

全国家計構造調査における ウエイトの推定方法について

令和元年 8 月 30 日

総務省統計局

2014年全国消費実態調査におけるウエイト

二人以上の世帯

$$i\text{市区町村、}j\text{世帯の集計用乗率}\beta_{ij} : \beta_{ij} = C_{qk} \times \sum_{m \in B} (\tilde{\alpha}_{im} \times M_{ijm})$$

ここで (q, k) は i 市区町村, j 世帯が該当する地方、世帯人員階級

C_{qk} : q 地方、 k 世帯人員の世帯分布補正係数

q 地方区分 : 北海道、東北、関東、北陸、東海、近畿、中国、四国、九州、沖縄
(計10区分)

k 世帯人員 : 2人、3人、4人、5人以上 (計4区分)

$\tilde{\alpha}_{im}$: i 市区町村、第 m 月目の調整済調整係数 (抽出率と回収率による世帯数の復元倍率)

M_{ijm} : i 市区町村、 j 世帯、第 m 月目の集計可能な調査票の有無 (1又は0)

B : 調査票を調査する第 m 月目の集合 ({1, 2, 3}、{2} 又は {3})

補正係数 C_{qk} の作成方法

$$C_{qk} = \frac{W_{qk}}{\sum_{(i,j) \in K_{qk}} \sum_{m \in B} (\tilde{\alpha}_{im} \times M_{ijm}) \times \frac{1}{|B|}}$$

W_{qk} : q 地方、 k 世帯人員の二人以上の世帯数 (労働力調査2014年平均)

K_{qk} : q 地方、 k 世帯人員に属する二人以上の世帯の集合

$|B|$: 調査票を調査する月数 (B の要素数)

2014年全国消費実態調査におけるウエイト

単身世帯

$$i\text{市区町村、}j\text{世帯の集計用乗率}\beta'_{ij} : \beta'_{ij} = D_{q'h} \times \sum_{m \in B'} (\tilde{\alpha}'_{im} \times M_{ijm})$$

ここで (q', h) は i 市区町村、 j 世帯が該当する**地方、男女・年齢階級**

$D_{q'h}$ ： q 地方、 h 男女・年齢階級の世帯分布補正係数

q 地方区分：北海道・東北、関東、北陸・東海、近畿、中国・四国、九州・沖縄
(計6区分)

h 男女・年齢階級区分：男女×{35歳未満、35～59歳、60歳以上} (計6区分)

$\tilde{\alpha}'_{im}$ ： i 市区町村、第 m 月目の調整済調整係数

M_{ijm} ： i 市区町村、 j 世帯、第 m 月目の集計可能な調査票の有無 (1又は0)

B' ：調査票を調査する第 m 月目の集合 ({1、2}、 {1} 又は {2})

補正係数 $D_{q'h}$ の作成方法

$$D_{q'h} = \frac{W'_{q'h}}{\sum_{(i,j) \in H_{q'h}} \sum_{m \in B'} (\tilde{\alpha}'_{im} \times M_{ijm}) \times \frac{1}{|B'|}}$$

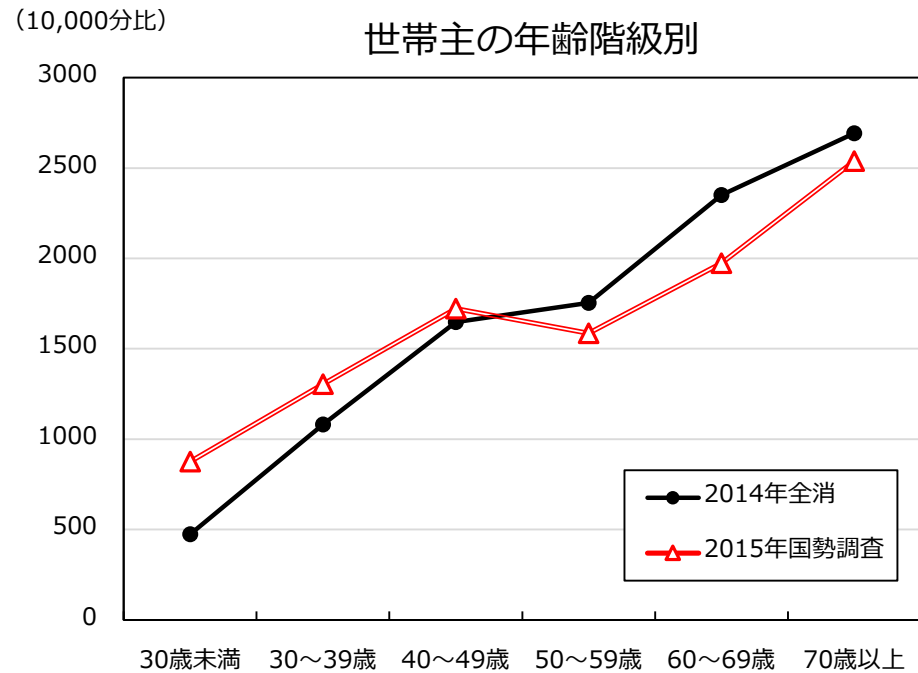
$W'_{q'h}$ ： q 地方、 h 男女・年齢階級の単身世帯数 (**労働力調査2014年平均**)

$H_{q'h}$ ： q 地方、 h 男女・年齢階級に属する単身世帯の集合

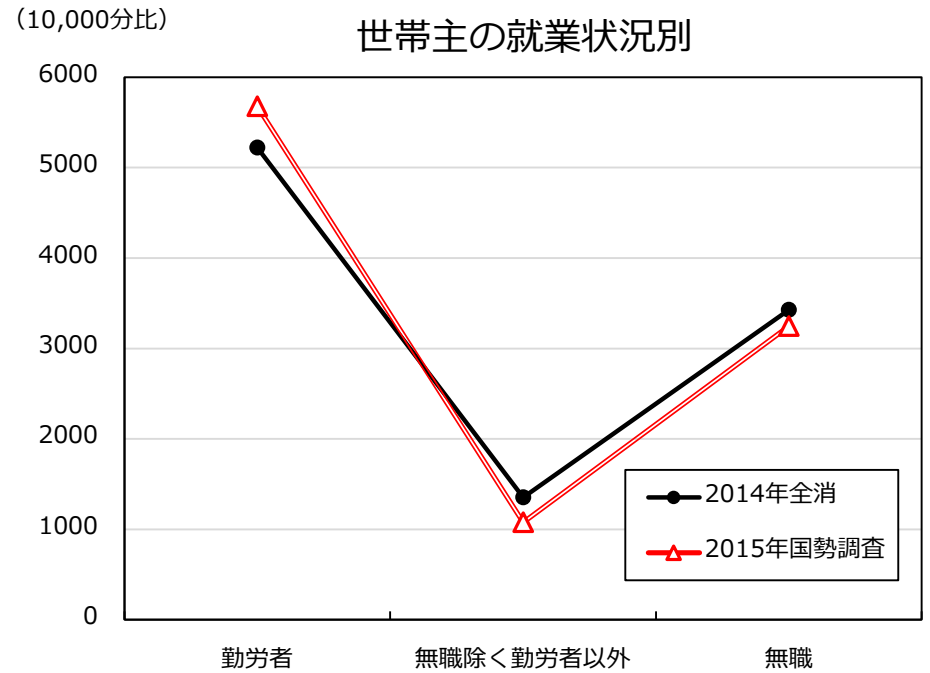
$|B'|$ ：調査票を調査する月数 (B' の要素数)

全国消費実態調査結果の世帯分布における特徴

世帯主の属性別世帯分布
(2014年全国消費実態調査結果, 2015年国勢調査結果)



※国勢調査においては一般世帯のみ。年齢「不詳」を除く



※国勢調査においては一般世帯のみ。全消の定義に合わせ特別集計を行った。

全国消費実態調査の結果は、国勢調査と比べ、以下のような特徴が見られる。

- 世帯主に若年者が少なく、高齢者が多い
- 世帯主に勤労者が少なく、勤労者以外が多い

全国家計構造調査における世帯の抽出方法

2014年全国消費実態調査からの主な変更点は以下のとおり：

- ① 1つの調査単位区から抽出する世帯の数を、二人以上の世帯については11から10へ減らし、単身世帯については1から2へ増やす。
- ② 家計調査世帯特別調査を実施し、家計調査世帯のうち約6,000世帯を集計に使用する。
- ③ 同時期に実施する全国単身世帯収支実態調査による結果を、バイアス補正を行った上で集計に使用する。
- ④ 調査単位区内の世帯抽出に使用する世帯名簿の作成について、調査員の聞き取り以外に、住民基本台帳を補助的に使用することも可能とする。

