

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

Trends of Imputation Methods for Statistical Surveys in Foreign Countries  
and the Scheme of Methodologies

坂下 信之

統計研究研修所統計研修研究官

Nobuyuki Sakashita

SRTI Senior Researcher for Statistical Training

平成 30 年 7 月

July 2018

総務省統計研究研修所

Statistical Research and Training Institute (SRTI)

Ministry of Internal Affairs and Communications

本ペーパーは、総務省統計研究研修所統計研修研究官が、その責任において行った統計研究の成果を取りまとめたものであり、その内容については、総務省統計局及び統計研究研修所の見解を表したものではありません。本ペーパーの内容については、執筆者に問い合わせ願いたい。

# 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

坂下 信之

## 概要

政府統計の精度維持・向上が喫緊の課題となる中で、欠測値や外れ値への対応はその重要な要素である。世界的にも 1980 年代半ばから今日でも参照される文献が現れ、今世紀に入ってから、国連などの場で盛んに議論されるようになってきている。

このため、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、必ずしも共通の了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。

キーワード：データ・エディティング、欠測値補完、インピュテーション、精度、ホット・デック法

# Trends of Imputation Methods for Statistical Surveys in Foreign Countries and the Scheme of Methodologies

Nobuyuki Sakashita

## Abstract

While maintenance and enhancement of accuracy in official statistics are emerging as urgent issues, treatment of missing data or outliers, is their substantial element. Looking around the world, those literatures referenced until today appear from the mid-1980s, and since the beginning of this century, the matter has been actively discussed at the United Nations and other places.

In this concern, we reviewed literatures, from the perspective of how the scheme of methodologies in coping with missing data has been organized, as well as the subsequent trends succeeding the last year's research.

We found that methods of imputation are still mostly conventional, such as hot-deck or ratio imputation, in practical adoption, although advanced methods are continuously under examinations, and, in a case, hot-deck method is adopted after examining advanced methods. Systems for editing and imputation are being updated in Canada and Austria, and are adopted in some other countries.

Common agreement in the scheme of methodologies is not necessarily prevalent. There, however, exists abundant stock of literatures since 1980, and, especially from the end of the 1990s, statistical institutions are increasingly disseminating and exchanging their own opinions.

Keywords: Data Editing, Imputation of Missing Data, Accuracy, Hot-deck Method

## 0. はじめに

政府統計の精度維持・向上が喫緊の課題となる中で、欠測値や外れ値への対応はその重要な要素である。世界的にも 1980 年代半ばから今日でも参照される文献が現れ、今世紀に入ってから、国連などの場で盛んに議論されるようになってきている。

昨年度は、その時点での諸外国におけるデータ・エディティング、特に欠測値補完の状況について、公開情報として入手可能な文献を調査し、現時点で行われている欠測値補完<sup>1</sup>はホット・デック法、比率代入法などの伝統的な手法が多く、回帰代入法や平均値代入法を用いている国（機関）も多い一方で、高度な分布モデルに基づいた手法は米国やオーストリアなどで検討されているなどの結果を得た。その結果は坂下（2017）としてまとめたが、入手可能な文献をとりあえず集めることを重視したため、方法の整理、訳語の定型化などは行っていない。

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。

以下はその結果であり、その構成は、1. がアメリカ合衆国及びカナダ、2. が欧州その他、3. が手法の体系、4. が実施者による解説、5. がまとめとなっている。

## 1. アメリカ合衆国及びカナダ

米国では、センサス局に属する The Center for Statistical Research & Methodology (CSRM) などで、欠測値補完に関するさまざまな研究が続けられている。また、カナダは相対的には小国ながら、公的統計の分野の研究及び開発においてさまざまな発信を行っている。

### （センサス局）

米国の人口センサスについては、CSRM の業務報告（CSRM (2016), CSRM (2017)）によると、2010 年センサスのデータを用いて、無回答<sup>2</sup> のフォローアップ (Nonresponse follow-up, NRFU) として、無回答世帯への接触方法を見直すとともに、行政情報を利用することで、聞き取り調査 (proxy response) より正確な結果を得られるかどうかの検討を行った。このプロジェクトは 2017 年に完結したが、Mulry and Keller (2017) によると、どちらがより正確かを決めるのは簡単ではないこと、行政データが歳入庁 (IRS) やメディケアから来ているにもかかわらず、聞き取り調査の住戸でのカウントが正しい位置にある割合が行政情報よりも高いこと、評価できない情報の割合は行政情報の方が高いことが結論となっている。

