

リサンプリングによる労働力調査推定精度評価

椿 広計[†]會田 雅人[‡]

Estimation Accuracy Evaluation of Labor Force Survey by Resampling Methods

TSUBAKI Hiroe
AIDA Masato

本報告は、労働力調査の推定精度を2通りのリサンプリング技法で評価した。労働力調査は、層化2段階抽出調査であり、調査区の抽出は近似的な人口に関する確率比例抽出、調査区内住戸の抽出は単純無作為抽出である。筆者らは、この層化2段階抽出の確率比例抽出を再現したリサンプリングと、調査区抽出については、単純無作為抽出で近似した方法を2017年9月に実施された労働力調査個票データに適用し、その結果を比較した。2つのリサンプリング技法で評価した推定量のブートストラップ分布は有意に異なるが、標準誤差の推定値は殆ど変わらない。従って、推定量の分布や信頼区間を求めるには、調査区の確率比例抽出が必要だが、標準誤差の評価には、単純な無作為調査区抽出を用いることが可能である。

キーワード：ブートストラップ

The authors evaluate the accuracy of the estimation of the labor force survey by resampling methods. Labor force survey uses stratified 2 stage sampling, the first stage is sampling from the census blocks with probability approximately proportional to the population of a block and the second is simple random sampling from the dwelling units of the sampled block. Two resampling methods are applied to the microdata from the labor force survey carried out in September 2017. The first method reproduces the original 2 stage sampling faithfully as possible and the second method resamples the blocks simply at random. The results show their bootstrap distributions of the estimates is significantly different, but their standard errors hardly change. Therefore the faithful resampling may be necessary to evaluate sampling distribution of the estimates or their confidence interval, however the simple resampling should be recommended only to evaluate their standard errors.

Key Words : Bootstrapping

[†] 独立行政法人統計センター[‡] 公益財団法人統計情報研究開発センター

1. 研究目的

総務省の労働力調査は、日本の就業状態を明らかにする基礎資料を作成するために、実施されている。労働力調査は、全国約 100 万国勢調査区を第一次抽出単位とし、抽出された 2912 調査区内の住戸を第 2 次抽出単位とし、11 地域別の層化 2 段抽出が行われている。また、調査区はその特性によって層化されたうえで、調査区内換算世帯数をウェイトとする近似的な人口比例抽出が行われていることが特徴的である。更に、線型推定の推定精度を向上させるために、比推定が用いられている。このような比較的複雑な標本設計がなされているため、これまで標本誤差推定には副標本法などが用いられてきた。本報告は、Efron(1979)、Funaoka et al. (2006)らのリサンプリング技法に基づき産業別就業者数の集計の推定精度を評価し、現行の副標本法と比較したものである。現在、労働力調査は、標本設計が比較的複雑なため、後述するように標本を調査開始月の別等により 8 分割した、副標本法により推定精度を評価されているからである。

さて、本報告の第 1 の目的は、労働力調査からのリサンプリング、特に調査区抽出（第一次抽出）を標本設計に準拠して、厳密に行う必要性が統計実務的にあるか否かを検討することである。労働力調査個票データには、第一次抽出、すなわち調査区確率比例抽出に用いたウェイトは含まれていない。仮に、第一次抽出に関わるリサンプリングを単純無作為抽出で行っても、確率比例抽出した場合と精度評価が、大きく変わらないのならば、労働力調査個票データ分析を行う利用者が、別途調査区ウェイト情報を入手し、個票データとリンケージさせる必要性は少ない。本報告のもう一つの目的は、現行の副標本法による精度評価とリサンプリングによる評価を比較することである。やはり、その推定精度評価に大きな差がなければ、単純な副標本法を今後も続けることが实际的だからである。

Efron (1979) によって提案された Bootstrap 法は、推定量の標本分布を標本からのリサンプリングに基づいて行う汎用性の高い方法である。Bootstrap 提唱以降、それを層化抽出標本に適用することの大標本論的正当性は、Bickel et al. (1984)、Rao and Wu (1988)で示された。また、有限母集団補正を意識した、特に層の大きさが小さい層化抽出調査に対する Bootstrap の修正なども多く提案されており、Shao (2003)は標本調査分野での総合報告となっている。

Bootstrap による推定精度評価は、公的統計に限らず多くの分野で既に行われてきた。日本の公的統計分野でも、標本設計の複雑な労働力調査に対しては、本報告と全く同じ問題意識の下、古橋・岩永(1991)が、1988 年の労働力調査の就業者数と完全失業者に対して、副標本法、Bootstrap、理論分散とで推定精度を比較評価している。ただし、古橋・岩永(1991)では、調査区の確率比例抽出（第一次抽出）だけをリサンプリングし、調査区内住戸抽出（第二次抽出）のリサンプリングは行っていない。このため、調査区内分散が無視されるため Bootstrap による推定誤差は過小評価されたと考察している。また、馬場・土屋・中村・小林(1997)も 1995 年の労働力調査に対して、標本調査区をリサンプリングの対象とする場合と標本世帯の世帯員をリサンプリングの対象とする場合のブートストラップによる精度評価を層別抽出の場合と層別抽出しない場合とで実施し、ブートストラップ法の実用可能性を論じている。

