崩壊後に本格的にとられた本邦金融政策は経済を安定化させ成長軌道に補正する意味で大きな役割を果たした。一方でそれらの政策の副次的効果が銀行経営やこれまでのビジネスモデルに負の影響を与えている。低金利下で預貸金利ギャップが縮小し、事業性融資や住宅ローンの利鞘が縮小している状況下で、銀行はこれまで以上に顧客への理解を深め採算性確保の可能性が残る無担保の個人ローンに対する事業を研究し拡大する必要がある。

独立行政法人統計センターの作成した「教育用疑似マイクロデータ」を分析および仮説と検証の結果、カードローン等無担保ローン利用者は20代前半から50代に至るまで借入需要は旺盛であるという結果が得られた。特に教育や自動車関連の用途だけでなく、交際費や教養娯楽への支出も順相関が高いことから若年層のローン利用を牽引していると考えられる。これらの分析結果は銀行の個人ローン事業の進むべき道の糸口として示唆となるだろう。

## 1. バブル崩壊後の金融政策と銀行の対応

### (1) バブル崩壊後の金融政策と市場金利

1990年代初頭から日本経済はバブル経済の崩壊による不良債権の増加、デフレスパイラルおよび円高が相互に進行する悪循環へ突入しインフレ率や経済成長率は低下した。これに対応し日本銀行は1990年8月時点では6%であった政策金利を段階的に引き下げ、1998年9月には0.25%とゼロ近傍へ引き下げることとなった。さらに1999年に無担保コールレート（オーバーナイト）をゼロ近傍へ誘導するという「ゼロ金利政策」を導入した。時の判断から2000年8月にはゼロ金利政策を解除するも、米国ITバブル崩壊の余波を受け2001年3月には操作目標を日本銀行当座預金残高とする「量的緩和政策」を導入した。その後、危機的経済情勢の後退に伴い2006年には「ゼロ金利政策」および「量的緩和政策」は解除された。

2010年にリーマンショックの発生に伴い、「包括緩和政策」の開始により社債、ETF、J-REITなどの金融商品も買い入れ対象とすることでイールドカーブの押し下げとリスクプレミアムの圧縮を図った。

現安倍政権下では、いわゆる3本の矢の一つ「デフレ脱却と持続的な経済成長の実現のための政府・日本銀行の政策連携について」が示され2013年4月に異次元緩和と「量的・質的金融緩和」の導入、次いで2016年1月には「マイナス金利付き量的・質的金融緩和」を導入した。これにより現



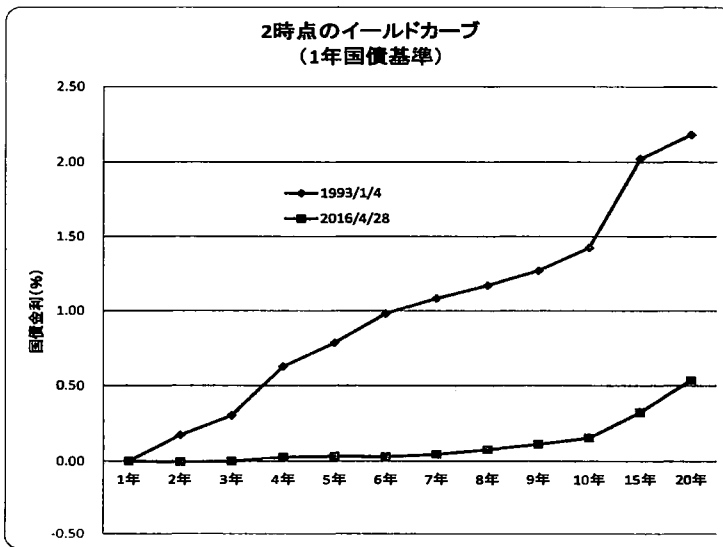
在も続く日本銀行当座預金▲0.1%、インフレ率 2%の「物価安定の目標」とマネタリーベースの拡大による大規模な金融緩和が行われており、1993年初の国債5年国債金利4.239%および10年国債金利4.877%は2016年4月28日時点の5年同-0.206%および10年同-0.082%までに下落している（図1-1）。また、両時点の1年国債の金利をゼロとした場合のイールドカーブを比較すると、大きくフラットニングしていることがわかる（図1-2）。

（図1-1）



データ出典：財務省国債金利情報

（図1-2）



データ出典：財務省国債金利情報を基に加工

(2) ゼロ金利政策の導入およびマイナス金利付き量的、質的金融緩和と邦銀に与えた影響  
バブル経済の崩壊以後、前述の低金利政策下において貸出金利と預金金利の差である利鞘が縮小傾

向にある中で邦銀は発生した多くの不良債権を償却すべくバランスシート調整に主要行・地方銀行とに経営資源の傾斜配分を余儀なくされる(図1-3)。近年では1~3%台での推移となり、銀行を取り組まなく不良債権の問題は解決したと考えられる。

(図1-3)