また、当初のすべての行政情報を用いる計画から、最良の行政情報をモデルによって特定することに変更しており、Morris et al. (2016) 及び Morris (2017) において、回答の得られ

<sup>1</sup> 「欠測値（の）補完」、「インピュテーション」、「補定」などの類義語があるが、ここでは欠測値への対処法一般を「欠測値（の）補完」とし、具体的な手法に対して「インピュテーション」の語を用いることとする。

<sup>2</sup> 坂下 (2017) では、「無回答」、「非回答」の用語が混在していたが、高橋 (2017) に従って調査した結果としての欠測については「無回答」で統一した。

なかった世帯について、代替に用いることのできる行政データをモデルによって特定する手法を研究している。それによると、米国には、単一の登録システムは存在しないので、このための情報は、センサス局の行政情報研究応用センター (Center for Administrative Records Research and Applications, CARRA) が IRS、メディケア&メディケイド (CMS)、TARGUSinfo などから行政情報を集めている。

坂下 (2017) で触れた月次卸売調査 (Monthly Wholesale Trade Survey, MWTS) の欠測値補完手法に関する研究も継続されている。経緯を見ると、複数の期にわたるプロジェクトとなっており、前期のプロジェクトについては、CSRSM (2014) に新たな補完手法を試すためのシミュレーション研究を行っており、現行の比率に基づいた単一代入法に対し、モデルに基づいた、あるいはモデルに支援された、あるいはベイズ的アプローチを考慮した研究を行うことを提言するとの記述があり、その具体的な検討として Klein et al. (2014) が合同統計会議 (Joint Statistical Meetings, JSM) で発表されている。今期の研究は、CSRSM (2016) 及び CSRSM (2017) が記述しており、現実的な合成 (simulated 又は artificial) を構成することによる MWTS の推定値の改善を研究している。補完手法としては、ランダムフォレストを条件モデルとした連鎖方程式による多変量インピュテーション (2016 年度)、ランダムフォレストを条件モデルとする順次回帰 (sequential regression) インピュテーション (2017 年度) が使用されている<sup>3</sup>。

米国の経済センサスの欠測値補完については、坂下 (2017) で、分類体系の変更により、分野によって異なっている現行の手法が使えなくなることを報告した。CSRSM (2016) によると、経済センサスの生産データについて、expansion (比率按分) 法、(ランダム及び最近隣) ホット・デック法及び順次回帰多変量法 (sequential regression multivariate imputation, SRMI) を検討しており、その結果である Bechtel et al. (2015) では、検討した手法の中ではいずれかのホット・デック法が最良の妥協案 (best compromise) と思われるが、2種類のホット・デック法のどちらが良いのかは、分野によって異なるとして、分類木を用いた判断によりインピュテーション・セルを作成することを提案している<sup>4</sup>。

地域社会調査 (American Community Survey, ACS) については、CSRSM (2016) によると、CoreLogic 社の商用ビッグデータから得られた情報を用いて、「性向が同じ層 (equal-propensity strata)」を構成し、抵当権の予測値の精度を上げることを検討している。

所得及びプログラム参加調査 (Survey of Income and Program Participation, SIPP) については、CSRSM (2017) によると、SIPP のような長期調査で見られる対数正規構造に基づいた、条件付きで独立な関係を用いたモデルにより、欠測値補完にも応用できる小地域推計を形成することについての論文 (Thibaudeau et al. (2017)) を公表している。なお、CSRSM (2016) では、同じ論文と思われるものが欠測値を予測することを目的とするものとして作成中と

<sup>3</sup> 論文になっていないため詳細不明だが、価格が収集されていない対象をインピュテーションで補うことにより、母集団を擬似的に再構成しているものと思われる。

<sup>4</sup> 坂下 (2017) では比率による手法と SRMI を検討していると報告したが、結局ホット・デック法に落ち着いたことになる。

されているので、小地域統計にも応用できるものとして目的を変更したものと思われる。なお、2013年のJSMで報告されたMcMillan(2013)によると、SIPPの欠測値補完は、候補値を一行に並べた完全に確定的なホット・デック法によっているが、この方法は、変動を過小評価するとして、多重代入法やSRMIとの比較を行っており、以後SIPPの欠測値補完を巡ってさまざまな研究が行われている模様である<sup>5</sup>。

