

欠測値への対応に関する各府省等職員研究発表の紹介

(各府省等ホームページに掲載されているものを収集、内容や意見は執筆職員個人のものであり、所属機関の見解を示すものではない)

令和4年4月現在

研究機関	研究成果			作成職員		掲載URL
	名称()内は発表媒体・発表機会	発表年度	要旨抜粋等	所属等(研究実施時点)	氏名	
総務省	建築着工統計調査の補正調査に関する標本設計移行期における推定方法の提案 (総務省統計委員会担当室ワーキングペーパー 2019-WP04)	平成31年度	<p>総務省が実施している統計精度向上の取組対象と調査の一つとして、国土交通省が実施している建築着工統計調査の補正調査がある。この補正調査は、建築着工前事費予定額と竣工後に確定した工事実施額について、そのか離率(補正)を把握するため標本調査である。</p> <p>その調査結果推定精度向上に向けて、総務省は新しい標本設計を提案している。本稿は標設計を変更した際に、旧来の方法で提案され際に、旧来の方法で提案されている。本稿は標本設計を変更した際に、旧来の方法で抽出された標本と新しい方法で抽出された標本が混在するために複雑な計算が必要となる移行期において、偏りなく推定を行う二つの方法を示す。一つは、偏りなく復元できる厳密な推定を行う計算方法であり、もう一つは簡易に推定を行う計算方法である。なお、ここで提案する簡易な方法は経験上の方法であるため、国土交通省より過去に実施された建築着工統計調査の補正調査のデータ提供を受けて、リサンプリングシミュレーションを実施し概ね偏りが無いことを示している。</p>	元総務省統計委員会担当室 総務省統計委員会担当室 総務省統計委員会担当室	澤谷 秀之 増成 敬三 上田 聖	http://www.soumu.go.jp/main_content/000631509.pdf
総務省	労働力調査における東日本大震災に伴う補完推計 (統計研究研修所 統計研究彙報 第70号 2013年3月)	平成25年度	<p>本稿は、労働力調査における東日本大震災に伴う補完推計方法とその結果について内容を概説するものである。</p> <p>労働力調査では、東日本大震災の影響により、岩手県、宮城県及び福島県(以下「被災3県」という。)において調査の実施が一時困難となったため、2011年の3月から8月までの間においては、被災3県を除く全国結果の集計・公表を行った。そこで、これら3県の同期間の結果について別途補完推計を行い、それを基に参考値として全国結果を算出した。</p> <p>この補完推計は、被災3県について一定の仮定の下に利用可能な他の関連するデータに基づいて可能な限り実態を反映するように行った。被災3県を含む全国の補完推計の結果は、労働力調査における平成24年の3月から8月までの前年同月比較や長期時系列比較などで参考値として用いることとしている。</p>	総務省労働力人口統計室	高橋 伸一	http://www.stat.go.jp/training/2kenkyu/2-2-703.html
総務省	労働力調査における東日本大震災に伴う補完推計について (日本統計学会 第80回大会(平成24年9月9日~12日))	平成24年度	<ul style="list-style-type: none"> 平成23年3~8月の各月を被災3県分について推計した上で、これを既公表の被災3県を除く全国の結果数値に加算することで、被災3県を含む全国結果を算出。 被災3県分の推計は、一定の仮定の下、利用可能な他の関連するデータを基に、可能な限り実態を反映するように推計。 被災3県の各県ごとに、「就業者数」と「完全失業者数」の総数をそれぞれ別に推計。非労働力人口は15歳以上人口からこれらを差し引いて計算。 	総務省労働力人口統計室	高橋 伸一	http://www.stat.go.jp/training/2kenkyu/pdf/gakkai/toukei/2012/takahash.pdf

