

諸外国の公的統計における欠測値の対処法

— 集計値ベースと公開型マイクロデータの代入法 —

高橋将宜*

要旨

欠測値の処理は、公的統計の重要なプロセスの一部であるが、我が国においてその方法的意義はほとんど議論されていない。本稿は、UNECE (国連欧州経済委員会) における国際的動向を踏まえ、回帰代入法、比率代入法、平均値代入法、ホットデック法といった確定的単一代入法に加え、確率的単一代入法と多重代入法について調査し、我が国における政府統計マイクロデータの作成・提供に資するものである。公的統計における国際的な現状を把握するため、UNECE 参加国のサーベイ調査を行った。また、様々な状況を想定した一連のモンテカルロ実験を用いて、各種代入法の有用性を検証した。諸外国の公的統計における集計値ベースの代入法は、データの種類に応じて採用されており、経済データには比率代入法が、世帯データにはホットデック法が適していることを明らかにした。公開型マイクロデータでは、多重代入法を採用することが望ましいことも示した。

キーワード

多重代入法, 単一代入法, 不完全データ, 無回答, 政府統計

1. はじめに

社会調査では、平均して約半数もの回答者が1つ以上の質問項目に答えないとされる (King et al., 2001)。特に、個人の収入や企業の売上高といった機微な質問は、回答率が低くなる傾向がある (Schenker et al., 2006)。また、意図的ではなく、回答者が質問項目を見落したり、回答し忘れたりするなどのケースもある。あるいは、回答者が引越したり、企業が倒産したりすると、継続的な調査が行えず欠測となる場合もある (Allison, 2002; de

Waal et al., 2011)。

このように、調査データにおいてすべてのデータを回収することは非常に困難であり、欠測値の統計的処理は、公的統計機関における実務の基本的な過程の1つである。具体的には、公的統計における欠測値は、代入法 (imputation: 補完法, 補定法) によって処理されるが (de Waal et al., 2011, 第7章)、これまで我が国では研究対象としてその方法的意義が議論されることはほとんどなかった。一方、国際的には、公的統計における代入法に関する研究は1950年代まで遡ることができ (U.S. Bureau of the Census, 1957, p.XXIV)、数多くの研究蓄積が存在する。たとえば、政府

* 正会員, 東京外国語大学経営戦略情報本部
e-mail: mtakahashi@tufs.ac.jp

統計の個票データの作成という文脈において、国連欧州経済委員会 (UNECE: United Nations Economic Commission for Europe) の統計データエディティングに関するワークショップ (Work Session on Statistical Data Editing) といった国際会議で盛んに議論されている。

本稿は、「政府統計マイクロデータの作成・提供における方法的展望」の特集企画の一部として、UNECEにおける国際的動向を踏まえ、我が国における政府統計マイクロデータの作成・提供に関して、欠測値の対処法という観点から方法的展望を追究する。本稿の前半では、諸外国においてどのような手法が用いられているかを調査し、従来の集計値ベースの代入法について、経済調査や世帯調査といった調査の種類ごとの特徴を調べて検証する。

また、坂田 (2006, p.31) が指摘するように、マクロ集計値による分析が主流であった20世紀半ばとは異なり、21世紀初頭の今日では、マイクロレベルの個体行動に関する実証分析が重要となっている。このような状況において、公的統計によって収集された調査データを公開型マイクロデータとして一般に利用可能とする需要が高まっている。供給側についても、2014年3月に閣議決定された第II期「公的統計の整備に関する基本的な計画」において公的統計の二次的利用の促進が言及されており、2016年度中に試行的なオンサイト利用を開始する予定となっている (中村, 平澤, 2016, pp.36-37)。我が国も公開型マイクロデータへの道を歩み始めたところである。よって、本稿の後半では、マイクロデータ公開を目的とした場合の代入法は、現行の集計値ベースの代入法とどのように異なる必要があるか、将来的な課題を議論する。

なお、本稿の議論は、公開型マイクロデータにおける欠測値の取り扱いに特化しており、秘匿性の確保は十分に担保されているという前提に立って議論をしている¹。匿名データの

秘匿性と有用性に関する詳しい議論は、伊藤, 星野 (2014) を参照されたい。

本稿第2章では、代入法の考え方と現状を示す。具体的には、欠測データの問題点を例証し、集計値ベースの欠測値処理方法として、確定的単一代入法 (deterministic single imputation) を導入する。また、UNECE参加国へのサーベイ調査の結果を提示し、欠測値処理に関する国際的動向を確認する。第3章では、経済データと世帯データを模したモンテカルロシミュレーションによって集計値ベースの代入法の検証を行う。第4章では、公開型マイクロデータにおける欠測値対処法の展望を議論し、公開型マイクロデータを模したモンテカルロシミュレーションによって確定的単一代入法、確率的単一代入法 (stochastic single imputation)、多重代入法 (multiple imputation) の検証を行う。第5章では、代入法とマイクロ分析の問題として、適合性を議論する。第6章において締めくくりとする。

2. 代入法の考え方と現状

2.1 欠測データの問題点

表1は4名の収入と年齢のシミュレーションデータである。黒数字は観測された値、灰色セルの白抜き数字は欠測値 (欠損値) の本来あるべき真値とする。また、表2では年齢が質的データとして、表3では収入が質的データとして記録されている。表1と表2の集計対象は収入の平均値であり、表3の集計対象は収入の最頻値だとする。

表1において、もしデータがすべて観測されているならば、4人の収入の平均値は(1)式のとおり453.75万円として簡単に計算できる。

$$\begin{aligned} \overline{\text{収入}}_{\text{真値}} &= \frac{1}{4} \sum_{i=1}^4 \text{収入}_i \\ &= \frac{239 + 421 + 505 + 650}{4} \quad (1) \\ &= 453.75 \end{aligned}$$

表1 量的データ

ID	収入	年齢
1	239	26
2	421	38
3	505	47
4	650	54

表2 量的・質的データ

ID	収入	年齢
1	239	1
2	421	1
3	505	2
4	650	2

表3 質的・量的データ

ID	収入	年齢
1	1	26
2	2	38
3	3	47
4	3	54

注：収入の単位は万円，年齢の単位は年である。表2の年齢では，1=40歳未満，2=40歳以上である。表3の収入では，1=0円以上250万円未満，2=250万円以上500万円未満，3=500万円以上である。表2と表3は，2.2節以降にて用いる。

一方，表1において，データが1つでも欠測している場合，(2)式のとおり平均値を計算することができない。平均値を計算できないということは，標準偏差，相関係数，回帰係数，標準誤差などの統計分析が不可能であることを意味している。つまり，欠測データの第1の問題点は，そのままでは統計分析ができないという問題である。

$$\begin{aligned}\overline{\text{収入}}_{\text{欠測}} &= \frac{1}{4} \sum_{i=1}^4 \text{収入}_i \\ &= \frac{239 + 421 + 505 + \text{収入}_4}{4} \quad (2) \\ &= \frac{1155 + \text{収入}_4}{4} = ?\end{aligned}$$

そこで，SAS，SPSS，STATAなど多くの統計ソフトウェアにおけるデフォルト設定では，1箇所でもセルの欠測している行を除去し，見かけ上の「完全データ」を構築することにより統計データ分析を可能としている。これをリストワイズ除去（完全ケース分析，ケースワイズ除去）と呼ぶ（Baraldi & Enders, 2010, p.10）。すなわち，表1において，ID4は存在していないことにし，(3)式として収入の平均値が388.33万円と求められる。しかし，真の平均値は453.75万円であり，欠測データに偏りがあるため過少推定していることが示唆される。また，年齢₄=54歳という貴重な情報を活用せずに捨て去ってしまっている。欠測データの第2の問題点は，偏りと効率性の問題であるⁱⁱ。

$$\begin{aligned}\overline{\text{収入}}_{\text{リストワイズ}} &= \frac{1}{3} \sum_{i=1}^3 \text{収入}_i \\ &= \frac{239 + 421 + 505}{3} \quad (3) \\ &= 388.33\end{aligned}$$

欠測データのメカニズムについて，Little & Rubin (2002) は，欠測が完全にランダムなMCAR (Missing Completely At Random)，欠測が条件付でランダムなMAR (Missing At Random)，欠測がランダムではないNMAR (Not Missing At Random) という分類を考案している。NMARは，MNAR (Missing Not At Random)とも表記されることがあるが，同じ概念である。

欠測がMCARの場合，欠測データは母集団のサブサンプルであり，偏りは発生しないが，効率性は下がる。欠測がMARの場合，欠測データは偏っている。Allison (2002, p.5) が指摘するとおり，MCARとMARでは欠測メカニズムのパラメータを無視することができる (Ignorable)。結果，代入法によって欠測データの偏りを是正できる。一方，欠測がNMARの場合，欠測メカニズムを無視できない (Non-Ignorable) ため，選択モデルやパターン混合モデルを用いて欠測のメカニズムに関する非常に強い仮定を置いた上で分析を行う (Allison, 2002, ch.7; Enders, 2010, ch.10)。後述するとおり，これらの手法は感度分析 (sensitivity analysis)ⁱⁱⁱとして有用である。

真の欠測メカニズムは不明であることが多いが，計画的な欠測データデザイン (Enders,

2010)によって欠測のメカニズムが明らかになる場合もある。たとえば、公的経済統計では、一般的に大企業の売上高は照会や督促によって必ず実測値を収集し、中小企業のデータのみ統計的に処理することが通例である(de Waal et al., 2011, pp.245-246)。この場合、売上高の欠測率は従業者数など企業の規模に応じて変化しており、MARだと考えられる。なお、Scheuren(2005)によれば、公的統計の欠測値におけるMCARの割合は約10%~20%、MARの割合は約50%、NMARの割合は約10%~20%とのことである。