#### (経済分析局 (BEA))

Xu et al. (2017) によると、経済分析局 (Bureau of Economic Analysis, BEA) では、多国籍企業についての諸調査のデータ・エディティングと欠測値補完を従来手作業で行っていたが、負荷が大きいとしてカナダ統計局が開発した自動エディティングとインピュテーションのシステム Banff<sup>6</sup> の適用を検討した。その結果、(1) Banffのエディティングは「ハード」(日本で「要訂正」と称しているもの)だが BEA の従来のエディットは多く「ソフト」(同じく「要確認」と称しているもの)であること、(2) Banffのエディティングは線形だが BEA のものは非線形のものが多いこと、(3) Banffのエディティングは数量データが対象だが BEA のものはカテゴリ変数が多いことなどの制約があるが、これらの対策を取れば、自動エディティングにかけたデータの統計量は従来の手作業によるものに極めて近いことが示されたとの結果を得て、部分的に採用している。

#### (農業統計サービス (NASS))

Miller (2017) は、全米農業統計サービス (National Agricultural Statistics Service, NASS) の農業センサス (COA) の最近隣法によるインピュテーション<sup>7</sup> において、新たに加わった調査項目に対しては以前の調査から得られるドナーが存在しないため、これを新たに作り出す手法を論じている。その手法として、Raghunathan et al. (2001)<sup>8</sup> で解説されている完全条件付指定 (Fully Conditional Specification, FCS) 法<sup>9</sup> に基づいてミシガン大学で開発された IVEware<sup>10</sup> を用いているとしている。この IVEware は最近 NASS の他の複数の調査で適用されており、具体例として2014年の農地の借地権、所有権及び遷移調査 (Tenure, Ownership, and Transition of Agricultural Land survey, TOTAL)、2016年の地域食品販売活動調査 (Local Food Marketing Practices Survey) を挙げ、2017年の有機食品調査 (Organic Food Survey) での適用も計画中としている。

#### (カナダ)

<sup>5</sup> 一部は坂下 (2017) にも記述している。

<sup>6</sup> Banffについては、野村総合研究所 (2013) 第4章に解説がある。

<sup>7</sup> COAのインピュテーション法については坂下 (2017) を参照のこと。

<sup>8</sup> Raghunathan et al. (2001) は、変数間の制約条件を満たしつつインピュテーションを行う手法を論じた論文で、後出のDe Waal (2017) でも比較的詳しく紹介されている。

<sup>9</sup> 高橋・伊藤 (2014)、高橋・渡辺 (2017) に解説がある

<sup>10</sup> <https://www.src.isr.umich.edu/software/> より入手できる。

Thomas (2017) によると、カナダでは、BEA の所でも言及した Banff の改良が続けられている。改良は 2018 年から 2020 年にかけて実装される予定であり、現在はカナダ統計局と他の省庁との協同体制を形成しているところである。この論文は、現在の Banff の各モジュールとユーザーについての簡単な解説を含む。

## 2 . 欧州その他

### ( オーストリア )

オーストリアの欠測値可視化と補完のための R パッケージ VIM については、坂下 (2017) でも報告したところであるが、Kowarik and Meraner (2017) によると、その機能のうち、k-最近隣法について、データ間の距離を計算する際の各変量につけるウェイトの決定の自動化を検討している。ランダムフォレスト法とラッソ回帰を評価し、一般にはウェイト付けした方がしないより良いとの結果を得ているが、同時にまだ初期段階の結果だとしている。

また、Koenig (2017) によると、オーストリアでは、2011 年以降、レジスタ・ベースの人口センサスを実施しているが、人口レジスタにおける職業の欠測が多く精度が低いため、新たな行政記録を用いることによる精度向上を検討している。2011 年センサスでは欠測値の補完はホット・デック法で行われたが、次回センサスでの手法は未定である。

### ( オランダ )

オランダ統計局が欧州統計家会議 (CES) に提出した De Waal (2017) は、変数間の関係や既知の総計など制約条件下の補完について論じたものだが、欠測値の補完全般についての良い文献調査となっている。

