

労働力調査における二重比推定値の精度評価—国勢調査マイクロデータを用いた実証研究

小林 良行*

Accuracy evaluation of a double ratio estimate in the Labor Force Survey: an empirical study using microdata of the Population Census

KOBAYASHI Yoshiyuki

労働力調査は、標本抽出方法として層化二段抽出を用い、推定方法として比推定を用いている。また、推定値の精度評価には副標本法で算出される標準誤差を用いている。完全失業率は2つの比推定値の比で表される二重比推定値である。本研究では、国勢調査の個別データから労働力調査と近似した方法により1000組の独立な層化二段抽出標本を作成し、各標本の完全失業率の推定値の精度を(1)理論式法、(2)副標本法、(3)ブートストラップ法で評価した。評価結果から、推定値の精度の分布では、理論式法は平均と散布度が3方法中最も小さく、副標本法の散布度はブートストラップ法の散布度の2~3倍の範囲の大きさであることがわかった。実務への適用容易性を考えると、毎月の精度計算は簡便な副標本法を用い、年平均結果を算出する際に精度の推定が安定的なブートストラップ法で再計算した精度を統計利用者に提供するのが実用的で推奨される。

キーワード：公的統計、完全失業率、マイクロデータ

The labor force survey adopts the stratified two-stage sampling and the ratio estimation methods. The accuracy of the estimates is evaluated by using subsamples to calculate standard errors for estimates. The unemployment rate of the labor force survey is measured as a double ratio estimate (defined as a ratio of 2 ratio estimates). The author extracted 1000 mutually independent samples from the population census microdata, using the stratified two-stage sampling method, similar to that adopted for the labor force survey. The author also assessed the accuracy of the estimated unemployment rate of each sample by using 3 different evaluation methods, i.e., (1) the method based on the sampling theory, (2) the subsample method, and (3) the bootstrap method. The calculation results indicate that the dispersion by the subsample method is 2-3 times that by the bootstrap method on the distribution of accuracy of estimates. Although the subsample method is suitable for estimating the accuracy of the monthly report, considering its application ease, it is recommended that official statisticians employ the bootstrap method once a year, at the time of the annual report dissemination, for added accuracy.

Keywords: official statistics, unemployment rate, microdata

* 元総務省統計研究研修所 e-mail: xiaolin.aura@gmail.com

1. 研究目的

公的統計の標本調査の中には、たとえば労働力調査、社会生活基本調査、家計調査などのように理論面と実務面の双方を勘案して複雑な標本設計となっている調査がある。このような標本調査では、推定値の実績精度を求めるのに様々な工夫がなされている。

中でも統計調査としての歴史が長い労働力調査(以下、「労調」)では、標本交代を考慮して互いに独立に抽出した8組の副標本を用いて推定値の誤差計算を行っている。しかし、この方法は情報処理技術が今日ほど発達していない時代に考えられたものであり、今日の計算機能力からするとより良い方法があり得るのではないかと、あるとすればその方法を提案して公的統計の改善に資することができるのではないかとというのが、本研究の前提となる問題意識である。本研究では、労調の完全失業率や平均週間就業時間のような二重比推定値¹に着目し、推定値の精度評価方法について実際の適用容易性など実務面も踏まえた検討を行うことを目的としている。

労調の推定値の精度評価方法としては、

- (1) 標本設計の背景となる標本理論に基づく方法(以下、「理論式法」)
- (2) 副標本法
- (3) ブートストラップ法

の3つが考えられる。石井(1988)は、層化二段抽出法による線形推定値の分散の推定値から比推定値及び二重比推定値の分散の推定値を求める式を導き、労調の月次データを用いて(1)と(2)の方法により就業者数、完全失業者数、完全失業率などの推定値の精度の比較を行っている。古橋・岩永(1991)は労調の月次データを用いて(1)と(3)の方法を適用し、就業者数及び完全失業者数の推定値の精度の比較を行っている。馬場他(1997)は、労調の月次データを用いて標本調査区と標本世帯のそれぞれの再抽出で層別抽出をする場合としない場合について、副標本法とブートストラップ法の精度の比較を行い、精度推定におけるブートストラップ法の有効性と実用可能性を論じている。また、椿・會田(2020)は労調の月次データを用いて、ブートストラップ法での標本調査区再抽出時に単純な無作為抽出と調査区ウェイトを考慮した確率比例抽出の2つの方法により求めた産業別就業者数の推定値の精度を比較し、実務的な観点からも単純な無作為抽出を精度評価に用いることが可能であるとしている。これらはいずれも労調の個別データを用いた実証研究である。

一方、これらの実証研究に対して、本研究の特徴は、国勢調査(以下、「国調」)の個別データを用いて、労調の標本抽出方法に従って層化二段抽出標本を複数作成し、完全失業率等の推定値について精度評価方法の違いによる精度の比較評価を行っている点である²。本論文では、完全失業率の評価結果を中心に、主として記述統計的な分析結果と考察を述べている。

以下、第2節では労調の標本設計の概略について述べ、次いで第3節では推定精度評価の実験に使用したデータの調製方法について述べる。第4節では精度評価に用いた3つの方法を説明する。第5節では3つの精度評価方法の比較結果を示し、第6節で結果の考察とまとめを示す。

2. 労働力調査の標本設計

2.1 標本抽出方法

労調の母集団は、国勢調査の調査区(以下、「国調調査区」)のうち駐留軍区域を除く全地域に常住する人である。ただし、駐留軍区域外であっても外国の外交官、軍人・軍属及びその家族は対象外である。

労調の標本抽出方法は、すべての国調調査区を国調の産業別結果等に基づく特性により分類する

¹ 労調の集計結果は比推定値であり、完全失業率は完全失業者数を労働力人口で除して得られる。完全失業率のような2つの比推定値の比は二重比推定値(the double ratio estimate)と呼ばれている(Cochran(1977):183)。

² 本研究は筆者が在職中に独立行政法人統計センターが運営するオンライン施設を利用して行ったもので、本論文は研究成果の一部をまとめたものである。本研究では、統計法の規定に基づき、国勢調査に係る調査票情報を使用した。

層化基準(表1及び表2参照)を用いた層化二段抽出法である。第1次抽出単位は、国調調査区または特定の特性³を持つ国調調査区を分割して設けた標本単位区(以下、両者を併せて「調査区」と総称。標本単位区については総務省統計局(2007:107, 109)を参照)である。調査区の抽出は、全国を11地域に分け、地域と調査区の層を組合せた抽出層⁴別に、調査区ウェイトの大きさに基づいて行う確率比例抽出である(地域、層別標本調査区配分については表3を参照)。調査区ウェイトは、世帯数を人口規模に比例するように変換した換算世帯数をもとに、換算世帯数/15(小数点以下を切り上げ)で算出される(換算世帯数の算式は総務省統計局(2007:109)を参照)。ただし、換算世帯数が15以下の調査区(表1の01及び02層の調査区)のウェイトは1としている。また、第2次抽出単位は調査区内の住戸であり、標本住戸は調査区ウェイトの逆数を用いて抽出する。なお、実務上は、調査区の抽出と住戸の抽出には系統抽出法を用いている(抽出手順の詳細は総務省統計局(2008)を参照)。毎月の標本調査区数は2,912、調査世帯は約4万世帯で世帯内の15歳以上人員約10万人が調査対象となっている

なお、標本交代を行うため及び推定値の標本誤差の算出に用いるため、標本調査区は8組の独立した副標本として抽出している。

2.2 標本の交代方式

月次結果や年平均結果の精度、月々及び年間の変化を見る場合の精度及び前年結果との比較の精度を確保するため、一つの標本調査区は、4か月間継続して調査した後他の標本調査区に交代し、翌年の同時期に再び4か月継続して調査を行う。また、標本調査区内では、標本住戸(調査対象は当該住戸に居住する世帯)は、標本調査区の調査期間の前半2か月と後半2か月とで交代する。標本住戸は調査区が再度調査対象となる翌年の同期に再び2か月間継続して調査を行う。

2.3 推定方法

毎月の全国結果は、男女、年齢階級⁶、地域⁷別目的項目人口の線形推定値とベンチマーク人口(毎月1日現在の推計人口)を用いて算出した比推定値を組合せることにより求めている。