2.2 確定的単一代入法

従来の公的統計では、調査データの合計値(平均値)を集計することを主目的とし、分布や分散に関する分析を行うことはまれである(de Waal et al., 2011, p.225)。そこで、平均値の点推定に関して不偏となる確定的な単一代入法を用いることが通例となっている。確定的な単一代入法とは、代入モデルから得られた予測値を欠測値の代替値として使用し、誤差項を加えたり多重化したりしない手法のことである。

実際のデータエディティングでは、電話や郵送による照会、信頼性の高い外部データによるコールドデッキ、論理的な処理などを行った後、それでも埋めることができなかった欠測値を統計的に処理する。中でも、回帰代入法、比率代入法、平均値代入法、ホットデッキ法がよく用いられるとされている(Hu et al., 2001; de Waal et al., 2011, ch.7)。本章では、これら4つの統計的欠測値処理手法のメカニズムについて簡単に紹介する。

2.2.1 回帰代入法

回帰代入法(regression imputation)では、観測データを用いて、最小二乗法によって(4)式のパラメータ β_0 と β_1 の推定を行う(高橋, 阿部, 野呂, 2015, pp.11-14)。なお、観測

データとは、リストワイズ除去済みデータのことを意味している。表1のデータでは、 $\beta_0 = -85.33$, $\beta_1 = 12.80$ と推定される。ID4の年齢の値は54であるので、(5)式のとおりID4の収入の値は605.87万円と推定できる。この値を(2)式の収入 $_i$ の値として用いれば、収入の平均値は442.72万円と集計される。リストワイズ除去と比べて集計値が改善している様子が示唆されている。

$$\widehat{\text{収入}}_i = \beta_0 + \beta_1 \text{年齢}_i \quad (4)$$

$$\text{収入}_4 = -85.33 + 12.80 \times 54 = 605.87 \quad (5)$$

2.2.2 比率代入法

比率代入法(ratio imputation)では、観測データを用いて、平均値の比率によって(6)式のパラメータ β_1 の推定を行う(高橋, 阿部, 野呂, 2015, pp.18-22)。表1のデータでは、観測データ部分における収入の平均値は388.33万円であり、観測データ部分における年齢の平均値は37歳である。これらは、リストワイズ除去による値である。よって、 $\beta_1 = 388.33/37 = 10.50$ と推定される。ID4の年齢の値は54であるので、(7)式のとおりID4の収入の値は567.00万円と推定できる。この値を(2)式の収入 $_i$ の値として用いれば、収入の平均値は433.00万円と集計される。リストワイズ除去と比べて集計値が改善している様子が示唆されている。

$$\widehat{\text{収入}}_i = \beta_1 \text{年齢}_i \quad (6)$$

$$\text{収入}_4 = 10.50 \times 54 = 567.00 \quad (7)$$

2.2.3 平均値代入法

平均値代入法(mean imputation)は、観測データの平均値を代入値として使用するものである。一般的に、平均値代入法は、極めて特殊な状況以外では有用性がない(高橋, 伊藤, 2013, pp.27-28; 高井, 星野, 野間, 2016, p.6)。しかし、表2のように年齢の値が数量

項目ではなく、質的なカテゴリーとして記録されていたとしよう。こういった場合、年齢のグループごとに平均値を求め、その値を代入値として採用するグループ平均値代入法 (group mean imputation) が用いられることがある (de Waal et al., 2011, pp.246-249)。年齢で層分けをすると、ID1とID2のグループ1とID3とID4のグループ2に分けることができる。ID4の収入の値を推定するために、グループ2の平均値505万円を用いる。したがって、この値を(2)式の収入₄の値として用いれば、収入の平均値は417.5万円と集計される。単純な平均値代入法とは異なり、グループ平均値代入法では、リストワイズ除去と比べて集計値が改善している様子が示唆されている。

2.2.4 ホットデッキ法

表3のように、年齢は数量項目だが、収入が質的なデータだとしよう。このように集計すべき項目が質的なデータである場合には、共変量の値が似ているデータをドナーとして選び、そのドナーの値を代入値とするホットデッキ法 (hot deck imputation) を用いる。ID4の年齢の値は54であり、ID3の年齢47に近いので、ID3がID4のドナーである。ID3の収入の値である3をID4の収入の代入値として採用する。今回の場合は、収入の最頻値は3であり、完全データの最頻値と一致することが見て取れる。

実際のデータにおいて適切なドナーを探すには、距離関数を定義して最近隣法 (nearest neighbor) を用いることが多い。この手法は本質的にはマッチングと同じである。ホットデッキ法とマッチングの詳細については、阿部 (2016, pp.57-59)、高井、星野、野間 (2016, pp.110-113)、栗原 (2015) も参考にされたい。具体的には、RパッケージHotDeckImputationによって実行できる (Joenssen, 2015)。なお、ホットデッキ法は、すべての項目が質的デー

タの場合にも使用できるノンパラメトリックな手法である。

2.3 現状の国際的動向：UNECE加盟国の調査結果

統計データエディティングに関するワークショップは、UNECE (国連欧州経済委員会) により1年半の周期で定期的に開催され、欧州を中心に米国、カナダ、オセアニアなどの各国統計機関が参集し、公的統計における欠測値やエラーの処理に関して意見交換を行う国際会議である^{iv}。UNECE参加国の中で、過去3回のいずれかの会合において研究報告を行った国を対象として下記の要領で調査を行った。

調査対象の母集団：23の国家統計機関

調査実施時期：2016年7月～9月

調査方法：データエディティングの専門職員に対して、メールにて調査票を送付
回収率：87.0% (2016年9月6日現在)

以下は、調査協力を得た20国家機関である：イタリア国家統計局、英国国家統計局、エストニア統計局、オーストラリア統計局、オーストリア統計局、オランダ統計局、カナダ統計局、スイス連邦統計局、スウェーデン統計局、スペイン統計局、スロバキア統計局、スロベニア統計局、デンマーク統計局、ドイツ連邦統計局、ニュージーランド統計局、ノルウェー統計局、フィンランド統計局、フランス国立統計経済研究所、米国センサス局、リトアニア統計局。いずれも国際的に公的統計をリードしている国家機関である。調査の結果は表4にまとめたとおりである。

問1では、回答の得られた20機関の実務において、4種類の代入法のほぼすべてが導入されていることが確認され、予想以上に平均値代入法が利用されていることが示された。問2では、比率代入法 (60.0%) とホットデッキ法 (65.0%) が重視されていることが分

表4 UNECEの調査結果(重複回答あり)

	回帰代入法	比率代入法	平均値代入法	ホットデック法
問1	95.0%	95.0%	95.0%	100.0%
問2	40.0%	60.0%	35.0%	65.0%
問3	30.0%	80.0%	35.0%	30.0%
問4	10.0%	10.0%	25.0%	80.0%

問1：貴機関の実務では、4つの手法のどれを用いていますか？

問2：一般的に、貴機関の実務では、4つの手法のどれがよく用いられていますか？

問3：事業所・企業を単位とする経済データにおいて、貴機関の実務では、4つの手法のどれがよく用いられていますか？

問4：世帯データにおいて、貴機関の実務では、4つの手法のどれがよく用いられていますか？

表5 UNECEの調査結果(重複回答あり)

	確率単一代入法	多重代入法	フラクショナル
問5	70.0%	40.0%	5.0%

問5：貴機関の実務では、確率的単一代入法、多重代入法、フラクショナル代入法のいずれかを用いていますか？ その場合、どの手法ですか？

かった。問3では、経済データにおいて比率代入法(80.0%)がよく用いられることが示され、回帰代入法はあまり利用されていないことも浮き彫りとなった^v。問4では、世帯データにおいてホットデック法(80.0%)がよく用いられていることが明らかとなり、世帯データにおける数量項目はグループ平均値代入法(25.0%)によって処理される場合があることも分かった。

表5に記したとおり、問5では、現行の集計値ベースの代入法として、20機関のうち14機関において確率的単一代入法(70.0%)が導入されており、8機関において多重代入法(40.0%)が導入されており、1機関においてのみフラクショナル代入法(5.0%)が導入されていることが判明した。フラクショナル代入法^{vi}については本稿では扱わないが、de Waal et al. (2011, pp.271-272)を参照されたい。

3. 集計値ベースの代入法

2.3節で調査したとおり、諸外国の公的統計機関では、平均値代入法、比率代入法、回帰代入法、ホットデック法の4種類すべてが利用されている。これら4つの手法について、

以下の3つのケース^{vii}を想定して、モンテカルロシミュレーションにより有用性を実験する。

- (1) 経済データ：対数正規分布の量的データ
- (2) 質的経済データ：対数正規分布の量的データと質的な共変量のデータ
- (3) 世帯データ：質的な集計項目と量的な共変量のデータ

モンテカルロシミュレーションとは、乱数を用いて繰り返し抽出を行う分析方法である。観察データをもとにして確率分布を仮定し、確率分布に従う確率変数を定量的に分析するために、コンピュータによって擬似乱数を生成して分析する(大野, 井川, 2015)。すなわち、モンテカルロシミュレーションとは、コンピュータを実験ラボとして使用する方法である。ラボと同様に、実験を完全にコントロールでき、ラボの環境設定を様々に変化させた結果を観測することで、効果を測定する(Carsey & Harden, 2014)。具体的には、以下の5つの手順により実行するものである(Mooney, 1997)。なお、本稿のすべての分析

