

経済物理学 2017 – 新たな領域との融合  
京都大学 基礎物理学研究所  
湯川記念館 Panasonic 国際交流ホール  
2017年12月7日（木）～ 8日（金）

# ツイート拠点の空間分布による人口分布推計

藤本 祥二 （金沢学院大）  
石川 温 （金沢学院大）  
水野 貴之 （国立情報学研究所, CIGS）

# イントロダクション

- 国勢調査や国民生活調査
  - 大規模な調査（日本でも5年に一度）
  - 頻度を上げたい、出来ればリアルタイムに
  - 発展途上国などでは調査が難しい（インフラ）
- 今回の研究
  - 日本の位置情報付きツイッター投稿データを用いて、人の流れを把握
  - 国勢調査や生活調査の結果と比較
  - 公的な統計調査をどの程度再現できるのか検討

# 使用ツイッターデータ

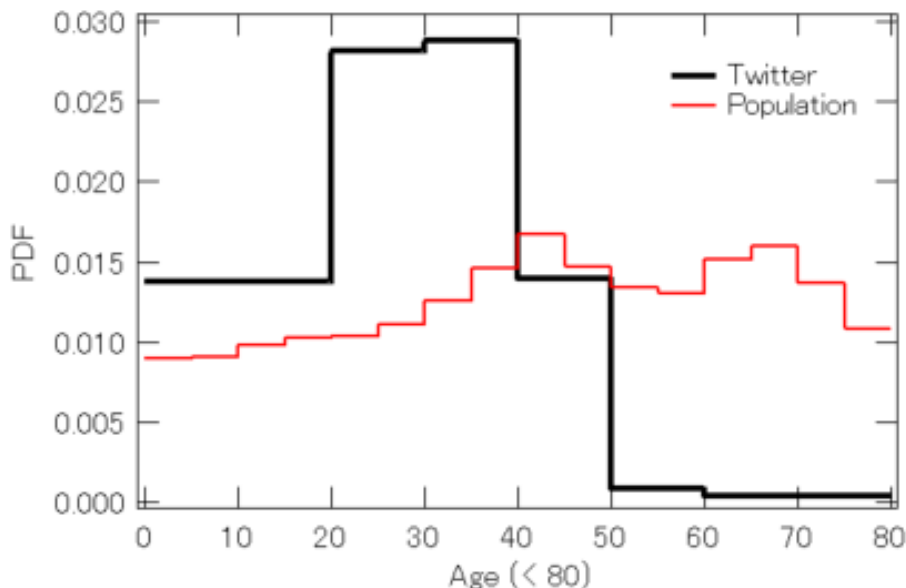
- 緯度・経度の位置情報が付いた  
ツイッターデータ  
(位置情報付きのものだけを入手)
  - ユーザID, 年月日, 時分秒, 緯度, 経度の項目を分析に用いる
  - 日本の領土を全て含んだ長方形の領域  
北緯 : 20.4~45.6                      東経 : 122.9~154.0
  - 期間  
2014/03/04~2015/04/26                      (419日間)

# 40万ユーザーのプロファイル

位置情報付きユーザーの  
約80%程度をカバー

ツイート内容からのプロファイリング情報

- ユーザーID, 都道府県, スクリーンネーム, 年齢 (7分類), 結婚, 年収 (8分類), 性別, 飲酒, 喫煙, カメラ等30種類の趣味の有無, 業界 (16分類), 職種 (10分類), 在住地方 (8分類), 車移動等6種類の移動手段の有無, 職業 (12分類), 役職 (9分類)



| 年齢\性別 | 男      | 女      | 不明    | 合計     |
|-------|--------|--------|-------|--------|
| _19   | 51229  | 18651  | 11133 | 81013  |
| 20_29 | 39517  | 31503  | 11622 | 82642  |
| 30_39 | 57243  | 17755  | 9567  | 84565  |
| 40_49 | 31281  | 5878   | 3916  | 41075  |
| 50_59 | 1708   | 519    | 254   | 2481   |
| 60_   | 1551   | 222    | 218   | 1991   |
| 不明    | 60393  | 30916  | 14924 | 106233 |
| 合計    | 242922 | 105444 | 51634 | 400000 |