上記の変更により、全国家計構造調査結果における世帯分布は、従来の全国消費実態調査結果における世帯分布に対し、新たなバイアスを発生させる可能性がある。

今までのような状況に対処するために、集計用乗率の分布補正の手法を改善し、全国家計構造調査及び過去の全国消費実態調査結果に対して適用することにより、精密な分布補正及び時系列比較可能性の確保を行う。

参考：全国家計構造調査の設計概要

市町村調査 (市:793 町村:215)

都道府県調査

簡易調査
(ショートフォーム)

基本調査
(ロングフォーム)

単身世帯
ミタ調査

家計調査世帯
特別調査

個人収支
状況調査

所得資産集計体系

44,000世帯

40,000世帯

2,000世帯

6,000世帯

900世帯

世帯票

世帯票

世帯票

特別
調査票

世帯票

年収・貯蓄等調査票

年収・貯蓄等調査票

年収・
貯蓄等
調査票

年間収入
調査票

年間収入
調査票

貯蓄等
調査票

家計簿

家計簿

家計簿

個人
収支簿

家計総合集計体系

個人収支
集計体系

ウェイト補正方法改善の方向性

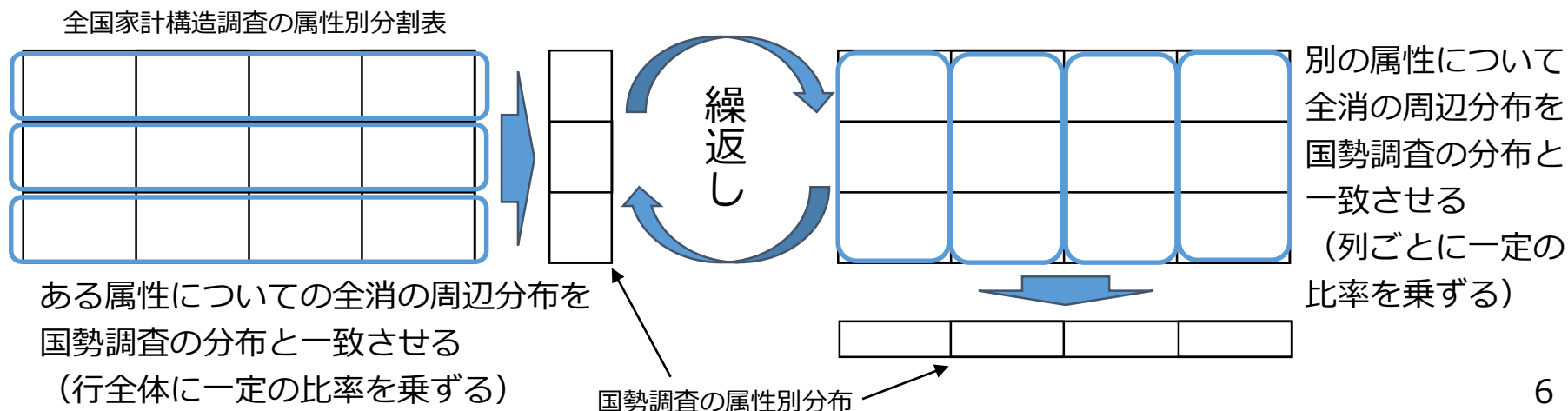
世帯の収支、資産等の分布に影響を与える世帯属性のうち、以下の基本的な属性の分布を全て都道府県別に補正したい。

- 世帯人員階級
- 世帯主の年齢階級
- 世帯主の就業状況（勤労、勤労以外、無職）

全国家計構造調査の世帯分布について、3つの属性による同時分布を、都道府県ごとに一致させるのは困難（調査世帯数が少ないため）

➡ 繰返し比例補正法（IPF法）により、3つの属性に対する周辺分布を満たす世帯分布を推定する

IPF法の計算イメージ（2次元の場合）



第9回消費統計研究会における議論①

以下の属性分布を補正する。

- 世帯主の年齢：25歳未満、25～34歳、…、75～84歳、85歳以上
- 世帯主の性別：男性、女性（単身世帯のみ）
- 世帯人員：1人、2人、3人、4人、5人以上
- 世帯主の就業状況：勤労世帯、無職世帯、その他
- 地域：都道府県

以下の5つの周辺分布が国勢調査に合うようにウエイトを作成する。

- 0層目** 調整済調整係数
- 1層目** 世帯人員×性別×世帯主の年齢
- 2層目** 単身世帯×就業状況
- 3層目** 都道府県×就業状況
- 4層目** 単身世帯×都道府県
- 5層目** 世帯主の年齢×就業状況