馬場・土屋・中村・山崎(1996)は、1992 年国民生活基礎調査に対して、Funaoka et al. (2006)は、1997 年の全国物価統計調査に対して、多段の不等確率のリサンプリングを実現するベルヌーイ型 Bootstrap を適用し、精度評価を試みている。

本報告は、Funaoka et al. (2006)に従い、第一次抽出を完全無作為抽出とした場合と古橋・岩永(1991)が試行した確率比例抽出した場合を行うと共に、第二次抽出は無作為抽出を行ったリサンプリングを行い、労働力調査の産業別就業者数の精度を評価したリサンプリング実験結果を示す。なお、Shao(2003)等で示された、有限母集団に関わる Bootstrap の補正、例えば標本の大きさの小さい層に対する補正技術などについては、リサンプリング実験の検討の対象としていない。

本報告では、第 2 章でこの労働力調査の標本設計について概説した上で、第 3 章で推定精度評価のために用いたリサンプリング実験の方法を紹介し、第 4 章でその評価結果を示し、第 5 章に考察

と今後の課題を示す。

2. 労働力調査の標本設計と推定方法の概略

ここでは、総務省統計局(2018)に基づき、リサンプリング技法の前提となる労働力調査の標本設計について、本報告の実験に必要な情報を紹介する。

調査区は、直近の国勢調査結果より国勢調査区の産業、従業上の地位に関する特性に注目して表1のように層化されている。表1の層符号03にある「換算世帯」は、調査区の人口規模に比例するように世帯数を変換したもので、(1)のように算出される。

$$\begin{aligned} \text{換算世帯数} = & \text{世帯人員が2人以上の一般世帯数} \\ & + (\text{世帯人員が1人の一般世帯数} + \text{施設等の世帯人員}) / 3 \end{aligned} \quad (1)$$

調査区の抽出は、この換算世帯数に近似的に比例するように行われている。すなわち、(2)を調査区ウェイトとした確率比例抽出がなされている。

$$\text{調査区ウェイト} = (\text{換算世帯数} / 15) \text{の切り上げ} \quad (2)$$

ただし、表1の層2については全て調査区ウェイトを1としている。

なお、労働力調査では、毎月の全国集計以外に11地域別の四半期集計を行っており、その推定精度を一定にするために地域ごとに標本配分も表2のように設計されている。標本調査区の抽出は、表2の地域ごとに表1の各層で、層内調査区のウェイト(2)を用いて確率比例抽出される。また、抽出された調査区からの住戸の抽出には、この調査区ウェイト(2)の逆数が用いられる。

推定に当たっては、標本抽出確率の逆数を乗率とする線型推定の精度を改善するために、国勢調査結果などから推定した毎月末日現在の人口をベンチマーク人口とする比推定方式が用いられている。すなわち、先ず男女(2区分)、年齢階級(16区分)、地域(11区分)毎に人口の線型推定値を算出し、

$$\text{比推定用乗率} = \text{ベンチマーク人口} / \text{人口の線型推定値}$$

が算出されている。なお、実際の労働調査個票データには、抽出標本毎に集計用乗率(3)が付与されている。

$$\begin{aligned} \text{集計用乗率} = & \text{標本の属する調査区の線型推定用乗率} \\ & \times \text{標本の属する男女、年齢階級、地域の比推定用乗率} \end{aligned} \quad (3)$$

このように労働力調査は精緻な標本設計と比推定が適用されていたため、標本誤差推定に当たっては、標本を調査開始月・1年目調査か2年目調査かに応じて8分割した副標本による推定値のばらつきを用いて推定精度が評価されてきた。本報告は、労働力調査産業別就業者数の集計値の標本分布をEfron(1979)、Funaoka et al.(2006)に基づくリサンプリング技法で、標本抽出や比推定操作を繰り返し再現し、推定精度の評価を試みたものである。

3. 推定精度のリサンプリング評価の実験方法

本報告の推定精度評価の対象とした2017年9月実施の労働力調査個票データは、統計研究研修所のオンサイト環境で分析することを前提に申請入手したものである。評価対象とした個票データは、自衛官、受刑者を除く15歳以上住民79678名の就業状態などからなっている。

推定精度評価のために、以下の2つの方法に基づく抽出標本からのリサンプリングを1000回実施し、推定量のブートストラップ分布を導いた。なお、リサンプリングの繰り返しについては、500回以上で、標準誤差推定やリサンプリング分布のヒストグラム形状が、単峰かつ対称な分布形で安定したので、1000回の結果を示している。

(1) 調査区抽出を確率比例抽出としたリサンプリング（以下、忠実なりサンプリング）

調査区リサンプリングについては、次のように行う。先ず、実際に第一次抽出された2912調査区から、表1の層 s ($s=1, \dots, S$)、表2の地域 r ($r=1, \dots, R$)ごとに、当該層・地域の抽出調査区 ($k=1, \dots, K_{sr}$)に対して(2)式の調査区ウェイト w_{srk} を参照する。次に、抽出調査区から、層 s 、地域 r に属する抽出調査区 k の抽出確率を $w_{srk} / \sum_{k=1}^{K_{sr}} w_{srk}$ として、 K_{sr} 調査区を復元抽出する。