⇔ 20代30代が最も多い

# 月別アクティブユーザー数(MAU)

位置情報付きデータの、月別ツイート数と、MAU

| 年/月     | ツイート数     | MAU     | 年/月     | ツイート数      | MAU     |
|---------|-----------|---------|---------|------------|---------|
| 2014/03 | 8,653,616 | 247,589 | 2014/10 | 8,103,683  | 268,674 |
| 2014/04 | 9,055,384 | 243,923 | 2014/11 | 7,445,254  | 268,025 |
| 2014/05 | 9,715,331 | 262,150 | 2014/12 | 10,571,060 | 363,151 |
| 2014/06 | 9,041,571 | 278,228 | 2015/01 | 9,607,290  | 346,343 |
| 2014/07 | 8,648,575 | 250,079 | 2015/02 | 8,924,144  | 377,909 |
| 2014/08 | 9,303,617 | 266,631 | 2015/03 | 11,745,895 | 446,305 |
| 2014/09 | 7,610,038 | 245,515 | 2015/04 | 9,151,611  | 382,600 |

MAU (Monthly Active User) : 1か月に1回以上ツイートしたユーザーのこと

Twitter JP 公式発表MAU

- 2015/12 3500万人
- 2016/11/01 4000万人突破

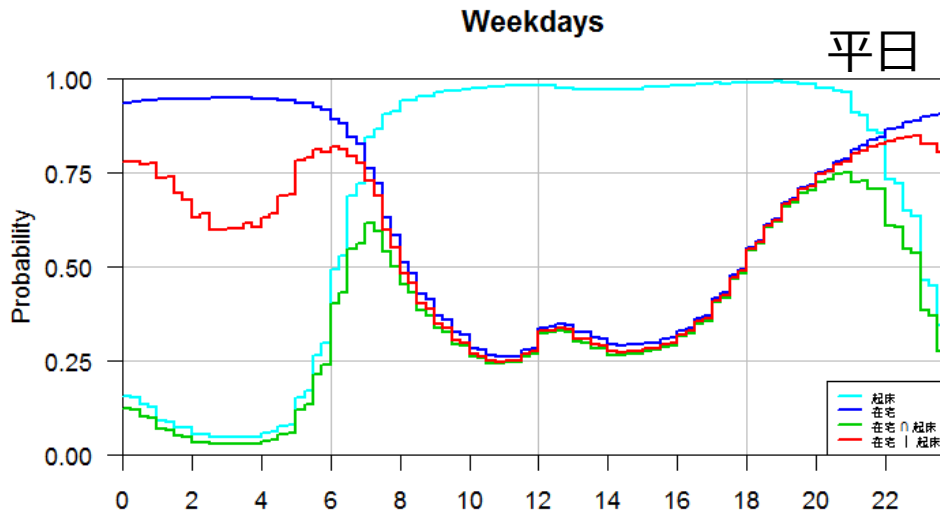
MAUの1~2%程度が位置情報を提供

# 国民生活時間調査

- NHK放送文化研究所（世論調査部）
  - 1960年から，5年に1度の調査
  - 2015年10月13日(火)～26日(月)の調査
    - 住民基本台帳から層化無作為2段抽出  
12,600人（12人×150地点×7回）
    - サンプル数

|    | 指定サンプル数 | 調査有効数  | 率     |
|----|---------|--------|-------|
| 平日 | 18,000  | 11,056 | 61.4% |
| 土曜 | 3,600   | 2,195  | 61.0% |
| 日曜 | 3,600   | 2,170  | 60.3% |