は、R 3.2.4を用いて実行した。

- (1) コンピュータにおいて擬似母集団を定義する。
- (2) 擬似母集団から標本を抽出する。
- (3) パラメータ推定値を計算する。
- (4) 上記の2と3を繰り返す(1,000回ほど)。
- (5) パラメータ推定値の相対頻度を集計する。

実験の評価は、(8)式の平均平方誤差 (Mean Squared Error: MSE) を用いて行う。推定値 $\hat{\theta}$ の MSE は、真値 θ のベクトルを生成し、 $\hat{\theta}$ のベクトルとの差を取り、差の二乗和をシミュレーション回数で割ることにより計算できる (Mooney, 1997; Carsey & Harden, 2014)。MSE が小さな値の手法ほど、相対的によい手法だといえる。実際には、Di Zio & Guarnera (2013, p.549) にならい、真値で正規化して平方根を取った(9)式の RRMSE (Relative Root Mean Squared Error) を用いた。

$$MSE = E[(\hat{\theta} - \theta)^2] \quad (8)$$

$$RRMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(\frac{\hat{\theta} - \theta}{\theta} \right)^2} \quad (9)$$

シミュレーションの設定は、以下のとおりである。母集団モデルを(10)式とし、 y_i の平均値の推定を目的とする。モンテカルロシミュレーションの繰り返し回数 T は1,000回とし、各々のシミュレーションにおいて $n=1000$ の標本データを生成した。 y_i の欠測は、第2章で言及した計画的な欠測データデザイン (Enders, 2010) を模した。具体的には、 $u_i \sim U(0, 1)$ とし、 $\text{med}(x_{1i})$ を x_{1i} の中央値とする。 $x_{1i} < \text{med}(x_{1i})$ かつ $u_i < 0.6$ の場合、 y_i の値が欠

測することで、 x_{1i} を条件とする MAR として生成し、欠測率は約30%に設定した。Schenker et al. (2006, p.925) によると、1997年から2004年までの National Health Interview Survey における収入と所得の欠測率はいずれも平均して約30%であり、この設定は現実的な数字である。また、誤差項 ε_i の分散は、 x_{1i} に比例して増大するものとし、分散は不均一である。 β_1 の値は $U(1.1, 2.0)$ からの無作為抽出によって設定し、 σ の値は $U(1.0, 2.0)$ からの無作為抽出によって設定した。値を変化させた他のシミュレーション結果においても、本稿の結果とはほぼ同じ内容の結果が得られている。 $LN(\cdot)$ は R 関数 `rlnorm`、 $N(\cdot)$ は R 関数 `rnorm`、 $U(\cdot)$ は R 関数 `runif` によってそれぞれ生成した。

$$y_i = \beta_1 x_{1i} + \varepsilon_i$$

ここで、

$$x_{1i} \sim LN(\logmean = 0, logsd = 1) \quad (10)$$

$$\varepsilon_i \sim N(\text{mean} = 0, \text{sd} = \sigma \sqrt{x_{1i}})$$

表6は、経済データの欠測値処理を模したもので、データのイメージは表1と同じである。分散が不均一な対数正規分布のデータにおける平均値の推定では、リストワイズ除去 (RRMSE = 0.302) と比べていずれの代入法も改善しているが、回帰代入法 (RRMSE = 0.050) やホットデック (RRMSE = 0.050) と比べて、比率代入法 (RRMSE = 0.048) によるパフォーマンスが最もよい。Cochran (1977, p.158) および Takahashi et al. (2017) に示されているとおり、比率代入法は $\varepsilon_i \sim N(0, \sigma \sqrt{x_{1i}})$ という不均一分散の場合に最良線形不偏推定量になるためである。

表7は、質的項目を含む経済データを模したもので、データのイメージは表2と同じで

表6 経済データの欠測値処理の RRMSE

	完全データ	リストワイズ	回帰代入法	比率代入法	ホットデック
経済データ	0.047	0.302	0.050	0.048	0.050

表7 質的経済データの欠測値処理のRRMSE

	完全データ	リストワイズ	平均値代入法
経済データ2	0.043	0.081	0.055

表8 世帯データの欠測値処理のRRMSE

	完全データ	リストワイズ	回帰代入法	比率代入法	ホットデック
世帯データ	0.038	0.123	0.381	0.381	0.056

ある。式(10)の x_{1i} の平均値を 0 と 1 の 2 つのグループに分けてデータを生成し、 x_{1i} を二値のデータに変換してデータを 1 つに統合し、 x_{1i} を所属グループとして、平均値と欠測の確率を変化させた。その他の設定は、表 6 の経済データと同じである。もし質的な共変量しか利用できない場合、グループ平均値代入法 (RRMSE=0.055) はリストワイズ除去 (RRMSE=0.081) よりも高い精度を示している。

表 8 は、世帯データの欠測値処理を模したもので、データのイメージは表 3 と同じである。 y_i の値を 3 つの順序のないカテゴリーに変換した。なお、 x_{1i} は数量項目のままである。 y_i の最頻値に分類される値の割合を推定することを目的とする。その他の設定は、表 6 と同じである。集計すべきデータが質的な変数の場合、ホットデック法 (RRMSE=0.056) のパフォーマンスが最もよく、回帰代入法 (RRMSE=0.381) と比率代入法 (RRMSE=0.381) は、このような状況において役に立たない。

4. 公開型マイクロデータにおける代入法の展望

ここまで、合計値(平均値)を集計することを前提とした欠測値の対処法について見てきた。2.3節で見たとおり、諸外国の公的統計では、回帰代入法、比率代入法、グループ平均値代入法、ホットデック法が用いられている。3章で検証したとおり、データの特성에応じて正しく使い分けられている。これらの手法

は、いずれも確定的単一代入法として知られているものである。その長所は、平均値(合計値)の点推定に関しては不偏であることだが、短所として、分布や分散の推定が不正確となることが挙げられる(阿部, 2016, p.55)。公開型マイクロデータを用いた分析における推測対象(estimand)は、平均値や合計値の算出だけとは限らない。

4.1 公開型マイクロデータと多重代入法

欠測データにおいて、平均値だけではなく、分散や標準誤差を妥当なものとするためには、多重代入法を用いる必要がある(Schafer & Graham, 2002; Donders et al., 2006; Baraldi & Enders, 2010; Cheema, 2014)。多重代入法は、複数の代入値を欠測データの分布から無作為抽出するものである。しかし、欠測データは観測されないため、欠測データの分布自体も観測できない。そこで、観測データからベイズ手法によって欠測値の事後分布を推定し、その事後分布から平均値と分散共分散の無作為抽出を行うことにより、代入モデルのパラメータ推定を考慮に入れた代入法を実行する(King et al., 2001)。

多重代入済みデータの具体例は、表 9 のとおりである。収入の空欄は欠測値であり、収入 1、収入 2、収入 3 の灰色セルの白抜き数字は、それぞれ 1 回目、2 回目、3 回目の多重代入を行った際の代入値である。収入 1 の平均値は 388.25 万円、収入 2 の平均値は 439.75 万円、収入 3 の平均値は 457.25 万円と

表9 多重代入済みデータの例 ($M=3$)

ID	収入	年齢	収入 1	収入 2	収入 3
1	239	26	239	239	239
2	421	38	421	421	421
3	505	47	505	505	505
4		54	388	594	664

集計される。収入の平均値の点推定は、428.42万円として集計される（3つの平均値の平均）。2.2節における単一代入法では、代入値を確定的に1つの値として算出し、推定を確実視していた。しかし、表9の例では、多重代入を行うごとに代入値が大きく変化しており、推定の不確実性が示されている。これにより、標準誤差を妥当なものとする事ができる。

多重代入法のメカニズムに関して、日本語による詳細な解説は、岩崎（2002，第10章）、高橋，伊藤（2014）、高橋，阿部，野呂（2015）、

阿部（2016，第5章）を参照されたい。

Rubin(1987)は、データ提供者によって M 個 ($M>1$) の多重代入済みデータが公開されれば、分析者の統計リテラシーに関わらず、分析者は欠測を気にすることなく様々な統計分析を行うことができると提案しており、多重代入法は公開型マイクロデータに適していることが示唆されている。付録に示すコードを貼り付けて使用することで、多重代入済みデータの統合といった実務上のわずらわしさに悩まされることなく分析を行うことができる。

たとえば、米国政府によって公開されているマイクロデータにおいて多重代入法を適用した例は、表10のとおりである。

2.3節のUNECE調査において問1～問5に回答した20機関に追加で調査票を送ったところ、18機関から回答を得て、回収率は90.0%であった（2016年9月6日現在）。結果は表11のとおりである。

表10 多重代入法による公開型マイクロデータの実例（米国政府）

調査実施者	調査名	対象データ	多重代入の数	公開日
Centers for Disease Control and Prevention ¹	2015 National Health Interview Survey	収入，所得	$M=5$	2016/6/30
Federal Reserve System ²	2013 Survey of Consumer Finances	ほぼすべての欠測変数	$M=5$	2014/9/25
Department of Transportation ³	2014 Fatality Analysis Reporting System	血中アルコール濃度	$M=10$	2015/12/1
Bureau of Labor Statistics ⁴	2014 Consumer Expenditure Survey	収入	$M=5$	2015/9/3