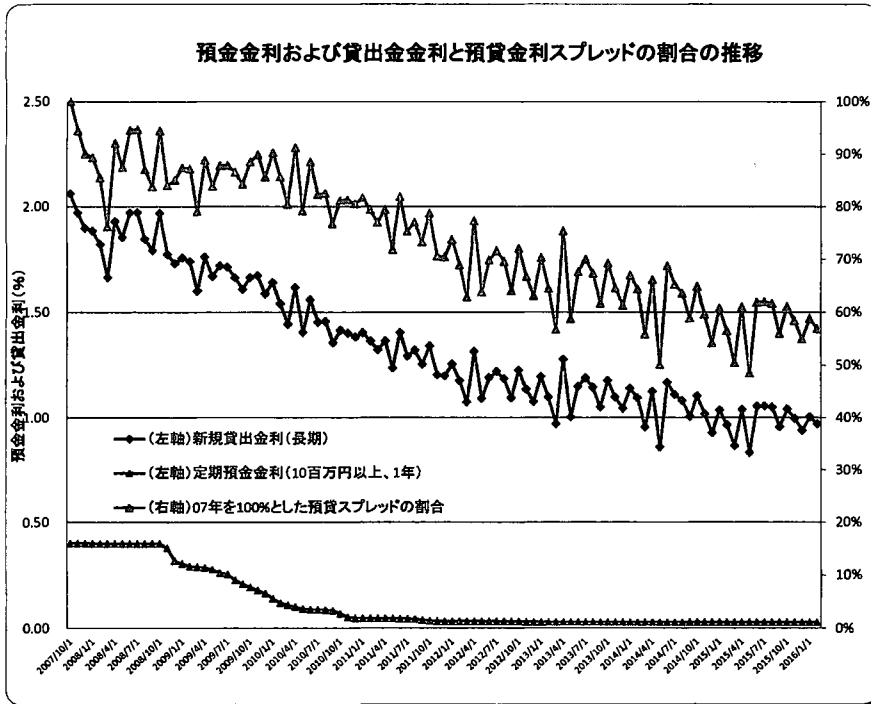


データ出典：金融庁 『平成 27 年 3 月期における金融再生法開示債権の状況等』

不良債権の問題と向き合う途上で、その処理に充てる原資としての収益が圧迫される状況が進行していた。それは、市中金利が下落するのに合わせて貸出金利も下落傾向にあるが、本来平行であるべき経費としての預金金利が下限(ゼロ)に近傍し推移せざるを得ない(預金がマイナス金利とならない)という状況である。つまり、銀行は両金利を平行に推移させたい意図に反し、貸出金利だけが下落を続け、預金金利がゼロ近傍に張り付いたままとなることで、本来の利益の源泉である貸出金利 - 預金金利(以下、預貸金利スプレッドと呼ぶ) = 利鞘を圧迫する状況が続いているのである。2007年10月の預貸金利スプレッドを100%とすると2016年2月には56%となっており、銀行の基礎的な利鞘は約10年間で概ね半減している状況にある(図1-4)。

日本銀行でも上述の通り「マイナス金利付き量的・質的金融緩和」政策の継続でマイナス金利の幅を今後も拡大する可能性がある状況下であり、利鞘の縮小はトレンドとして続くことを想定せざるをえない。

(図1-4)



データ出典：日本銀行『時系列統計データ』

### (3) 地銀のビジネスモデルの支柱である個人ローンスタンス

銀行として利益の源泉である預貸金利ギャップの縮小という困難な状況下にあつて、リテール事業、とりわけ個人ローンが存在感を増している。個人ローンは住宅ローンに代表される有担保ローンと、カードローン、教育ローン、自動車ローンなどの無担保ローンに大別される。

住宅ローンは低金利の影響を受け一般的な法人向け貸し出しと同様に低金利での運用を迫られている。一方でカードローンをはじめとした無担保ローンは、個人の利便性を追求していることから、銀行としては発行にできるだけ顧客の手間を省く仕組みを構築するために金利を確保できる、つまり顧客ニーズに弾力的に対応しているがゆえに利鞘が確保できる商品であると換言できる。

平成21年の割賦販売法<sup>1</sup>および平成22年の貸金業法等<sup>2</sup>の改正により銀行カードローンの果たすべき役割は一層高まっただけでなく、より顧客に密着したリテール金融を行う契機ともなった。銀行の貸出金におけるカードローンの割合はごく少ないが、利便性およびリスクの対価としての金利が相応ために収益に与えるインパクトは少なくない。

「マイナス金利付き量的・質的金融緩和」が継続され、一層のマイナス幅拡大が行われる可能性がある限り、市場金利は押し下げられイールドカーブはフラットニングを続けるだろう。この状況下

<sup>1</sup> 個別信用購入あっせん業を行う場合は経済産業省へ登録を受ける必要があり、各種要件を満たし同省の審査を経なければならない。

<sup>2</sup> 改正以前は多くの貸金業者が出資法の上限である年利29.2%と利息制限法の上限である年15~20%の間(グレーゾーン金利)で貸し付けを行うことが問題視されていた。平成22年に貸金業法等が改正され出資法の上限金利は20%に引き下げられた。また貸し付け残高が年収の1/3を超えられないとする総量規制もその際導入された。

では、いかに個人向けカードローンのニーズをつかむか、個人のライフイベントを金融行動から解明し資金需要を解明するか、さらに「マイナス金利付き量的・質的金融緩和」政策下で単体での利鞘確保が難しくなった住宅ローンの収益的な補完としてカードローンのクロスセルをいかに追求するかは銀行のビジネスモデルを探る上でも目下の課題となる。