# 国民生活時間調査の比較対象項目

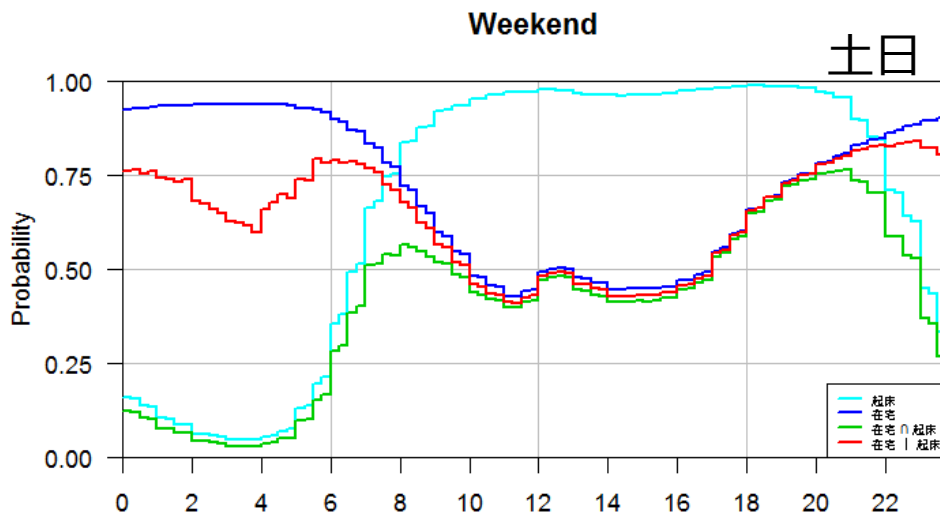


水色 :  $P(\text{起床})$   
 $1 - P(\text{睡眠})$

青色 :  $P(\text{在宅})$

緑色 :  $P(\text{在宅} \cap \text{起床})$

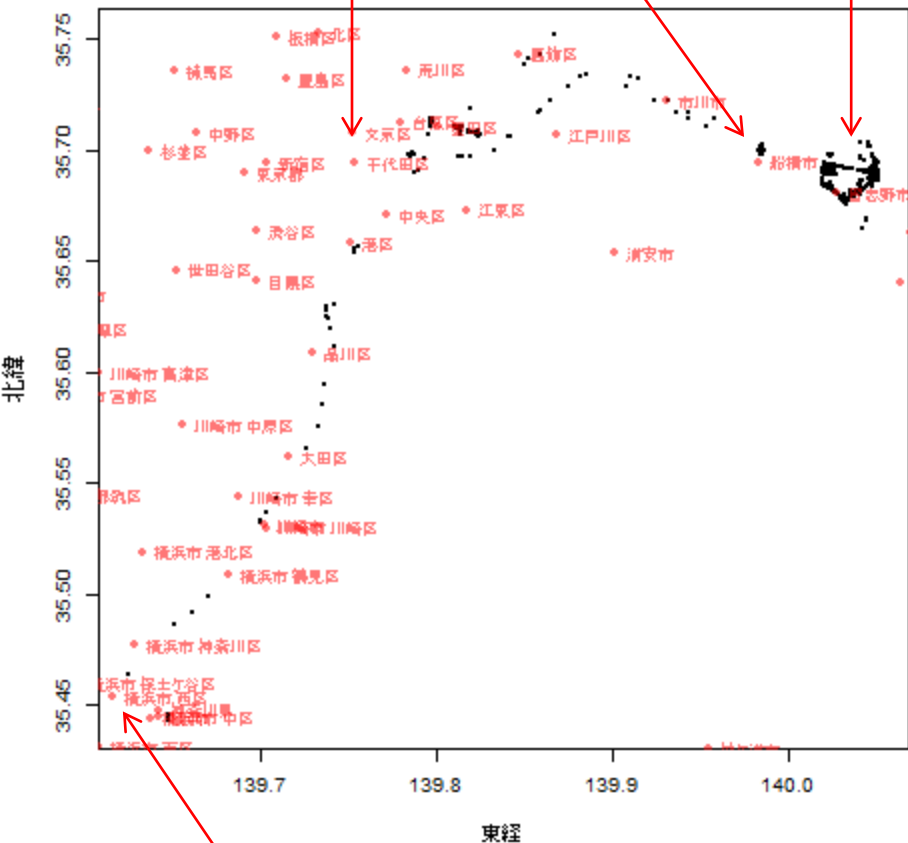
赤色 :  $P(\text{在宅} | \text{起床})$   
 $= \text{緑色} / \text{水色}$