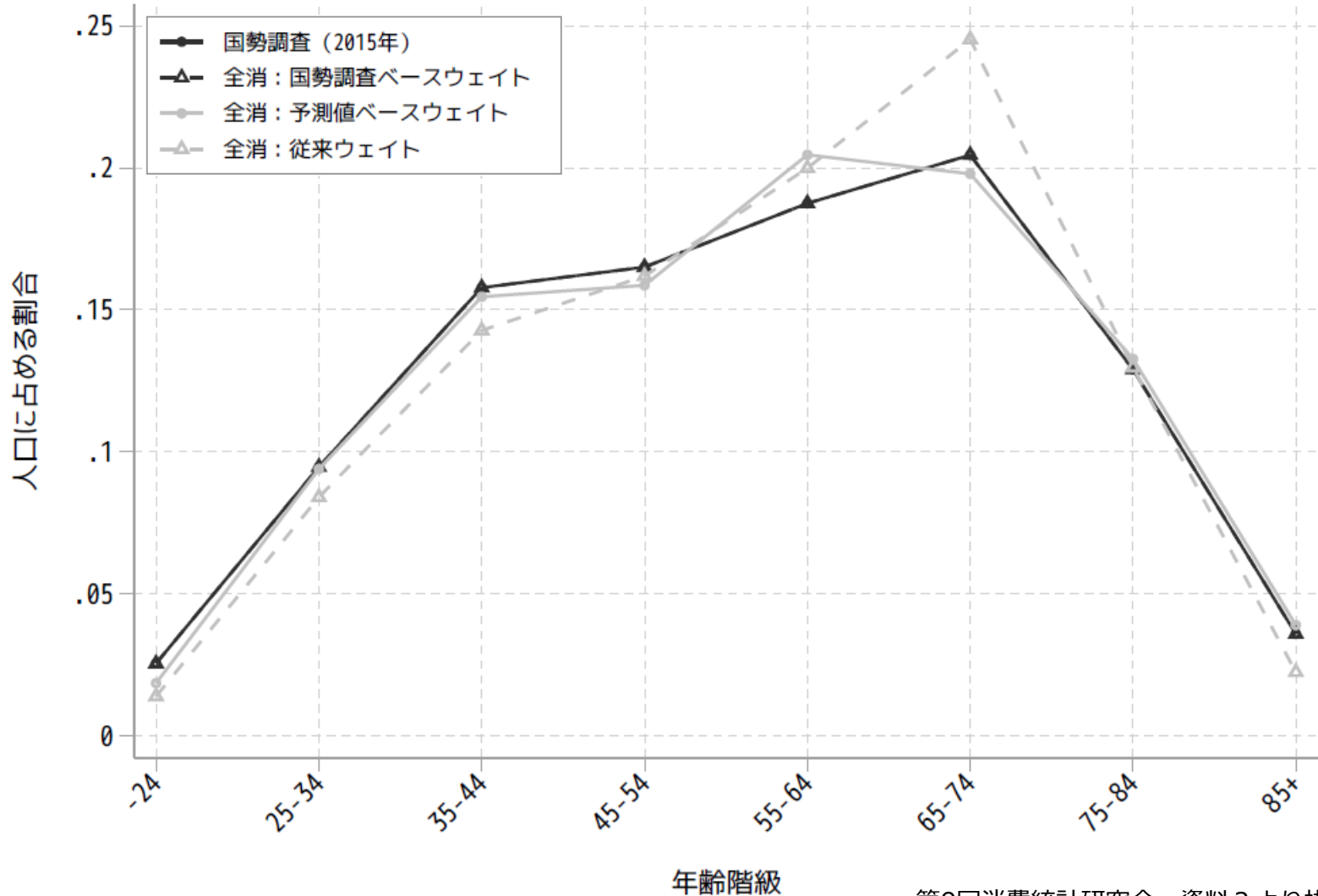


世帯主の年齢分布や年齢別就業率を国勢調査に合わせることはできた。一方で、住居の所有関係や共働き世帯の割合などは更なる調整が必要。

第9回消費統計研究会における議論②

世帯主の年齢階級別分布

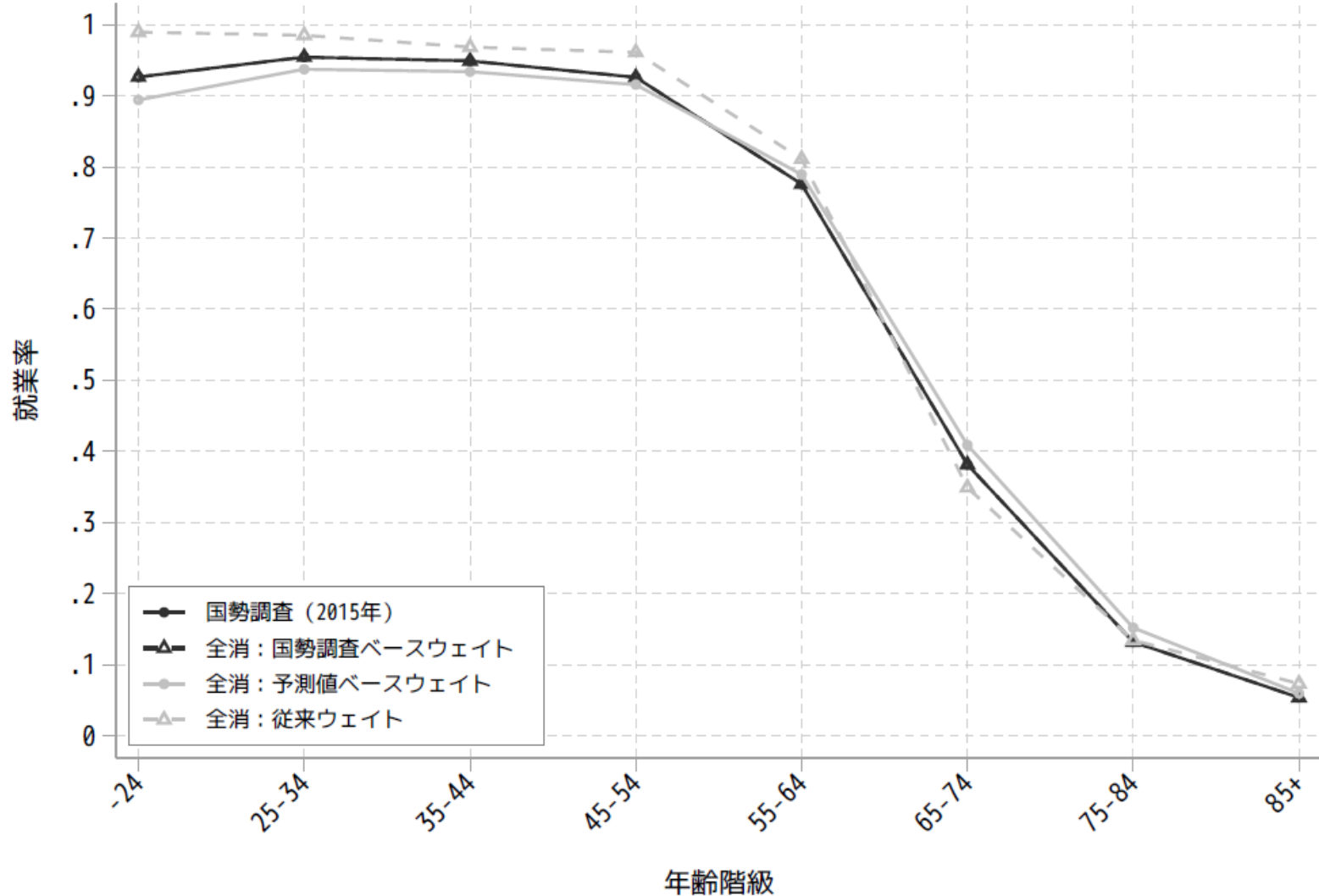
国勢調査と全国消費実態調査の比較（2014年）



第9回消費統計研究会における議論③

世帯主の年齢階級別就業率

国勢調査と全国消費実態調査の比較 (2014年)



第9回消費統計研究会における議論④

住居の所有関係別世帯分布

ウェイトの種類	従来ウェイト	全国消費実態調査		国勢調査
		国勢調査ベース	予測値ベース	
持ち家	0.754	0.740	0.748	0.676
民営の賃貸住宅	0.161	0.174	0.170	0.219
公営の賃貸住宅	0.039	0.038	0.037	0.040
都市再生機構・公社等	0.013	0.013	0.012	0.015
社宅・公務員住宅	0.023	0.025	0.023	0.025
借間	0.006	0.007	0.007	0.009
寮・寄宿舍	0.003	0.003	0.003	0.017

前回の議論からの変更点①

使用するデータセットを、全国のものから都道府県別のものに変更

前は全国のデータがそろってから計算することを前提としていたが、演算管理の関係上、都道府県別のデータで計算する必要

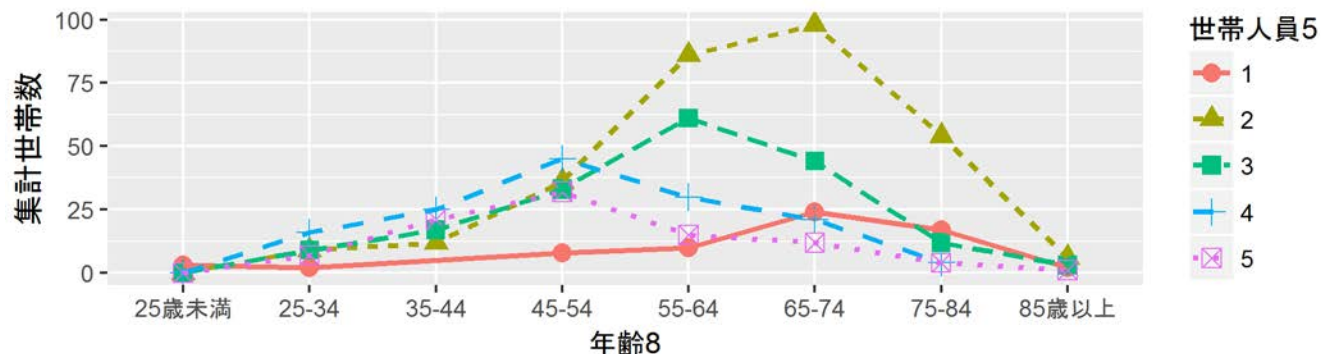
前回と同じ階級の定義及び層の定義では、全国家計構造調査の調査世帯で当てはまるものがない部分ができる可能性

例えば

- 世帯人員別に見た場合、世帯人員が多くなるほど、世帯主が25歳未満や85歳以上の世帯の調査世帯数が少なくなる

IPF法に用いる初期分布のうち、世帯数が0となるセルの割合が多くなると、計算の収束は保証されず、収束も遅くなる (Norman, 1999)