## 2. 疑似マイクロデータ分析による銀行ビジネスモデルの補完

### (1) 個人ローン利用者の基礎調査

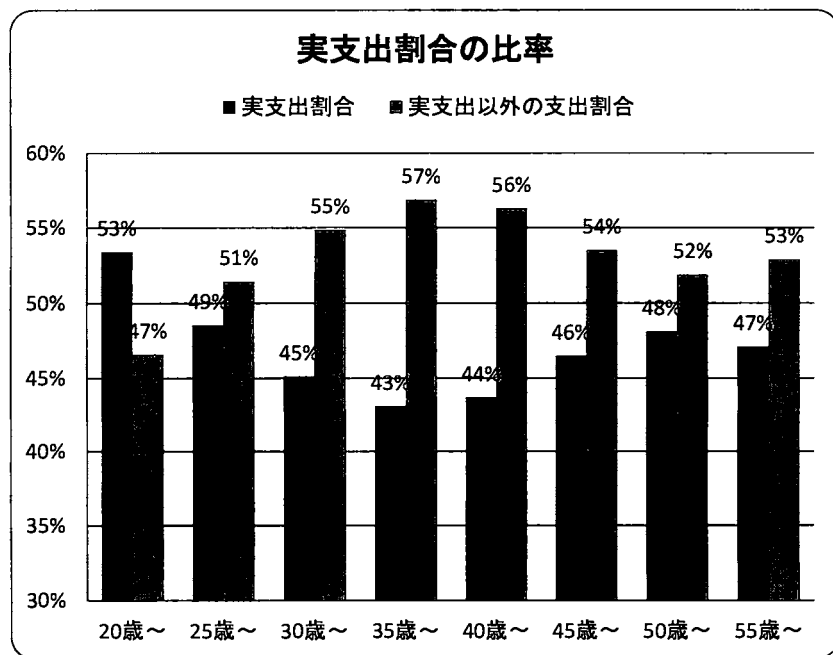
本編においては独立行政法人統計センターの作成した「教育用疑似マイクロデータ」(以下、マイクロデータ)を使用した。なお集計用乗率を重みとして使用している。このマイクロデータ上で示される「土地家屋借金返済」>0を住宅ローン等利用者、「他の借入金返済+分割払・一括払購入借入金返済」>0をカードローン等無担保ローン利用者と便宜的にみなす。また資産形成層の動向に焦点を当てるため、60歳以上は基本的に分析から除外している。

銀行の顧客として個人ローン契約の主体となるのは20代~50代までの資産形成層が中心である。個人ローンの利用を伸ばすためには、資産形成層の動向をよく理解しなければならず、その年代毎に傾向を探る必要もある。

### (2) マイクロデータの実支出割合からの分析

母集団全体の支出の状況を概観してみると、食料、住居など生活するうえで必要な消費支出および税金などの非消費支出をまとめた実支出は20代前半では高いものの、それ以降は下落し、30代後半では43%となっている。一方、預貯金、借入金の返済、保険掛金などの実支出以外の支出は逆の割合となることから30代後半で最高の57%となる(図2-1)。

(図2-1)

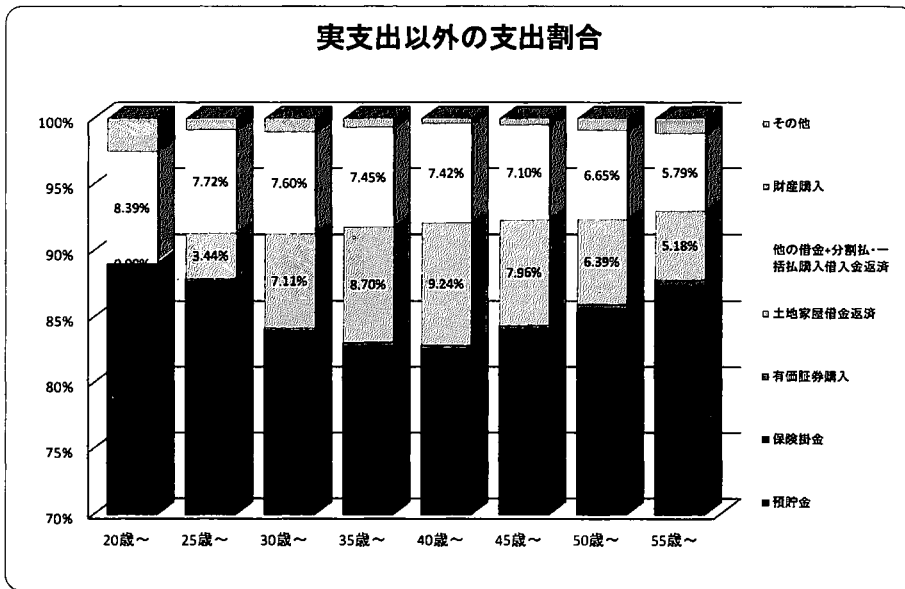


n=23,383 60歳以上は除く、また不詳は集計から除外している。集計用乗数利用有

(3) 実支出以外の支出の内訳

実支出以外の支出の大部分は各年代とも大部分を預貯金が占めるが、実支出とそれ以外の割合で乖離が大きくなる 30～40 代は土地家屋借金返済＝住宅ローン等の返済が多くなっていることがわかる。また、特徴的であるのは他の借入金返済+分割払・一括払購入借入金返済＝カードローン等無担保ローンがどの年代を通して、一定の割合を占めており住宅ローン以上に身近なファイナンスとして定着していることがわかる（図 2-2）。

(図 2-2)