¹ http://www.cdc.gov/nchs/nhis/nhis_2015_data_release.htm

² <http://www.federalreserve.gov/econresdata/scf/scfindex.htm>

³ <http://www.nber.org/data/fars.html>

⁴ <http://www.bls.gov/cex/csxmicrodoc.htm>

表11 UNECEの調査結果（重複回答あり）

	不完全	確定単一	確率単一	多重代入	該当なし
問6	22.0%	50.0%	61.1%	44.4%	22.2%

問6：仮に、調査データを「公開型マイクロデータ」としてオープンにするとしたら、いずれの手法を用いるべきですか？

問6の自由記述では、Rubin(1987)が提唱しているとおり、統計リテラシーに関わらずすべての国民が欠測を気にすることなく同じ分析を行えるように、公開型マイクロデータは代入済みデータであることが望ましいという意見があった。しかし、具体的に代入済み公開型マイクロデータの作成に際して、確定的単一代入法(50.0%)、確率的単一代入法(61.1%)、多重代入法(44.4%)のいずれを用いるべきか意見が分かれており、コンセンサスが形成されていないことが示された。なお、欠測値にフラグを立てるべきであるという点については一致した見解が見られた。そうすることで、統計学上級者は、自らの望む欠測値対処法によって公開型マイクロデータの分析を実行できるからである。

4.2 公開型マイクロデータのための代入法

4.1節で見たとおり、諸外国の公的統計機関では、公開型マイクロデータの欠測値処理手法として、リストワイズ除去、確定的単一代入法、確率的単一代入法、多重代入法の4種類が提案されている。これら4つの手法を用いた平均値と回帰分析の精度について、モンテカルロシミュレーションによって評価を行う。確率的単一代入法とは、確定的単一代入法に誤差項を追加し、データのばらつき具合を調整する欠測値処理手法である(高橋、阿部、野呂, 2015, pp.15-18)。なお、多重代入法の繰り返し回数 M は5に設定した。また、多重代入法は、RパッケージAmelia IIによるものである(Honaker et al., 2011)。

4.2.1 説明変数が欠測している場合の回帰分析

シミュレーションの設定は、以下のとおりである(詳しい設定は3章も参照されたい)。母集団モデルを(1)式とし、 x_{1i} の平均値の推定および β_1 の推定を目的とする。モンテカルロシミュレーションの繰り返し回数 T は1,000回とし、各々のシミュレーションにおいて $n=1000$ の標本データを生成した。 x_{1i} の欠測は、 y_i を条件とするMARとして生成し、欠測率は約30%に設定した。具体的には、3章と同様に、 $y_i < \text{med}(y_i)$ の場合、 $\Pr(x_{1i} = \text{missing}) = 0.6$ である。対数正規分布の経済データを分析するために、データを自然対数に変換したことを想定し、データは正規分布で生成した。 β_1 の値は $U(1.1, 2.0)$ からの無作為抽出によって設定し、 σ の値は $U(1.0, 2.0)$ からの無作為抽出によって設定した。値を変化させた他のシミュレーション結果においても、本稿の結果とほぼ同じ内容の結果が得られている。

$$y_i = \beta_1 x_{1i} + \varepsilon_i$$

ここで、

$$x_{1i} \sim N(\text{mean} = 0, \text{sd} = 1)$$

$$\varepsilon_i \sim N(\text{mean} = 0, \text{sd} = \sigma)$$
(11)

表12は、説明変数が欠測している場合の \bar{x}_1 に関するRMSE、回帰係数 β_1 に関するRRMSE、名目95%信頼区間のカバー率を示している。

\bar{x}_1 のRMSEについて、単一代入法と多重代入法は不偏であるのに対して、リストワイズ除去には偏りがある。確定単一代入法(RMSE = 0.085)、多重代入法(RMSE = 0.087)、確率

表12 説明変数が欠測している場合の \bar{x}_1 と β_1 の推定

	完全データ	リストワイズ	確定単一	確率単一	多重代入法
RMSE(\bar{x}_1)	0.076	0.618	0.085	0.090	0.087
RRMSE(β_1)	0.026	0.062	0.139	0.031	0.030
95%CIカバー率	94.9	61.8	0.1	90.5	94.7

注：真の \bar{x}_1 が0であるため、RRMSEではなくRMSEを用いた。CIは信頼区間である。95%CIカバー率は、1,000回のモンテカルロ実験のうち、95%信頼区間に真の β_1 が含まれた割合である。

単一代入法 (RMSE = 0.090) の精度はほぼ同じだが、リストワイズ (RMSE = 0.618) の精度は非常に悪い。

β_1 のRRMSEについて検討する。多重代入法 (RRMSE = 0.030) の精度が最もよく、ついで確率単一代入法 (RRMSE = 0.031), リストワイズ (RRMSE = 0.062) となっており、確定単一代入法 (RRMSE = 0.139) の精度は最も悪い (Allison, 2002, p.53; Carpenter & Kenward, 2013, p.28)。

β_1 の名目95%信頼区間について検討する。多重代入法による信頼区間は、94.7%の確率で真のパラメータを捕らえることができおり、非常に正確である。確率単一代入法による信頼区間は、90.5%の確率で真のパラメータを捕らえている。これは、名目5%の第一種の過誤率が約2倍の9.5%になっているということであり、深刻な問題である (Enders, 2010, pp.53-54)。リストワイズによる信頼区間は、61.8%の確率で真のパラメータを捕らえている。これは、名目5%の第一種の過誤率が8倍近い38.2%になっているということであり、非常に深刻な問題である。確定単一代入法による信頼区間は、0.1%の確率で真のパラメータを捕らえている。これは、名目5%の第一種の過誤率が20倍近い99.9%になっているということであり、極めて深刻な問題である。

説明変数が欠測している回帰分析と基本統計量を研究の目的とする場合、多重代入法を実行するべきであることが分かる。

4.2.2 被説明変数が欠測している場合の回帰分析

シミュレーションの設定は、以下のとおりである。母集団モデルを(11)式とし、 y_i の平均値および β_1 の推定を目的とする。 y_i の欠測は、 x_{1i} を条件とするMARとして生成し、欠測率は約30%に設定した。その他の設定は4.2.1項に準じている。

表13は、被説明変数が欠測している場合の \bar{y} に関するRMSE、回帰係数 β_1 に関するRRMSE、名目95%信頼区間のカバー確率を示している。

\bar{y} のRMSEについて、単一代入法と多重代入法は不偏であるのに対して、リストワイズ除去には偏りがある。確定単一代入法 (RMSE = 0.073), 多重代入法 (RMSE = 0.074), 確率単一代入法 (RMSE = 0.075) の精度はほぼ同じだが、リストワイズ (RMSE = 0.609) の精度は非常に悪い。

β_1 のRRMSEについて検討する。リストワイズ (RRMSE = 0.027), 確定単一代入法 (RRMSE = 0.027), 多重代入法 (RRMSE = 0.028), 確率単一代入法 (RRMSE = 0.029) のいずれの値もほぼ同じである。被説明変数が欠測しており、回帰分析が目的の場合、代入法を行ってもリストワイズ除去と変化がないことが分かる。これは、被説明変数の欠測がMARの場合、不完全なケースは回帰係数の推定に寄与していないためである (Little, 1992; Carpenter & Kenward, 2013, pp.24-28; Raghunathan, 2016, p.99)。

β_1 の名目95%の信頼区間について検討す

表13 被説明変数が欠測している場合の \bar{y} と β_1 の推定

	完全データ	リストワイズ	確定単一	確率単一	多重代入法
RMSE (\bar{y})	0.067	0.609	0.073	0.075	0.074
RRMSE (β_1)	0.021	0.027	0.027	0.029	0.028
95%CIカバー率	94.8	95.0	80.0	83.9	94.2

注：真の \bar{y} が0であるため、RRMSEではなくRMSEを用いた。CIは信頼区間である。95%CIカバー率は、1,000回のモンテカルロ実験のうち、95%信頼区間に真の β_1 が含まれた割合である。

る。リストワイズによる信頼区間は、95.0%の確率で真のパラメータを捕らえることができている。非常に正確である。多重代入法による信頼区間は、94.2%の確率で真のパラメータを捕らえることができている。こちらも非常に正確である。確率単一代入法による信頼区間は、83.9%の確率で真のパラメータを捕らえている。これは、名目5%の第一種の過誤率が3倍以上の16.1%になっているということであり、深刻な問題である。確定単一代入法による信頼区間は、80.0%の確率で真のパラメータを捕らえている。これは、名目5%の第一種の過誤率が4倍の20.0%になっているということであり、深刻な問題である。つまり、被説明変数が欠測しているとき、単一代入法を行ってはならないことが示されている。

被説明変数が欠測している回帰分析と基本統計量を研究の目的とする場合、各々のケースでは次点ではあるが、総合的に多重代入法が最もよいといえる。

5. 代入法とマイクロ分析：適合性の問題

代入モデルと分析モデルが同一の変数を持ち、同じ数のパラメータを推定する場合、2つのモデルは適合 (congenial: 融和) しているという (Enders, 2010, p.227; 阿部, 2016, p.118; 高井, 星野, 野間, 2016, p.123)。これまで見てきたモデルは、すべて適合しているものであった。しかし、現実的には、代入モデルと分析モデルが一致しない場合が考えられ、このとき代入モデルは適合性を持たず、理論的には多重代入法のパラメータ推定値の一致性は保証されない。