例：岩手県の年齢階級別集計世帯数



前回の議論からの変更点②

都道府県別計算を行うにあたり、以下の条件で計算が収束するかを確認した。

① 前回研究会の計算と同じ階級及び層の定義を使用した場合

➡ 計算が収束しない都道府県が存在した。

層の定義条件が細かくなると、全消の調査世帯が含まれないセルが増加し全体の収束を悪くするためと考えられる。

② 層の定義を以下の条件に変更した場合（0層目は同じ）

1層目 世帯人員

2層目 世帯主の年齢

3層目 世帯主の就業状態

➡ 单身及び二人以上の世帯についての分布が再現できなかった。

計算は収束し、総世帯についてはよく再現できていた。世帯人員と他の条件を別々の層に入れてしまったことが原因と考えられる。

上記2点の問題を解決するために、以下の方法を採用した。

- 層の定義により全消の世帯が含まれないセルについても、ごく小さなウエイトを割り当てることで、収束しない問題を回避する（Choupani et al, 2016）。
- 層の定義条件を工夫し、二人以上と单身をそれぞれ再現できるようにする（次のスライドを参照）。

前回の議論からの変更点②

階級の定義及び層の定義を以下のように変更

階級の定義（全ての都道府県で同じ・前回の計算から変更なし）

- 世帯主の年齢：25歳以下、25～34歳、…、75～84歳、85歳以上
- 世帯主の性別：男性、女性（単身世帯のみ）
- 世帯人員：1人、2人、3人、4人、5人以上
- 世帯主の就業状況：勤労世帯、無職世帯、その他

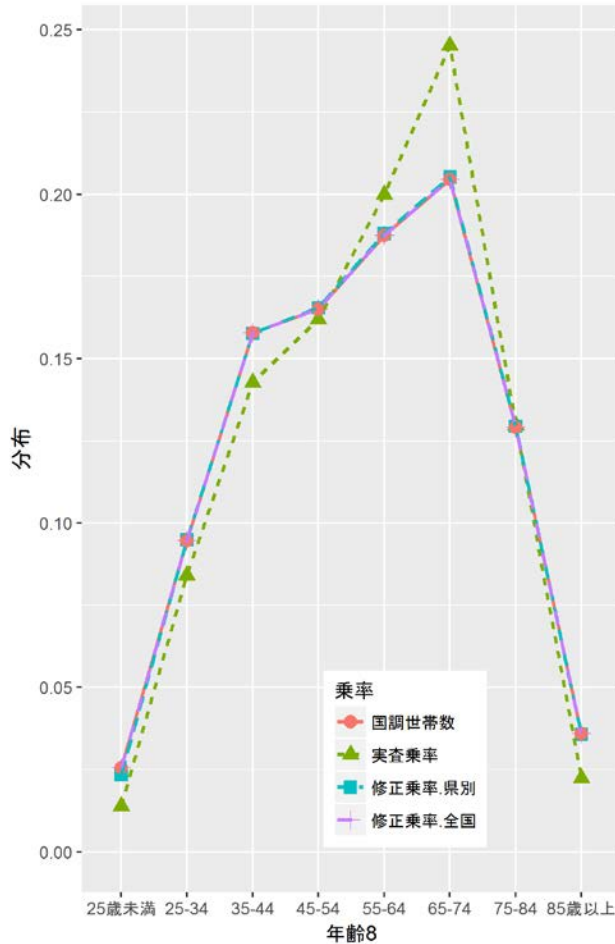
層の定義（全ての都道府県で同じ）

- 0層目** 調整済調整係数
- 1層目** 世帯人員（二人以上の世帯のみ）
- 2層目** 世帯主の年齢（二人以上の世帯のみ）
- 3層目** 世帯主の就業状況（二人以上の世帯のみ）
- 4層目** 世帯主の性別×単身と二人以上の別（総世帯）
- 5層目** 世帯主の年齢（総世帯）
- 6層目** 世帯主の就業状況（総世帯）

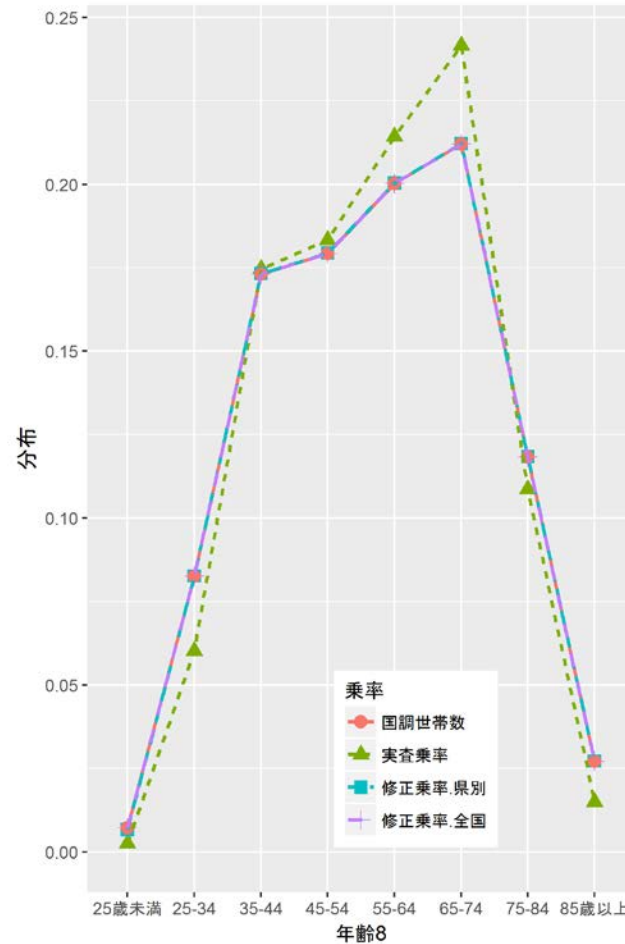
二人以上の世帯と総世帯を別々に合わせることで、間接的に単身世帯の分布を合わせることを目指す。