住戸リサンプリングについては、リサンプリングされた調査区 k で抽出された住戸から、実際に抽出された住戸数と同数の住戸を無作為復元抽出する¹。

(2) 調査区抽出を単純無作為抽出としたリサンプリング（以下、単純なりサンプリング）

調査区リサンプリングについては、上記(1)の忠実なりサンプリングの代わりに、層 s 、地域 r に、実際に抽出された調査区数 K_{sr} と同数の調査区を確率 $1/K_{sr}$ で、無作為復元抽出する。住戸リサンプリングは、(1)と同じように行う。

なお、労働力調査で抽出されている調査区は、5章の副標本法との比較でも紹介するように調査開始時期、1年目調査・2年目調査に8分割可能である。この8分割は独立な抽出標本と考えられるので、リサンプリングに際し所与と考え層とみなすこともあり得る。しかし、このリサンプリング実験では、この8分割を層とは考えず調査区を抽出している。

また、実際の労働力調査の標本抽出は、系統抽出であるが、このリサンプリング実験は全て、無作為抽出ないしは確率比例抽出で行っており、ここでのリサンプリング実験は系統抽出との推定精度には大きな差が無いという仮定で行っている。

また、上記のリサンプリング技法は1章で述べたように、有限母集団に関わる補正を無視している。実際、非復元抽出の影響が大きくなる、調査区数の小さい層、ないしは調査区内住戸数が小さい調査区に対しては、Shao(2003)などが指摘しているように推定精度を過大評価することとなる。従って、調査区内住戸数が3未満の10調査区、並びに表1の04層の一部(0401、0402、0403および0404層)及び調査区ウェイト1となる02層(人口0の調査区)、03層(換算世帯15以下の調査区)に属する調査区並びに調査区内住戸は、形式的に悉皆調査区、悉皆層として扱った。すなわち、リサンプリングの対象とせず、リサンプリング標本に常に含めることとした。

比推定値のリサンプリング標本での導出については、次のように行った。

Step 1: リサンプリング前の全ての個票データに付随する集計用乗率を用いて、男女2区分、年齢階級16区分、地域11区分ごとにベンチマーク人口 P_{ijk} (添え字は、 i が男女、 j が年齢階級、 k が地域に相当し、 $i=1, 2$ 、 $j=1, \dots, 16$ 、 $k=1, \dots, 11$)をクロス集計する。

Step 2a: リサンプリングを基に得られた $l=1, \dots, 1000$ の標本に対して、男女、年齢階級、地域区分ごとに個票データに付随する線型推定乗率を用いてクロス集計を行い人口の線型推定値 p_{ijk} と21産業別就業者数の線型推定値 y_{ijklm} (添え字 m は、産業分類に相当し、 $m=1, \dots, 21$)を算出する。

¹ 実際の労働力調査では、地域ごとに、層をあらかじめ合併した層から調査区を抽出している。

Step 2b: Step 2a に引き続き、男女、年齢階級、地域区分ごとに産業別就業者数集計の比推定値 $Y_{ijklm} = y_{ijklm} \times P_{ijk} / p_{ijkl}$ を算出する。

Step 3: Step 2b で得られた比推定値 Y_{ijklm} を男女、年齢階級、地域区分について集計し、産業 m の就業者数の第 l リサンプリングに対する推定値 $Y_{+++lm} = \sum_{i=1, 2} \sum_{j=1, \dots, 16} \sum_{k=1, \dots, 11} Y_{ijklm}$ を算出する。

Step1、Step 2、Step 3 を通じて、産業 m の就業者数の比推定値 Y_{+++lm} ($l=1, \dots, 1000$) のリサンプリングに基づく経験分布が導出される。

4. リサンプリング技法の適用結果

19 産業別就業者数のブートストラップ分布のパーセント点、推定値の平均、標準偏差（推定値の標準誤差）、標準誤差率を忠実なりサンプリングと単純なりサンプリングについてそれぞれ求めた結果を表3に示す。表3にはリサンプリング技法による推定値の平均値の差に関する Welch の t 統計量並びに 2017 年 9 月調査に関する総務省統計局の公表数値も示した²。また、これらの数値計算には R version 3.4.2 を用いた。

表3の t 統計量の絶対値が3を超えているものについては、2つのリサンプリング技法によって推定値の標本分布の期待値が有意にずれていると考えるとすれば、19 産業中有意差が無い産業は、漁業、建設業、製造業、宿泊飲食業、生活関連サービス業、医療福祉産業の6産業だけである。また t 統計量の絶対値が10を超えているのは、その大きさの順に、(他に分類されない) サービス業、運輸郵便業、公務、鉱業採石業、教育、不動産業の6産業である。