本章では、代入モデルが分析モデルを内包している場合と分析モデルが代入モデルを内包している場合の2つのケースにおいて、適合性の問題を検証する。

5.1 代入モデルが分析モデルを内包する場合

シミュレーションの設定は、以下のとおりである。代入モデルを(12)式とし、分析モデルを(13)式として、 x_{1i} の平均値および β_1 の推定を目的とするⁱⁱⁱ⁾。(12)式は、(11)式に X の2変量分布が加わったものである。モンテカルロシミュレーションの繰り返し回数 T は1,000回とし、各々のシミュレーションにおいて $n=1000$ の標本データを生成した。 x_{1i} の欠測は、 y_i を条件とするMARとして生成し、欠測率は約30%に設定した。 $MN(\cdot)$ はR関数mvrnormによって生成した。 β_1 の値は $U(1.1, 1.5)$ からの無作為抽出によって設定し、 σ の値は $U(1.1, 1.5)$ からの無作為抽出によって設定した。値を変化させた他のシミュレーション結果においても、本稿の結果とほぼ同じ内容の結果が得られている。

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (12)$$

$$y_i = \beta_1 x_{1i} + \varepsilon_i \quad (13)$$

ここで、

$$X \sim MN(\text{mean} = 0, \text{sd} = 1)$$

$$X = (x_{1i}, x_{2i})$$

$$\text{cor}(X) = \begin{pmatrix} 1.0 & 0.6 \\ 0.6 & 1.0 \end{pmatrix}$$

$$\varepsilon_i \sim N(0, \sigma)$$

代入モデルが分析モデルを内包する形の場合、厳密には代入モデルと分析モデルは適合していないものの、表14から明らかのように多重代入法の精度に問題は発生せず4.2.1項の結果とほぼ一致していることが伺える (Enders, 2010, pp.228-229; Carpenter & Kenward, 2013, pp.64-65)。これが意味していることは、データ提供者である公的統計機関は、できるだけ多くの利用可能な変数を代入モデルに含めることができ^{ix)}、その中から機微な情報を含む変数を省いた上で公開することが許されるということである (高井, 星野, 野間, 2016, p.124)。

表14 代入モデルが分析モデルを内包するケース

	完全データ	リストワイズ	確定単一	確率単一	多重代入法
RMSE (\hat{x}_1)	0.074	0.633	0.080	0.083	0.081
RRMSE (β_1)	0.026	0.058	0.084	0.029	0.028
95%CIカバー率	95.6	64.8	14.4	91.5	95.6

注：真の x_1 が0であるため、RRMSEではなくRMSEを用いた。CIは信頼区間である。95%CIカバー率は、1,000回のモンテカルロ実験のうち、95%信頼区間に真の β_1 が含まれた割合である。

5.2 分析モデルが代入モデルを内包する場合

シミュレーションの設定は、以下のとおりである。代入モデルを(14)式とし、分析モデルを(15)式として、 x_{1i} の平均値および β_1 の推定を目的とする。その他の設定は、5.1節と同じである。

$$y_i = \beta_1 x_{1i} + \varepsilon_i \quad (14)$$

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (15)$$

分析モデルが代入モデルを内包する形の場合、表15から明らかなように、いずれの代入法による推定も著しく悪い結果となっている。すなわち、代入モデルよりも大きな分析モデルを用いてはならない。しかしながら、4.2節の場合と同様に、リストワイズ除去は推奨されない。 x_{1i} の単変量に関して、平均値の推定では単一代入法と多重代入法は不偏であるのに対して、リストワイズ除去には偏りがある。確定単一代入法 (RMSE=0.093)、多重代入法 (RMSE=0.094)、確率単一代入法 (RMSE=0.098)の精度はほぼ同じだが、リストワイズ (RMSE=0.739)の精度は非常に悪い。代入モデルが分析モデルよりも制約的な場合には、

ベストな代入手法は存在しないおそれがあり、避けるべきケースだとされる (Enders, 2010, p.229; Carpenter & Kenward, 2013, p.64)。

6. おわりに

本稿は、現在の公的統計における集計値ベースの代入法において、データの種類に応じた手法が採用されていることを示した。具体的には、経済データには比率代入法が用いられ、世帯データにはホットデック法が用いられることを明らかにした。また、現行の集計値ベースの代入法とは異なり、公開型マイクロデータでは、その性格上、分析の推測対象は平均値や合計値だけとは限らず、回帰係数や標準誤差を用いた統計分析を行うには多重代入法が望ましいことも示した。

現行の手法を勘案すると、経済マイクロデータではTakahashi(2017a)およびTakahashi(2017b)にて提唱されている多重比率代入法を活用することが望ましく、世帯マイクロデータではCranmer & Gill(2013)にて提唱されている多重ホットデック代入法を活用することが望ましいだろう。

多重代入法に関する近年の研究 (Graham et

表15 分析モデルが代入モデルを内包するケース

	完全データ	リストワイズ	確定単一	確率単一	多重代入法
RMSE (\hat{x}_1)	0.087	0.739	0.093	0.098	0.094
RRMSE (β_1)	0.036	0.063	0.119	0.117	0.115
95%CIカバー率	95.3	82.0	5.6	8.9	13.7

注：真の x_1 が0であるため、RRMSEではなくRMSEを用いた。CIは信頼区間である。95%CIカバー率は、1,000回のモンテカルロ実験のうち、95%信頼区間に真の β_1 が含まれた割合である。

al., 2007; Bodner, 2008) では、多重代入済みデータの数は多いほど望ましいことが示されているが、実務的には多くの多重代入済みデータの公表は難しい。実際に、米国政府機関によって公開されているマイクロデータ(表10)では、5~10個の多重代入済みデータを公開している。また、今回のシミュレーションは5個の多重代入済みデータに基づいて実行したが、その結果は非常に良好であった。よって、実務上の便宜も考慮すると、5個の多重代入済みデータを公開すれば十分だと考えられる。

代入済みデータを用いた統計解析では、適合性の確保が必須事項であるため、公開型マイクロデータにおいては代入モデルを明示しなければならない。公開型マイクロデータを用いる分析者は、代入モデルを前提とした分析を行うことができる。これは、坂田(2006, pp.36-38)が指摘する公開型マイクロデータにおける様々な制約に類するものである。つまり、公開型マイクロデータといえども、分析者の望むすべての分析が可能となるわけではなく、調査データ特有の作成事情と性格を意識した上で分析を行う必要がある⁵。それでも既存の集計表だけに依存した分析と比べれば、公開型マイクロデータでは計り知れない可能性が広がるであろう。

もともとRubin(1987)が提案していたとおり、多重代入済みデータが公開されれば、統計リテラシーに関わらず、すべての分析者は欠測にわずらわされずに統計解析を行うことができる。一方、統計リテラシーの高い分析者が、最尤法など自らの望む形で欠測値の処理を行った上で統計解析を行う場合も想定して、欠測値にはフラグを立てるべきである。最尤法による欠測データ解析は、高井、星野、野間(2016, pp.23-101)に詳しい。

最後に、本稿では紙面の都合上、詳細な議論は省くが、Ragunathan(2016, p.182)が指摘するとおり、仮定(assumption)を必要としな

い欠測値処理手法は存在しない。多重代入法は、最も汎用的な欠測値処理手法の1つであるが、あらゆる状況下において常にベストな手法というわけではない。とりわけ、通常の高多重代入法は、他の欠測値処理手法と同様に欠測のメカニズムをMARと想定している。したがって、データ提供者は、欠測のメカニズムがMARの仮定から逸脱した場合に備えて、感度分析を実行する必要がある。多重代入法にパターン混合モデルを適用した感度分析については、Carpenter & Kenward(2013, pp.229-241)を参考にされたい。日本語での解説は、阿部(2016, pp.163-166)を参照されたい。また、近年では、観測データの情報を最大限に活用して、代入法に関する診断手法も提案されているので、これらも合わせて活用されたい(Abayomi et al., 2008; Honaker et al., 2011)。日本語での解説は、高橋、伊藤(2013, pp.64-74)を参照されたい。

付録：多重代入済みデータの作成と分析に関するコード例

本付録では、RパッケージAmelia II (Honaker et al., 2011)によって多重代入済みデータを生成し、RパッケージZelig (Imai et al., 2008)によって統計分析を行うためのコードを示す。

まず、代入者は、多重代入法($M=5$)により欠測値を処理する(高橋、伊藤, 2013, pp.48-49)。下記の例では、5個の多重代入済みデータのファイルが生成される。

```
library(Amelia)
set.seed(6997582)
a.out <- amelia(data, m = 5)
write.amelia(obj = a.out, file.stem = "out-
data", orig.data = F, separate = T, row.
names = F)
```

次に、代入者は下記のコードを準備し、上

記で作成した5個の多重代入済みデータと一緒に公開する。分析者は、5個の多重代入済みデータのファイルをダウンロードし、下記のコードをRのコンソールに貼り付けるだけでよい。ただし、分析者はRパッケージhot.deck (Cranmer & Gill, 2013) をインストールする必要がある。なお、この方法は、Rの内部で多重代入済みデータを保持して分析するのではなく、公開型マイクロデータを想定して、いったん多重代入済みデータを掃き出し、再度データの読み込みを行った上で分析する際に必要な手順である。この方法は、AmeliaとZeligの仕様書には書かれていない。

```
data1<-read.csv("outdata1.csv",header=T)
data2<-read.csv("outdata2.csv",header=T)
data3<-read.csv("outdata3.csv",header=T)
data4<-read.csv("outdata4.csv",header=T)
data5<-read.csv("outdata5.csv",header=T)
```

```
idata<-list(imp1=data1,imp2=data2,imp3=
data3,imp4=data4,imp5=data5)
idata<-list(imputations=idata)
library(hot.deck)
midata<-hd2amelia(idata)
```