ウェイト推定結果：全国・年齢階級別世帯分布

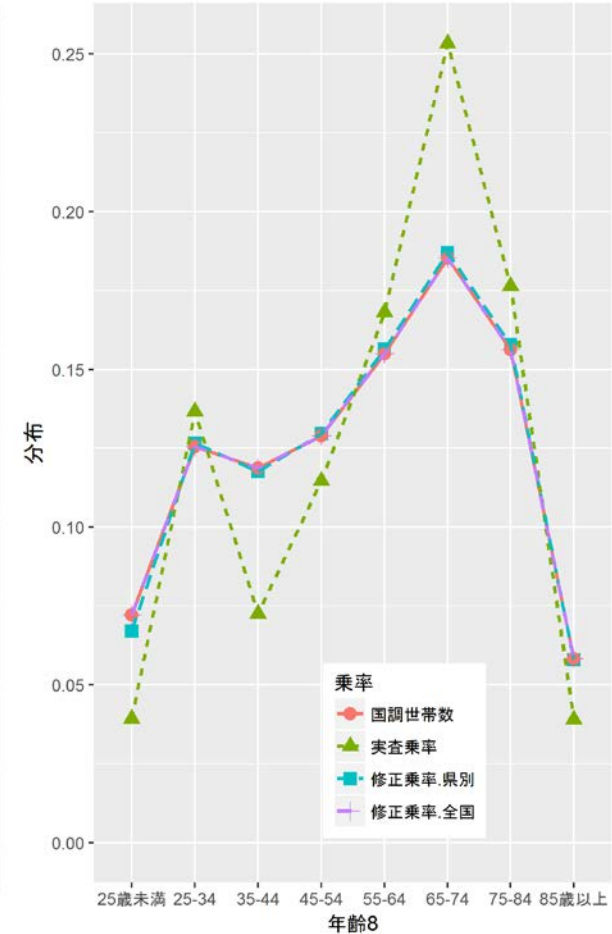
総世帯



二人以上の世帯



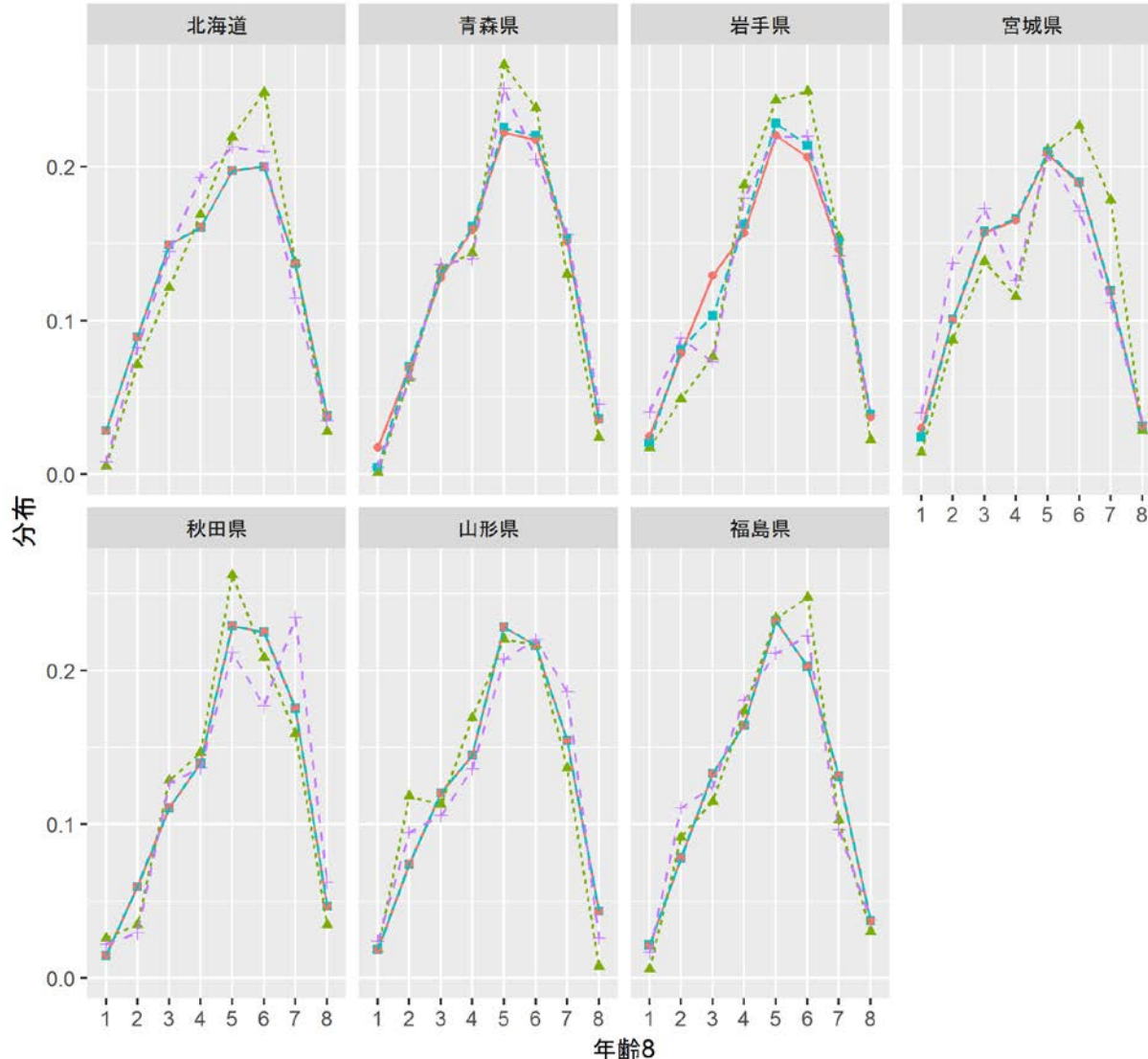
単身世帯



全国データでの推定（前回と同じ計算）と都道府県別データでの推定に大きな違いは見られない。どちらも国勢調査の分布をよく再現できている。

ウエイト推定結果：都道府県別①

乗率 — 国調世帯数 — 実査乗率 — 修正乗率 県別 — 修正乗率 全国



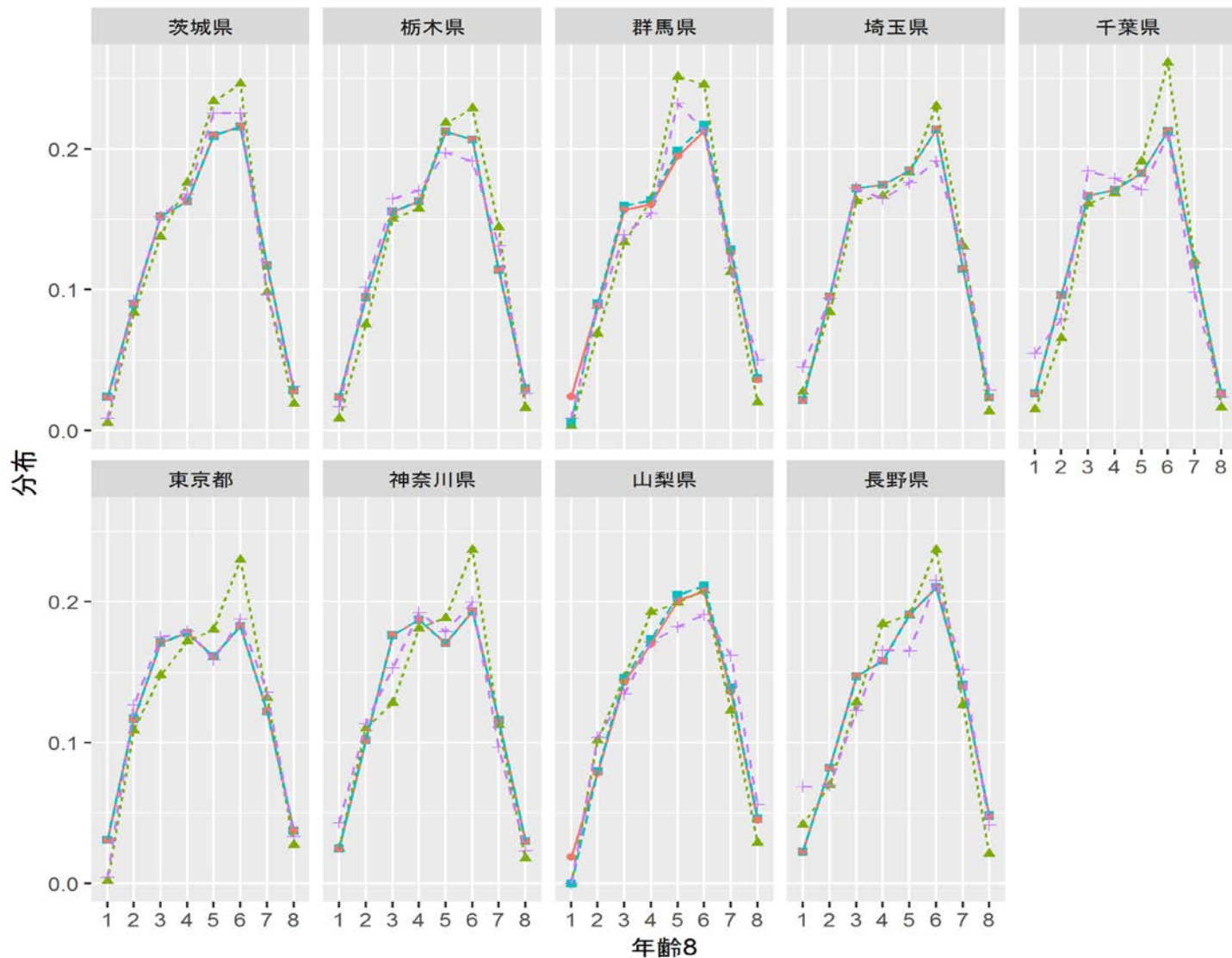
都道府県別年齢階級分布 (北海道・東北)

全国のデータで推定した結果もおおむね国勢調査の分布に合っているが、違いが大きくなる部分も存在する。都道府県別のデータで推定した結果の方は、国勢調査の分布とほぼ一致している。