2.4 精度評価方法

毎月の全国結果の標本誤差は、全標本と8組の副標本から推定した全国結果を用いて、下記の式で求めている。全標本による推定値を \bar{X}_0 とし、第 t 副標本($t = 1, 2, \dots, 8$)による推定値を \bar{X}_t とすると、 \bar{X}_0 の標本誤差は、

$$\sqrt{\frac{1}{8(8-1)} \sum_{t=1}^8 (\bar{X}_t - \bar{X}_0)^2} \quad (1)$$

で求められる。労調の月次結果及び年平均結果の標本誤差は、調査項目ごとの標準誤差率を曲線の当てはめにより平均的に評価し、得られた平均的な標準誤差率に推定値を乗じて標準誤差を算出したものが『労働力調査年報』に掲載されている。

3. 使用したデータとその調製

3.1 母集団データの調製方法

本研究で利用したデータは、平成17年国調個別データ及び平成17年国調標本調査基礎資料デー

³ 50人以上の世帯からなる学生の寮・寄宿舎等。表1の層大分類符号04の国調調査区が該当する。

⁴ 標本抽出時は北海道、東北、南関東、北関東・甲信、北陸、東海、近畿、中国、四国、九州及び沖縄の11区分、四半期ごとの地域別結果、年平均結果などの公表時は沖縄を九州に含めた10区分としている。

⁵ 実際の抽出では、地域内で調査区ウェイト計が著しく小さい層は特性が類似している層と合併して標本調査区抽出用の抽出層としている。ただし、層符号が02、03、0401、0402、0403、0404の各層は全く異質の特性を持つ層なので、どの地域でも層の合併は行わない。

⁶ 年齢階級は2007年から15区分(15~19歳から80~84歳までの5歳階級及び85歳以上)別の推計に変更されている。

⁷ 平成17年国調を母集団とする標本設計(総務省統計局(2008))では地域は大都市部・非大都市部の2区分であったが、2012年から標本抽出と同じ11区分に変更されている。

タである。労調の標本設計は国調の産業別集計結果に基づいて行うため、タイムラグはほぼ2年である。本研究では、そのタイムラグの間に東日本大震災のような社会の様相に大きな変化が生じ、その影響が長期に渡るような事態が生じていない時期に実施された平成17年国調の個別データを用いた。

標本調査基礎資料は、調査区を抽出単位として国、地方公共団体、民間企業などが行う世帯や個人に関する各種標本調査の設計に資するため、国調の集計結果に基づき作成、提供されているものである。標本調査基礎資料には、地域(都道府県、市区町村)、国調調査区番号、層符号、人口、世帯数、換算世帯数、調査区ウェイト、標本単位区数等の情報が含まれている。本研究では平成17年国調の標本調査基礎資料データ(総務省統計局(2007)参照)を用いた。

国調データを用いてできるだけ労調の標本抽出に近い形で標本抽出を行うため、次のデータ調製を行った。

- (1) 聞取り世帯を除外及び調査区内の無人分割区を除外する。
- (2) 世帯概念を国調の一般世帯・施設等の世帯から労調の普通世帯・準世帯に組み替える(世帯概念の対応については総務省統計局(2007:107)を参照)。
- (3) 特別な層(04層)の標本抽出は労調と同様に標本単位区で行うため、04層の調査区内個別データを標本単位区へ分割する。
- (4) 標本調査基礎資料データから層符号などを付与する。
- (5) 労調と同様に01層(刑務所・拘留所がある区域、自衛隊区域及び駐留軍区域の調査区並びに水面調査区)のデータを除外する。

上記の調整済みデータ(31755調査区、1680537世帯)が標本抽出用の母集団データとなる。

3.2 標本データの調製

今回の研究では、まず労調の集計結果を公表する地域の中で標本調査区配分が最小(152調査区)である四国ブロックを分析対象として実験を行うこととした。実験に用いる標本データは、労調と同様の抽出方法で作成した互いに独立な1000組の標本から構成されている。

- (1) 標本調査区は抽出層別に調査区ウェイトの大きさに基づき確率比例抽出する。その際、標本調査区が8組の独立な副標本となるよう抽出する。
- (2) 世帯を第2次抽出単位とし、標本世帯は調査区ウェイトの逆数を用いて抽出する。
- (3) 上記の方法で得られた国調個別データを1組の標本とする。

4. 精度評価方法

母集団の変数(大きさ、特性値など)を英大文字で、標本の変数を英小文字で表す。推定値は母特性値(総計、分散など)を表す変数に $\hat{}$ を付ける。また、線形推定値と比推定値を区別して表記する場合、後者には $\tilde{}$ を付けることにする。

4.1 理論式法

石井(1988)⁸及びCochran(1977)⁹によれば、比推定値の分散・共分散は線形推定値の分散・共分散

⁸ 石井(1988)の方法を用いるに当たっては原著の数式の変数名、添字等の表記を本稿のものに合わせるよう修正するとともに原著の表記誤り等を訂正した上で適用している。

⁹ Cochran(1977:184)では、2つの母比率 $R = Y/X$ と $R' = Y'/X'$ の推定値を $\hat{R} = \bar{y}/\bar{x}$ と $\hat{R}' = \bar{y}'/\bar{x}'$ とし、二重比推定値 \hat{R}/\hat{R}' の分散を、相対分散 $C_{\hat{R}\hat{R}'} = V(\hat{R})/R^2$ などを使って表す式を示している。なお、Cochran(1977:183)では、 $C_{\hat{R}\hat{R}'}$ を与える式を $C_{RR} = C_{\bar{y}\bar{y}} + C_{\bar{x}\bar{x}} + 2C_{\bar{y}\bar{x}}$ と記載しているが、この式のもとになる式(Cochran(1977:154))の第3項の符号は+でなく-になっているので、上式の第3項の符号は誤植であろう。

をもとに求められ¹⁰、二重比推定値の分散は比推定値の分散・共分散から求められる。また、これらの統計量の推定値も同様の手順で求めることができる。以下、Step1 及び Step2 で完全失業者数及び労働力人口の比推定値、比推定値の標準誤差及び標準誤差の推定値、標準誤差率及び標準誤差率の推定値を求める。また、Step3 で完全失業率の二重比推定値、二重比推定値の標準誤差及び標準誤差の推定値、標準誤差率及び標準誤差率の推定値を求める。算出の手順は以下の通りである。(Step1) 層化二段抽出法による線形推定値の分散・共分散およびその不偏推定値を求める。このステップでは比推定値の分散・共分散及びそれらの推定値を求めるのに必要となる統計量を求める。母集団におけるある目的属性を持つ人口(以下、「目的属性人口」)を X で表すと、

$$\begin{aligned} X &= \sum_{h=1}^H X_h \\ &= \sum_{h=1}^H \sum_{i=1}^{M_h} \sum_{j=1}^{N_{hi}} X_{hij} \end{aligned} \quad (2)$$

である。ここで、 $h(=1, 2, \dots, H)$ は地域と国調調査区の層符号を組合せた標本調査区抽出層の番号を表す。 $i(=1, 2, \dots, M_h)$ は調査区の番号、 M_h は第 h 層の母集団調査区数を表す。また、 $j(=1, 2, \dots, N_{hi})$ は世帯の番号、 N_{hi} は第 h 層、第 i 調査区の母集団世帯数を表す。(2) 式の X_h は第 h 層の目的属性人口を表し、 X_{hij} は第 h 層、第 i 調査区、第 j 世帯内の目的属性人口を表す。

世帯員の男女の別を a で、年齢階級を b で表すと、世帯の目的属性人口は、下式のように

$$X_{hij} = \sum_{a=1}^2 \sum_{b=1}^{15} X_{hijab} \quad (3)$$

と世帯員の男女、年齢階級別の目的属性人口の和で表せる。式(2)、(3)より、 X は

$$\begin{aligned} X &= \sum_{h=1}^H \sum_{i=1}^{M_h} \sum_{j=1}^{N_{hi}} \sum_{a=1}^2 \sum_{b=1}^{15} X_{hijab} \\ &= \sum_{h=1}^H \sum_{a=1}^2 \sum_{b=1}^{15} X_{hab} \\ &= \sum_{a=1}^2 \sum_{b=1}^{15} X_{ab} \end{aligned} \quad (4)$$

と表すことができる。

X の線形推定値 \hat{X} は、第一段抽出が層ごとに独立して行われていることから、層別の線形推定値 \hat{X}_h の和で表される。第 h 層、第 i 標本調査区、第 j 標本世帯内のある属性を有する調査人口を x_{hij} とすると、

$$\begin{aligned} \hat{X} &= \sum_{h=1}^H \hat{X}_h \\ &= \sum_{h=1}^H \frac{1}{m_h} \sum_{i=1}^{m_h} \frac{W_h}{w_{hi}} \frac{N_{hi}}{n_{hi}} \sum_{j=1}^{n_{hi}} x_{hij} \\ &= \sum_{h=1}^H \frac{W_h}{m_h} \sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}} x_{hij} \quad (\text{標本世帯の抽出方法から } n_{hi} = N_{hi}/w_{hi}) \end{aligned} \quad (5)$$