### ( ドイツ )

ドイツの人口センサスについては坂下 (2017) でも報告したが、Spies (2017) によると、次回センサスに向けて異なった欠測値補完手法を評価している。現在検討されているのは、ドイツのインピュテーションツールと CANCEIS の 2 種類の近隣法、及び多重代入法の適用可能性である。CANCEIS については、2011 年センサスでも検討されたが不採用となり、同年の住宅センサスでは使用されたとしている。ただし、2011 年に CANCEIS が採用されなかったのは技術的理由であり、今回は欠測値の補完業務が連邦統計局で行われるため、費用対効果では CANCEIS に有利と述べている。この論文は、それぞれの手法を解説し長所と短所を論じている。

### ( フランス )

フランスの国立統計経済研究所 (INSEE) が CES に提出した Deroyon and Favre-Martinoz (2017) では、外れ値への対処法として、Kokic and Bell 法 (Kokic and Bell (1994) ) と Conditional Bias 法 (Beaumont et al. (2013) ) による winsorization を比較シミュレーションし、両者は推



計値の平均二乗誤差を小さくすることができ、ほぼ同じ結果となるとしている。シミュレーション用データとしては、社会保障のデータから、賃金構造及び労働コスト調査 (Enquête sur le coût de la main d'oeuvre et la structure des salaires, Ecmoss) の手法により標本抽出したものをを用いている。

( 欧州統計システム (ESS) )

欧州統計システム (ESS) がデータ妥当性検証方法の一般化を目指す ValiDat Foundation プロジェクトについては、坂下 (2017) に記したが、その後を受けて設立された ValiDat Integration について Weichert (2017) が報告している。ValiDat Foundation プロジェクトでは、その成果として各国で一般的な手法を集めたデータ検証手法についてのハンドブック Di Zio et al. (2016) を作成しており、ValiDat Integration はその改良を目的のひとつとしている。

また、Gramaglia and Tronet (2017) は、加盟国から欧州統計局 (Eurostat) に送られるデータの妥当性検証を近代化するために、複数の側面からのアプローチを取っていると報告している。具体的には、手法 (Methodology)、処理 / ガバナンス (Processes/Governance)、情報基準 (Information Standards)、情報技術 (Information Technology)、人材 (Human Resources) の 5 側面を挙げ、手法について共通の定義、分類、尺度を提供し、更なる発展のための枠組みとするために上記ハンドブックを作成したことを記し、各国の妥当性検証処理をどのように近代化できるかの実地の調査のために、ValiDat Integration が資することを期待されているとしている。

( イスラエル )

イスラエルは 2008 年に人口の 100% をカバーする行政データと約 17% の世帯に対する標本調査を結合した人口センサスを実施したが、行政データは完全ではなく、エディティングと欠測値の補完が必要であった。Klejman (2017) は、収入データの欠測値補完について解説しており、それによると、軍人と管理人については正規分布などを仮定した統計的なインピュテーション、それ以外については CANCEIS を用いた最近隣法が用いられている。

### 3 . 手法の体系

「0 . はじめに」に述べたとおり、坂下 (2017) では、宇都宮・園田 (2001) に倣って、用語などは基本的に原文に従い、方法の整理、訳語の定型化などは行わなかった。一方で、宇都宮・園田 (2001) では、代表的な欠測値補完方法を“ Mean Imputation”、“ Hot Deck”、“ Cold Deck”、“ Multiple Imputation”、“ Substitution”として、それぞれ簡単に解説しており、また、その後に公表された平川・鳩貝 (2012) は、代表的な欠測値補完方法を、「欠側項目のデータのみ」を使うものと「欠側項目のデータ + 欠側項目以外のデータ」を使うものに分け、更に前者をそのデータを回答企業から取得するもの (ランダム・ホットデック・ドナー補完、平均値補完など) 、欠側企業から取得するもの (横置き補完、伸び率補完など) に分け

ている。欠側項目のデータ + 欠側項目以外のデータを使うものの具体例としては、項目間比率補完を挙げている。ただし、ここでも宇都宮・園田 (2001) を引用する形で欠測値補完の方法の整理についてははっきりとしたコンセンサスはないとしており、理論的・体系的というよりは実務的な整理となっている。

実際に、日本における公的統計の現場でも、必ずしも欠測値の補完について理論的・体系的に整理した上で実務が行われているわけではない。欠測値の補完について公開し、外部と議論することが少なかった時代にはそれでも実務上の不都合はあまりなかったと考えられるが、政府統計の精度維持・向上が喫緊の課題となる中では、共通の用語の欠如が支障となるおそれがある。このため、以下では海外の文献が、欠測値の補完手法をどのように整理しているかを概観する。