横軸の1～8はそれぞれ年齢階級の25歳未満～85歳以上に対応

ウエイト推定結果：都道府県別②

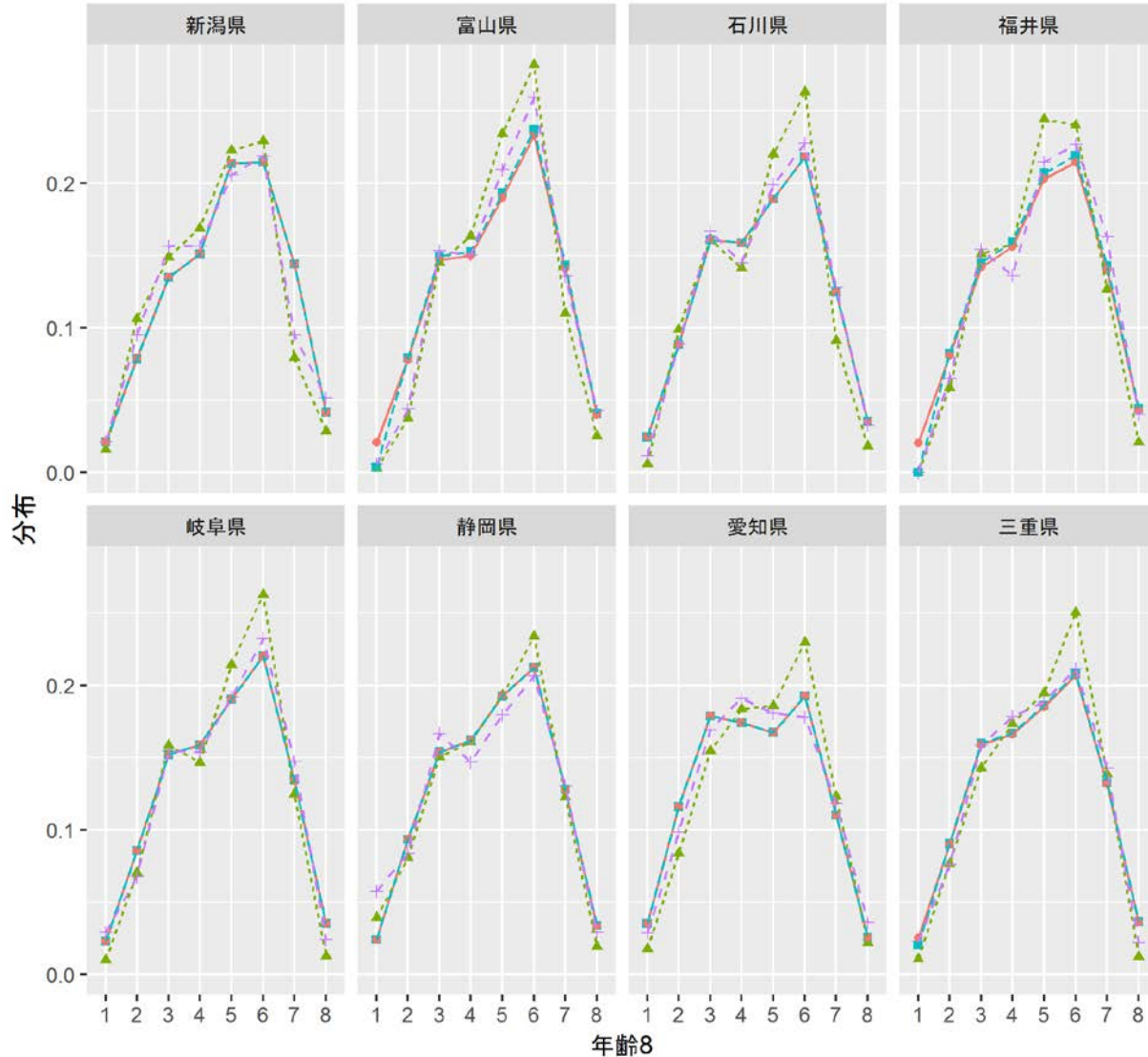
乗率 国調世帯数 実査乗率 修正乗率.県別 修正乗率.全国



都道府県別
年齢階級分布
(関東地方)

ウエイト推定結果：都道府県別③

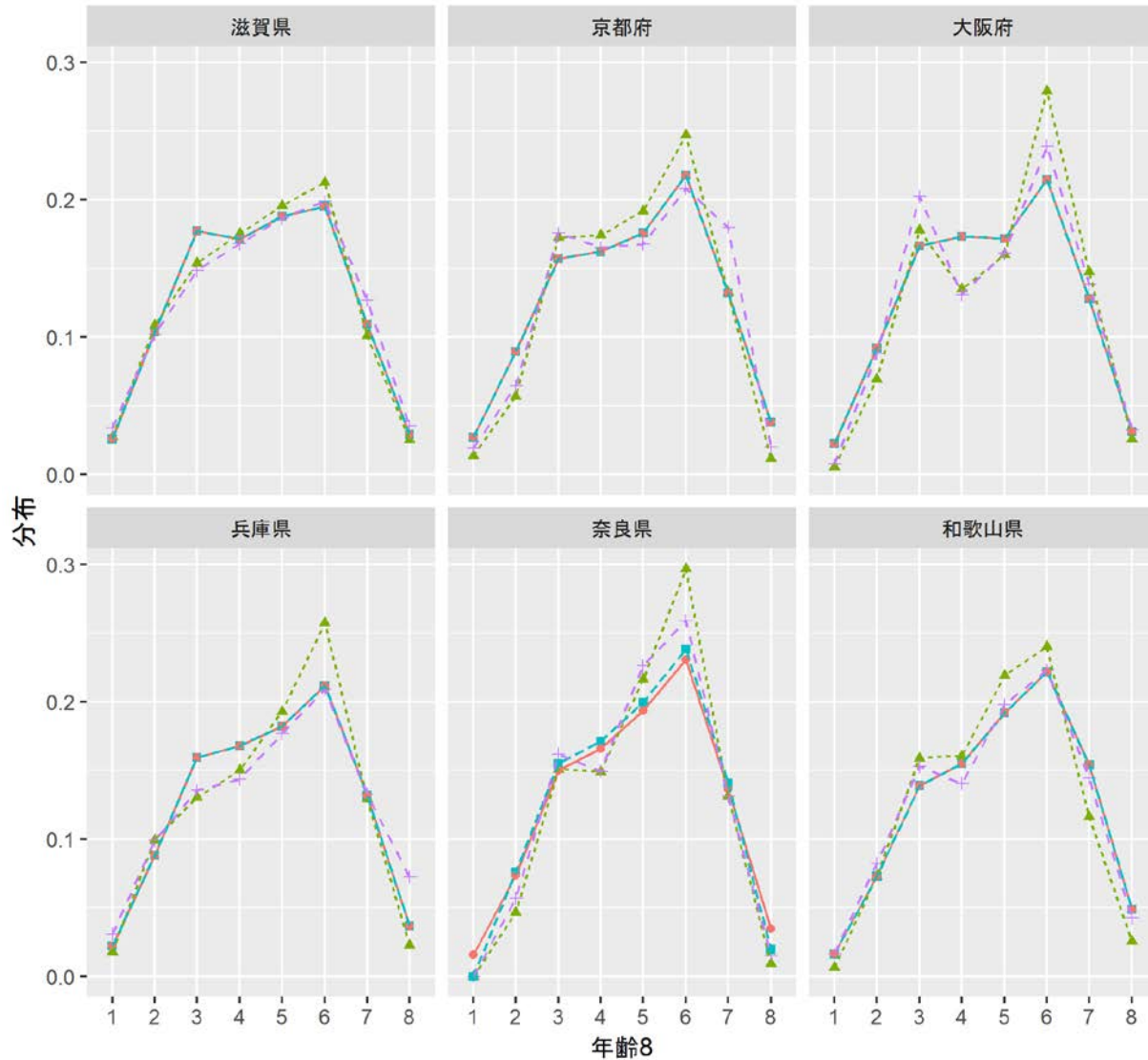
乗率 — 国調世帯数 — 実査乗率 — 修正乗率 県別 — 修正乗率 全国



都道府県別年齢階級分布
(北陸・東海地方)