である。ここで、 m_h は第 h 層の標本調査区数を表し、 n_{hi} は第 h 層、第 i 標本調査区の標本世帯数を表す。また、 \hat{X}_h は第 h 層の線形推定値を表し、 W_h は第 h 層の母集団調査区のウェイト計を表し、 w_{hi} は第 h 層、第 i 標本調査区の調査区ウェイトを表す。

x_{hij} についても式(3)と同様の関係

$$x_{hij} = \sum_{a=1}^2 \sum_{b=1}^{15} x_{hijab} \quad (6)$$

が成り立つので、式(5)、(6)より、 \hat{X} は

¹⁰ 津村・築林(1986:199-204, 210-213)は、層化二段抽出法の線形推定値の分散・共分散及びそれらの推定値の算出式を示しているが、補助量とそれを整数化したウェイトを用い、有限母集団修正項を含めたかなり複雑な式になっている。理論的に厳密ではあるものの実務に適用するのは難しいと考えて、本研究では採用しなかった。

$$\begin{aligned}\hat{X} &= \sum_{h=1}^H \frac{W_h}{m_h} \sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}} \sum_{a=1}^2 \sum_{b=1}^{15} x_{hijab} \\ &= \sum_{h=1}^H \sum_{a=1}^2 \sum_{b=1}^{15} \hat{X}_{hab} \\ &= \sum_{a=1}^2 \sum_{b=1}^{15} \hat{X}_{ab}\end{aligned}\quad (7)$$

と表すことができる。

(Step1-1) 線形推定値の分散・共分散の算出

以下、 $\sum_{h=1}^H$ を \sum_h と、 $\sum_{i=1}^{m_h}$ を \sum_i と、 $\sum_{j=1}^{n_{hi}}$ を \sum_j と、 $\sum_{a=1}^2 \sum_{b=1}^{15}$ を $\sum_{a,b}$ と略記する。また、15歳

以上人口を \hat{P} で表す。式(5)、(7)及び石井(1988)から、 $V(\hat{X}_{ab})$ 及び $COV(\hat{X}_{ab}, \hat{P}_{ab})$ は、

$$V(\hat{X}_{ab}) = V(\sum_h \hat{X}_{hab}) = \sum_h V(\hat{X}_{hab}) \quad (8)$$

$$COV(\hat{X}_{ab}, \hat{P}_{ab}) = COV(\sum_h \hat{X}_{hab}, \sum_h \hat{P}_{hab}) = \sum_h COV(\hat{X}_{hab}, \hat{P}_{hab}) \quad (9)$$

$$V(\hat{X}_{hab}) = \frac{W_h}{m_h} (W_h \cdot \sigma_{Bhabx}^2 + \sum_i N_{hi} \cdot \frac{N_{hi} - n_{hi}}{N_{hi} - 1} \cdot \sigma_{Whiabx}^2) \quad (10)$$

$$COV(\hat{X}_{hab}, \hat{P}_{hab}) = \frac{W_h}{m_h} (W_h \cdot \sigma_{Bhabxp} + \sum_i N_{hi} \cdot \frac{N_{hi} - n_{hi}}{N_{hi} - 1} \cdot \sigma_{Whiabxp}) \quad (11)$$

で求められる。

ここで、

$$\sigma_{Bhabx}^2 = \frac{1}{W_h} \sum_i W_{hi} \left(\frac{X_{hiab}}{W_{hi}} - \bar{X}_{hab} \right)^2 \quad (12)$$

$$\bar{X}_{hab} = \frac{1}{W_h} \sum_i X_{hiab} \quad (13)$$

$$\sigma_{Whiabx}^2 = \frac{1}{N_{hi}} \sum_j (X_{hijab} - \bar{X}_{hiab})^2 \quad (14)$$

$$\bar{X}_{hiab} = \frac{1}{N_{hi}} \sum_j X_{hijab} \quad (15)$$

$$\sigma_{Bhabxp} = \frac{1}{W_h} \sum_h \left(\frac{X_{hiab}}{W_{hi}} - \bar{X}_{hab} \right) \left(\frac{P_{hiab}}{W_{hi}} - \bar{P}_{hab} \right) \quad (16)$$

$$\sigma_{Whiabxp} = \frac{1}{N_{hi}} \sum_j (X_{hijab} - \bar{X}_{hiab}) (P_{hijab} - \bar{P}_{hiab}) \quad (17)$$

である。式(10)及び(11)を同一地域の層について合計し、比推定値の分散・共分散算出に用いる地域別の $V(\hat{X}_{ab})$ 及び $COV(\hat{X}_{ab}, \hat{P}_{ab})$ を得る。

(Step1-2) 分散・共分散の推定値の算出

以下、 $\sum_{h=1}^H$ を \sum_h と、 $\sum_{i=1}^{m_h}$ を \sum_i と、 $\sum_{j=1}^{n_{hi}}$ を \sum_j と略記することにし、 \hat{X} の分散の不偏推定量を $\hat{V}(\hat{X})$ で表し、 \hat{X} と \hat{P} の共分散の不偏推定量を $\widehat{COV}(\hat{X}, \hat{P})$ で表すことにする。式(5)、(7)及び石井(1988)から、 $\hat{V}(\hat{X}_{ab})$ 及び $\widehat{COV}(\hat{X}_{ab}, \hat{P}_{ab})$ は、

$$\hat{V}(\hat{X}_{ab}) = \hat{V}(\sum_h \hat{X}_{hab}) = \sum_h \hat{V}(\hat{X}_{hab}) \quad (18)$$

$$\widehat{COV}(\hat{X}_{ab}, \hat{P}_{ab}) = \widehat{COV}(\sum_h \hat{X}_{hab}, \sum_h \hat{P}_{hab}) = \sum_h \widehat{COV}(\hat{X}_{hab}, \hat{P}_{hab}) \quad (19)$$

$$\hat{V}(\hat{X}_{hab}) = \frac{S_{Bhabx}^2}{m_h} \quad (20)$$

$$\widehat{COV}(\hat{X}_{hab}, \hat{P}_{hab}) = \frac{S_{Bhabxp}}{m_h} \quad (21)$$

で求められる。

ここで、

$$\begin{aligned}
S_{Bhabx}^2 &= \frac{1}{m_h-1} \sum_i \left(\frac{W_h}{W_{hi}} \cdot \frac{N_{hi}}{n_{hi}} \sum_j x_{hijab} - \hat{X}_{hab} \right)^2 \\
&= \frac{1}{m_h-1} \sum_i (W_h \sum_j x_{hijab} - \hat{X}_{hab})^2
\end{aligned} \tag{22}$$

$$S_{Bhabxp} = \frac{1}{m_h-1} \sum_i (W_h \sum_j x_{hijab} - \hat{X}_{hab})(W_h \sum_j p_{hijab} - \hat{P}_{hab}) \tag{23}$$

である。式(20)及び(21)を同一地域の層について合計し、比推定値の分散・共分散の推定値算出に用いる地域別の $\hat{V}(\hat{X}_{ab})$ 及び $\widehat{COV}(\hat{X}_{ab}, \hat{P}_{ab})$ を得る。線形推定値の分散・共分散の推定値は、1000組の標本について算出する。

(Step2) 比推定値の分散・共分散及びそれらの推定値を求める。

労調では、比推定値 \tilde{X} は男女、年齢階級別の比推定値 \tilde{X}_{ab} の和で求めているので、

$$\tilde{X} = \sum_{a,b} \tilde{X}_{ab} \tag{24}$$

である。

男女、年齢階級別の比推定値 \tilde{X}_{ab} は、人口 P を X に対する補助量として、

$$\tilde{X}_{ab} = \hat{X}_{ab} \cdot \frac{B_{ab}}{\hat{P}_{ab}} = B_{ab} \cdot \frac{\hat{X}_{ab}}{\hat{P}_{ab}} \tag{25}$$

で求められる。

ここで、 \hat{X}_{ab} は男女、年齢階級別目的属性人口の線形推定値、 \hat{P}_{ab} は男女、年齢階級別15歳以上人口の線形推定値、 B_{ab} は男女、年齢階級別ベンチマーク人口である。なお、本研究では補助量 P は就業状態不詳を含まない15歳以上人口とし、ベンチマーク人口は就業状態不詳を含む15歳以上人口を用いた¹¹。