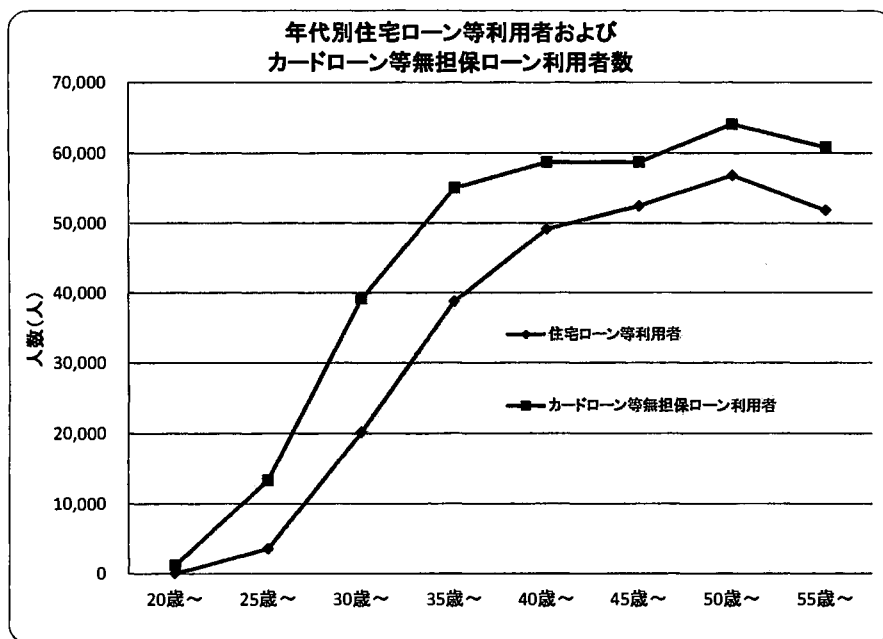


n=23,383 60歳以上は除く、また不詳は集計から除外している。集計用乗数利用有

(1) 住宅ローン等利用者とカードローン等無担保ローン利用者の状況について

一般的にカードローンなどの無担保ローンは住宅ローンに比べ少額であることに加え、審査が早く手軽に申込できることから、20代より利用者が多く借入機会の増加に伴い、35歳以上～40歳未満までにはカードローン等無担保ローンの利用者はピーク近くまで一気に上昇する。一方住宅ローン等利用者は年齢を重ねるに従った、結婚や子育てなどのライフイベントに伴い、40歳以上45歳未満までの間に徐々に増加している。住宅ローンの場合は金額が大きく年間の返済額もまとまった額となることから相応の年収が必要となることも影響しているといえよう。（図 2-3）

(図 2 - 3)



住宅ローン等利用者 : n=272,807 60歳以上は除く、また不詳は集計から除外している。集計用乗数利用有

カードローン等無担保ローン利用者 : n=351,151 60歳以上は除く、また不詳は集計から除外している。集計用乗数利用有

## (2) 住宅ローン等利用者像の仮説

ローン利用者の属性を意識し商品性やプロモーション、マーケティングに生かすことは、販売を促進するうえで最も重要である。今般の分析に当たり、より焦点を絞った分析とするため事前に仮説を立てそれを証明し利用者の行動特性を解明する方法を採用した。

### 【住宅ローン等利用者像の仮説】

- ① 住宅ローン利用者については収入に順相関がある
- ② 住居が賃貸であることと逆相関となる。
- ③ 居住者は複数であることが想定されるので食費や公共料金の支払が一定程度順相関する。

## (3) 住宅ローン等利用者像の仮説の検証

住宅ローン利用者の行動特性を探索するため、住宅ローン等利用者(土地家屋借金返済)と直接または間接的に因果関係の存在が推定される複数の収支項目に対し相関分析を行った。下の図は各年代の相関係数を算出し、年代毎に相関係数を配置したものであり各収支項目に対し年代を追うごとに相関係数がどのように推移するのかを図示したものである。相関係数の検定結果はいずれも有意水準を0.05とし帰無仮説を棄却しうる値のみを使用した。したがって棄却できないデータに関しては同系列でもプロットしていない。

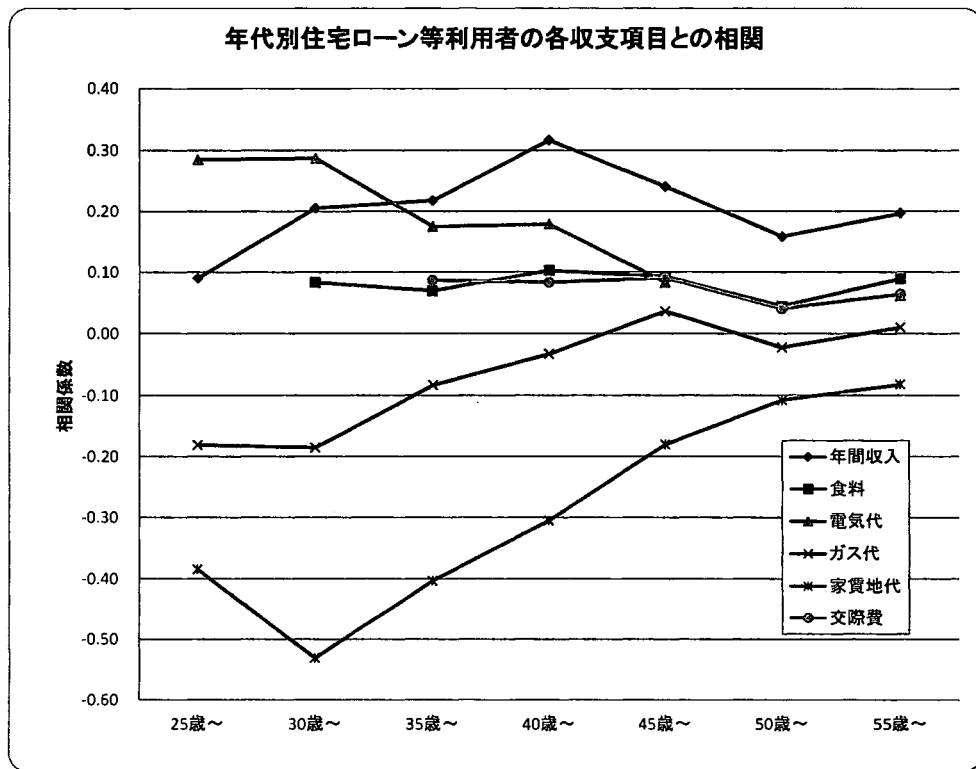
#### (4) 住宅ローン利用者の傾向

住宅ローン等利用者の傾向として、2点があげられる。1点目の傾向は20～40代では相関係数の絶対値が大きいのにに対し、50代後半に近づくにつれトレンドとしてゼロへ近づいているということである。他の項目に関して有意な相関がみられないことから年を経るにつれて年収が増加すると同時に、ライフスタイルや家族構成の多様化により画一的な支出が減少する傾向にあるといえる。