ウエイト推定結果：都道府県別④

乗率 — 国調世帯数 — 実査乗率 — 修正乗率 県別 — 修正乗率 全国

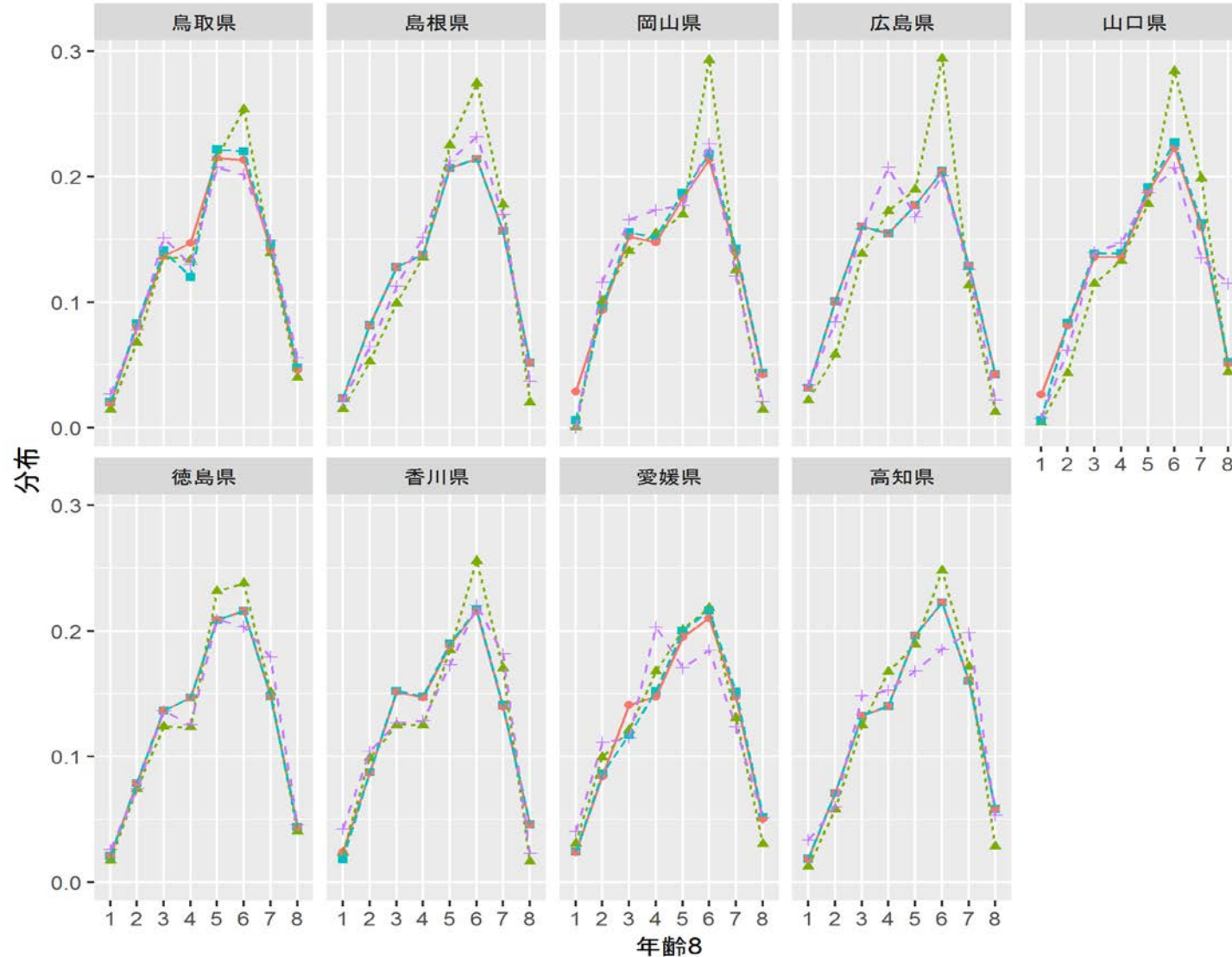


都道府県別年齢階級分布
(近畿地方)

ウエイト推定結果：都道府県別⑤

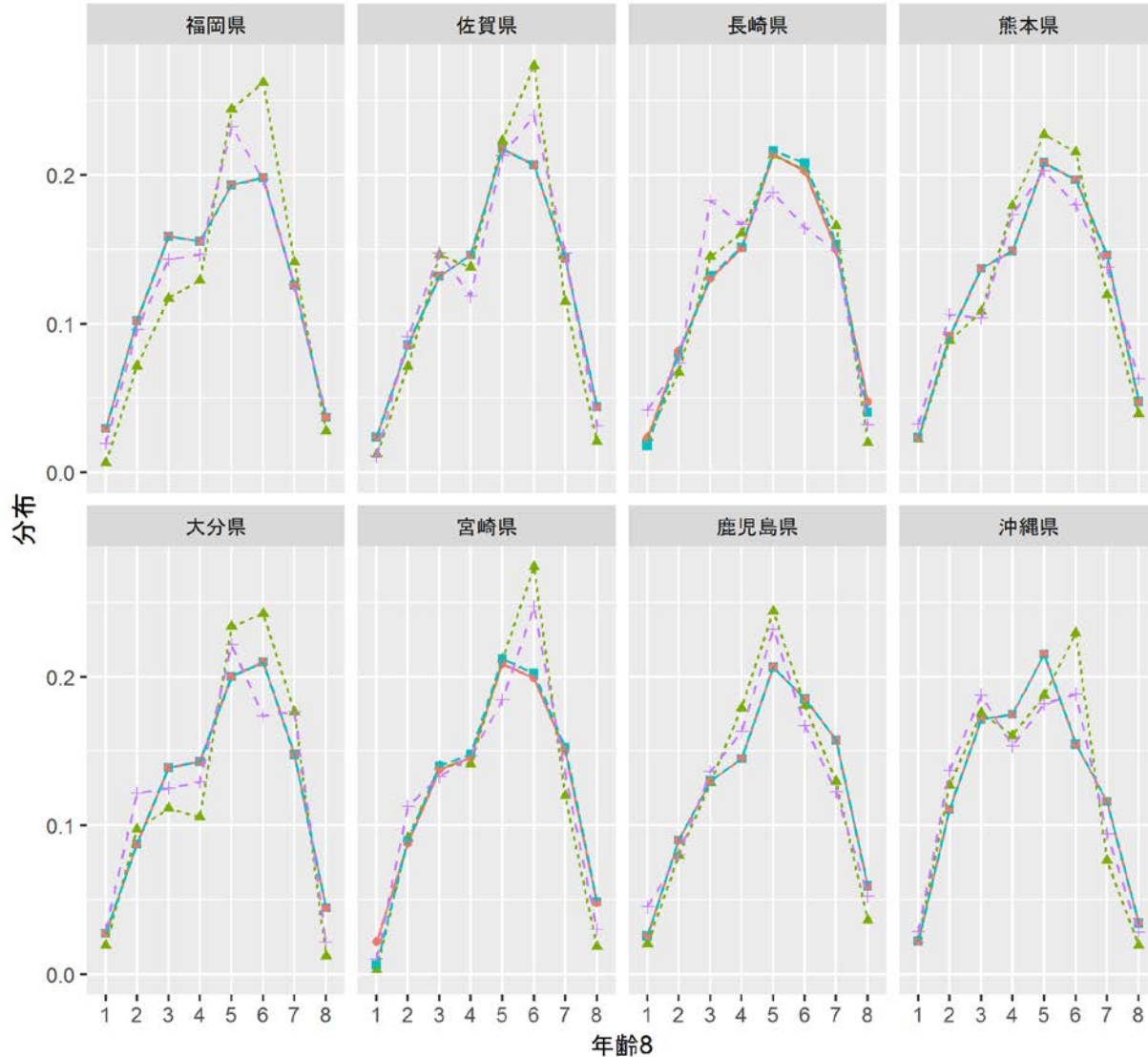
乗率 — 国調世帯数 — 実査乗率 — 修正乗率. 県別 — 修正乗率. 全国

都道府県別
年齢階級分布
(中国・四国地方)