研究機関	研究成果			作成職員		掲載URL
	名称() 内は発表媒体・発表機会)	発表年度	要旨抜粋等	所属等 (研究実施時点)	氏名	
統計研究研修所 (総務省)	近年の諸外国の統計調査における欠測値補完の動向について (統計研究研修所 Research Paper No.51)	令和3年度	本稿では、アメリカ合衆国及び欧州各国の近年の動向を把握するとともに、アメリカ合衆国については人口センサスに関して継続して行われているプロジェクトの状況を調査した。その結果、調査対象に接触できないことの多さが問題となっているアメリカ合衆国の人口センサスでは、行政記録の利用が調査対象への接触の判断とデータ補完の両面で課題となっており、その際、インピュテーションが実査時点の課題にもなっていること、カナダや欧州でも、コロナ禍などの影響で研究発表は低調になっているものの、インピュテーションに関する行政記録の利用、機械学習の適用、システムの構築・共有などさまざまなプロジェクトが進められていることが分かった。	統計研究研修所統計研修研究官	坂下 信之	https://www.stat.go.ip/training/2kenkyu/research/pdf/research51.pdf
統計研究研修所 (総務省)	統計調査の欠測値補完方法に関する研究動向について (主に米国とオランダ) (統計研究研修所 Research Paper No.48)	令和2年	本リサーチペーパーでは、各国で継続して研究されてきたテーマに焦点を当て、テーマを特定した継続的な研究の見られるアメリカ合衆国及びオランダについて重点的に論文を収集した。また、一般用マイクロデータを用いて、諸外国では頻りに用いられているが我が国では適用例の少ないホット・デック法の数値シミュレーションを試みた。その結果、少なくとも今回対象にした2カ国については、インピュテーションシステム、ベイズモデル、制約条件下のインピュテーションなどの継続的な課題を設定し、豊富な過去の蓄積の上で新たな検討を行っていることが分かった。また、我が国の統計データにおける欠測値への対処にホット・デック法を適用することは可能であると考えられるが、ドナーの選び方などの具体的な方法は、適用する調査に応じて子細かつ実務的に検討する必要がある。	統計研究研修所統計研修研究官	坂下 信之	https://www.stat.go.ip/training/2kenkyu/research/pdf/research48.pdf
統計研究研修所 (総務省)	統計調査の欠測値補完方法に関する基本的文献と諸外国の動向について (統計研究研修所 Research Paper No.44)	令和元年度	今までの調査を受け、引用されることが多く、欠測値補完についての基本的文献と思われる書籍の収集・調査及び各国の最新動向の調査を行った。その結果、リサーチペーパー第43号で見たように共通の体系的な理解が必ずしも存在しているわけではなく、また出版物には多重代入法などの理論的な書籍が多い傾向があるものの、1980年以降、議論の根拠となるような多数の文献が提供されていることが分かった。また、近年の動向としては、個別のインピュテーション手法よりも、行政情報の利用に伴うインピュテーションの必要性の発生への対応、総合システムの開発、商用あるいは他国で開発したシステムの利用に関する話題が多くなっており、国によっては機械学習についての検討も進んでいることが分かった。	統計研究研修所統計研修研究官	坂下 信之	http://www.stat.go.ip/training/2kenkyu/2-1-44.html
統計研究研修所 (総務省)	諸外国における統計調査の欠測値補完方法の動向と手法の体系について (統計研究研修所 Research Paper No.43)	平成30年度	昨年度の調査のあとを受けて、今年度は、その後の動向に加えて、欠測値補完の方法の体系がどのように整理されてきたかの観点からの文献収集を行った。その結果、高度な手法の検討は続けられているものの、実務においてはなおホット・デック法、比率代入法などの伝統的な手法が採用されることが多く、先進的な手法を検討したうえでホット・デック法を採用した例もある。カナダやオーストリアでは、システムの改良が続けられており、他国での採用例も見られる。手法の体系については、必ずしも共通の理解が存在しているわけではないが、1980年以降豊富な文献の蓄積があり、特に90年代末からは、統計を作成している機関自ら発信することも増え、意見の交換が行われていることが分かった。	統計研究研修所統計研修研究官	坂下 信之	http://www.stat.go.ip/training/2kenkyu/2-1-43.html