# ツイート拠点の判定

あるユーザーの位置情報付きツイート

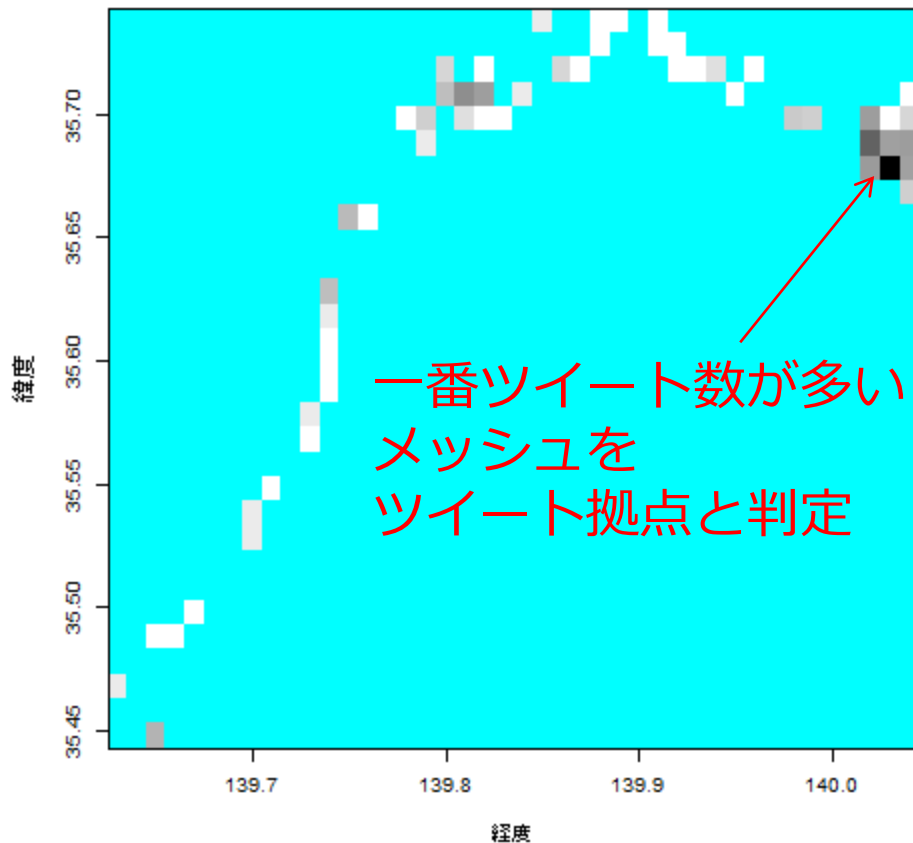
千代田区 船橋市 習志野市



横浜市

約1km四方

メッシュ (3次メッシュ) に切って  
メッシュ内のツイート数をカウント

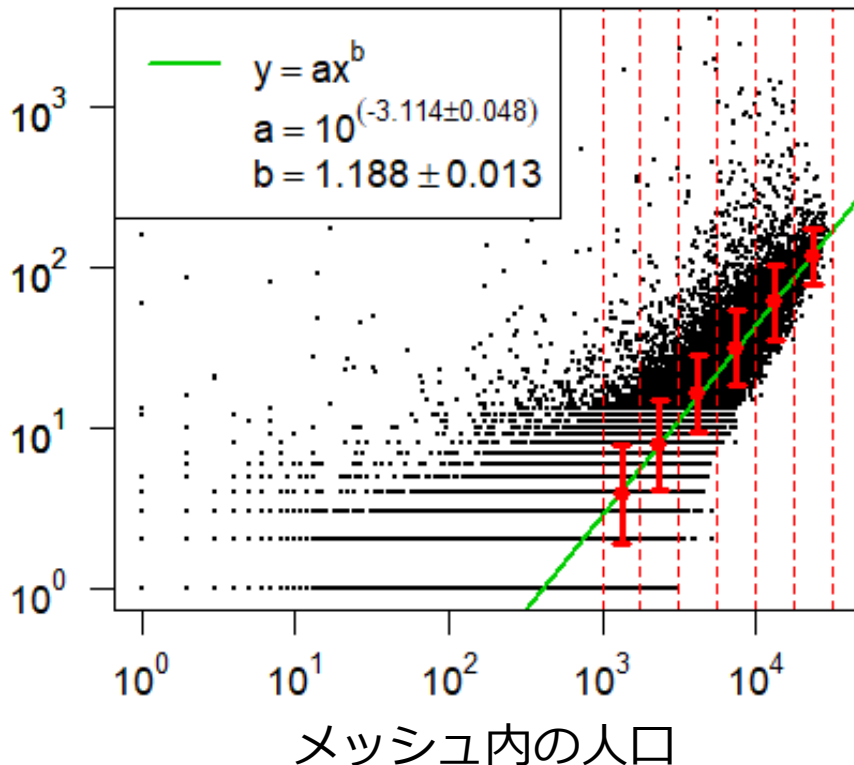


一番ツイート数が多い  
メッシュを  
ツイート拠点と判定

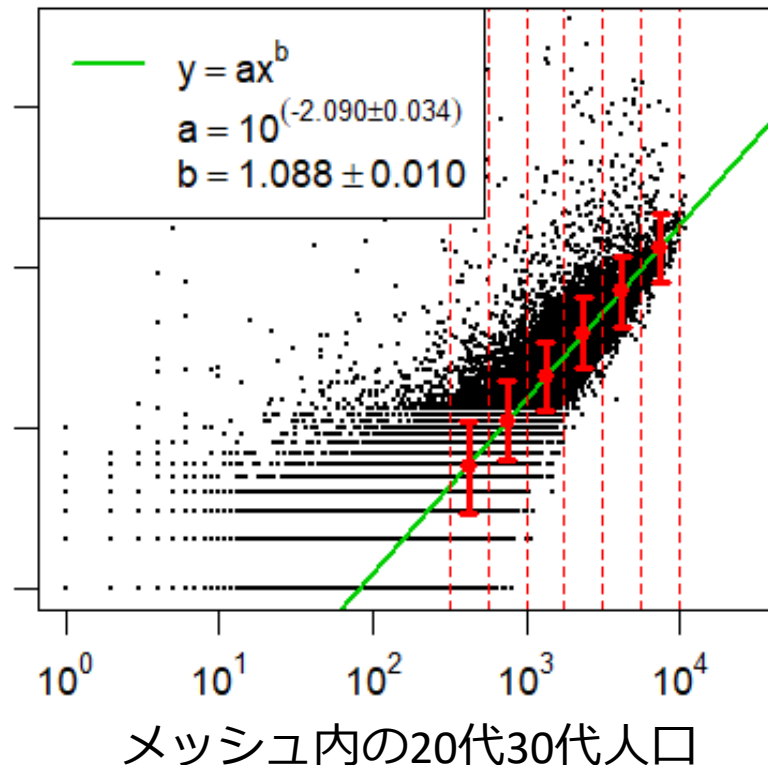


# 人口vsツイート拠点ユーザー数

相関係数 : 0.46  