( Sande (1982) )

今回調査した文献で最も古い Sande (1982) は、欠測値の補完手法として、(1) アドホックな値の使用 (手作業又はルールによる) (2) 事後層化した後の各層の平均値又は他の代表値 (カテゴリー変数における最頻値等) の使用、(3) 変量間の関係のモデル化、(4) 履歴データの使用 (前月値や前年値等) (5) 他の情報源 (行政データなど) の使用、(6) 同じ調査の個別データをマッチング<sup>11</sup> して使用 (シーケンシャル・ホット・デック又はランダムな選択) (7) 同じ調査の類似した性質の個別データを使用、(8) ハイブリッド方式 を挙げ、各手法に比較的細かい説明を加えている。注目されるのは、(6) の中で「この手法はホット・デックと呼ばれることが多いが、文献においてホット・デックの定義についての合意はない。」とした上で、「私はこれをランダム又はランダムに準ずる手順により、同じ調査から得た値を用いて欠測値に提供することとする。」と述べていることで、実際、(7) は通常はホット・デック法の一種の「最近隣ホット・デック法」と呼ばれているものであるが、ここでは別掲され、「数量データでマッチングする時に適した手法」とされている。

( Kalton and Kasprzyk (1986) )

海外の文献で引用されることの多い Kalton and Kasprzyk (1986)<sup>12</sup> では、通常、全無回答 (Total Nonresponse) にはウェイト修正、項目無回答にはインピュテーションで対応するとした上で、欠測値の補完手法として、(a) 演繹的 (deductive) インピュテーション、(b) 全体の平均による (Overall mean) インピュテーション、(c) 分類内の平均による (Class mean) インピュテーション、(d) 全体でのランダムな (Random overall) インピュテーション、(e) 分類内でのランダムな (Random within classes) インピュテーション、(f) シーケンシャルなホット・デック (Sequential hot-deck) インピュテーション、(g) 階層的なホット・デック (Hierarchical hot-deck) インピュテーション、(h) 回帰による (Regression) インピュテーション

<sup>11</sup> 英語の動詞は “match” だが、「マッチする」では日本語として不自然なので、「マッチングする」で統一する。

<sup>12</sup> 概要を附録 1 として採録する。

ン、(i) 距離関数によるマッチング (Distance function matching) を挙げた上で<sup>13</sup> 手法を分類する視点を論じている。その視点は、残差を加えるか (加えるならばどのような形の残差か) と補助情報の用い方は分類を表すダミー変数の形かそれとも直接に回帰する形か、である。また、この論文では、回帰によるインピュテーションの変種として、予測平均値マッチング (Predictive mean matching) に触れている。これは、回帰による推定値の代わりに、実際のデータが取った値の中で推定値に最も近いものを代入することにより、有り得ない値となることを避ける手法で、ホット・デック法の一つとされる<sup>14</sup>。その後、欠測値の補完が及ぼす変数間の関係への影響と多重代入法が論じられている<sup>15</sup>。

( Longford (2005) )

Springer 社から出版されている Longford (2005) は、単一値によるインピュテーションとして、平均値インピュテーション、他の値によるインピュテーション (時系列データの前期値の延長など)、最近隣インピュテーション、ホット・デック法、ウェイト補正、回帰インピュテーション、専門家による判断を挙げている。ただし、同書は分析を主眼とし、調査設計を分析に「正直に」反映するために多重代入法を推奨する内容となっており、単一インピュテーション (代入法) について必ずしも体系的な分類を行っているわけではない。また、ここでも Sande (1982) と同様に最近隣法をホット・デック法とは (近接したものと はしているが) 別に扱っている。

( Andridge and Little (2010) )

Andridge and Little (2010) は、ホット・デック法について「実地には広範囲に用いられているが、他の手法に比べて背景の理論が開発されていない」とした上で、詳しい解説を行っている。ここではホット・デック法を「無回答者 (レシピエント) の一つないしそれ以上の変数を、双方で観察された特性において無回答者 (レシピエント) と類似する回答者 (ドナー) から観察された値で置き換えること」とし、例としてドナーをランダムに選ぶ「ランダム・ホット・デック法」と何らかの尺度により一つのドナー (通常は「最近隣」) を特定してそこから代入する「確定的ホット・デック法」を挙げている。一方、平均値などの統計量を代入する手法は「いくつかの共通の性質を持つが、われわれはこれをホット・デック法とは考えない」としている。また、ホット・デック法の名称は、「データ・ストレージとしてパンチカードを用いたことに由来し、無回答者に対してドナーとなりうるカー