(Step2-1) 比推定値の分散・共分散の算出

式(24)、(25)、石井(1988)及びCochran(1977)から、比推定値の分散及び2つの比推定値の共分散は、(Step1)で求めた男女、年齢階級別の線形推定値の分散・共分散、15歳以上人口の線形推定値の期待値等を用いて、

$$\begin{aligned}
V(\tilde{X}) &= \sum_{a,b} B_{ab}^2 \cdot V\left(\frac{\hat{X}_{ab}}{\hat{P}_{ab}}\right) \\
&\cong \sum_{a,b} \{B_{ab}/E(\hat{P}_{ab})\}^2 \cdot \{V(\hat{X}_{ab}) + R_{1ab}^2 \cdot V(\hat{P}_{ab}) - 2R_{1ab} \cdot COV(\hat{X}_{ab}, \hat{P}_{ab})\}
\end{aligned} \tag{26}$$

$$\begin{aligned}
COV(\tilde{X}, \tilde{Y}) &= \sum_{a,b} B_{ab}^2 \cdot COV\left(\frac{\hat{X}_{ab}}{\hat{P}_{ab}}, \frac{\hat{Y}_{ab}}{\hat{P}_{ab}}\right) \\
&\cong \sum_{a,b} \{B_{ab}/E(\hat{P}_{ab})\}^2 \cdot \{COV(\hat{X}_{ab}, \hat{Y}_{ab}) + R_{1ab} \cdot R_{2ab} \cdot V(\hat{P}_{ab}) \\
&\quad - R_{2ab} \cdot COV(\hat{X}_{ab}, \hat{P}_{ab}) - R_{1ab} \cdot COV(\hat{Y}_{ab}, \hat{P}_{ab})\}
\end{aligned} \tag{27}$$

で求められる。

ここで、 $E(\hat{P}_{ab}) = P_{ab}$ 、 $R_{1ab} = E(\hat{X}_{ab})/E(\hat{P}_{ab}) = X_{ab}/P_{ab}$ 、 $R_{2ab} = E(\hat{Y}_{ab})/E(\hat{P}_{ab}) = Y_{ab}/P_{ab}$ とする。

$V(\tilde{X})$ から比推定値の標準誤差 σ を求め、 σ と X から \tilde{X} の標準誤差率を求める。

(Step2-2) 分散・共分散の推定値の算出

比推定値の分散・共分散の推定値 $\hat{V}(\tilde{X})$ 、 $\widehat{COV}(\tilde{X}, \tilde{Y})$ 等は、式(26)及び(27)の $V(\hat{X}_{ab})$ 、 $COV(\hat{X}_{ab}, \hat{P}_{ab})$ 等の代わりに(Step1-2)で求めた $\hat{V}(\hat{X}_{ab})$ 、 $\widehat{COV}(\hat{X}_{ab}, \hat{P}_{ab})$ 等を用いて求める。また、 $E(\hat{P}_{ab}) \cong \hat{P}_{ab}$ 、 $\hat{R}_{1ab} \cong \hat{X}_{ab}/\hat{P}_{ab}$ 、 $\hat{R}_{2ab} \cong \hat{Y}_{ab}/\hat{P}_{ab}$ とする。

¹¹ 労調の比推定では、ベンチマーク人口は毎月1日現在の推計人口を用いている。推計人口では、男女、年齢階級はわかるが、就業状態はわからない。本研究のベンチマーク人口は労調と同じ考えに基づくものである。

$\hat{V}(\tilde{X})$ から比推定値の標準誤差の推定値(以下、「推定標準誤差」) $\hat{\sigma}$ を求め、 $\hat{\sigma}$ と \tilde{X} から \tilde{X} の標準誤差率の推定値(以下、「推定標準誤差率」)を求める。比推定値の分散・共分散の推定値は、1000組の標本について算出する。

(Step3) 二重比推定値の分散及び二重比推定値の分散の推定値を求める。

\tilde{X} と \tilde{Y} を X と Y の比推定値とすると、母集団の比率 $R = X/Y$ の二重比推定値を

$$\tilde{R} = \tilde{X}/\tilde{Y} \quad (28)$$

で表す。

(Step3-1) 二重比推定値の分散の算出

二重比推定値の分散は、式(28)、石井(1988)及びCochran(1977)から、

$$\begin{aligned} V(\tilde{R}) &\equiv \{1/E(\tilde{Y})\}^2 \cdot \{V(\tilde{X}) + R^2 \cdot V(\tilde{Y}) - 2R \cdot COV(\tilde{X}, \tilde{Y})\} \\ &= \sum_{a,b} \{1/E(\tilde{Y})\}^2 \cdot \{V(\tilde{X}_{ab}) + R^2 \cdot V(\tilde{Y}_{ab}) - 2R \cdot COV(\tilde{X}_{ab}, \tilde{Y}_{ab})\} \end{aligned} \quad (29)$$

で求められる。ここで、 $E(\tilde{Y}) = \sum_{a,b} B_{ab} \cdot E(\frac{\tilde{Y}_{ab}}{\tilde{P}_{ab}}) \equiv \sum_{a,b} B_{ab} \cdot \frac{E(\tilde{Y}_{ab})}{E(\tilde{P}_{ab})} = \sum_{a,b} B_{ab} \cdot \frac{Y_{ab}}{P_{ab}}$ 、 $R \equiv E(\tilde{X})/E(\tilde{Y})$ とする。

$V(\tilde{R})$ から二重比推定値の標準誤差 σ_R を求め、 σ_R と R から \tilde{R} の標準誤差率を求める。

(Step3-2) 分散の推定値の算出

二重比推定値の分散の推定値 $\hat{V}(\tilde{R})$ は、式(29)の $V(\tilde{X}_{ab})$ 、 $COV(\tilde{X}_{ab}, \tilde{Y}_{ab})$ 等の代わりに(Step2-2)で求めた $\hat{V}(\tilde{X}_{ab})$ 、 $\widehat{COV}(\tilde{X}_{ab}, \tilde{Y}_{ab})$ 等を用いて求める。また、 $E(\tilde{Y}) \equiv \tilde{Y}$ 、 $\hat{R} \equiv \tilde{X}/\tilde{Y}$ とする。

$\hat{V}(\tilde{R})$ から二重比推定値の推定標準誤差 $\hat{\sigma}_R$ を求め、 $\hat{\sigma}_R$ と \tilde{R} から \tilde{R} の推定標準誤差率を求める。二重比推定値の分散は、1000組の標本について算出する。

4.2 副標本法

全標本による推定値を \tilde{X}_0 とし、第 t 副標本($t = 1, 2, \dots, 8$)による推定値を \tilde{X}_t とすると、 \tilde{X}_0 の推定標準誤差 $\hat{\sigma}$ は、

$$\hat{\sigma} = \sqrt{\frac{1}{8(8-1)} \sum_{t=1}^8 (\tilde{X}_t - \tilde{X}_0)^2} \quad (30)$$

で求められる。完全失業率のような二重比推定値の推定標準誤差 $\hat{\sigma}_R$ を求める場合は、上式の \tilde{X}_0 及び \tilde{X}_t を \tilde{R}_0 及び \tilde{R}_t に置き換えて算出する。 $\hat{\sigma}$ と \tilde{X}_0 から \tilde{X}_0 の推定標準誤差率を求め、 $\hat{\sigma}_R$ と \tilde{R}_0 から \tilde{R}_0 の推定標準誤差率を求める。上記の統計量を1000組の標本について算出する。

4.3 ブートストラップ法

ブートストラップ法は、標本から再抽出標本を作成し、再抽出標本から推定量を算出する操作を反復することにより得られる推定量の分布の分散から推定値の分散を推定する方法である(Efron(1979))。今日まで、標本から再抽出する際に復元単純無作為抽出する方法以外に、標本規模や標本抽出方法を考慮した修正法が提案されてきている(標本誤差推計研究会(1998))。しかし、実証研究の中には、ブートストラップ法によるバイアスを調整した結果や標本抽出方法に忠実な再抽出をした結果が、単純なブートストラップ法による結果とほとんど差異がないとの報告もある(山口(2016)、椿・會田(2019))。椿・會田(2019)は、単純な調査区リサンプリングに基づき、推定精度を評価することには実務上問題はなく、むしろ個票データに含まれていない標本設計情報を参