研究機関	研究成果			作成職員		掲載URL
	名称() 内は発表媒体・発表機会)	発表年度	要旨抜粋等	所属等 (研究実施時点)	氏名	
統計研究研修所 (総務省)	諸外国の公的統計における欠測値補完 (インプテーション) の現状 ~文献調査~ (統計研究研修所 Research Paper No. 40)	平成29年度	日本における検討の端緒とすることを目的として、諸外国におけるデータ・エディティング、特に欠測値補完 (インプテーション) の状況について、ネットなどで公開情報として入手可能な文献を調査した。調査対象としたのは、各国統計機関が公表した論文や国際会議用のプロシーディングなどで、近年のものを中心に、必要に応じて過去に遡る方法を取った。	統計研究研修所統計研修研究官	坂下 信之	http://www.stat.go.jp/training/2kenkyu/2-1-40.html
統計研究研修所 (総務省)	諸外国の公的統計における欠測値補完 (インプテーション) の現状 (経済統計学会 2017年度全国研究大会 (平成29年9月10日~13日))	平成29年度	諸外国における欠測値補完の状況について文献調査を行った。調査対象は公開情報として入手可能なもの。公表論文や国際会議用のプロシーディングなど近年のものが中心。	統計研究研修所統計研修研究官	坂下 信之	http://www.stat.go.jp/training/2kenkyu/pdf/gakkai/keizai/2017/sakashita.pdf
独立行政法人統計センター	公的統計における欠測値補完の研究：多重代入法と単一代入法 (統計センター 製表技術参考資料30)	平成27年度	経済センサスは、経済統計における母集団情報を整備する基幹統計と位置づけられる重要な統計調査である。しかし、調査データでは、すべての項目についてデータが得られる保証はなく、欠測により発生する偏りに対処しなければならない。そこで、統計センターでは、多重代入法を始めとする欠測値補完の研究を実施してきた。本稿は、我が国の公的統計における統計作成に資する材料として、その研究成果を記録として残すものである。また、統計実務者にとっての手引書としての性格上、専門的な数式を一切用いず、数値例と図表によって欠測値補完についての直感的な解説を試みたものである (数式は脚注に記載している)。本稿では、確定的回帰補完、確率的回帰補完、比率補完といった単一代入法とEMアルゴリズムによる多重代入法に関して、各々の補完手法のメカニズム及びそれぞれの手法の長所と短所を紹介している。また、統計分析のためのフリーソフトウェア環境Rによる実行方法も示した。擬似的に欠測値を発生させた平成24年経済センサス - 活動調査のデータ等を使用して、欠測に起因する誤差の数値評価を多重代入法によって行う。	統計センター統計情報・技術部統計技術研究課上級研究員 一橋大学経済研究所附属社会科学統計情報研究センター助教 内閣府大臣官房政府広報室世論調査専門官	高橋将宜 阿部穂日 野呂竜夫	https://www.nstac.go.jp/services/pdf/sankousiryu2706.pdf
独立行政法人統計センター	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~ (統計研究研修所 統計研究彙報 第71号 No. 3)	平成26年度	様々な多重代入法アルゴリズムのメカニズムを示し、経済センサス - 活動調査の速報データとシミュレーションデータを用い、公的経済統計における欠測値補完に関して、いずれのアルゴリズムが優れているかを検証	統計センター統計情報・技術部統計技術研究課 統計センター製表部管理企画課経済センサス業務推進室	高橋 将宜 伊藤 孝之	http://www.stat.go.jp/training/2kenkyu/2-2-713.html
独立行政法人統計センター	経済調査における売上高の欠測値補完方法について ~多重代入法による精度の評価~ (統計研究研修所 統計研究彙報 第70号 No. 2)	平成24年度	EDINETデータを使用し、個別データの補完方法として多重代入法 (Multiple Imputation) を評価した。様々な欠測値対処法とその限界を示し、代替法として多重代入法を導入し、フリーソフトウェアRの汎用多重代入法パッケージAmeliaを利用して、多重代入法による欠測値補完の精度評価を行う	統計センター情報技術部統計技術研究課	高橋 将宜 伊藤 孝之	http://www.stat.go.jp/training/2kenkyu/2-2-702.html