<sup>13</sup> それぞれの解説は先行する Kalton and Kasprzyk (1982) の方が詳しいが、分類が多少異なっている。また、おそらく Little (1986)、Rubin (1986) の発表を受けて、予測平均値マッチングについての解説が詳しくなっている。

<sup>14</sup> 坂下 (2017) でも予測平均値マッチングの実例がいくつか存在していたが (ドイツの農業センサス、イタリアのビジネス構造統計)、今回元論文とされた Little (1986)、そこでオリジナルの提案とされている Rubin (1986)、それを展開した Little (1988) は未入手である。

<sup>15</sup> 多重代入法については、 $m=2$  程度の簡単なものに留まっているが、これは当時 (1986 年) の計算機の能力の限界のためと思われる。

ドの束 (deck) を指している。」、「現在処理しているものであるため束が “hot” であると言う。これに対して、ドナーとして予め用意した値を用いることが “cold deck” である。」としているのが注目される<sup>16</sup>。

また、同論文では、古典的なホット・デック法はセンサス局で人口動態調査 (Current Population Survey, CPS) のために開発され、同局や他の政府統計部局で用いられていると述べ、具体例として教育省の全米教育統計センター (National Center for Education Statistics, NCES) で最近行われた 20 の調査のうち、11 がホット・デック法 (シーケンシャル又はランダム)、残り 9 が確定的インピュテーション (平均値など)、コールド・デック法あるいはベイジアン法を用いているとしている。適用されているものは、シーケンシャル・ホット・デック法が多いが、最近はより複雑なマッチング手法や多変量の欠測を扱うためのより良い手法が使われているとして、ドナー・プールの作成方法、ウェイト付きのホット・デック法などについて論じている。

(国内の文献)

なお、我が国のホット・デック法に関する文献としては、統計センターで行われた研究をとりまとめた川崎・小林 (1988) があり、ここではホット・デック法を「同じ標本の中の欠測でない値によりこれを補う方法」、コールド・デック法を「予め外から与えられた一定の値により欠測の決め付けを行う方法」としている。ホット・デック法のいくつかの種類についても言及されているが、この研究はその年度限りのものと見られ、その後継承された形跡がない。

#### 4. 実施者による解説

統計の実施者が欠測値の補完手法を解説したもので、一般に公開されているものとしては、米国の労働統計局とセンサス局で作成した U. S. Bureau of Labor Statistics and U. S. Census Bureau (2006)、医療研究・品質調査機構 (AHRQ) の Agency for Healthcare Research and Quality (2007) がある。また、若干古くなるが、オランダ統計局から Schulte Nordholt (1997)、De Waal (2000) などの資料が公開されている。

(米国労働統計局及びセンサス局)

U. S. Bureau of Labor Statistics and U. S. Census Bureau (2006) は、前記 CPS の手法全体を解説しており、その中で欠測値の補完については、(1) 関係性によるインピュテーション (演繹的インピュテーションと同じもの。この呼称も文献により異なる。)、(2) 前期も用いられた標本であれば、時系列的インピュテーション、(3) ホット・デック法 を用いているとしている。ホット・デック法によるインピュテーションはセル内でシーケンシャルに行われ

<sup>16</sup> 坂下 (2017) にも記したように、ホット・デック、コールド・デックの用語の用法は世界的にも必ずしも一定していず、日本の現場でも混乱が見られるが、このように同じ調査の他のデータから代入する手法をホット・デックと呼ぶのが主流である。なお、Andridge and Little (2010) の概要を附録 2 として採録する。

ており、対象となる変量の順序は、世帯、人口属性、労働力属性などとなっている。

(米国医療研究・品質調査機構 (AHRQ))

Agency for Healthcare Research and Quality (2007) は、AHRQ で実施している医療費支出パネル調査 (MEPS) の欠測値補完手法についての報告である。インプューションはセルごとにウェイト付きのホット・デック法で行われ、ドナーが対象データより少ないときは両者が等しくなるまでセル統合が行われる。この文献では具体的な例をもとにプロセスを詳しく解説している。