2点目の傾向は表示した各収支項目は直接または間接的に因果関係の理由付けが可能であるということである。例えば住宅ローンの返済がある人は持ち家であるから、一般的な傾向として地代家賃はないと考えられるので、顕著な逆相関となっている。また、年間収入が高ければ住宅ローンを組む動機や審査結果にも影響するだろう。電気代とガス代がゼロ付近を軸と考えると対照となっている点もオール電化やIHの普及が影響していると考えられる。住宅ローンがあり、世帯が3人以上となり子供を持つ世帯となると広い世代で食費も弱い相関を維持することが当然と考えられる(図2-4)。

#### (5) 仮説検証の結論

このように20～40代は相対的に相関が確認でき、かつ住宅ローン利用との因果関係が一定程度解明できる支出であることから、顕著な相関は表れこそしないが、概ね方向性は正しいことが分かった。(図2-4)



n=23,383 60歳以上は除く、また不詳は集計から除外している。集計用乗数利用有。有意でない相関はプロットしていない。

#### (6) カードローン等無担保ローン利用者像の仮説

住宅ローン利用者と同様、より焦点を絞った分析とするため事前に仮説を立てそれを証明し利用者の行動特性を解明する方法を採用した。

##### 【カードローン等無担保ローン利用者像の仮説】

- ① 住宅ローンに比べカードローン等無担保ローンは自由に利用できることから年間収入とは逆相関にある。
- ② 教育ローン、自動車ローンなど目的型の無担保ローンが存在するため教育や自動車関連を用途とする支出と順相関がある。
- ③ 加齢に伴い、世帯のライフイベントや交際にかかる支出も多くなることで自由になる資金がなくなる 40 代以上でカードローン等無担保ローンの利用が多くなるかもしれない。

#### (7) カードローン等無担保ローン利用者像の仮説の検証

カードローン等の無担保ローン利用者の各収支項目としても 2 点の傾向がある。1 点目の特徴は年間収入と順相関が確認できるという点である。無担保ローンは通常の日常生活と関係のある消費支出に充てられ収入が支出に追いつかないため、年間収入との逆相関が想像された。実際には相対的に高い順相関を 50 代以上も維持している。

2 点目の傾向は支出に関して相対的に突出した相関がみられないことである。特に金融機関で一般的に用意している教育ローンや自動車ローンが基本的な使途としている教育費用やそれに準じた仕送り金、自動車関連費用が一定の順相関を維持していることから商品設定の妥当性が確認できる。

3 点目の傾向は、各項目の相関が 20 代から 50 代まで住宅ローンに比べ大きく変化がない点である。住宅ローン等利用者の場合、相関係数はゼロ付近へ収束していくのに対し、カードローン等無担保ローンは 0.15 付近へ収束する。つまり年齢を重ねても需要が存在し続けているということである。また前掲利用者数の図 2-5 で明らかとなった「20 代から傾きが住宅ローンに対して急である、言い換えれば若年から需要が多い」という点から考えてみても符合する。教養娯楽に関して 20 代後半では有意水準が 0.11 となったことからプロットしていないものの、0.601 という強い相関係数が算出された。また交際費に関しては住宅ローン等利用者でも一定の順相関は見られたが、カードローン等無担保ローン利用者の場合にはより若年が高い相関があることも特徴的であることから若年のカードローン等無担保ローンの利用者を牽引していると考えられる。

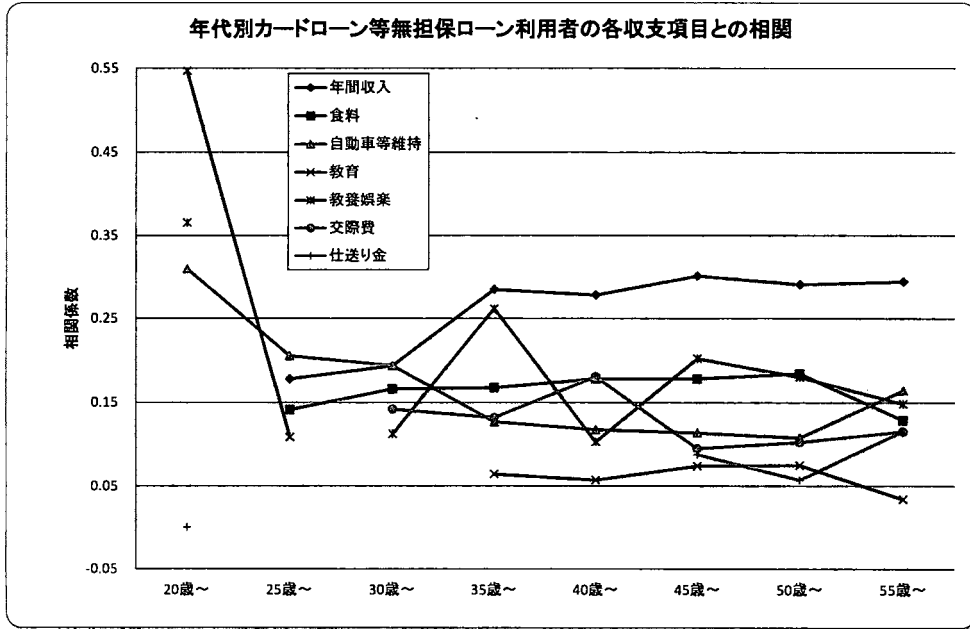
なお、教育に関して 20 代前半で高い順相関となっていることについては統計的な P 値は有意だが本稿での説明が難しく別に検討を要すると考えられる。

#### (8) 仮説検証の結論

年間収入とは順相関にあり、高収入であれば借入をしないと結論付けるのは早計である。用途はそれぞれであり、ライフイベントに合わせた借入を行っている。また、40 代以上よりも若年の方が新たに借入を開始する傾向がある。