最後に、分析者はRパッケージZeligを利用して統計分析を行う(高橋, 伊藤, 2013, p.49)。分析に使用する変数「 $x_1 \sim x_2 + x_3$ 」を指定し、分析を行うモデル「 $model = "ls"$ 」を指定するだけでよい。多重代入済みデータによる複数の分析結果の統合は、Zeligによって自動で行われる。

```
library(Zelig)
z.out <- zelig(x1~x2+x3, data = midata,
model = "ls", cite = F)
summary(z.out)
```

謝辞

本稿は、経済統計学会関東支部例会(2016年7月)、経済統計学会第60回全国研究大会(2016年9月)における報告に加筆・修正したものである。各学会における参加者の方々からは、有益なコメントをいただいた。また、2名の査読者から有益なコメントをいただき本稿を改善することができた。ここに深く感謝の意を表したい。ただし、本稿にあり得べき誤りはすべて執筆者に属する。

注

- i 本稿における公開型マイクロデータは、特定のマイクロデータ提供方法を限定的に意味してはいない。従来の集計値を利用する立場とは異なり、分析者の裁量によって分析が行える環境を想定している。つまり、本稿における「公開型マイクロデータ」は、「一般公開型マイクロデータ」、「匿名化マイクロデータ」、「調査票情報(個票データ)」のすべてを内包する大きな概念として使用している。また、本来的には、マイクロデータにおいて「公開」という用語は、オープンデータのような一般向けに提供する場合に用いるものであり、学術研究のために利用要件を満たした研究者に対して提供する場合には「公開」という言葉は用いないのが一般的である。しかし、本稿では、代入を行う者(調査機関)と分析を行う者(一般市民, 研究者)が別であることが特に重要な論点であり、利用者が一般市民か研究者かという区別はしていない。ゆえに、上記のようなケースも、本稿では「公開型マイクロデータ」に含めている。
- ii 偏りとは推定量の期待値と真のパラメータ値との差である。効率性とは推定量の分散の大きさであり、分散は n の値が小さくなるにつれて大きくなる。
- iii NMARの仮定が正しいとした場合に、MARの仮定に基づく結果にどれだけ影響があるかを評価する手法である(阿部, 2016, p.160)。結果が大きく異なる場合、MARの仮定に基づく分析結果

- の信頼性は高いとみなせる。一方、結果が大きく異なる場合、結果の信頼性が低く、MARの仮定をより妥当なものとするために補助変数を多く組み入れるなどの対処が必要である。
- iv 筆者は、2012年9月のノルウェー会合、2014年4月のフランス会合、2015年9月のハンガリー会合に出席した。
 - v 重回帰モデル、多項モデル、ロジスティックモデルなど、回帰代入法は比率代入法よりも守備範囲が広く、そういった場面で活用されることがある (de Waal et al., 2011, pp.233-235)。
 - vi フラクショナル代入法は、多重代入法と同様、代入を繰り返す手法である。多重代入法とは以下の3点で異なっている (de Waal et al., 2011, p.272) : (1)頻度論に基づく「不適切な」多重代入法 (improper multiple imputation) とみなすことができる ; (2)多重代入法における分散の肥大化を最小化することを目的としている ; (3)ホットデックを利用し、質的なデータへの対応を可能としている。
 - vii この3つの類型は網羅的ではないが、経済データと世帯データについて経験的に知られている一般的な性質をシミュレーションしたものである。経済データは売上高・資本金・従業者数など数量項目が多く、その分布は右に歪んでいるという特徴がある。また、世帯データは住居の種類・配偶者の有無・勤め先など質的項目が多いという特徴がある。
 - viii x_{1i} に欠測が発生しているため、厳密な意味での代入モデルは $x_{1i} = \gamma_0 + \gamma_1 y_i + \gamma_2 x_{2i} + \varepsilon_i$ であり、 $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ は y_i の母集団モデルである。
 - ix 現行の集計値ベースの代入法では、欠測値の代入のために用いられる変数はマイクロデータに含まれ得る変数の一部に過ぎないが、適合性の問題により、公開型マイクロデータに含まれる利用可能な変数のすべてを用いて欠測値の代入を行う必要がある。
 - x 無数の分析者が無数の分析モデルを構築し、代入者はもはや無数の分析モデルを事前に考慮して代入モデルを作ることができないこと、これが公開型マイクロデータにおける代入法の難しさである。よって、代入モデルが先に固定のものとしてあり、その代入モデルの制約が分析者の方に発生する。したがって、マイクロデータの利用者の視点から調査データ特有の作成事情と性格を意識した上で分析を行う必要がある。

参考文献

- [1] 阿部貴行 (2016) 『欠測データの統計解析』, 朝倉書店。
- [2] 伊藤伸介, 星野なおみ (2014) 「国勢調査マイクロデータを用いたスワッピングの有効性の検証」, 『統計学』第107号, pp.1-16。
- [3] 岩崎学 (2002) 『不完全データの統計解析』, エコノミスト社。
- [4] 大野薫, 井川孝之 (2015) 『モンテカルロ法入門』, 一般社団法人金融財政事情研究会。
- [5] 栗原由紀子 (2015) 「統計的マッチングにおける推定精度とキー変数選択の効果 — 法人企業統計調査マイクロデータを対象として —」, 『統計学』第108号, pp.1-15。
- [6] 坂田幸繁 (2006) 「個票データと統計利用」, 『統計学』第90号, pp.31-42。
- [7] 高井啓二, 星野崇宏, 野間久史 (2016) 『欠測データの統計科学 — 医学と社会科学への応用』, 岩波書店。
- [8] 高橋将宜, 阿部穂日, 野呂竜夫 (2015) 「公的統計における欠測値補定の研究 : 多重代入法と単一代入法」, 『製表技術参考資料』 No. 30, pp.1-95。
- [9] 高橋将宜, 伊藤孝之 (2013) 「経済調査における売上高の欠測値補定方法について ~ 多重代入法による精度の評価 ~」, 『統計研究彙報』第70号, No. 2, pp.19-86。
- [10] 高橋将宜, 伊藤孝之 (2014) 「様々な多重代入法アルゴリズムの比較 ~ 大規模経済系データを用いた分析 ~」, 『統計研究彙報』第71号, No. 3, pp.39-82。
- [11] 中村英昭, 平澤鋼一郎 (2016) 「公的統計の二次的利用の促進に関するわが国の取組状況」, 『経済統計学会第60回 (2016年度) 全国研究大会報告要旨集』, pp.36-37。
- [12] Abayomi, K., Gelman, A. & Levy, M. (2008) “Diagnostics for Multivariate Imputations”, *Applied Statistics* Vol. 57, No. 3, pp.273-291.
- [13] Allison, P. D. (2002) *Missing Data*, Sage Publications, Thousand Oaks.

- [14] Baraldi, A. N. & Enders, C. K. (2010) “An Introduction to Modern Missing Data Analyses”, *Journal of School Psychology* Vol. 48, No. 1, pp.5-37.
- [15] Bodner, T. E. (2008) “What Improves with Increased Missing Data Imputations?” *Structural Equation Modeling* Vol. 15, pp.651-675.
- [16] Carpenter, J. R. & Kenward, M. G. (2013) *Multiple Imputation and its Application*, A John Wiley & Sons Publication, West Sussex.
- [17] Carsey, T. M. & Harden, J. J. (2014) *Monte Carlo Simulation and Resampling Methods for Social Science*, Sage Publications, Thousand Oaks.
- [18] Cheema, J. R. (2014) “Some General Guidelines for Choosing Missing Data Handling Methods in Educational Research”, *Journal of Modern Applied Statistical Methods* Vol. 13, No. 2, pp.53-75.
- [19] Cochran, W. G. (1977) *Sampling Techniques*, third edition, John Wiley & Sons, New York.
- [20] Cranmer, S. J. & Gill, J. (2013) “We Have to Be Discrete About This: A Non-Parametric Imputation Technique for Missing Categorical Data”, *British Journal of Political Science* Vol. 43, No. 2, pp.425-449.
- [21] de Waal, T., Pannekoek, J., & Scholtus, S. (2011) *Handbook of Statistical Data Editing and Imputation*, John Wiley & Sons, Hoboken.
- [22] Di Zio, M. & Guarnera, U. (2013) “A Contamination Model for Selective Editing”, *Journal of Official Statistics* Vol. 29, No. 4, pp.539-555.
- [23] Donders, A. R. T., van der Heijden, G. J. M. G., Stijnen, T. & Moons, K. G. M. (2006) “Review: A Gentle Introduction to Imputation of Missing Values”, *Journal of Clinical Epidemiology* Vol. 59, pp.1087-1091.
- [24] Enders, C. K. (2010) *Applied Missing Data Analysis*, The Guilford Press, New York.
- [25] Graham, J. W., Olchowski, A. E. & Gilreath, T. D. (2007) “How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory”, *Prevention Science* Vol. 8, No. 3, pp.206-213.
- [26] Honaker, J., King, G. & Blackwell, M. (2011) “Amelia II: A Program for Missing Data”, *Journal of Statistical Software* Vol. 45, No. 7, pp.1-47.
- [27] Hu, M., Salvucci, S., & Lee, R. (2001) *A Study of Imputation Algorithms*, Working Paper No. 2001-17. U.S. Department of Education, National Center for Education Statistics.
- [28] Imai, K., King, G. & Lau, O. (2008) “Toward a Common Framework for Statistical Analysis and Development”, *Journal of Computational and Graphical Statistics*, Vol. 17, No. 4, pp.892-913.
- [29] Joenssen, D. W. (2015) *Hot Deck Imputation: Hot Deck Imputation Methods for Missing Data*, Version 1.1.0, <https://cran.r-project.org/web/packages/HotDeckImputation/index.html>
- [30] King, G., Honaker, J., Joseph, A. & Scheve, K. (2001) “Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation”, *American Political Science Review*, Vol. 95, No. 1, pp.49-69.
- [31] Little, R. J. A. (1992) “With Missing X’s: A Review”, *Journal of the American Statistical Association* Vol. 87, No. 420, pp.1227-1237.
- [32] Little, R. J. A. & Rubin, D. B. (2002) *Statistical Analysis with Missing Data*, second edition, John Wiley & Sons, Hoboken.
- [33] Mooney, C. Z. (1997) *Monte Carlo Simulation*, Sage Publications, Thousand Oaks.
- [34] Raghunathan, T. (2016) *Missing Data Analysis in Practice*, CRC Press, Boca Raton.
- [35] Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.
- [36] Schafer, J. L. & Graham, J. W. (2002) “Missing Data: Our View of the State of the Art”, *Psychological Methods* Vol. 7, No. 2, pp.147-177.
- [37] Schenker, N., Raghunathan, T. E., Chiu, P.-L., Makuc, D. M., Zhang, G. & Cohen, A. J. (2006) “Multiple Imputation of Missing Income Data in the National Health Interview Survey”, *Journal of the American Statistical Association* Vol. 101, No. 475, pp.924-933.
- [38] Scheuren, F. (2005) “Multiple Imputation: How It Began and Continues”, *The American Statistician*