対数データの相関係数 : 0.79



相関係数 : 0.49  
対数データの相関係数 : 0.83



人口（住宅人口、夜間人口）と、ツイート拠点ユーザー数の比較結果

- 人口の大きなところは傾き1の直線
- エラーバーにx依存性はない
- 20代30代のツイッター利用者の多い人口でよりよく合う

# 時間の分割

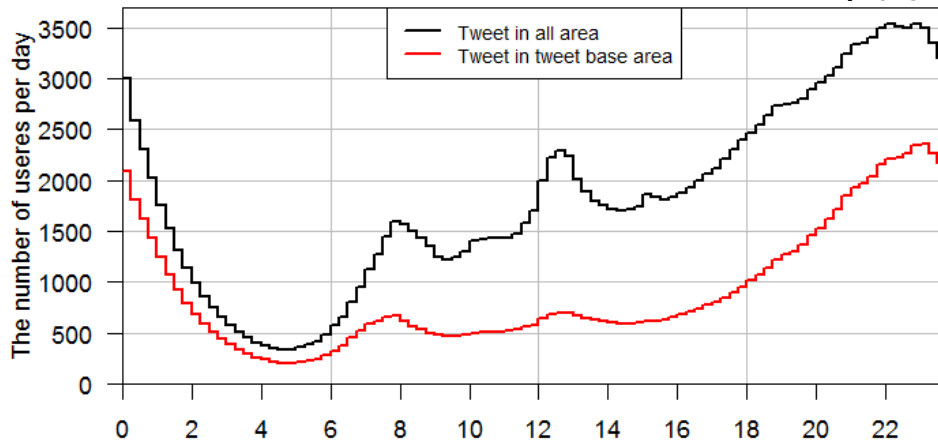
- 1時間を4等分（15分のビン）する
  - 比較対称の国民生活調査の結果に合わせた
- 15分間に一度でも発言したアクティブユーザー数をカウント
- このアクティブユーザー数を地域や曜日等の条件を付けて集計する