(図 2-5)



n=23,383 60歳以上は除く、また不詳は集計から除外している。集計用乗数利用有。有意でない相関はプロットしていない。

### (7) まとめ

バブル経済の崩壊以後の経済情勢に対応したゼロ金利政策量的緩和政策、および近時のマイナス金利付き量的・質的緩和政策等は、日本経済を安定化させ成長軌道に補正する意味で大きな役割を果たした。一方でそれらの政策余波が図らずも銀行経営やこれまでのビジネスモデルに負の影響を与えている。

低金利下で預貸金利ギャップが縮小し、事業性融資や住宅ローンの利鞘が縮小している状況下で銀行ではこれまで以上に顧客への理解を深め採算性確保の可能性が残る無担保の個人ローンに対する事業を拡大する必要がある。

住宅ローン等利用者は土地家屋を保有していることと、年間収入・家賃地代などの因果関係が直接・間接的に順・逆相関が読み取れる。

一方カードローン等無担保ローン利用者は20代前半から50代に至るまで借入需要は旺盛である。教育や自動車関連の用途だけでなく、交際費や教養娯楽への支出も順相関が高いことから若年層のローン利用を牽引していると考えられる。

これらの分析結果は銀行の個人ローン事業の進むべき道の糸口として示唆となるだろう。

(了)

(参考文献)

「非伝統的金融政策の実践と理論」日本銀行 2014

付録 規定課題

1) 産業符号・性別 クロス表

表 1

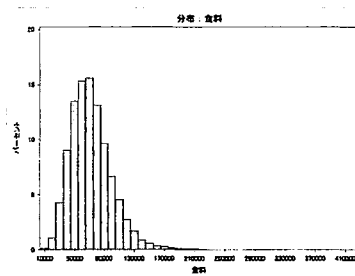
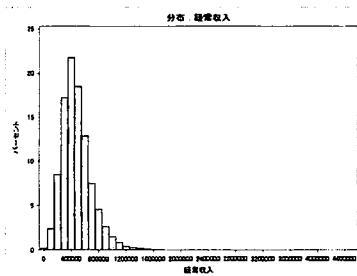
a)単純クロス表 b)10万世帯比でのクロス表

| 産業符号(改)              | 性別     |       | 合計<br>(ALL) | 性別                  |                     |
|----------------------|--------|-------|-------------|---------------------|---------------------|
|                      | 1      | 2     |             | 1                   | 2                   |
|                      |        |       |             | Sum                 | Sum                 |
|                      | N      | N     | N           | 100k(intg)<br>集計用乗率 | 100k(intg)<br>集計用乗率 |
| 04_鉱業                | 3      | .     | 3           | 7                   | .                   |
| 05_建設業               | 2,134  | 9     | 2,143       | 6,346               | 31                  |
| 06_製造業               | 5,668  | 173   | 5,841       | 17,837              | 510                 |
| 07_電気・ガス・熱供給・水道業     | 222    | .     | 222         | 697                 | .                   |
| 08_情報通信業             | 547    | .     | 547         | 2,020               | .                   |
| 09_運輸業               | 1,235  | 3     | 1,238       | 3,852               | 6                   |
| 10_卸売・小売業            | 2,083  | 201   | 2,284       | 6,644               | 589                 |
| 11_金融・保険業            | 472    | 39    | 511         | 1,447               | 104                 |
| 12_不動産業              | 40     | .     | 40          | 144                 | .                   |
| 13_飲食店・宿泊業           | 50     | 32    | 82          | 158                 | 107                 |
| 14_医療・福祉             | 371    | 270   | 641         | 1,022               | 810                 |
| 15_教育・学習支援業          | 1,059  | 89    | 1,148       | 3,110               | 269                 |
| 16_複合サービス事業          | 321    | 3     | 324         | 981                 | 11                  |
| 17_サービス業(他に分類されないもの) | 1,816  | 106   | 1,922       | 5,842               | 326                 |
| 18_公務(他に分類されないもの)    | 3,656  | 130   | 3,786       | 10,691              | 430                 |
| VV_不詳                | 9,704  | 1,591 | 11,295      | 31,204              | 4,806               |
| 合計 (ALL)             | 29,381 | 2,646 | 32,027      | 92,000              | 8,000               |

2) 産業符号・性別 クロス表

表 2

|     | 経常収入    | 食料     |
|-----|---------|--------|
| P20 | 310,531 | 49,701 |
| P40 | 404,883 | 63,146 |
| P60 | 501,181 | 76,055 |
| P80 | 635,875 | 93,036 |



3) 男女別・経常収入および食料費の5分位分布

経常収入

|                   | Sum    |       |
|-------------------|--------|-------|
|                   | 男      | 女     |
| 経常収入 5 分位フラグ【文字列】 |        |       |
| 1 310,531 未満      | 15,494 | 4,506 |
| 2 310,531 以上      | 18,617 | 1,377 |
| 3 404,883 以上      | 19,095 | 911   |
| 4 501,181 以上      | 19,359 | 640   |
| 5 635,875 以上      | 19,435 | 567   |
| 合計 (ALL)          | 92,000 | 8,000 |

食料品

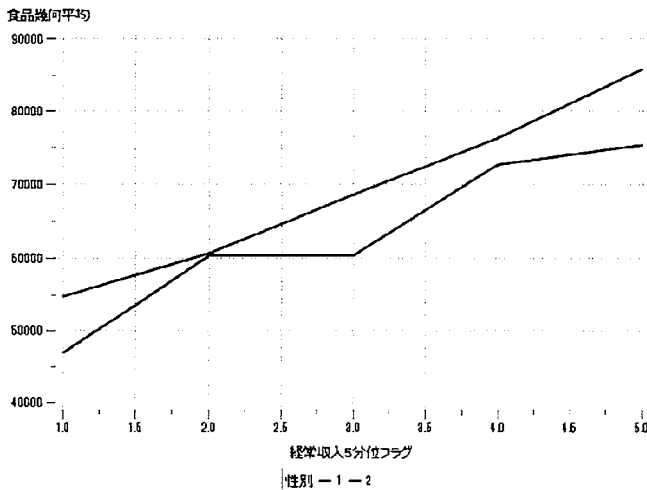
|              | Sum    |       |
|--------------|--------|-------|
|              | 男      | 女     |
| 食料 5 分位【文字列】 |        |       |
| 1 49,701 未満  | 16,803 | 3,196 |
| 2 49,701 以上  | 18,097 | 1,903 |
| 3 63,146 以上  | 18,734 | 1,268 |
| 4 76,055 以上  | 19,049 | 948   |
| 5 93,036 以上  | 19,317 | 685   |
| 合計 (ALL)     | 92,000 | 8,000 |