そこで、 t 統計量の絶対値が最も大きく、忠実なりサンプリング技法と単純なりサンプリング技法とで推定量のブートストラップ分布の乖離が最も大きい、(他に分類されない) サービス業と t 統計量の絶対値が最小の生活関連サービス業について、実際に 1000 回のリサンプリングで求めた推定量のヒストグラムを描いたのが、それぞれ図1、図2である。推定量の推定標本分布自体は単峰で正規分布に近い形状を示している。詳細は省略するが、その他の産業についてのヒストグラムも全て分布の形状は同様である。但し、図1では単純なりサンプリング技法による推定値の分布は、忠実なりサンプリング技法による推定量の分布と比較して、平均値同様右側にシフトしている。表3に示したブートストラップ分布の2.5%を下限、97.5%を上限とする、両産業の就業者数推定値の信頼率95%の両側信頼区間も、忠実なりサンプリングと単純なりサンプリングとの場合で、それぞれ次のようになる。

(他に分類されない) サービス業就業者数推定値 95%信頼区間

忠実なりサンプリング技法：[4101881, 4563087]

単純なりサンプリング技法：[4154410, 4647957]

2017 年 9 月労働力調査推定公表値：443 万人

生活関連サービス業就業者数推定値 95%信頼区間

忠実なりサンプリング技法：[2244823, 2606077]

² 表3の表頭は産業名を一部省略している。正式には、農業・林業、漁業、鉱業・採石業・砂利採取業、建設業、製造業、電気・ガス・熱供給・水道業、情報通信業、運輸業・郵便業、卸売業・小売業、金融業・保険業、不動産業・物品賃貸業、学術研究・専門・技術サービス業、宿泊業・飲食サービス業、生活関連サービス業・娯楽業、教育・学習支援業、医療・福祉、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(他に分類されるものを除く)である。

単純なりサンプリング技法：[2233396, 2619682]

2017年9月労働力調査推定公表値：242万人

このように、忠実なりサンプリング技法を単純なりサンプリング技法で近似することは、ブートストラップ分布自体にはバイアスが生じる可能性が高いことが分かる。しかし、推定量の標準誤差率の評価については、表3を見る限り有効数字2桁目が最大2ずれるだけである。

また、2つのリサンプリング分布の平均値と公表数値の誤差率を比較すると、必ずしも忠実なりサンプリング分布の期待値の誤差率が小さいわけではない。実際、鉱業採石業就業者数の公表値は50000名で、単純なりサンプリング分布の平均値48521名の誤差率は-3.0%、忠実なりサンプリング分布の平均値の43887名の誤差率は-12.2%である。ただし、他の産業での忠実なりサンプリング技法の誤差率絶対値の最大値が、電気ガス業の3.4%であることや、単純なりサンプリング技法の誤差率絶対値の最大値が公務の-4.1%と比較して、鉱業採石業就業者数の忠実なりサンプリング分布の平均値は、突出して大きな誤差率である。

5. 考察と今後の課題

4章の検討から、調査区抽出が近似的に換算世帯数に比例する確率比例抽出となっていることを無視して、調査区リサンプリングを行うことで、推定値のブートストラップ分布並びにその期待値は有意にシフトする機会が多いことが分かった。しかし、ブートストラップ分布の平均値と公表数値とを比較する限り、忠実なりサンプリング技法と単純なりサンプリング技法とで、どちらが優越しているとは言えない。リサンプリングが公表数値の標準誤差を求めるために行うという観点では、単純な調査区リサンプリングに基づいて、推定誤差を評価することには、実務上問題はない。むしろ単純なりサンプリング技法の方が、調査区ウェイトといった個票データには含まれていない標本設計情報を参照する必要がなく、簡便な方法として推奨できる。

一方、現行の副標本法による標本誤差については、「総務省統計局(2018)」表4-1に2017年平均結果の標準誤差率が14産業について記載されている³。なお、現行の副標本は、「総務省統計局(2018)」に記述されている通り、

- A 1 …… 1月、5月又は9月に調査開始の1年目調査区
- A 2 …… 1月、5月又は9月に調査開始の2年目調査区
- B 1 …… 2月、6月又は10月に調査開始の1年目調査区
- B 2 …… 2月、6月又は10月に調査開始の2年目調査区
- C 1 …… 3月、7月又は11月に調査開始の1年目調査区
- C 2 …… 3月、7月又は11月に調査開始の2年目調査区
- D 1 …… 4月、8月又は12月に調査開始の1年目調査区
- D 2 …… 4月、8月又は12月に調査開始の2年目調査区

といった「同質」の標本による8分割である。すなわち、地域間変動、層間変動、地域ごとの層内調査区変動成分がほぼ無作為に副標本に配分されている。これから、8副標本毎に比推定値を求め、その標準誤差（推定値の標準偏差の $1/8^{1/2}$ 倍）を全標本からの推定値の標準誤差としているのである。

表3の忠実なりサンプリング技法に基づく標準誤差率を横軸に副標本法の標準誤差率を縦軸にプロットしたのが、図3である。両者の関係について切片の無い単回帰式を当てはめると

$$\text{副標本法の標準誤差率} \approx 0.353 \times \text{忠実なりサンプリング技法による標準誤差率}$$

³ 当初、「総務省統計局(2018)」は、表題に「全国の主な項目の月別結果数値の標本誤差」との誤りがあり、これを基に分析を進めた。このため、図3では副標本法の標準誤差率をそのまま縦軸にプロットしている。

