

1. ワーキングペーパーの作成（目次案に関しては別紙参照）
（役割分担）
 - 目次・文章案作成、全体まとめ 事務局
 - 図表・基礎データの提供 統計センタ・慶應大学
 - 確認 全員（スケジュール）
 - ～2/28:文書案作成
 - ～3/15:メールにてコメント、意見交換
 - ～3/29:文書の完成
 - 4月に総務省HP掲載（ペーパー掲載メンバ）
 - 資料作成・確認に担務したメンバの内、希望者全員
2. 来年度の計画(未定)
 - 実施予定であるが、内容に関しては検討要
3. その他
 - ビッグデータ連携会議で内容の報告(2/28)
 - 企画部会(統計委員会配下)で実施報告(3/6)

民間が保有する人口データの有効性の検証（目次・シナリオ案）

0 要旨

（主旨）

- 民間が保有するメッシュ人口は国勢調査と一定の相関があり、差率の標準偏差から判断すると、××人以上では国勢調査と同レベルの信頼性がある
- メッシュ人口が一般に活用されることは、社会のために有効である

1 はじめに

（主旨）

- 公に人口分布を把握できるものは国勢調査しかないが、民間からは基地局や GPS を使用した人口データが提供されており、マーケティングや公共施策の評価などに使用されているが、その有効性・信頼性の評価は行われていない
- これらを公的統計と同様に広く利活用していく仕組みを作っていくことは有効なことであり、そのために、そのデータの有効性・信頼性の評価を行うことは必要なプロセスである
- 今回はその中でも GPS によるメッシュ人口データに焦点をあてて、国勢調査や基地局データとの比較することにより、その有効性・信頼性の評価・検証を行う

2 評価・検証方法

(1) 使用するデータ

- (a) Agoop 流動人口データ（GPS）
- (b) モバイル空間統計データ（基地局）
- (c) 国勢調査

(2) 評価・検証する内容

- (a) 国勢調査夜間人口（メッシュ単位）との比較
- (b) 国勢調査昼間人口（市町村単位）との比較
- (c) モバイル空間統計（メッシュ単位）①との比較
- (d) モバイル空間統計との比較（メッシュ単位）②
- (e) モバイル空間統計との比較（メッシュ単位）③

（説明）

- 本文にはデータの特長（ex：24 時間平均しか無いため休日は平日よりも、夜間帯に近いものとして取り扱った）や、検証の目的を記入し、比較対象データ・その抽出条件・検証方法（相関分析・差率分析）などの詳細は別紙に

記述

3 検証結果

- (1) 国勢調査夜間人口（メッシュ単位）との比較
- (2) 国勢調査昼間人口（市町村単位）との比較
- (3) モバイル空間統計との比較①
- (4) モバイル空間統計との比較②
- (5) 人口変動が大きいエリアでの分析

（説明）

- まず散布図と差率グラフを並べ、次に相関係数と差率の平均・標準偏差の比較表と単純な比較コメントを記す。
- この章では極力、考察は記述せず、結果コメントのみとしたい
 - グラフや数値から見た一般論としての推察（ここでは昼間帯の外出）は当章に記述し、検証目的に関わる考察は4章に記述する
 - 特異値を示したメッシュに対する結果は当章に記述し、それに帯する考察は極力4章に記述する
- 必要に応じて特異値を示したメッシュを除いて、再度、散布図・差率グラフ・相関係数・差率の平均・標準偏差を記す（以下、同様）

4 考察

5 結びにかえて

（補論）

統計センタの市区町村の分析

以上

民間が保有する人口データの有効性の検証（記入例）

0 要旨

（主旨）

- 民間が保有するメッシュ人口は国勢調査と一定の相関があり、差率の標準偏差から判断すると、××人以上では国勢調査と同レベルの信頼性がある
- メッシュ人口が一般に活用されることは、社会のために有効である

（用語定義）

- 省略時はメッシュ単位
- 差率＝

1 はじめに

（主旨）

- 公に人口分布を把握できるものは国勢調査しかないが、民間からは基地局や GPS を使用した人口データが提供されており、マーケティングや公共施策の評価などに使用されているが、その有効性・信頼性の評価は行われていない
- これらを公的統計と同様に広く利活用していく仕組みを作っていくことは有効なことであり、そのために、そのデータの有効性・信頼性の評価を行うことは必要なプロセスである
- 今回はその中でも GPS によるメッシュ人口データに焦点をあてて、国勢調査や基地局データとの比較することにより、その有効性・信頼性の評価・検証を行う

2 評価・検証方法

(3) 使用するデータ

- (d) Agoop 流動人口データ（GPS）
- (e) モバイル空間統計データ（基地局）
- (f) 国勢調査

(4) 評価・検証する内容

- (f) 国勢調査夜間人口（メッシュ単位）との比較
 - (f) 国勢調査夜間人口との相関・差率のバラツキを、“モバイル空間統計⇔国勢調査”を基準として評価・検証
- (g) 国勢調査昼間人口（市町村単位）との比較
 - (f) 国勢調査昼間人口（市町村単位）との相関を、時間帯への依存性の視点で評価・検証