照する必要がない単純なりサンプリング技法の方が簡便な方法として推奨できるとしている。

古橋・岩永(1991)は、標本抽出時の層により標本を層化し、調査区の再抽出を層ごとに行っている。しかし、当時の計算機能力の制約のためか現在からみると反復回数は100回と少なく感じる。馬場他(1997)は、調査区を再抽出の単位とした場合、層別リサンプリングと単純無作為リサンプリングの差が大きいとしている。両者の標準誤差率を比較すると、層別リサンプリングの方が小さくなっており、層別の効果と思われる。本研究では、古橋・岩永(1991)及び馬場他(1997)と同じく、調査区は層ごとに復元単純無作為抽出で再抽出するが、世帯は再抽出せずに精度推定の計算に用いている。

再抽出標本を作成する際には、椿・會田(2019)と同様にして、02層(人口0の調査区)、03層(換算世帯数が15以下の調査区)及び04層の一部(0401、0402、0403及び0404層)の調査区(以下、「固定層」)は再抽出の対象とせず常に再抽出標本に含めている。ただし、椿・會田(2019)と異なる点は、本研究では国調データから標本を抽出しているため、02層に属する調査区の個別データが存在しないことである。なお、標本から再抽出する回数は、最初の100組の標本について各々1000回の再抽出を行い再抽出標本の分布が単峰でほぼ左右対称になったことが確認できたので、この基準を1000組の標本すべてに適用した。

固定層以外の抽出層については下記の手順で求めた。

(Step1) 標本の抽出層ごとに層別標本調査区数 m_h と同数の調査区を復元単純無作為抽出し再抽出標本とする。再抽出標本内の標本世帯及び固定層の標本世帯を合わせて、男女、年齢階級、抽出層別目的属性人口(完全失業者数及び労働力人口)及び15歳以上人口の線形推定値を求める。これらを \hat{X}_{hab}^* 及び \hat{P}_{hab}^* とする。

(Step2) 男女、年齢階級別ベンチマーク人口 B_{ab} 、 $\hat{X}_{ab}^*(=\sum_h \hat{X}_{hab}^*)$ 及び $\hat{P}_{ab}^*(=\sum_h \hat{P}_{hab}^*)$ を用いて男女、年齢階級別目的属性人口の比推定値 \tilde{X}_{ab}^* を求める。

(Step3) 比推定値 $\tilde{X}^* = \sum_{a,b} \tilde{X}_{ab}^*$ 及び二重比推定値 $\tilde{R}^* = \tilde{X}^*/\tilde{Y}^*$ を求める。

(Step4) Step1 から Step3 までを1000回繰り返す。

(Step5) t 回目($t = 1, \dots, 1000$)の再抽出標本から得られた比推定値及び二重比推定値を θ_t^* で表すと、標本組別のブートストラップ法による推定値(以下、「BS推定値」)及び推定標準誤差は、

$$\bar{\theta}^* = \frac{1}{1000} \sum_t \theta_t^* \quad (31)$$

$$\hat{\sigma}^* = \sqrt{\frac{1}{(1000-1)} \sum_t (\theta_t^* - \bar{\theta}^*)^2} \quad (32)$$

で求められる。 $\hat{\sigma}^*$ と $\bar{\theta}^*$ から $\bar{\theta}^*$ の推定標準誤差率を求める。

1000組の標本について、上記のStep1～Step5を行い、BS推定値、推定標準誤差及び推定標準誤差率を求める。

5. 計算結果

表4は、母集団から抽出した1000組の標本から得られた推定値の分布と各標本からブートストラップ法により求めた推定値の分布を、完全失業率、完全失業者数及び労働力人口について、基本統計量等で比較したものである。なお、本研究で行った計算と結果の整理には、SAS、R及びExcelを用いた¹²。

標本組別推定値の分布(以下、「標本組別推定値分布」)と標本組別BS推定値の分布(以下、「標本組別BS推定値分布」)の平均値と母集団値を比較すると、完全失業率ではどちらの分布の平均値も

¹² SAS University Editionのサポートが研究の途中で切れてしまったため、SASデータをRの環境に移行し、計算と分析をRで継続した。標本組別推定値等の分布の基本統計量などはExcelを用いて計算、整理している。

母集団値とほぼ同水準であり、完全失業者数と労働力人口では平均値の方が母集団値より約 2%程度上回っている。標本組別推定値分布と標本組別 BS 推定値分布を最小値、最大値、中央値、パーセント点、変動係数、範囲などの基本統計量で比較すると、完全失業率、完全失業者数及び労働力人口のいずれにおいてもほぼ同水準となっている。さらに詳細に見てみると、完全失業率と完全失業者数では標本組別 BS 推定値分布は標本組別推定値分布に比べて全体的にわずかに右にシフトしている一方で、労働力人口では標本組別 BS 推定値分布の方が全体的にわずかに左にシフトしている。本論文では、分布の散布度として変動係数、十分位範囲、四分位範囲及び範囲の 4 つを用いている。完全失業率、完全失業者数及び労働力人口のいずれについても、標本組別推定値分布と標本組別 BS 推定値分布の散布度は、ほぼ同水準となっている。また、標本組別推定値と標本組別 BS 推定値の間には、完全失業率、完全失業者数及び労働力人口のいずれにおいても相関係数が 0.996 と非常に強い相関がみられる。以上のことから、標本組別推定値分布と標本組別 BS 推定値分布はほとんど同じ形状の分布になっていると言える(完全失業率のみ図 1 に示す)。また、ブートストラップ法により分布のシフトが起こるものの、推定値の分布が必ずしも常に一定の方向にシフトするわけではないことがわかる。

表 5 は、精度評価方法(理論式法、副標本法及びブートストラップ法)別に求めた標本組別推定標準誤差率の分布(以下、「標本組別推定標準誤差率分布」)を、完全失業率、完全失業者数及び労働力人口について、基本統計量で比較したものである。精度評価方法別の標本組別推定標準誤差率分布をみると、完全失業率、完全失業者数及び労働力人口のいずれも、程度の違いはあるものの右に裾を引いた分布となっている(完全失業率のみ図 2 に示す)。

表 5 より、精度評価方法別の標本組別推定標準誤差率分布の平均値を標準誤差率(推定標準誤差率の真値)と比較すると、完全失業率の理論式法ではほぼ同水準であるものの、副標本法では約 1.13 倍、ブートストラップ法では約 1.15 倍の水準と過大推定になっている。次いで完全失業者数に関してみると、理論式法では標準誤差率の約 0.98 倍とやや過小推定になっている。一方、副標本法とブートストラップ法では完全失業率の場合と同様に過大推定になっており、副標本法で約 1.16 倍、ブートストラップ法で約 1.11 倍の水準である。また、労働力人口に関してみると、理論式法では約 0.95 倍の水準と過小推定になっている。副標本法では標準誤差率の約 3 倍と突出して過大推定になっており、ブートストラップ法では約 1.11 倍と過大推定となっている。

精度評価方法別に標本組別推定標準誤差率分布の平均値を比較すると、完全失業率では副標本法とブートストラップ法の差は約 0.09%ポイントと小さくほぼ同水準となっている。完全失業者数においては副標本法の方がブートストラップ法に比べやや大きく、その差は約 0.36%ポイントとなっている。また、労働力人口では、理論式法とブートストラップ法で分布の平均値がそれぞれ約 1%となっているのに対して、副標本法は約 3%とやや大きくなっている。完全失業率、完全失業者数及び労働力人口のいずれにおいても標本組別推定標準誤差率分布の平均値は、理論式法が最も小さくかつ最も真値に近い。

標本組別推定標準誤差率分布の散布度は精度評価方法間の差異が顕著である。散布度を完全失業率、完全失業者数及び労働力人口の各項目について精度評価方法別にみると、変動係数の大きさはいずれの項目でも理論式法が最も小さく、次いでブートストラップ法であり、副標本法が最も大きくなっている。また、十分位範囲、四分位範囲及び範囲でみても同様の傾向がみられる。完全失業者数と労働力人口の副標本法による変動係数は共に約 28%程度であり、一方ブートストラップ法による変動係数は共に約 10%程度である。精度評価方法の違いによる差異はみられるが、どちらの方法によっても完全失業者数と労働力人口の大きさによる差異はほとんど見られない。しかし、理論式法では完全失業者数と労働力人口との間で変動係数に約 1%ポイント程度の差異が見られている。

散布度を精度評価方法間で比較すると、完全失業率では変動係数、十分位範囲、四分位範囲及び範囲のいずれで見ても、副標本法はブートストラップ法の約 2~3 倍程度になっている。また、ブ