(オランダ)

Schulte Nordholt (1997) はオランダの所得構造調査 (Structure of Earnings Survey, SES) の欠測値補完手法を紹介している。そこでは、SES では変数の分布が重要な関心事項であるが、確定的インプューションでは分布を歪めてしまい、統計的回帰インプューションは簡単な解決策だが必ずしも可能性のある (feasible) 値とならない、それ故ホット・デック法がより良い代替案であると述べている。

De Waal (2000) は、オランダの統計調査をビジネス調査と社会調査に分け、それぞれに用いられる欠測値の補完手法を解説している。

ビジネス調査については、演繹的 (論理的) インプューション、各種的回帰インプューション (平均値など簡素なものが多い)、比率インプューション、履歴によるインプューションが多く、時系列によるインプューション (移動平均法など) も用いられるが珍しく、(他のデータセットから得た同じ企業の値を代入する) コールド・デック法はよく用いられ、ホット・デック法は珍しいと述べている。

社会調査については、Schulte Nordholt (1998) (未入手であるが、SES については Schulte Nordholt (1997) と共通)、Hilbink et al. (1995) (オランダ語の内部資料) に沿って欠測値の補完手法を解説している。具体的には、住宅需要調査 (Housing Demand Survey) では、変数がいくつかのグループに分けられ、インプューションは各グループの中で行われ、グループの中では欠測値の割合が最も少ない変数にまずインプューションを行う。離散変数には主にランダム・ホット・デック、連続変数にはランダム・ホット・デック又は予測平均値マッチングによるインプューションを行い、異なったグループの関連する変数には同じドナーから得た値を用いると述べている。また、SES で用いるホット・デック法は、ランダム・ホット・デック法は用いるレコードの数が多すぎて効率的でないとして、シーケンシャルなホット・デック法を用い、潜在的なドナーをランダムな順序に並べることにより、ランダムな要素を導入しているとしている。更に、ニューラル・ネットワークを用いたインプューションも検討しているが、その精度が十分かどうかは疑いがあるとしている。欧州共同体世帯パネル調査 (European Community Household Panel Survey, ECHP) では、パネル調査という特性からインプューションは複雑であり、毎回別々にインプューション

を行う方法では収入が奇妙な動きになるので同時にインピュテーションした方が良いが、この方法は新たな回が入ってくると以前の回がすべて変化するため、より単純に、新たな回のインピュテーションでは以前の回を考慮しつつ、以前の回は修正しないとしている。初回のインピュテーションには、分類内のランダム・ホット・デック法が用いられている。雇用と給与に関する調査 (Annual Survey on Employment and Wages) で用いられているランダム・ホット・デック法においては、処理速度などの実用的な理由からすべての潜在的なドナーが用いられているわけではなく、各分類にランダムに選ばれたドナーのストックを作成しているとしている。社会調査の連続変数には線形回帰によるインピュテーションがよく用いられているが、非線形回帰はほとんど用いられていない。

(その他)

なお、未入手であるが、John Wiley & Sons から出版されている Rubin (1987)、Little and Rubin (2002)、De Waal et al. (2011) などがよく引用されている。

## 5. まとめ

以上の調査から得られる傾向をまとめると、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、米国の経済センサスのように多変量の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。また、カナダの Banff、オーストリアの VIM でシステムの改良が続けられており、Banff や同じくカナダの CANCEIS は他国での採用例も見られる。

また、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ることが分かった。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。わが国としても、この分野における先進諸国と共通の土台に立った上で、情報発信、交換を行う必要がある。

## 附録 1 Kalton and Kasprzyk (1986) の概要

著者（当時の肩書き）

Graham Kalton, ミシガン大学調査研究センター (Survey Research Center, University of Michigan)

Daniel Kasprzyk, センサス局人口課 (Population Division, U. S. Bureau of the Census)

掲載誌

Survey Methodology, June 1986, Vol. 12, No. 1, pp. 1-16, カナダ統計局 (Statistics Canada) で発行する査読付き論文誌。

主な内容は以下のとおり。

### 第 1 節 序

無回答には全無回答 (Total Nonresponse) と項目無回答 (Item Nonresponse) があるとして、前者には乗率補正 (Weighting Adjustments)、後者にはインピュテーションを使用すると記述している。