- (h) モバイル空間統計（メッシュ単位）①との比較
- (ア) モバイル空間統計との相関・差率のバラツキを、“国勢調査2010⇔2015”を基準として評価・検証
 - (イ) モバイル空間統計との相関・差率のバラツキを、経年変化（2015→2017）の視点で評価・検証
 - (ウ) モバイル空間統計との相関・差率のバラツキを、平日・休日差異の視点で評価・検証
 - (エ) モバイル空間統計との相関を、メッシュサイズの視点で評価・検証
 - (オ) モバイル空間統計との相関を、年齢帯への依存性の視点で評価・検証
- (i) モバイル空間統計との比較（メッシュ単位）②
- (ア) モバイル空間統計との差率とそのバラツキを人口規模の視点で分析
- (j) モバイル空間統計との比較（メッシュ単位）③
- (カ) 日周期の人口変動が大きいエリアを対象に分析
 - (キ) 長期的（2015→2017）な人口変動が大きいエリアを対象に分析
- （説明）
- 本文にはデータの特性（ex：24時間平均しか無いため休日は平日よりも、夜間帯に近いものとして取り扱った）や、検証の目的を記入し、比較対象データ・その抽出条件・検証方法（相関分析・差率分析）などの詳細は別紙に記述

3 検証結果

(6) 国勢調査夜間人口（メッシュ単位）との比較

（主旨）

- 国勢調査との相関において、基地局データよりも相関関係が弱い、差率のバラツキは同程度である。相関が弱いのは人口規模が大きいメッシュ（約3000人以上）である
- 基地局データ、GPSデータとも、散布図のグラフ（何て言う？）の傾きは1以下、差率平均はマイナス値になっているのは、休日昼間帯の外出を意味していると推察できる（本来は夜間データとの比較を行うべきところ24時間平均になっていることの制約）
- 特異値メッシュは都心エリア（ターミナル駅周辺等、昼間帯に人口が集中するエリア）であり、休日昼間帯での過度な集中を意味している→考察候補（集中度合いがMAで違うかも、メッシュ人数としてどちらが正しいかを言うこ

とは難しい)

(説明)

- まず散布図と差率グラフを並べ、次に相関係数と差率の平均・標準偏差の比較表と単純な比較コメントを記す。
- この章では極力、考察は記述せず、結果コメントのみとしたい
 - グラフや数値から見た一般論としての推察（ここでは昼間帯の外出）は当章に記述し、検証目的に関わる考察は4章に記述する
 - 特異値を示したメッシュに対する結果は当章に記述し、それに帯する考察は極力4章に記述する
- 必要に応じて特異値を示したメッシュを除いて、再度、散布図・差率グラフ・相関係数・差率の平均・標準偏差を記す（以下、同様）

(7) 国勢調査昼間人口（市町村単位）との比較

(主旨)

- 10時～16時の全ての時間帯において、国勢調査昼間人口と高い相関関係があるが、10時での相関性が最も強い
- 特異値メッシュを除くことによりXXになる（港区の位置が変わる）
- （市区町村毎の時間別分析をどこまで表現するか）→時間帯別データを使用しており、分析レベルが異なるので、補論として切り出したい

(8) モバイル空間統計との比較①

(主旨)

- 国勢調査（2010：2015）間の相関性と比べて、相関関係が弱く、バラツキも大きい
- 平日・休日とも相関関係があり、休日は特に強い相関性を示すが、差率のバラツキに関しては平日・休日で差が無い
- 休日は経年による相関係数・バラツキの変化が無いが、平日は経年により相関性が弱くなっている
- 特異値メッシュ（平日データでGPSの人口が過剰）が2点あり、この2点を除くと平日の相関性は休日と同程度となり、経年による相関性の変化も無くなる
- 経年により平日は相関関係が弱くなっていくが、休日は相関関係の強さが変わらない。平休日ともバラツキは変化がない
- メッシュサイズを大きくすることにより相関性は強くなるが、バラツキに関

しては変わらない

- 10代・70代の基地局データよりも、20代～60代の基地局データと強い相関を示す

(9) モバイル空間統計との比較②

(主旨)

- 人口10000人以上のメッシュで差率は0.3に収束している(どちらが多い?)
-
-

(10) 人口変動が大きいエリアでの分析

- ① 日周期
- ② 長期(2年間)

(主旨)

-

4 考察

(主旨)

- GPSデータは国勢調査や基地局データとの相関関係が見られる
- データの特性として・・・(モバイル空間統計との比較における、基地局データとの方式の違いと、標本母数・特性の違い)ので、・・・は留意する必要がある
 - GPSはエリア分解能が高いが、基地局よりも標本の偏りがある
 - 限定したエリア内で極端に人口が集中するケース(ターミナル駅周辺等)
 - 汐留・新橋エリア
 - 住宅街、高齢者が集中している
 - 等
- 500mメッシュ人口が約3000人を超えるメッシュは国勢調査と同等のレベルの信頼がある

5 結びにかえて

(主旨)

- メッシュ人口が一般に活用できるようにされることは、社会のために有効である
- ・・・・(残検討事項)に関しては合理的な推定が行えていないので、それを明らか

にしていくことにより有用性が増す

(補論)

統計センタの市区町村の分析

(検討事項)

- 主旨は GPS データの有効性の検証か、それとも基地局も含めた民間メッシュ人口データの有効性か (目的からすると後者だが、検証内容からすると前者である)
→はじめにの表現で工夫する
- 533936902・533936904 の扱い。相関係数等の数値は該当メッシュを除いたものを算出するが、①理由の表現、②全体 (他メッシュ) への影響をどう考えるのか (全てが無効になるとは思わないが、どこまで有効とするか)
→

以上