ートストラップ法と理論式法を比較すると、範囲で見た場合が約3.5倍とやや大きいのを除き、ブートストラップ法が理論式法の約2倍となっている。同様なことが完全失業者数に関しても見て取れる。労働力人口では、副標本法の散布度がブートストラップ法に比べて、変動係数以外で約8～9倍と突出して大きい。それに対して、ブートストラップ法と理論式法とを比較すると、範囲で見た場合が約4倍とやや大きいのを除き、どの散布度でも約3倍程度となっている。

6. 考察とまとめ

本論文では、国調の四国ブロックの個別データを使用し、労調の完全失業率に着目して二重比推定値の精度評価方法を比較した実験の結果を報告した。

副標本法は、標本設計が複雑だったり、集計項目が多い場合でも簡便に精度計算ができるという実務上の利点があり、実用的である。

表6などを見る限り、副標本法による労働力人口の標本組別推定標準誤差率分布の平均値が他の方法の場合に比べて大きいのは、労働力人口の標本組別推定標準誤差の分布(以下、「標本組別推定標準誤差分布」)が大きく右に裾を引いた分布となっていることに起因していると考えられる。労働力人口は就業者数と完全失業者の和である。完全失業者も右に裾を引いた分布となっているが、母集団の規模から考えて労働力人口の標本組別推定標準誤差分布の散布度が大きくなる要因は、就業者数の標本組別推定標準誤差分布の散布度の大きさにあると考えられる。副標本法では、たまたま就業者数が非常に多いいくつかの標本調査区が一つの副標本に入り、その他の副標本ではそうならなかった場合、8組の副標本間で推定値に大きく差異が生じ、推定標準誤差が大きくなることあり得る¹³。

副標本法は、標本の大きさによってあまり副標本の数を増やせない場合や各副標本の大きさに差がある場合、また組分けのしかたによって、標準誤差率の推定値のばらつきが大きく、1組の標本から得られる1回だけの標準誤差率の推定結果は不安定である¹⁴。

理論式法は、他の方法と比較して標準誤差率の推定値は安定的で真値にも近いことが期待され、今回の実験でも総じて良い結果となっている。

今回用いたデータは労調の実際の標本設計に基づき、第一段抽出は抽出層内の平均ウェイトを抽出間隔とする系統抽出、第二段抽出は調査区ウェイトを抽出間隔とする系統抽出として得られたものである。一方、今回の精度評価の算式は、標本抽出法に層化二段抽出法を仮定した推定値の誤差分散の算式を用いている。実際に得られたデータの標本抽出方法と精度評価の算式が前提としている標本抽出方法に差異がある場合の影響の検証については今後の検討課題である。

理論式法では、比推定値や二重比推定値の精度計算のために、完全失業者などの項目ごとに順を追っていくつかの必要な統計量を作成していく必要があり、他の方法に比べて計算に必要なアプリケーションシステムの設計が複雑になる。今回の実験では精度評価方法別の計算時間の計測ができなかった¹⁵、理論式法が他の2つの方法に比べて計算時間がかかるか否かは判断できなかった。

¹³ 標本調査基礎資料データを見ると、四国ブロックだけでも世帯数が150以上となる調査区が04層では78、04層以外では83あり、それらのうち200以上となるものは04層では33、04層以外では16ある。一般的には世帯数が多いからと言って必ずしも就業者数が多くなるわけではないが、本研究で用いた標本の中で筆者が確認したところ、標本調査区の世帯数が150以上あり、ほとんどを就業者が占めるという事例があった。世帯は普通世帯・準世帯でとらえたものである。

¹⁴ 山口(2016)は、社会生活基本調査を用いて、副標本の組数を4,6,8とした場合の副標本法とブートストラップ法で求めた標準誤差の推定値を比較し、副標本の組数を増やしても必ずしもばらつきが小さくなるとは限らず、副標本の大きさ、同質性などを考慮した副標本を作成する必要性を指摘している。標本誤差推計研究会編(1998)では、土地基本調査世帯調査を用いて、理論式法、副標本法及び2種類のバイアス修正したブートストラップ法の間で標準誤差率の推定値の分布を比較しており、副標本法では組分けの違いによるばらつきが大きく、1回だけの推定は不安定であると述べている。

¹⁵ 当初、各精度評価方法別に個別データから推定値等を得るまでの計算時間を計測する予定であったが、実際には精度評価方法別に計算時間を分離することが困難であった。

理論式法では、実際の標本設計を反映した算式を解析的に導出する必要があるが、労調のように標本設計が複雑であると、算式の導出が非常に複雑になる場合や導出自体が困難になる場合がある(馬場他(1997)、標本誤差推計研究会編(1998))。

ブートストラップ法は、理論式法のように精度計算の算出式を標本設計に従って導出する必要がなく、理論式法ほど精度計算の手順が複雑ではない。精度計算の算出式は形式的には副標本法と似ているので実務家も理解しやすいため、実務での適用容易性は十分あると思われる。また、推定精度のばらつきが理論式法に比べると大きいものの、副標本法に比べると安定的である。本研究では、調査区単位の再抽出のみ行っており、世帯単位の再抽出は行っていない。世帯単位のばらつきを考慮すると、今回の結果に比べ推定標準誤差は大きくなると考えられる。また、世帯単位の再抽出処理を行うとすると、計算量が増大するため計算時間は増加するが、コンピューターの計算能力が今後も進歩していくことを考えれば解決可能な課題と考えられる。

古橋・岩永(1991)は、就業者数と完全失業者数の標準誤差率の推定値についてブートストラップ法が理論式法を下回るのは、標本の再抽出を調査区単位で行っており、世帯間のばらつきが考慮されていないためとしている。また、完全失業者数の標準誤差率の推定値について変動係数でばらつきを見ると、ブートストラップ法の方が理論式法よりかなり下回っている。この結果は、本研究における実験結果と整合しない。古橋・岩永(1991)では12か月分(昭和63年1月～12月調査分)の標準誤差率をブートストラップ法と理論式法で算出し、その月別標準誤差率を用いて平均、標準偏差、変動係数を算出している。古橋・岩永(1991)と本研究の結果が整合しなかった理由の1つとして、使用したデータが、12か月分と少なく、たまたまブートストラップ法より理論式法で算出した変動係数が大きく出る期間のデータであったことが考えられる¹⁶。しかし、副標本法の変動係数が理論式法とブートストラップ法に比べかなり大きくなっている点は、本研究の実験結果と整合的である。

統計利用者に統計を的確、有効に利用してもらうには、より正確な精度情報を提供することが必要である。労調の月次結果及び年平均結果の標本誤差は、調査項目ごとの標準誤差率を曲線の当てはめにより平均的に評価し、得られた平均的な標準誤差率に推定値を乗じて標準誤差を算出している。そうであるなら、標準誤差率の推定値のばらつきが小さい精度評価方法を採用する方が、より安定的な回帰曲線を得ることが期待できる点で望ましいと言える。一方、方法の選択に当たっては、実務の観点も考慮する必要がある。実務者にとって、採用しようとする方法のわかりやすさ、適用しやすさは重要な観点である。図2で見ると、ブートストラップ法は、今回の理論式法に比べ若干の不安定さはあるものの副標本法に比べると安定的である。ブートストラップ法は、実務家にとって、理論式法のように実際の標本設計に従って精度評価の算式を導出する必要なく適用できるという点で、理論式法に比べて実用上の利点が大きいと考えられる。

実務的な適用容易性を考えると、集計日程が非常にタイトな月次集計の精度計算は簡便で迅速に結果を得られる副標本法で行い、推定値の精度を公表している年次報告書をまとめる際にはブートストラップ法で精度を再計算し、より安定的な精度を提供することが実用的で推奨される。また、理論式法は、5年ごとの標本設計の変更後の一定期間に他の方法と並行して精度計算を行い、その結果をメルクマールとして利用することも考えられる。

コンピューターの計算能力が今後も進歩していくことを考えれば、標本調査の精度評価方法に副標本法以外の方法の導入を検討することは必要であろう。

¹⁶ 1000組の標本から一定の標本数のグループを複数作り、各グループの変動係数を算出したところ、理論式法の変動係数がブートストラップ法を上回る事例が出現し得ることがわかった。また、グループの大きさが小さいと出現の頻度は大きいことがわかった。