残差標準偏差=0.0028

という関係式が得られる。

仮に、標本の大きさが12倍となったために、1か月ごとの標準誤差率は、 $12^{1/2}=3.5$ 倍程度になると簡便のために考えれば、副標本法の各月の標準誤差率は、平均 $0.353 \times 12^{1/2}=1.22$ 倍程度大きくなっていると評価される。ただし、実際には12か月分のデータは、調査区のローテーションサンプリングのため、完全に独立とはいえず互いに正相関があるので、1.22よりは標準誤差率は小さくなっていると考えられる。古橋・岩永(1991)は、調査区のみをリサンプリングした場合の誤差評価が若干過小評価であるが、副標本法は、それに比して標準誤差率を若干過大評価していると指摘しており、本報告の評価と整合的ではある。いずれにせよ、副標本法はリサンプリング技法の観点からは、第一次抽出を抽出調査区数の1/8を8回だけリサンプリングしたものに相当し、その不安定性は無視できない。例えば図3によれば、製造業、卸小売業は、いずれも就業者数が約1000万人なのに、副標本法による標準誤差率は、約2倍の開きがある。一方本報告のリサンプリング技法では、両者の標準誤差率は殆ど変わらない。従って、今後本報告で示した単純なりサンプリング技法などによる精度評価を実装することには、実務的にも意味があるものと考えられる。

謝辞

本報告に当たっては、岩永琢磨氏から様々な助言を頂いた。また、第一次稿の不完全性に対して、査読者から丁寧な修正意見を提示いただいた。また、総務省統計局統計調査部労働力人口統計室には、著者らの照会に応じて、調査区ウェイトの提供、副標本法標準誤差率の意味などについて、種々ご尽力いただいた。ここに深甚の謝意を表したい。

参考文献

- [1] 総務省統計局(2018)労働力調査標本設計の解説, ISSN 2187-7882, <https://www.stat.go.jp/data/roudou/hyohon/pdf/hyohon.pdf>
- [2] 馬場康維・土屋隆裕・中村好宏・山崎伸彦(1996)ブートストラップ法による標準誤差推定の試み,第10回日本計算機統計学会大会, 68-71.
- [3] 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推定の試み(2),第11回日本計算機統計学会大会, 86-87.
- [4] 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,49号, 37-50.
- [5] Bickel, P. J. and Freedman D.A. (1984) Asymptotic Normality and the Bootstrap in Stratified Sample, *the Annals of Statistics*, Vol.12・No.2, pp.470-482.
- [6] Efron, B.(1979) Bootstrap Methods: Another Look at the Jackknife, *the Annals of Statistics*, Vol. 7・No.1, 1-26.
- [7] Funaoka, F., Saigo, H., Sitter, R. R. and Toida, T. (2006) Bernoulli Bootstrap for Stratified Multistage Sampling, *Survey Methodology*, Vol.32・No.2, 151-156.
- [8] Rao, J. N. K. and Wu, C. F. J. (1988) Resampling Inference with Complex Survey Data, *Journal of the American Statistical Association*, Vol.83・No. 401, 231-241.
- [9] Shao, J. (2003) Impact of the Bootstrap on Sample Surveys, *Statistical Science*, Vol.18・No.2, 191-198.

表1 調査区の種類とその基準（総務省統計局（2018）表2-3の抜粋）

層符号	層の特性
01	調査の範囲外・対象外区域（駐留軍、自衛隊、刑務所区域など）
02	人口0の調査区
03	換算世帯15以下の調査区
04	会社・官公庁などの単身者50人以上が居住する寮などがある調査区
05	漁業就業者比率が0.2以上の調査区
06	漁業就業者比率が0.1以上0.2未満の調査区
07	製造業・建設業の業主比率が0.1以上の調査区
08	卸売業、小売業、宿泊業、飲食サービス業の業主比率が0.1以上の調査区
09	情報通信業、運輸業、郵便業、金融・保険業、不動産業、物品賃貸業、学術研究、専門・技術サービス業、生活関連サービス業、娯楽業、教育、学習支援業、医療、福祉、複合サービス事業、サービス業の業主比率が0.1以上の調査区
10	農林業の就業者比率が0.3以上の調査区
11	農林業の就業者比率が0.1以上0.3未満の調査区
12	公務の就業者比率が0.1以上の調査区
13	金融・保険業、不動産業、物品賃貸業の雇用者比率が0.1以上の調査区
14	製造業の雇用者比率が0.3以上の調査区
15	建設業の雇用者比率が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	卸売業、小売業、宿泊業、飲食サービス業の雇用者比率が0.2以上の調査区
18	学術研究、専門・技術サービス業、生活関連サービス業、娯楽業、教育、学習支援業、複合サービス業、サービス業の雇用者比率が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、郵便業の雇用者比率が0.1以上の調査区
20	製造業の雇用者比率が0.2以上0.3未満の調査区
21	製造業の雇用者比率が0.1以上0.2未満の調査区
22	卸売業、小売業、宿泊業、飲食サービス業の雇用者比率が0.1以上0.2未満の調査区
23	学術研究、専門・技術サービス業、生活関連サービス業、娯楽業、教育、学習支援業、複合サービス事業、サービス業の雇用者比率が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 2017年9月における地域別標本調査区数（総務省統計局（2018）表2-4より抜粋）
：総計 2912 調査区