# 時間帯別

## 1日当たりのアクティブユーザー数

Weekdays

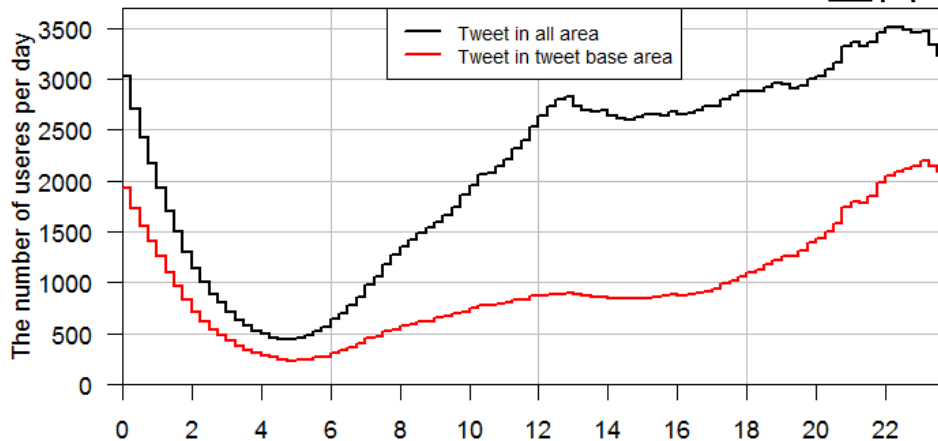
平日



黒線：どこでツイートしたか  
条件なしのユーザー数

Weekend

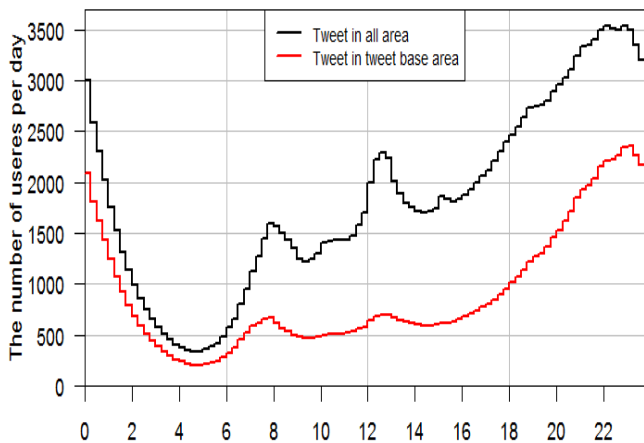
土日



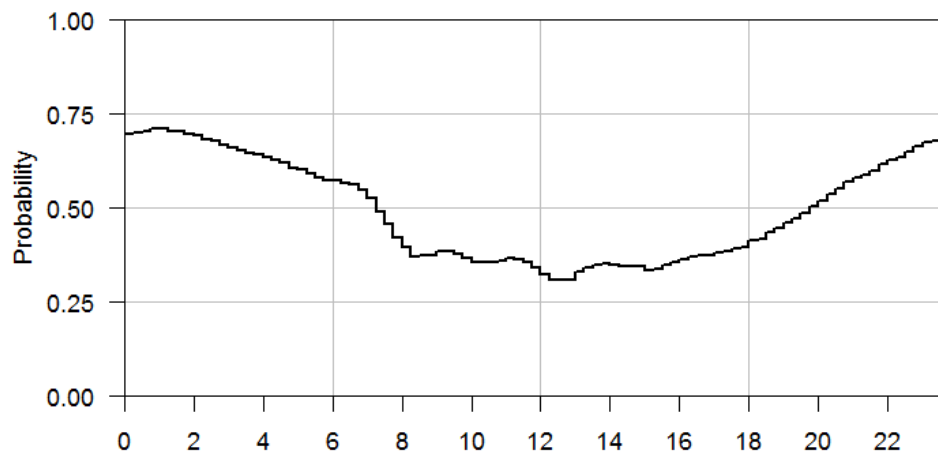
赤線：ツイート拠点でツイートした  
ユーザー数

# ツイート拠点でのツイート率

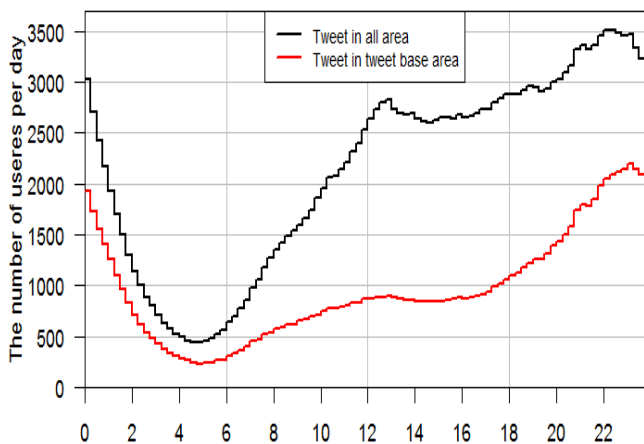
Weekdays



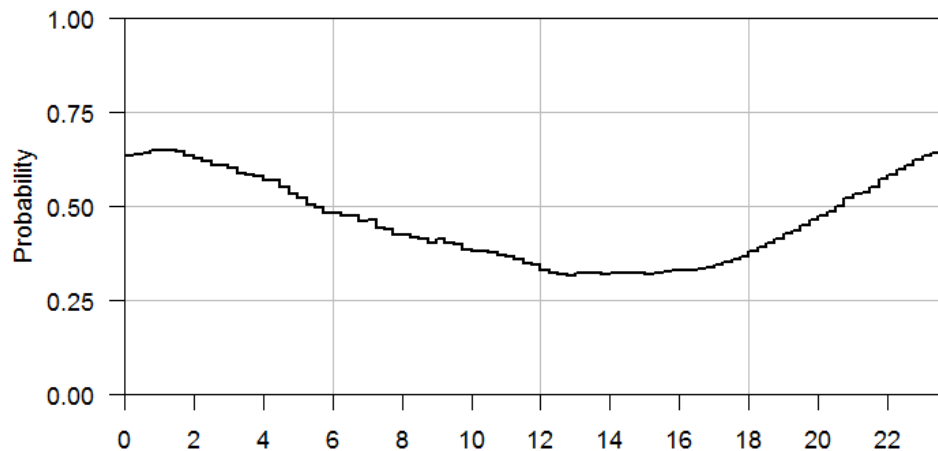
Weekdays



Weekend



Weekend