謝辞

2名の匿名の査読者の方からは、多くの有益な指摘及びコメントをいただき、本稿の内容を大きく改善することができた。ここに記して深く感謝の意を表したい。

参考文献

- [1]石井達男(1988), 「労働力調査の標本誤差の一試算」, 『統計局研究彙報』, 第46号, 総務庁統計局・統計センター, 55-78.
- [2]総務省統計局(2007), 『平成17年国勢調査 調査区関係資料 利用の手引』.
- [3]総務省統計局(2008), 『労働力調査 標本設計の解説 第3版』.
- [4]総務省統計局(2011), 『労働力調査の解説 第3版』.
- [5]椿広計・會田雅人(2019), 「リサンプリングによる労働力調査推定精度評価」, 『統計研究彙報』, 第76号, 総務省統計研究研修所, 39-50.
- [6]津村善郎・築林昭明(1986), 『標本調査法』, 岩波書店.
- [7]馬場康維・土屋隆裕・中村好宏・小林良行(1997), 「ブートストラップ法による標準誤差の推定の試み(2)」, 第11回日本計算機統計学会大会, 86-89.
- [8]標本誤差推計研究会編(1998), 『標本誤差の推計方法—最新時代の理論と実証—』, 財団法人統計情報研究開発センター.
- [9]古橋正宏・岩永琢磨(1991), 「労働力調査の標本誤差」, 『統計局研究彙報』, 第49号, 総務庁統計局・統計センター, 37-50.
- [10]山口幸三(2016), 「副標本による標本誤差の計測」, 『統計学』, 第111号, 経済統計学会, 17-26.
- [11]Cochran, W.G.(1977), “Sampling Techniques”, third edition, John Wiley & Sons.
- [12]Efron, B.(1979), Bootstrap methods:Another look at the Jackknife, *the Annals of Statistics*, Vol.7, No.1, 1-26.

表 1 国勢調査区の層別基準(総務省統計局(2007)「産業・従業上の地位別就業者数等による調査区分類(大分類・小分類)」より抜萃)

分類符号		分類基準
大分類符号	小分類符号	
01		刑務所・拘置所がある区域、自衛隊区域及び駐留軍区域の調査区並びに水面調査区
02		人口が0の調査区
03		換算世帯数が15以下の調査区
04		50人以上の世帯からなる学生の寮・寄宿舎等のある標本単位区
05		漁業の就業者の比が0.2以上の調査区
06		漁業の就業者の比が0.1以上0.2未満の調査区
07		建設業、製造業の業主の比が0.1以上の調査区
08		卸売・小売業、飲食店、宿泊業の業主の比が0.1以上の調査区
09		情報通信業、運輸業、金融・保険業、不動産業、医療、福祉、教育、学習支援業、複合サービス事業、サービス業の業主の比が0.1以上の調査区
10		農林業の就業者の比が0.3以上の調査区
11		農林業の就業者の比が0.1以上0.3未満の調査区
12		公務の就業者の比が0.1以上の調査区
13		金融・保険業、不動産業の雇用者の比が0.1以上の調査区
14		製造業の雇用者の比が0.3以上の調査区
15		建設業の雇用者の比が0.1以上の調査区
16		医療、福祉の雇用者の比が0.1以上の調査区
17		卸売・小売業、飲食店、宿泊業の雇用者の比が0.2以上の調査区
18		教育、学習支援業、複合サービス事業、サービス業の雇用者の比が0.2以上の調査区
19		電気・ガス・熱供給・水道業、情報通信業、運輸業の雇用者の比が0.1以上の調査区
20		製造業の雇用者の比が0.2以上0.3未満の調査区
21		製造業の雇用者の比が0.1以上0.2未満の調査区
22		卸売・小売業、飲食店、宿泊業の雇用者の比が0.1以上0.2未満の調査区
23		教育、学習支援業、複合サービス事業、サービス業の雇用者の比が0.1以上0.2未満の調査区
99		上記のいずれにも属さない調査区

表2 国勢調査区の層別基準—04層(総務省統計局(2007)「産業・従業上の地位別就業者数等による調査区分類(大分類・小分類)」より抜萃)

分類符号		分類基準
大分類符号	小分類符号	
04	01	学生の寮・寄宿舎(ただし、50人以上の世帯)のある標本単位区
04	02	病院・療養所(ただし、50人以上の世帯)のある標本単位区
04	03	社会施設(ただし、50人以上の世帯)のある標本単位区
04	04	後置番号が4のうち、上記のいずれにも属さない標本単位区
04	11	寮などに住む建設業の就業者が50人以上の標本単位区
04	12	建設業の世帯の比が0.2以上の標本単位区
04	21	寮などに住む製造業の就業者が50人以上の標本単位区
04	22	製造業の世帯の比が0.3以上の標本単位区
04	31	寮などに住む卸売・小売業、飲食店、宿泊業の就業者が50人以上の標本単位区
04	32	卸売・小売業、飲食店、宿泊業の世帯の比が0.3以上の標本単位区
04	41	寮などに住む金融・保険業、不動産業の就業者が50人以上の標本単位区
04	42	金融・保険業、不動産業の世帯の比が0.2以上の標本単位区
04	51	寮などに住む電気・ガス・熱供給・水道業、情報通信業、運輸業の就業者が50人以上の標本単位区
04	52	電気・ガス・熱供給・水道業、情報通信業、運輸業の世帯の比が0.3以上の標本単位区
04	61	寮などに住む医療、福祉、教育、学習支援業、複合サービス事業、サービス業の就業者が50人以上の標本単位区
04	62	医療、福祉、教育、学習支援業、複合サービス事業、サービス業の世帯の比が0.4以上の標本単位区
04	71	寮などに住む公務の就業者が50人以上の標本単位区
04	72	公務の世帯の比が0.4以上の標本単位区
04	91	後置番号が8の調査区のうち、上記のいずれにも属さない標本単位区
04	92	給与住宅に住む一般世帯数の比が0.5以上の調査区のうち、上記のいずれにも属さない標本調査区

表3 労働力調査の地域別層別標本調査区配分(出典：総務省統計局(2011) 付表6)
(地域別標本調査区数)

	分類符号		北海道	東北	南関東	北関東 甲信	北陸	東海	近畿	中国	四国	九州	沖縄	合計	
	大分類 符号	小分類 符号													
I	02		1	1	1	1	1	1	1	1	1	1	1	11	
	03		3	3	3	3	3	3	3	3	3	2	2	31	
	04	01	1	1	1	1	1	1	1	1	1	1	1	11	
	04	02	1	1	1	1	1	1	1	1	1	1	2	2	13
	04	03	1	1	1	1	1	1	1	1	1	1	1	1	11
	04	04	1	1	1	1	1	1	1	1	1	1	1	11	
II	05		16	40	8	32	16	16	8	24	24	32	8	232	
	06														
	11														
	10											8			
III	04	21	8	48	8	112	16	56	8	8	32	32	8	880	
	04	22													
	14														
	20														
	21				104		64	120	112	56					
IV	04	11	16	8	16	8	8	8	8	8	8	8	8	8	
	04	12													
	15														
	07														
V	04	31	8	8	24	8	8	16	16	8	48	16	64	944	
	04	32													
	08														
	17														
	22				72		64	208	32	32		48			120
VI	04	61	8	8	8	16	8	16	16	8	8	24	8	768	
	04	62													
	16														
	18														
	23				16		16	80	32	24		24			
09															
VII	04	41	8	8	8	8	8	8	16	16	8	8	24	768	
	04	42													
	13														
	04	51													
	04	52			56										
	19														
	04	71													
	04	72			8										
	12														
	04	91													
	04	92			16		24	56	16	8		8			56
99															
合計			176	232	592	240	176	304	400	208	152	288	144	2,912	

表4 標本組別推定値及びブートストラップ推定値の分布の基本統計量—完全失業率、完全失業者及び労働力人口

	完全失業率(%)		完全失業者(人)		労働力人口(人)	
	標本組別推定値	標本組別ブートストラップ推定値	標本組別推定値	標本組別ブートストラップ推定値	標本組別推定値	標本組別ブートストラップ推定値
平均	6.7278	6.7501	134723	135143	2002703	2002394
標準偏差	0.5888	0.5865	11746	11706	25640	25541
最小値	5.0526	5.0592	100482	100627	1919212	1918105
10%点	5.9561	5.9822	119216	119410	1968180	1968357
第1四分位	6.3218	6.3527	126678	127300	1984830	1984276
中央値	6.7247	6.7397	134406	135022	2002352	2002115
第3四分位	7.1143	7.1316	142637	142945	2021368	2020667
90%点	7.5264	7.5444	150777	151406	2035723	2035269
最大値	8.4884	8.4797	168800	168866	2084878	2083248
変動係数(%)	8.75	8.69	8.72	8.66	1.28	1.28
十分位範囲	1.5703	1.5622	31561	31996	67542	66912
四分位範囲	0.7925	0.7789	15959	15645	36538	36392
範囲	3.4358	3.4205	68318	68239	165666	165143
相関係数	0.996		0.996		0.996	
(母集団値)	6.7527		132234		1958229	