Vol. 59, No. 4, pp.315-319.

- [39] Takahashi, M. (2017a) "Multiple Ratio Imputation by the EMB Algorithm: Theory and Simulation", *Journal of Modern Applied Statistical Methods* Vol. 16, No. 1 (in press).
- [40] Takahashi, M. (2017b) "Implementing Multiple Ratio Imputation by the EMB Algorithm in R", *Journal of Modern Applied Statistical Methods* Vol. 16, No. 1 (in press).
- [41] Takahashi, M., Iwasaki, M. & Tsubaki, H. (2017) "Imputing the Mean of a Heteroskedastic Log-Normal Missing Variable: A Unified Approach to Ratio Imputation", *Statistical Journal of the IAOS* Vol. 33, No. 3 (in press).
- [42] U.S. Bureau of the Census (1957) *U.S. Census of Manufactures 1954*, Vol. II, Industry Statistics, Part 1, General Summary and Major Groups 20 to 28, U.S. Government Printing Office, Washington, D. C.

Missing Data Treatments in Official Statistics: Imputation Methods for Aggregate Values and Public-Use Microdata

Masayoshi TAKAHASHI*

Summary

The treatment of missing data is an important aspect of the production system in official statistics; however, the methodological significance has rarely been debated in Japan. This study builds on the practices of UNECE, assessing a variety of imputation methods, such as deterministic single imputation (regression imputation, ratio imputation, mean imputation, hot deck imputation), stochastic single imputation, and multiple imputation. The purpose of this study is to contribute to the development of the official statistics production system in Japan. In order to obtain the picture of the status quo in official statistics among the national statistical agencies around the world, this study conducts a survey on the UNECE member states. Also, the current study assesses the utility of each imputation method, using a series of Monte Carlo experiments with a variety of parameter settings. This study shows that, in the current practices of imputation among the national statistical institutes, imputation methods are adopted for each data type, such that ratio imputation is often used for economic data and hot deck imputation for household data. Furthermore, the current study demonstrates that multiple imputation is suited for public-use microdata.

Key Words

Multiple imputation, single imputation, incomplete data, nonresponse, government statistics

* IR Office, Tokyo University of Foreign Studies
e-mail : mtakahashi@tufs.ac.jp

機関誌『統計学』投稿規程

経済統計学会（以下、本会）会則第3条に定める事業として、『統計学』（電子媒体を含む。以下、本誌）は原則として年に2回（9月、3月）発行される。本誌の編集は「経済統計学会編集委員会規程」（以下、委員会規程）にもとづき、編集委員会が行う。投稿は一般投稿と編集委員会による執筆依頼によるものとし、いずれの場合も原則として、本投稿規程にしたがって処理される。

1. 総則

1-1 投稿者

会員（資格停止会員を除く）は本誌に投稿することができる。

1-2 非会員の投稿

- (1) 原稿が複数の執筆者による場合、筆頭執筆者は本会会員でなければならない。
- (2) 常任理事会と協議の上、編集委員会は非会員に投稿を依頼することができる。
- (3) 本誌に投稿する非会員は、本投稿規程に同意したものとみなす。

1-3 未発表

投稿は未発表ないし他に公表予定のない原稿に限る。

1-4 投稿の採否

投稿の採否は、審査の結果にもとづき、編集委員会が決定する。その際、編集委員会は原稿の訂正を求めることがある。

1-5 執筆要綱

原稿作成には本会執筆要綱にしたがう。

2. 記事の分類

2-1 研究論文

以下のいずれかに該当するもの。

- (a) 統計およびそれに関連した分野において、新知見を含む会員の独創的な研究成果をまとめたもの。
- (b) 学術的な新規性を有し、今後の研究の発展可能性を期待できるもので、速やかな成果の公表を目的とするもの。

2-2 報告論文

研究論文に準じる内容で、研究成果の速やかな報告をとくに目的とする。

2-3 書評

統計関連図書や会員の著書などの紹介・批評。

2-4 資料

各種統計の紹介・解題や会員が行った調査や統計についての記録など。

2-5 フォーラム

本会の運営方法や統計、統計学の諸問題にたいする意見・批判・反論など。

2-6 海外統計事情

諸外国の統計や学会などについての報告。

2-7 その他

全国研究大会・会員総会記事、支部だより、その他本会の目的を達成するために有益と

思われる記事。

3. 原稿の提出

3-1 投稿

原稿の投稿は常時受け付ける。

3-2 原稿の送付

原則として、原稿は執筆者情報を匿名化したPDFファイルを電子メールに添付して編集委員長へ送付する。なお、ファイルは『統計学』の印刷レイアウトに準じたPDFファイルであることが望ましい。

3-3 原稿の返却

投稿された原稿（電子媒体を含む）は、一切返却しない。

3-4 校正

著者校正是初校のみとし、大幅な変更は認めない。初校は速やかに校正し期限までに返送するものとする。

3-5 投稿などにかかわる費用

- (1) 投稿料は徴収しない。
- (2) 掲載原稿の全部もしくは一部について電子媒体が提出されない場合、編集委員会は製版にかかる経費を執筆者（複数の場合には筆頭執筆者）に請求することができる。
- (3) 別刷は、研究論文、報告論文については30部までを無料とし、それ以外は実費を徴収する。
- (4) 3-4項にもかかわらず、原稿に大幅な変更が加えられた場合、編集委員会は掲載の留保または実費の徴収などを行うことがある。
- (5) 非会員を共同執筆者とする投稿原稿が掲載された場合、その投稿が編集委員会の依頼によるときを除いて、当該非会員は年会費の半額を掲載料として、本会に納入しなければならない。

3-6 掲載証明

掲載が決定した原稿の「受理証明書」は学会長が交付する。

4. 著作権

4-1 本誌の著作権は本会に帰属する。

4-2 本誌に掲載された記事の発行時に会員であった執筆者もしくはその遺族がその単著記事を転載するときには、出所を明示するものとする。また、その共同執筆記事の転載を希望する場合には、他の執筆者もしくはその遺族の同意を得て、所定の書面によって本会に申し出なければならない。

4-3 前項の規定にもかかわらず、共同執筆者もしくはその遺族が所在不明のため、もしくは正当な理由によりその同意を得られない場合には、本会が承認するものとする。

4-4 執筆者もしくはその遺族以外の者が転載を希望する場合には、所定の書面によって本会に願い出て、承認を得なければならない。

4-5 4-4項にもとづく転載にあたって、本会は転載料を徴収することができる。

4-6 会員あるいは本誌に掲載された記事の発行時に会員であった執筆者が記事をウェブ転載するときには、所定の書類によって本会に申し出なければならない。なお、執筆者が所属する機関によるウェブ転載申請については、本人の転載同意書を添付するものとする。

- 4-7 会員以外の者、機関等によるウェブ転載申請については、前号を準用するものとする。
- 4-8 転載を希望する記事の発行時に、その執筆者が非会員の場合には、4-4, 4-5項を準用する。
1997年7月27日制定(2001年9月18日, 2004年9月12日, 2006年9月16日, 2007年9月15日, 2009年9月5日, 2012年9月13日, 2016年9月12日一部改正)

『統計学』創刊60周年記念特集掲載号発行規程

『統計学』創刊60周年記念特集論文(以下, 記念特集論文)の掲載号の編集・発行作業は, 経済統計学会2014年度会員総会の決議にもとづき『統計学』創刊60周年記念事業委員会(以下, 事業委員会)が行なう。記念特集論文の掲載号(以下, 記念特集掲載号)の発行は, 本規程にしたがって処理される。

1. 総則

1-1 テーマの確定及び原稿執筆者の選定と資格

特定テーマに関わる論文構成の確定及び執筆者の選定は, 企画案と執筆計画にもとづき, 事業委員会が行なう。

1-2 未発表

原稿は未発表ないし他に公表予定のない原稿に限る。

1-3 原稿の採否およびレフェリー制の導入について

提出された原稿の採否は, レフェリーによる厳格な審査の結果にもとづき, 事業委員会が決定する。レフェリーの選任は事業委員会が行なう。事業委員会は原稿の書換え, 訂正を求めることができる。