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
176	232	600	240	176	312	392	200	152	288	144

表3 産業別就業者数集計に関する推定値のブートストラップ分布
表上半分が単純なリサンプリング技法による推定値の分布、下半分が忠実なリサンプリング技法による推定値の分布

単純な リサンプリング	農林業	漁業	鉱業採石	建設	製造	電気ガス	情報通信	運輸郵便	卸小売	金融保険	不動産等	学術研究	宿泊飲食	生活関連 サービス	教育	医療福祉	複合 サービス	サービス (他に分類 されない)	公務
0.50%	1933709	101060	16712	4675259	10114007	189978	1947159	3145545	10389926	1566910	1093936	1965604	3535874	2180785	2798715	7641073	472210	4115241	1998006
2.50%	1999745	115683	22875	4736711	10222008	206325	2000206	3184691	10510154	1618015	1137217	2023415	3614430	2233396	2849721	7752778	489072	4154410	2036720
25%	2149382	159174	39082	4910327	10482063	243638	2130685	3317529	10751806	1727885	1224797	2143653	3771327	2356547	2983175	7935105	543585	4327134	2147395
50%	2221554	186709	47331	4994570	10613952	262433	2194094	3384705	10882982	1779719	1275741	2208238	3854760	2417812	3055518	8044633	572745	4410601	2216234
75%	2300133	211259	57349	5092233	10754334	283664	2257130	3457458	11016915	1843459	1328253	2273835	3938886	2480523	3122909	8144735	601628	4491564	2278686
97.50%	2444981	265317	78904	5258608	10968786	324883	2388045	3602123	11251938	1955155	1430883	2407566	4099984	2619682	3258343	8348424	664142	4647957	2401367
99.50%	2502895	287218	87736	5356597	11080936	352091	2467588	3667023	11368157	2013740	1477172	2471578	4172475	2672234	3316947	8425100	692458	4698659	2442905
平均	2223453	186433	48521	4998940	10612452	263775	2195368	3387150	10885979	1782856	1278361	2208802	3854600	2418424	3052951	8042060	572715	4406154	2215350
標準偏差 (推定値の標準誤差)	110764	38149	14071	133571	196411	29818	98995	104408	192291	85538	75524	98330	125615	96942	104615	156550	44102	123398	93083
標準誤差率	0.05	0.205	0.29	0.027	0.019	0.113	0.045	0.031	0.018	0.048	0.059	0.045	0.033	0.04	0.034	0.019	0.077	0.028	0.042
忠実な リサンプリング	農林業	漁業	鉱業採石	建設	製造	電気ガス	情報通信	運輸郵便	卸小売	金融保険	不動産等	学術研究	宿泊飲食	生活関連 サービス	教育	医療福祉	複合 サービス	サービス (他に分類 されない)	公務
0.50%	1921512	95806	17395	4685098	10170021	196772	1916414	3143870	10342194	1568538	1062675	1930632	3570402	2188333	2847522	7637705	479331	4035397	2022698
2.50%	1982319	109835	22106	4748473	10274447	212045	1978686	3221965	10495320	1611981	1119355	1996420	3620572	2244823	2896214	7743019	496873	4101881	2081544
25%	2130004	158237	35579	4913362	10495803	246639	2107978	3375725	10736045	1702556	1206835	2119468	3776342	2351973	3017272	7940954	550206	4255070	2196421
50%	2200110	183133	42941	5003722	10624982	266644	2175995	3447368	10854925	1760144	1256133	2186443	3857116	2418819	3086427	8050702	580913	4333333	2256914
75%	2272394	210043	51284	5092091	10761017	290855	2244750	3521259	10989494	1817095	1303723	2253630	3939377	2480883	3160017	8162430	611757	4420505	2321147
97.50%	2423190	267019	70339	5279470	11022389	332159	2372424	3656099	11225494	1933744	1395006	2384453	4096416	2606077	3307493	8359860	673518	4563087	2440589
99.50%	2494254	289790	78287	5342659	11126056	349804	2418155	3715058	11336553	1982598	1438782	2445721	4158284	2671045	3355873	8432397	695255	4622572	2485447
平均	2203489	185233	43887	5007547	10631895	268915	2174971	3446048	10860598	1761826	1254412	2187091	3856533	2418141	3090356	8049177	582000	4336289	2258747
標準偏差 (推定値の標準誤差)	109052	39990	12004	133557	195581	31227	101050	111390	186865	82920	71587	98746	119195	93999	103240	162773	44228	120036	89877
標準誤差率	0.049	0.216	0.274	0.027	0.018	0.116	0.046	0.032	0.017	0.047	0.057	0.045	0.031	0.039	0.033	0.02	0.076	0.028	0.04
平均値の差	-19964	-1200	-4634	8607	19443	5140	-20397	58898	-25381	-21030	-23950	-21711	1934	-283	37405	7117	9285	-69864	43397
t値	-5.51	-0.91	-11.45	1.94	3	4.98	-6.1	16.03	-4.08	-7.62	-10.04	-6.63	0.49	-0.09	10.91	1.32	6.33	-17.5	14.51
公表数値	2210000	190000	50000	5000000	10630000	260000	2210000	3400000	10920000	1820000	1290000	2210000	3860000	2420000	3070000	8070000	580000	4430000	2310000