(1) 十分位範囲、四分位範囲及び範囲の単位はパーセントポイント

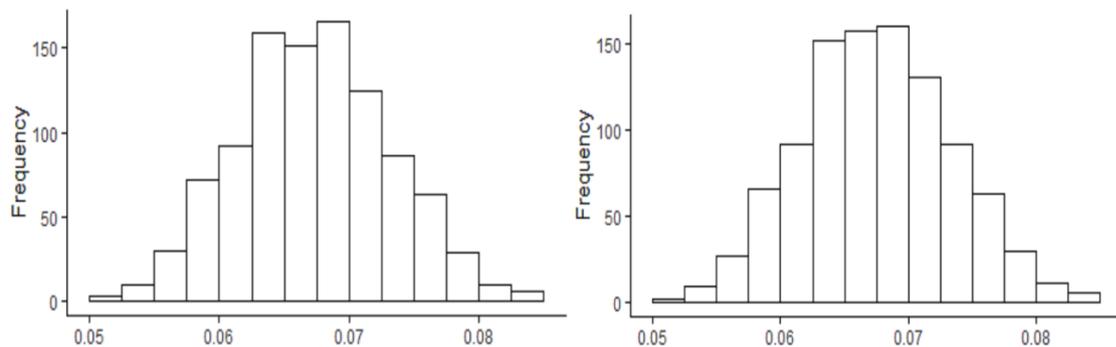


図1 完全失業率の標本組別推定値と標本組別ブートストラップ推定値の分布
 左図：標本組別推定値、右図：標本組別ブートストラップ推定値

表5 標本組別推定標準誤差率分布の基本統計量－完全失業率、完全失業者及び労働力人口

	完全失業率(%)			完全失業者(%)			労働力人口(%)		
	理論式法	副標本法	ブートストラップ法	理論式法	副標本法	ブートストラップ法	理論式法	副標本法	ブートストラップ法
平均	7.2703	8.1853	8.2707	7.3246	8.6400	8.2784	0.9667	3.0538	1.1337
標準偏差	0.3561	2.2678	0.8222	0.3549	2.3848	0.8302	0.0338	0.8553	0.1085
最小値	6.2858	2.0600	6.2507	6.3335	2.1828	6.3272	0.8804	0.6917	0.9124
10%点	6.8316	5.3072	7.3040	6.8889	5.5579	7.3037	0.9252	2.0086	1.0146
第1四分位	7.0129	6.6308	7.7347	7.0699	7.0276	7.7310	0.9438	2.4818	1.0666
中央値	7.2581	8.1005	8.1966	7.3118	8.4916	8.2061	0.9643	2.9924	1.1207
第3四分位	7.4964	9.5622	8.7177	7.5537	10.1019	8.7018	0.9874	3.5541	1.1871
90%点	7.7502	11.2026	9.2208	7.8106	11.8154	9.2504	1.0103	4.1760	1.2511
最大値	8.4352	19.4782	13.6941	8.5070	17.6487	14.0473	1.0894	7.6083	1.7629
変動係数(%)	4.90	27.71	9.94	4.85	27.60	10.03	3.50	28.01	9.57
十分位範囲	0.9186	5.8955	1.9168	0.9217	6.2575	1.9467	0.0850	2.1674	0.2365
四分位範囲	0.4835	2.9314	0.9830	0.4837	3.0743	0.9708	0.0436	1.0723	0.1205
範囲	2.1494	17.4181	7.4434	2.1735	15.4659	7.7201	0.2090	6.9166	0.8505
(標準誤差率)	7.2195			7.4490			1.0185		

(1) 十分位範囲、四分位範囲及び範囲の単位はパーセントポイント

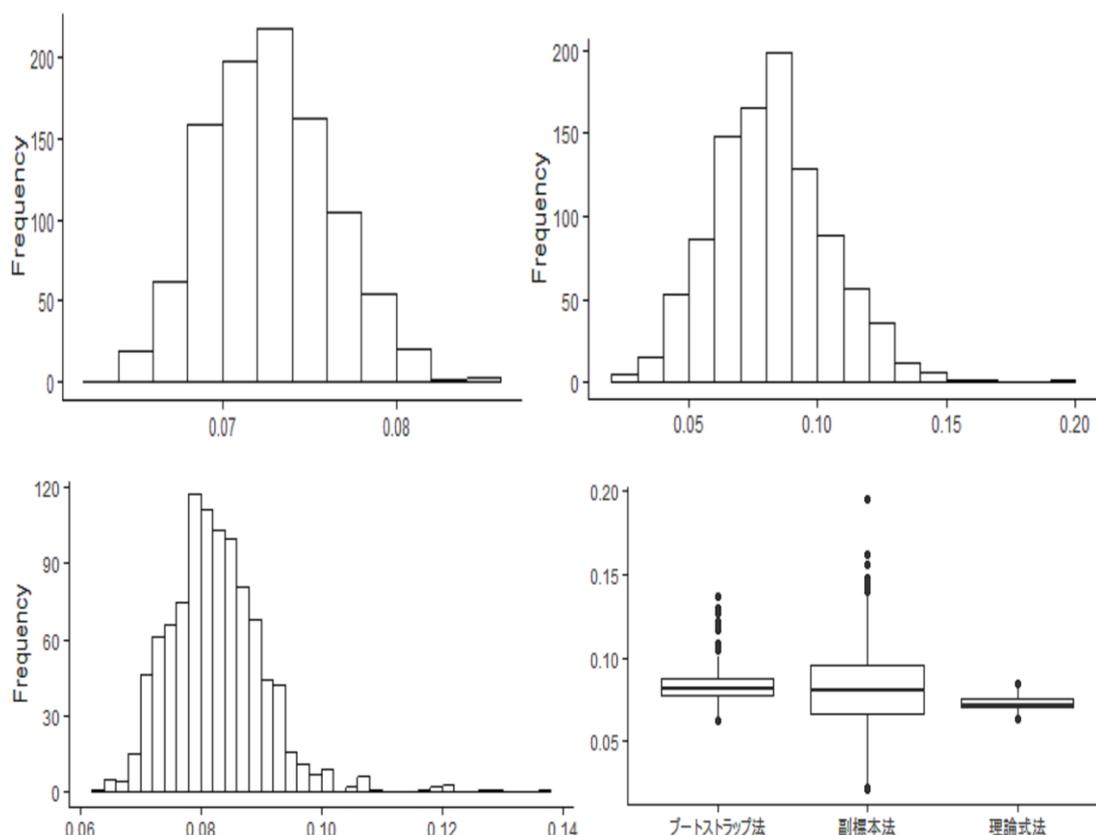


図2 完全失業率の標本組別推定標準誤差率分布

上段： 左図：理論式法、右図：副標本法

下段： 左図：ブートストラップ法、右図：精度評価方法別の分布比較

表6 標本組別推定標準誤差分布の基本統計量－完全失業率、完全失業者及び労働力人口

	完全失業率(%)			完全失業者(人)			労働力人口(人)		
	理論式法	副標本法	ブートストラップ法	理論式法	副標本法	ブートストラップ法	理論式法	副標本法	ブートストラップ法
平均	0.4874	0.5498	0.5573	9833	11623	11169	19356	61151	22697
標準偏差	0.0266	0.1564	0.0666	536	3297	1344	609	17135	2142
最小値	0.4130	0.1390	0.3927	8335	2629	7810	17828	13957	18320
10%点	0.4544	0.3546	0.4843	9161	7293	9682	18621	40229	20419
第1四分位	0.4695	0.4401	0.5122	9462	9206	10239	18942	49792	21334
中央値	0.4865	0.5420	0.5492	9815	11470	11019	19321	59764	22434
第3四分位	0.5060	0.6495	0.5924	10173	13921	11889	19723	71159	23658
90%点	0.5203	0.7536	0.6426	10523	15871	12871	20145	83667	24976
最大値	0.6135	1.2691	0.9262	12392	22584	18776	21831	150632	35735
変動係数(%)	5.47	28.44	11.96	5.45	28.37	12.03	3.15	28.02	9.44
十分位範囲	0.0659	0.3990	0.1583	1362	8578	3189	1524	43437	4558
四分位範囲	0.0366	0.2093	0.0802	711	4715	1650	781	21367	2324
範囲	0.2005	1.1301	0.5336	4057	19955	10966	4003	136676	17415
(標準誤差)	0.4875			9850			19945		

(1) 完全失業率の十分位範囲、四分位範囲及び範囲の単位はパーセントポイント