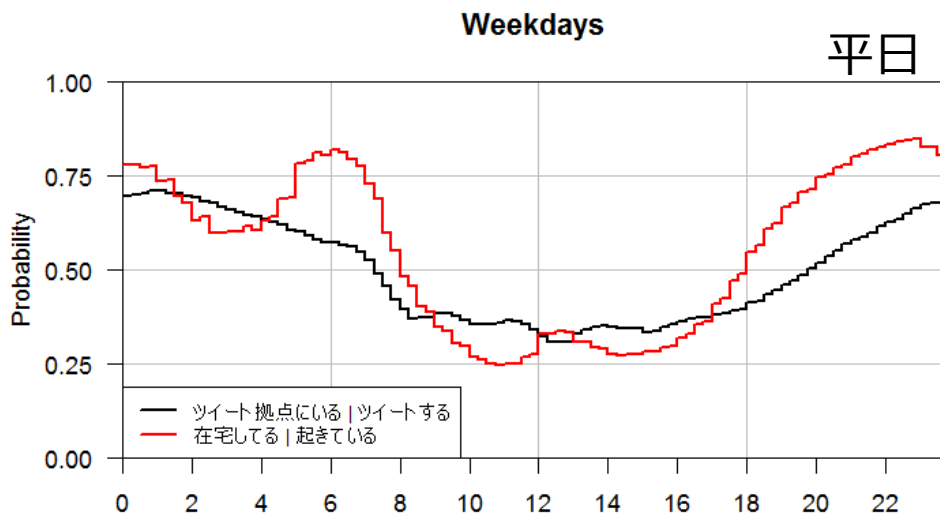


赤線

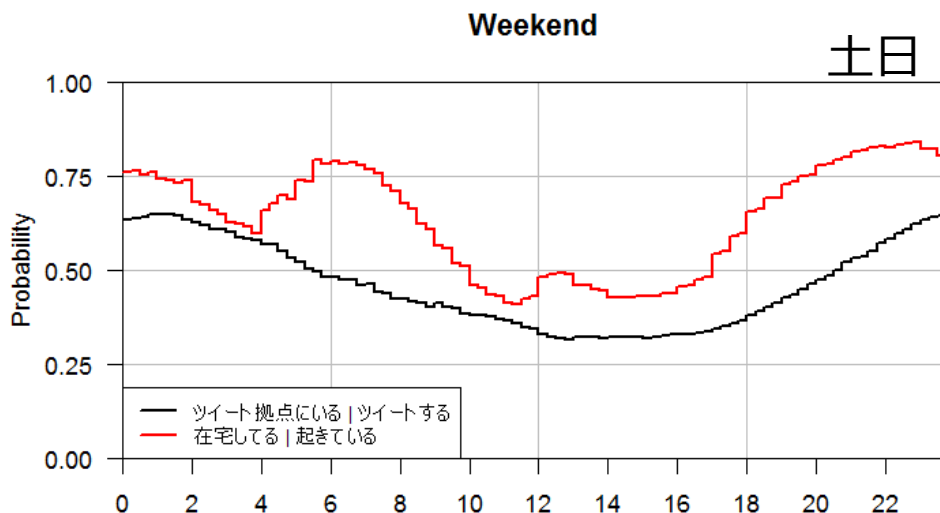
黒線



# ツイート率と生活時間調査の比較



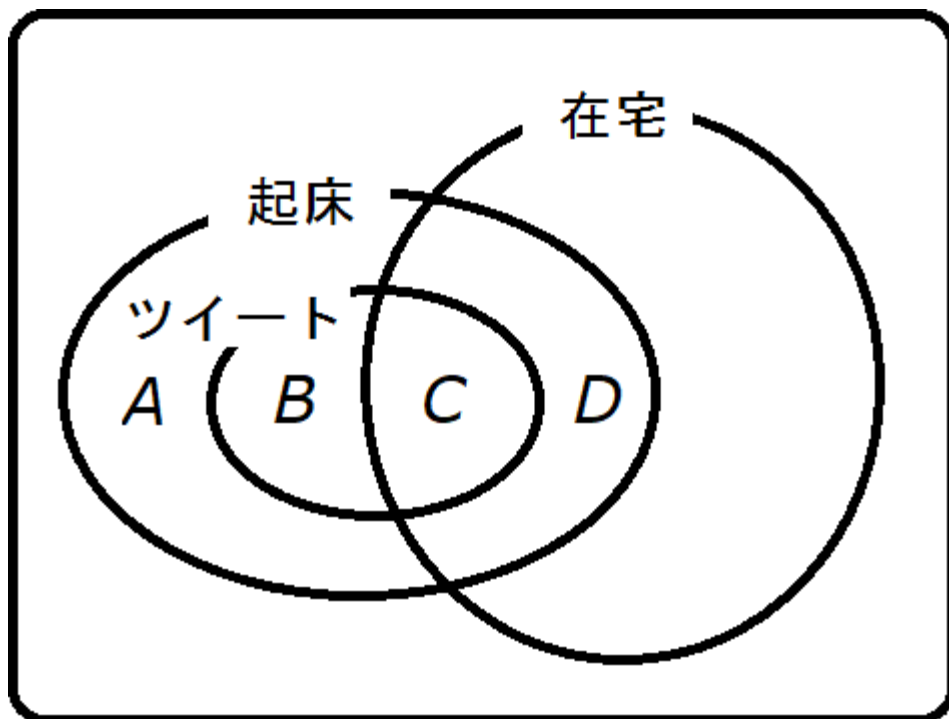
相関係数 : 0.864



黒色 :  $P(\text{ツイート拠点} | \text{ツイート})$   
赤色 :  $P(\text{在宅} | \text{起床})$

相関係数 : 0.781

# 前頁の結果の解釈



ツイート拠点を自宅だと仮定

$$\begin{aligned} \text{黒色} &: P(\text{在宅} \mid \text{ツイート}) \\ &= \frac{C}{B + C} \end{aligned}$$

$$\begin{aligned} \text{赤色} &: P(\text{在宅} \mid \text{起床}) \\ &= \frac{C + D}{A + B + C + D} \end{aligned}$$

赤色 > 黒色 の時 (平日の昼や土日) は

非在宅での  
ツイート/非ツイート

$$\frac{B}{A} > \frac{C}{D}$$

在宅での  
ツイート/非ツイート

# まとめと今後の課題

- ツイッターデータの位置情報を用いて、国勢調査の人口分布と国民生活時間調査の起床在宅率が再現可能な事を示した
  - 頻繁な調査の可能性
  - 発展途上国での可能性
- ツイッターの付加情報を利用する
  - 人々の意識の違いによる分析
  - 防災意識とハザードマップとの関係