1-4 執筆要綱

原稿作成は別に定める『統計学』創刊60周年記念特集掲載号執筆要綱にしたがう。

2. 原稿の提出

2-1 原稿の締切り

本誌発行の円滑のため, 締切り日を設ける。締切り日以降に原稿が到着した場合や, 訂正を求められた原稿が期日までに訂正されない場合, 掲載されないことがある。

2-2 原稿の送付

原稿は原則として, PDFファイル(『統計学』の印刷レイアウト)を電子メールに添付して事業委員会委員長へ送付する。

2-3 原稿の返却

提出された原稿は, 採否にかかわらず原則として返却しない。

2-4 校正

掲載が決定した原稿の著者校正は初校のみとし, 内容の変更を伴う原稿の変更は原則的に認めない。内容の変更を伴う変更の場合は, 事業委員会およびレフェリーの許可を必要とする。初校は速やかに校正し期限までに返送するものとする。

2-5 執筆などにかかわる費用

投稿料は原則として徴収しない。別刷は, 執筆者の希望により, 作成するが, 実費を徴収する。校正段階で原稿に大幅な変更が加えられた場合, 実費の徴収などを行うことがあ

る。

3. 著作権

記念特集論文の著作権は経済統計学会に帰属する。詳細は、『統計学』の投稿規程に準ずる。

『統計学』創刊60周年記念特集掲載号投稿原稿査読要領

1. 経済統計学会（以下、本会）の機関誌『統計学』創刊60周年記念特集掲載号に掲載する「論文」の査読制度について、この要領を定める。
2. 『統計学』創刊60周年記念事業委員会（以下「事業委員会」）委員長に送付された原稿については、事業委員会による第一次審査を行い、事業委員会が別に定める「執筆要綱」に準拠しているかどうかを判定する。
3. 「論文」の掲載にあたっては、第二次審査を必要とする。
4. 第一次審査を経た「論文」の原稿は、速やかに第二次審査へ付されるものとする。
5. 事業委員会は、次の事項を審議決定する。
 - (1) 第一次審査結果の確認
 - (2) 第二次審査を担当する2名のレフェリーの選任
6. 第二次審査にあたるレフェリーは会員から選任する。
7. 第二次審査にあたって、レフェリーについては匿名性を確保する。
8. 第二次審査における判定は、(1)論文として掲載可、(2)論文として条件付掲載可、(3)掲載不可とし、レフェリーはその理由を明示するものとする。
9. 第二次審査でレフェリー間での審査結果が異なる場合には、事業委員会はレフェリーと協議し、掲載の可否について最終的な判断を下すものとする。

編集委員会からのお知らせ
機関誌『統計学』の編集・発行について

編集委員会

2016年9月より、新しい規定にもとづいて、「研究論文」と「報告論文」が設定されました。皆様からの積極的な投稿をお待ちしております。また、本号より掲載が開始された『統計学』創刊60周年記念特集論文につきましては、本号の『統計学』創刊60周年記念特集掲載号関連諸規程』ならびに学会の公式ウェブサイトをご参照下さい。

1. 投稿は、常時、受け付けています。なお、書評、資料および海外統計事情等については、下記の[注記2]をご確認下さい。
2. 次号以降の発行予定日は、
第113号：2017年9月30日、第114号：2018年3月31日です。
3. 投稿に際しては、「投稿規程」、「執筆要綱」、および「査読要領」などをご熟読願います。最新版は、学会の公式ウェブサイトをご参照下さい。
4. 原稿は編集委員長(下記メールアドレス)宛にお送り下さい。
5. 原稿はPDF形式のファイルとして提出して下さい。また、紙媒体での提出も旧規程に準拠して受け付けます。紙媒体の送付先は編集委員長宛にお願いします(住所は会員名簿をご参照下さい)。
6. 原則として、すべての投稿原稿が査読の対象となります。
7. 通常、査読から発刊まで査読が順調に進んだ場合でも、2ヶ月から3ヶ月程度を要します。投稿にあたっては十分に留意して下さい。

編集委員会、投稿応募についての問い合わせは、
下記メールアドレス宛に連絡下さい。
また、編集委員長へのメールアドレスも下記になります。

editorial@jses.jp

来年度(2017年度)の編集委員は、つぎのとおりです。

編集委員長 藤井輝明(大阪市立大学)
副委員長 水野谷武志(北海学園大学)
編集委員 橋本貴彦(立命館大学)
小林良行(総務省統計研究研修所)
山田 満(東北・関東支部所属)

[注記1] 『統計学』の定期刊行に努めておりますので、できるかぎり早期のご投稿をお願いします。113号(2017年9月30日発行予定)への掲載を想定した場合、「研究論文」と「報告論文」の原稿は、2017年7月初旬を目途として、それまでにご投稿ください。

[注記2] 書評、資料および海外統計事情等について、執筆、推薦、および依頼等をお考えの会員がおられましたら、企画や思いつきの段階で結構ですので、できるだけ早い段階で、編集委員会にご一報下さい。以上

編集後記

研究成果を投稿下さいました執筆者の皆様、査読に関わって下さいました皆様、そして、書評の依頼をお引き受け下さいました皆様に、心からお礼申し上げます。また、本号より、『統計学』創刊60周年記念特集論文』の掲載も開始されました。特集論文を投稿下さいました皆様、そして、創刊60周年記念事業委員会(委員長：水野谷武志会員)の皆様にも、改めて感謝申し上げます。さて、次号113号より、藤井輝明編集委員長のもとで、本誌が編集されます。編集委員会では、機関誌『統計学』を充実させていくために、皆様からの率直なご意見と、そして、研究成果の積極的なご投稿をお待ちしております。今後ともよろしくお願ひ申し上げます。(朝倉啓一郎 記)

執筆者紹介

水野谷武志	(北海学園大学経済学部)	田添篤史	(京都大学経済学研究科)
金子治平	(神戸大学大学院農学研究科)	山口秋義	(九州国際大学)
福島利夫	(専修大学経済学部)	西村善博	(大分大学経済学部)
高橋将宜	(東京外国語大学経営戦略情報本部)		

支部名

事務局

北海道	062-8605	札幌市豊平区旭町 4-1-40 北海学園大学経済学部 (011-841-1161)	水野谷武志
東北・関東	980-8511	仙台市青葉区土樋 1-3-1 東北学院大学経済学部 (022-721-3417)	前田修也
関西	567-8570	茨木市岩倉町 2-150 立命館大学経営学部 (072-665-2090)	田中力
九州	870-1192	大分市大字旦野原 700 大分大学経済学部 (097-554-7706)	西村善博

『統計学』編集委員

朝倉啓一郎 (東北・関東) [長]	藤井輝明 (関西) [副]
前田修也 (東北・関東)	橋本貴彦 (関西)
山田満 (東北・関東)	

『統計学』創刊60周年記念事業委員会

水野谷武志 (北海道) [長]	大井達雄 (関西) [副]	伊藤伸介 (東北・関東)
池田伸 (関西)	村上雅俊 (関西)	杉橋やよい (東北・関東)
上藤一郎 (東北・関東)	朝倉啓一郎 (東北・関東)	西村善博 (九州)

統計学 No.112

2017年3月31日 発行	発行所	経済統計学会 〒112-0013 東京都文京区音羽1-6-9 音羽リスマチック株式会社 TEL/FAX 03(3945)3227 E-mail: office@jsest.jp http://www.jsest.jp/
	発行人	代表者 西村善博
	発売所	音羽リスマチック株式会社 〒112-0013 東京都文京区音羽1-6-9 TEL/FAX 03(3945)3227 E-mail: otorisu@jupiter.ocn.ne.jp 代表者 遠藤誠

STATISTICS

No. 112

2017 March

Articles

- Extended Childcare Time for Married Couples with Infants
..... Takeshi MIZUNOYA (1)
- Investigation on Financialization of Japanese Economy :
Focusing on the Character of Industrial Capital
..... Atsushi TAZOE (15)

Book Reviews

- Jun-ichi OKABE and Aparajita BAKSHI, *A New Statistical Domain in India :
An Enquiry into Village Panchayat Databases*, Tulika Books, New Delhi, 2016
..... Jihei KANEKO (30)
- I.I. ELISEEVA and A.L. DMITRIEV, *General Survey on History of Russian State
Statistics*, Rostok, St. Petersburg, 2016
..... Akiyoshi YAMAGUCHI (37)
- Akira NOZAKI ed., *Unequal Society*, Dobunkan Shuppan, Co., Tokyo, 2016
..... Toshio FUKUSHIMA (43)

Special Section : The 60th Anniversary of the *Journal*

- Introduction Takeshi MIZUNOYA (47)
- Special Topic A : Problems in Microdata Analysis of Official Statistics Based on
Probability Sampling Designs**
- The Reform of Population Census : French Rolling Census
..... Yoshihiro NISHIMURA (49)
- Special Topic B : Methodological Perspectives in the Creation and Release of Official
Microdata**
- Missing Data Treatments in Official Statistics :
Imputation Methods for Aggregate Values and Public-Use Microdata
..... Masayoshi TAKAHASHI (65)

Activities of the Society

- Activities in the Branches of the *Society* (84)
- Prospects for the Contribution to the *Journal* (89)

JAPAN SOCIETY OF ECONOMIC STATISTICS