研究機関	研究成果			作成職員		掲載URL
	名称() 内は発表媒体・発表機会)	発表年度	要旨抜粋等	所属等(研究実施時点)	氏名	
独立行政法人統計センター	収入額の欠測値補定にふさわしい層分けの提案 (統計センター 製表技術参考資料21)	平成24年度	<p>本稿は、売上(収入)金額の欠測値補定を層に分ける場合と分けられない場合の当てはまりの精度を取りまとめたものである。本検証で用いたデータは、平成16年サービス業基本調査データである。</p> <p>まず、層で分けることのない最大規模の層を「産業大分類×事業所形態×経営組織」とし、これを「基準層」とした。次に、この基準層から更に「都道府県」別、「産業中分類」別、「事業従事者規模」別と層を設け、これらの層から欠測値補定の推計にふさわしい層の組合せを見つけ出した。ただし、小さく層を区切ることは各々の層におけるデータ数が少なくなるため、データ数をある程度確保できる層レベルで検証を行う。以下は本稿の検証の条件である。</p> <ul style="list-style-type: none"> ・検証に使用したモデルは、比率の当てはめモデルである。 ・欠測値補定にふさわしい層の選定における、評価基準を「欠測のない本来の収入額の総和」と「欠測補定済の収入額の合計値」の差とした。 ・検証としては、以下の2パターンを実施した。 <ul style="list-style-type: none"> イ) 「基準層」と「基準層×都道府県」での比較 ロ) 「基準層」、「産業中分類」、「基準層×事業従事者規模」、「産業中分類×事業従事者規模」での比較 	統計センター情報技術部統計技術研究課	伊藤 孝之1、野呂 竜夫1、阿部 穂日	https://www.nstac.go.jp/services/pdf/sankousiryoku2503.pdf
独立行政法人統計センター	多変量外れ値の検出 ～繰返し加重最小二乗(IRLS)法による欠測値の補定方法～ (統計研究研修所 統計研究彙報 第69号 No. 2)	平成23年度	<p>企業財務データを用いて、売上高を従業者数で説明する単回帰モデルを作成し、通常最小二乗法に代えてロバスト回帰の一種で統計調査の集計実務に適用可能な簡便なアルゴリズムの繰返し加重最小二乗法(IRLS: Iteratively Reweighted Least Squares)を用いることにより、自動的に外れ値の影響を抑えて安定した補定値を得られることを示す。</p>	独立行政法人統計センター情報技術部研究主幹室	和田 かず美	http://www.stat.go.jp/training/2kenkyu/2-2-692.html
独立行政法人統計センター	ヨーロッパにおけるデータエディティング及び補定に関する調査報告 ～EDIMBUSプロジェクトを中心に～ (統計研究研修所 統計研究彙報 第66号 No. 4)	平成20年度	<p>近年のヨーロッパにおけるデータエディティング及び補定に関する取組の状況と2007年に取りまとめられた横断的経済調査のデータエディティング及び補定に関するプロジェクト(EDIMBUSプロジェクト)の推奨実践マニュアル(Recommended Practices Manual)の概要について報告</p>	独立行政法人統計センター情報技術部研究主幹	小林 良行	http://www.stat.go.jp/training/2kenkyu/2-2-664.html