ウエイト推定結果：都道府県別⑥

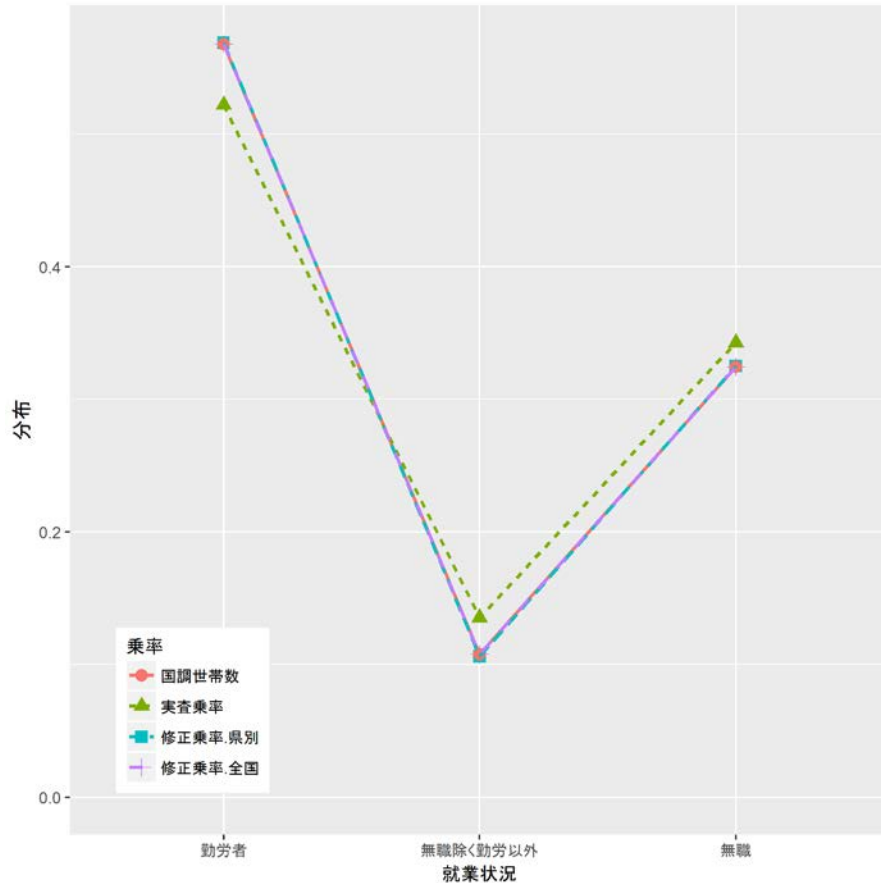
乗率 — 国調世帯数 — 実査乗率 — 修正乗率 県別 — 修正乗率 全国



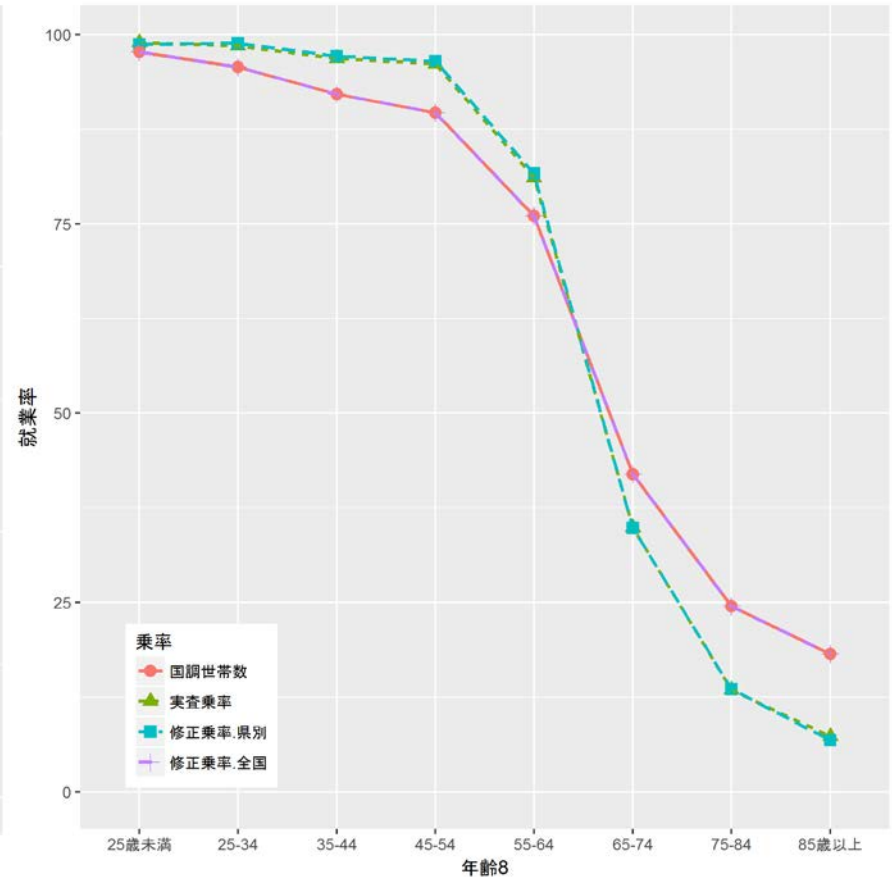
都道府県別年齢階級分布
(九州・沖縄地方)

ウエイト推定結果：全国・世帯主の就業状況関連

就業状況別世帯分布



年齢階級別就業率



就業状況別の世帯分布（左図）についても、国勢調査をよく再現できている。しかし、就業率（右図）については、県別での推定値は全く調整がされていない。これは層の定義に年齢と就業状況の同時分布を用いることで解消される。

結果のまとめ

- 全国家計構造調査におけるウエイトの推定方法について、前回までの研究会の議論を踏まえ検討した。
- 都道府県別のデータを使用し、世帯主の年齢階級など主要な世帯属性別の分布を国勢調査に合わせることに成功した。
- 一方で、検討に使用した全国消費実態調査における調査世帯数の制限により、全国のデータを使用した場合に比べて、複数の属性の同時分布を合わせることが難しくなった。
- 実際の集計にあたり、分布を調整する属性の選択について、今後さらに検討し、実装する。

課題①：収束しない問題の回避方法

サンプルサイズの小さな全国消費実態調査のデータについて、IPF法による細かな調整を可能とするために、今回は全消の調査世帯が存在しない属性の組合せについても、ごく小さなウエイトを割り当てた（ただし、周辺分布である国勢調査には必ず存在している）。この方法は簡便である一方、推定結果にバイアスを生む可能性が指摘されている（Guo, Bhat 2006）。

この問題への対処方法としては以下のものが挙げられる。

- ① 割り当てるウエイトを、任意の小さな値ではなく、地方全体や全国の分布から求められる値とする方法（Ye et.al., 2009）
- ② 全消の調査世帯が存在しない属性の組合せがなくなるように、属性階級をまとめる方法

ただし、②の手法の場合、属性階級の定義が粗くなるため、満足な調整ができない可能性がある。実際、2014全消のデータにおいては、若年の単身世帯が少ないことから、都道府県別に調整する場合、年齢階級を従来の3区分（～34、35～59、60～）としても、性別と組み合わせた場合には空白となるセルが存在する。

課題②：周辺分布として参照するデータ

今回の分析では2015年国勢調査のデータのうち、年齢を1歳若くした分布を用いたが、全国家計構造調査での使用には、以下のような問題点がある。

- ① 2019年全国家計構造調査の公表までに2020年国勢調査の結果は公表されないため、国勢調査については2015年以前のデータしか利用できない。

国勢調査データにおける時点調整の方法については、前回の消費統計研究会での議論を踏まえ、検討中である。

- ② 国勢調査と全国家計構造調査（全消を含む）における世帯主の定義の違いが、世帯主の属性を利用した分布補正の結果に影響する可能性がある。

全国消費実態調査における世帯主は、その世帯の家計上の主たる収入を得ている人、と定義されているが、国勢調査には明確な定義が存在しない。そのため、全消の分布を国勢調査に合わせることは、過剰な補正となる可能性がある。

ただし、世帯主に依存しない世帯属性（住居の所有関係、18歳未満や65歳以上の世帯員の人数など）を利用することによって、回避できる可能性がある。

参考文献

Choupani, A.A., Mamdoohi, A.R. (2016), "Population synthesis using iterative proportional fitting (IPF): A review and future research", *Transportation Research Procedia*, 17, 223-233

Guo, J.Y., Bhat, C.R. (2007), "Population Synthesis for Microsimulating Travel Behavior", *Transportation Research Record: Journal of the Transportation Research Board*, 2014, 92-101

Norman, P. (1999) "Putting Iterative Proportional Fitting on the Researcher's Desk", *School of Geography Working Paper*, 99/03

Ye, X., Konduri, K., Pendyala, R.M., Sana, B., Waddell, P. (2009) "A methodology to match distributions of both household and person attributes in the generation of synthetic populations", 88th Annual Meeting of the Transportation Research Board, Washington, D.C.