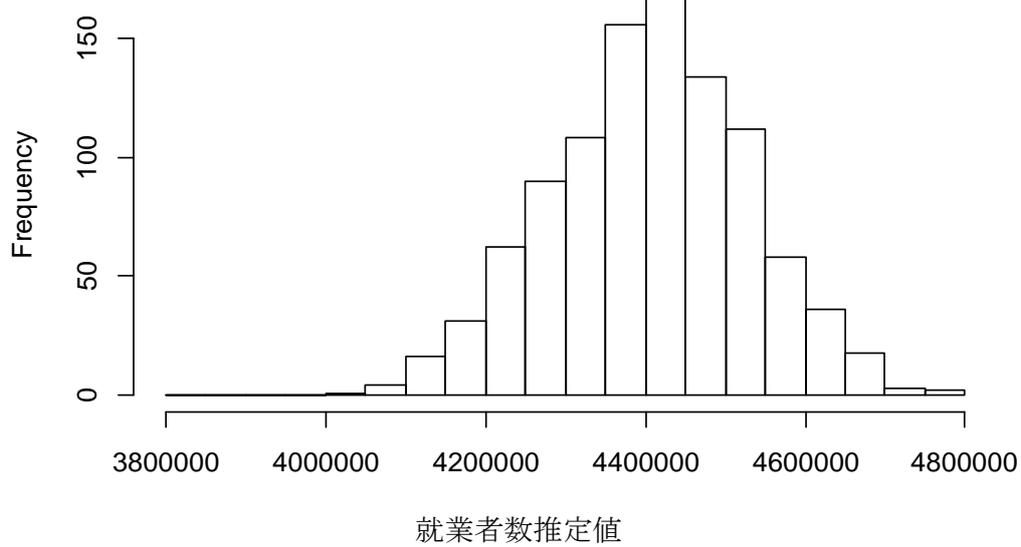
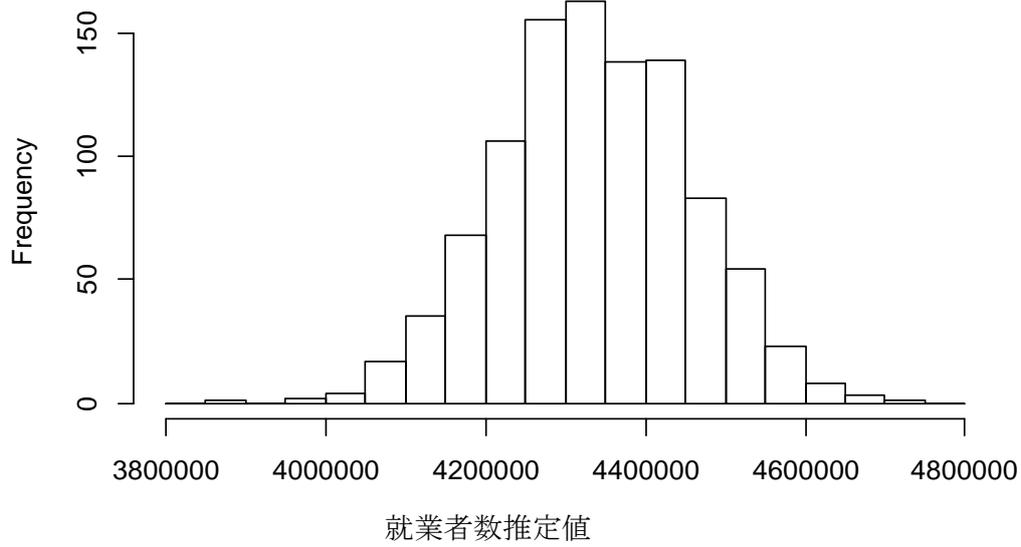