4) 男女別・経常収入および食料費の5分位分布  
表 4

| 性別       | 経常収入 5分位フラグ | Sum    |        |        |        |        |
|----------|-------------|--------|--------|--------|--------|--------|
|          |             | 食料 1   | 食料 2   | 食料 3   | 食料 4   | 食料 5   |
| 1        | 1           | 5,827  | 3,974  | 2,798  | 1,831  | 1,064  |
|          | 2           | 5,130  | 4,501  | 3,978  | 2,996  | 2,012  |
|          | 3           | 2,911  | 4,301  | 4,395  | 4,291  | 3,198  |
|          | 4           | 1,906  | 3,233  | 4,082  | 4,906  | 5,231  |
|          | 5           | 1,029  | 2,087  | 3,481  | 5,026  | 7,813  |
|          | 合計 (ALL)    | 16,803 | 18,097 | 18,734 | 19,049 | 19,317 |
| 2        | 1           | 2,462  | 994    | 565    | 314    | 172    |
|          | 2           | 371    | 399    | 235    | 211    | 161    |
|          | 3           | 243    | 225    | 216    | 164    | 63     |
|          | 4           | 68     | 134    | 156    | 143    | 140    |
|          | 5           | 53     | 152    | 96     | 117    | 149    |
|          | 合計 (ALL)    | 3,196  | 1,903  | 1,268  | 948    | 685    |
| 合計 (ALL) | 19,999      | 20,000 | 20,001 | 19,998 | 20,003 |        |

5) 男女別・経常収入別に食料費を常用対数変換した平均値の比較  
表 5

| 性別 | 経常収入 5分位フラグ | 世帯数    | 食料常用対数 | 幾何平均   |
|----|-------------|--------|--------|--------|
|    |             | Sum    | Mean   |        |
| 1  | 1           | 15,494 | 4.7389 | 54,819 |
|    | 2           | 18,617 | 4.7832 | 60,704 |
|    | 3           | 19,095 | 4.8368 | 68,672 |
|    | 4           | 19,359 | 4.8826 | 76,313 |
|    | 5           | 19,435 | 4.9335 | 85,793 |
|    | 合計 (ALL)    | 92,000 | 4.8395 | 69,103 |
| 2  | 1           | 4,506  | 4.6713 | 46,913 |
|    | 2           | 1,377  | 4.7811 | 60,405 |
|    | 3           | 911    | 4.7813 | 60,439 |
|    | 4           | 640    | 4.8617 | 72,730 |
|    | 5           | 567    | 4.8774 | 75,407 |
|    | 合計 (ALL)    | 8,000  | 4.7325 | 54,013 |



付録 規定課題の SAS プログラム

1) 産業符号・性別 クロス表

a)10 万人比 (男 92 千、8 千) で案分した集計用乗率「100k(intg)集計用乗率を算出

**PROC TABULATE**

DATA=WORK. QUERY\_FOR\_APPEND\_TABLE\_000D\_0008

```

;

VAR "男"n "女"n "100k(intg)集計用乗率"n;
CLASS "性別"n /ORDER=UNFORMATTED MISSING;
CLASS "産業符号(改)"n /ORDER=UNFORMATTED MISSING;
TABLE /* 行ディメンション*/
'産業符号(改)' n*F=COMMA12.
ALL={LABEL="合計 (ALL)"}*F=COMMA12.,
/* 列ディメンション */
'性別' n*F=COMMA12. *
N*F=COMMA12.
ALL={LABEL="合計 (ALL)"}*
N
'性別' n*F=COMMA12. *
Sum*F=COMMA12. *
'100k(intg)集計用乗率' n*F=COMMA12. ;
;

```

**RUN:**

2) 経常収入および食料費

**PROC TABULATE**

DATA=WORK. QUERY\_FOR\_APPEND\_TABLE\_000D\_0008

```

;

VAR "経常収入"n "食料"n;
TABLE /* 行ディメンション*/
P20*F=COMMA13.
P40*F=COMMA13.
P60*F=COMMA13.
P80*F=COMMA13.,
/* 列ディメンション */
'経常収入' n*F=COMMA13.
'食料' n*F=COMMA13. ;
;
WEIGHT "集計用乗率"n;

```

**RUN:**

ヒストグラム

**PROC UNIVARIATE** DATA=WORK. SORTTempTableSorted NOPRINT ;

VAR "経常収入"n "食料"n;

HISTOGRAM ;

**RUN: QUIT;**

ODS GRAPHICS OFF;

**RUN: QUIT;**

3) 男女別・経常収入および食料費の5分位分布

**PROC TABULATE**

DATA=WORK. QUERY\_FOR\_APPEND\_TABLE\_000D\_0009

FORMAT=COMMA12.

;

VAR "性別2"n "男"n "女"n;

CLASS "経常収入5分位フラグ【文字列】"n / ORDER=UNFORMATTED

MISSING:

CLASS "性別"n / ORDER=UNFORMATTED MISSING;

TABLE /\* 行ディメンション\*/

'経常収入5分位フラグ【文字列】'n\*F=COMMA20.