研究機関	研究成果			作成職員		掲載URL
	名称() 内は発表媒体・発表機会)	発表年度	要旨抜粋等	所属等 (研究実施時点)	氏名	
独立行政法人統計センター	サービス業基本調査における経理項目の補定法	独立行政法人統計センター	<p>本稿は、サービス業基本調査における給与支給総額、収入額等の経理項目を対象として、欠測値の補定法に関する検討を行った結果を取りまとめたものである。</p> <p>検討では、平成16年サービス業基本調査を対象とし、経理項目に対する回帰補定の改善の可能性を調べた。現行の補定法では、例えば、給与支給総額は、1雇用者当たり給与支給総額に当該事業所の雇用者数を乗じた値で補定する。つまり、雇用者数を説明変数とし、それに一定の比率(1雇用者当たり給与支給総額)を乗じる形であることから、定数項のない回帰モデルに相当する。ここで、1雇用者当たり給与支給総額は、事業所の産業、規模、地域などによる層ごとに算出される。</p> <p>まず予備的な分析において、定数項の適用や変数変換の必要性が示唆された。さらに、回帰モデルを構築するためには、完全データのサイズが小さい層が多いという問題があった。本研究では、線形回帰モデルを拡張した対数線形回帰モデルと1人当たり回帰モデルの当てはめを行い、回帰の予測残差に基づくモデル比較の方法を提案した。予測残差による評価や外れ値の検出に一定の効果があることを示したが、さらなる検討や業種業態などの背景領域の知識が必要であると思われる。また、層の構築については、層化されていない単回帰モデルの残差を多段階別分析によって分類する方法と、現行の補定法における層化変数を説明変数に入れた重回帰・数量化I類モデルによって層化因子の効果を確認する方法を提案した。どちらも回帰モデルの構築と統合的な方法であるが、2つの方法で得られた層化が一致しない場合がある。業種業態などの背景領域の情報を取り入れるといった専門的な判断が必要と考えられる。</p>	統計センター研究センター非常勤職員(シンフォニカ主任研究員) 統計センター研究センター	村田 磨理子 畠山 昌子 磯部 祥子 亀本 薫	https://www.nstac.go.jp/services/pdf/sankousiryoku2003.pdf
日本銀行	「分からない」という回答から分かること？ --短観「物価見通し」における無回答バイアス-- (日本銀行ワーキングペーパーシリーズ No.18-J-5 2018年6月)	平成30年度	<p>本稿は、短観の「物価見通し」における「分からない」という回答の中から、真の意味では「分かる」と推察される回答を機械学習の手法を用いて識別する。そのうえで、真の意味では「分かる」と識別された企業について、何らかの理由で回答されなかった反実仮想的な「物価見通し」を推定する。そして、その推定結果にもとづいて、無回答バイアスの大きさを評価する。分析の結果として、次の四点を指摘することができる。第一に、真の意味では「分かる」が、何らかの理由で「分からない」と回答した企業は、きわめて少ない。第二に、真の意味では「分かる」と識別された企業は、真の意味でも「分からない」と識別された企業に比べて、相対的に企業規模が小さく非製造業に多い。第三に、真の意味では「分かる」と識別された企業の反実仮想的な「物価見通し」は、公表集計値と統計的に有意に異なる。第四に、以上を踏まえると、短観の「物価見通し」の無回答バイアスは、統計的には無視できる。</p>	日本銀行統計調査統計局	宇野洋輔 安達孔	https://www.boj.or.jp/research/wps_rev/wps_2018/data/wp18i05.pdf

研究機関	研究成果			作成職員		掲載URL
	名称()内は発表媒体・発表機会)	発表年度	要旨抜粋等	所属等(研究実施時点)	氏名	
日本銀行	ビジネスサーベイにおける欠測値補完の検討—全国企業短期経済観測調査(短観)のケース— (日本銀行ワーキングペーパーシリーズ No.12-J-8 2012年8月)	平成24年度	短観における現行の欠測値補完方法とその代替方法について、2004年から2010年までのデータを基にシミュレーションを行い、統計精度を比較した。その結果、設備投資額、売上高、経常利益の主要3項目の全てにおいて、「現行の補完方法より精度が高いかほぼ同等であるような代替方法が存在する」という結論が得られた。	日本銀行統計調査統計局	平川貴大 鳩貝淳一郎	https://www.boj.or.jp/research/wps_rev/wps_2012/wp12i08.htm/
日本銀行	「全国企業短期経済観測調査」における欠測値補完の検討 (日本銀行ワーキングペーパーシリーズ No.12-J-8 2001年8月)	平成13年度	短観の代表的な計数項目である設備投資、売上高、経常損益について、①現行の方法(「平均値補完」、Mean Imputation)、②回答のある直近調査回の値の代入(「横置き補完」、Cold Deckの一手法)、③前回調査回の値に層内の回答者の前回調査比(3月調査時における計画の場合は、初回調査であることから前年比)伸び率を乗じた値の代入(「伸び率補完」、Hot Deckの一手法)の3つを選択肢として比較検討したところ、現行「平均値補完」よりも代替手法の方が精度は高いとの結果になった。	日本銀行調査統計局経済統計課	宇都宮 浄人 園田 桂子	https://www.boj.or.jp/research/wps_rev/wps_2001/cwp01j11.htm/