図1 他に分類されないサービス産業の就業者数推定値のブートストラップ分布
 上段：忠実なりサンプリング、 下段：単純なりサンプリング

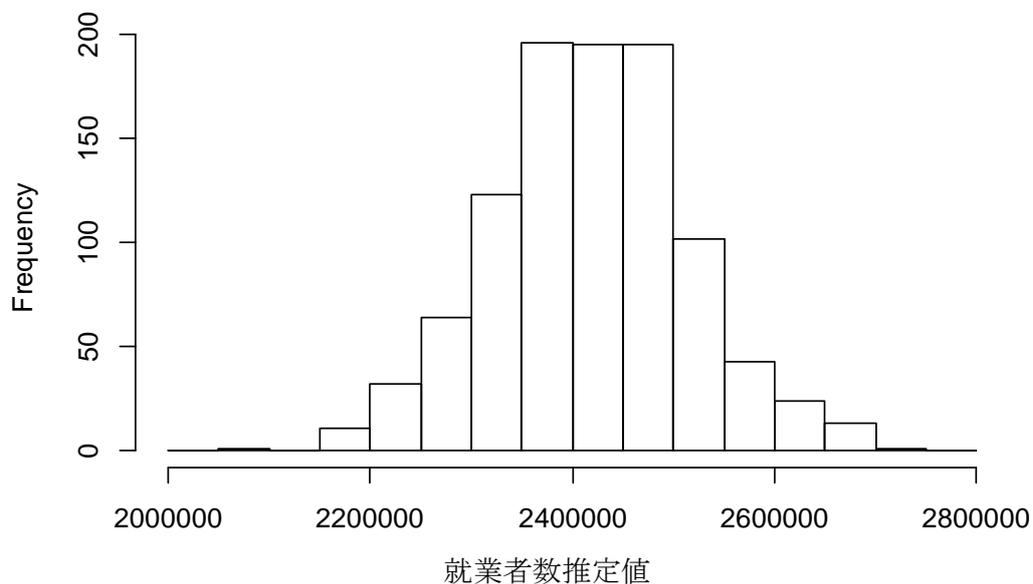
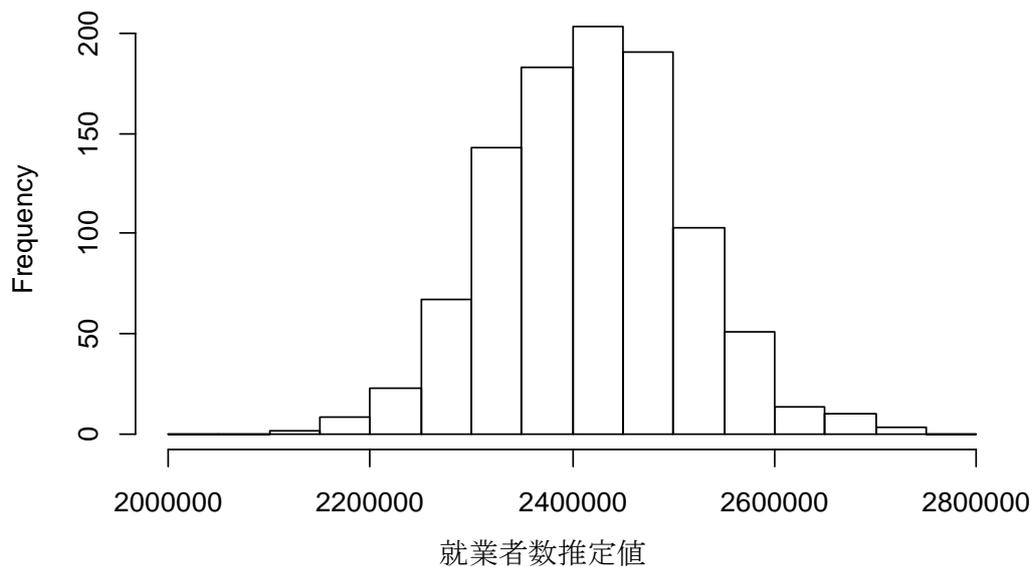


図2 生活関連サービス業の就業者数推定値のブートストラップ分布
 上段：忠実なりサンプリング、 下段：単純なりサンプリング

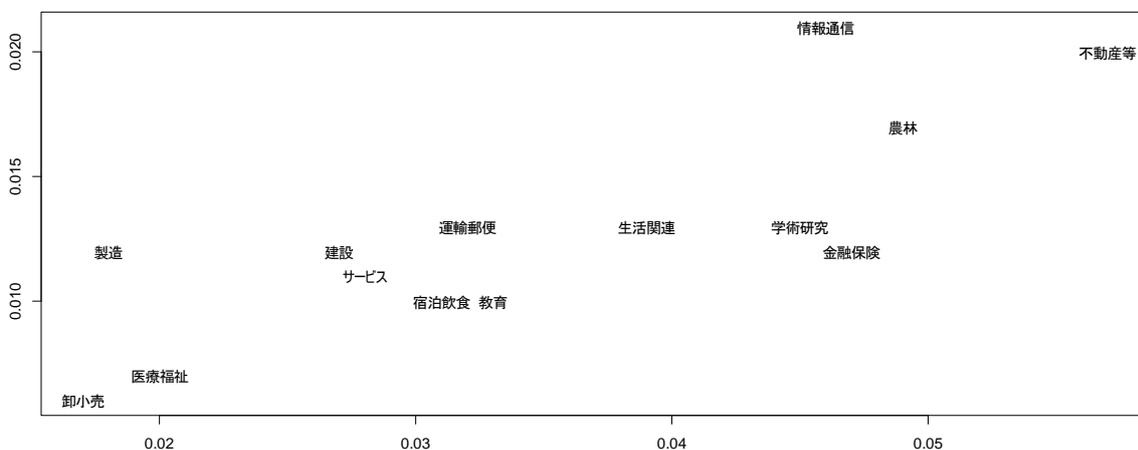


図3 忠実なりサンプリング技法と副標本法の標準誤差率の関係
 横軸：忠実なりサンプリング技法の標準誤差率、縦軸：副標本法の標準誤差率