ALL={LABEL="合計 (ALL)"}\*F=COMMA20.,

/\* 列ディメンション \*/

Sum\*F=COMMA20. \*(

'男' n\*F=COMMA20.

'女' n\*F=COMMA20.) ;

;

WEIGHT "100k(int g)集計用乗率"n;

**RUN;**

**PROC TABULATE**

DATA=WORK. QUERY\_FOR\_APPEND\_TABLE\_000D\_0009

;

VAR "男"n "女"n;

CLASS "食料5分位【文字列】"n / ORDER=UNFORMATTED MISSING;

CLASS "性別"n / ORDER=UNFORMATTED MISSING;

TABLE /\* 行ディメンション\*/

'食料5分位【文字列】'n\*F=COMMA12.

ALL={LABEL="合計 (ALL)"}\*F=COMMA12.,

/\* 列ディメンション \*/

Sum\*(

'男' n

'女' n) ;

;

WEIGHT "100k(int g)集計用乗率"n;

**RUN;**

4) 男女別・経常収入・食料費についての3重クロス表

食料常用対数を算出

**PROC SQL;**

CREATE TABLE WORK. QUERY\_FOR\_APPEND\_TABLE\_000D\_000C AS

SELECT /\* 食料常用対数 \*/

(LOG10(t1.'食料' n)) FORMAT=COMMA32.20 AS '食料常用対数' n,

食品幾何平均を算出

**PROC SQL;**

CREATE TABLE WORK. QUERY\_FOR\_STABSUMMARYTABLES\_0004 AS

SELECT /\* 食品幾何平均 \*/

(10 \*\* t1.'食料常用対数\_Mean' n) AS '食品幾何平均' n,

これを基に集計表を構築

**PROC TABULATE**

DATA=WORK. QUERY\_FOR\_APPEND\_TABLE\_000D\_0012

```

;
VAR "世帯数"n "食料常用対数"n "幾何平均"n;
CLASS "性別"n /ORDER=UNFORMATTED MISSING;
CLASS "經常収入5分位フラグ"n /ORDER=UNFORMATTED MISSING;
TABLE /* 行ディメンション*/
'性別' n*(
' 經常収入5分位フラグ' n
ALL={LABEL="合計 (ALL)"}),
/* 列ディメンション */
'世帯数' n*F=COMMA7. *
Sum*F=COMMA7.
'食料常用対数' n*F=COMMA32. 4*
Mean*F=COMMA32. 4 ;
;
WEIGHT "100k(int g)集計用乗率"n;

```

折れ線グラフ作成

**PROC SORT**

```

DATA=WORK. QUERY_FOR_STABSUMMARYTABLES_0004 (KEEP="經常収入5分位フラグ
"n "食品幾何平均"n "性別"n)
OUT=WORK. SORTTempTableSorted
;
BY "經常収入5分位フラグ"n;

```

**RUN;**

**SYMBOL1**

```

INTERPOL=JOIN
HEIGHT=10pt
VALUE=NONE
LINE=1
WIDTH=2

```

CV = \_STYLE\_

**SYMBOL2**

```

INTERPOL=JOIN
HEIGHT=10pt
VALUE=NONE
LINE=1
WIDTH=2

```

CV = \_STYLE\_

**Legend1**

```

FRAME
;

```

**Axis1**

```

STYLE=1
WIDTH=1
MINOR=
(NUMBER=1
)

```

;

```

Axis2
    STYLE=1
    WIDTH=1
    MINOR=
    (NUMBER=1
    )

;
TITLE;
TITLE1 "折れ線グラフ";
PROC GPLOT DATA = WORK.SORTTempTableSorted
;
PLOT '食品幾何平均'n * "経常収入5分位フラグ"n          ="性別"n
/
    VAXIS=AXIS1

    HAXIS=AXIS2

FRAME LHREF=34
    CHREF=LTGRAY
    AUTOHREF
    LVREF=34
    CVREF=LTGRAY
    AUTOVREF
    LEGEND=LEGEND1
;
RUN; QUIT;

```

(丁)

## 主催：SASユーザー会 世話人会

---

|          |        |                             |
|----------|--------|-----------------------------|
| 代表世話人    | 大橋 靖雄  | 中央大学                        |
| 世話人      | 伊藤 陽一  | 北海道大学                       |
| (氏名50音順) | 小野 潔   | 株式会社インテック                   |
|          | 岸本 淳司  | 九州大学 ARO次世代医療センター           |
|          | 堺 伸也   | イービーエス株式会社                  |
|          | 坂巻 英一  |                             |
|          | 周防 節雄  | 兵庫県立大学                      |
|          | 菅波 秀規  | 興和株式会社                      |
|          | 高橋 行雄  | BioStat 研究所株式会社             |
|          | 野原 賢一  | 独立行政法人 統計センター               |
|          | 松岡 淨   | 日本メディア株式会社/順天堂大学 臨床研究支援センター |
|          | 宮内 亨   | 独立行政法人 統計センター               |
|          | 八木 章   | 近畿大学 元教授                    |
|          | 山之内 直樹 | 第一三共株式会社                    |

## 協賛

(掲載50音順)

イービーエス株式会社  
エイツーヘルスケア株式会社  
FMD K&L Japan株式会社  
クリンククラウド株式会社  
スタットコム株式会社  
株式会社タクミインフォメーションテクノロジー  
TIS 株式会社  
DOTインターナショナル株式会社  
日本メディア株式会社

## 協力

SAS Institute Japan 株式会社

## SASユーザー総会事務局

---

〒160-0022 東京都新宿区新宿6-27-56 新宿スクエア5F  
TEL:03-5774-7018 受付時間:10:00~17:00(平日 12:00~13:00、土・日・祝日を除く)  
E-mail:sasuser\_entry@seminar.jp



## 論文集

2016年7月21日 初版第1刷発行  
発行:SASユーザー会 SAS Institute Japan株式会社