### 第 2 節 乗率補正

2.1 母集団乗率補正 (Population Weighting Adjustments, 層ごとの母集団の大きさが分かっている場合に使用、標本分布の母集団からのずれを補正する事後層化に似ているが目的が異なるとして区別している。) 2.2 標本乗率補正 (Sample Weighting Adjustments, 母集団の大きさを利用しない方法。) 2.3 レーキング比率補正 (Raking Ratio Adjustments, 複数の特性に対する母集団の同時分布が周辺分布についてのみ分かっている場合に使用。) 2.4 回答確率による乗率 (Weighting with Response Probabilities, さまざまな手法が提案されているが、まだあまり採用されていない。) について記述している。

### 第 3 節 インピュテーション

#### (a) 演繹的 (Deductive) インピュテーション

他のアイテムへの回答パターンから確実に演繹できる場合で、理想的なインピュテーションの形である。

(b) 全体の平均による (Overall mean) インピュテーション

回答者全体の平均をすべての欠測値に割り当てる手法。

(c) 分類内の平均による (Class mean) インピュテーション

全標本をインピュテーションに用いる補助変数により分類し、各分類の中で回答者の平均をすべての欠測値に割り当てる。

(d) 全体でのランダムな (Random overall) インピュテーション

回答者全体の中からランダムに標本を選び、その値を無回答者に割り当てる。同一の調査の回答者の値を無回答者に割り当てるホット・デック法のもっとも単純な形である。

(e) 分類内でのランダムな (Random within classes) インピュテーション

このホット・デック法では、各分類の中で選択が行われる。

(f) シーケンシャルなホット・デック (Sequential hot-deck) インピュテーション

CPS での労働力に関する事項で用いられる手順によって説明。この手順は、分類を定めることに始まり、各分類において、インピュテーションされる変数に（おそらく前回調査から取られた）一つの値が割り当てられる。その調査のレコードを順番に見て、そのアイテムに回答があればその値でインピュテーション用の値を置き換え、回答がなければインピュテーションにとって置かれた値を割り当てる。

（以下、ランダムなインピュテーションとの類似、同じドナーを複数回用いることによる短所などを論じている。）

(g) 階層的なホット・デック (Hierarchical hot-deck) インピュテーション

シーケンシャルなホット・デック法の短所を避ける手法。回答者と無回答者は補助変数によって構成された細かい分類に分けられるが、無回答者と回答者とのマッチングは階層的に行われる、すなわち当初のインピュテーション用分類において対応する回答者がいなければ分類が統合され、相対的に低い段階においてマッチングが行われる。

(h) 回帰による (Regression) インピュテーション

回答データをインピュテーションが必要な変数を補助変数から回帰することに用いる。代入する値は予測値か予測値に残差を加えたものである。残差を得る方法は何種類かある。

(i) 距離関数によるマッチング (Distance function matching)

「最も近い」回答者のデータを用いるホット・デック法。ここで「最も近い」は補助変



数に対する距離関数によって測られ、さまざまな距離関数の形が提案されている。

一見すると手法は多岐にわたるように見えるが、一般化すると次のように記すことができる。

$$\hat{y}_{mi} = b_{r0} + \sum b_{rj} z_{mij} + \hat{e}_{mi}$$

ここで、 $\hat{y}_{mi}$ はy値の欠測したi番目のレコードへの代入値、 $z_{mij}$ は補助変数を反映したそのレコードに対する値、 $b_{r0}$ と $b_{rj}$ は回帰係数、 $\hat{e}_{mi}$ は残差でその選び方はインプューテーションの手法による。 $\hat{e}_{mi}$ がゼロであれば予測値が代入され、さもなければ何らかの残差が加えられる。また、 $z_{mij}$ が分類を示すダミー変数で、残差がゼロであれば分類内の平均によるインプューテーションである。分類内でのランダムなインプューテーションは分類内の平均によるインプューテーションに、回答者の値の平均からのずれを示す残差を加えることで得られる等と記述している。(中略)

以下、

3.1 残差の選び方

3.2 インプューテーション用分類又は回帰によるインプューテーション

3.3 インプューテーションの関係性に及ぼす影響

3.4 多重代入法

について記述している。3.2の中で、回帰によるインプューテーションの変種として予測平均値マッチングについて記述し、距離関数によるマッチングに類似したホット・デック法の一つとしている。

#### 第4節 結論 (Concluding Remarks)

ウェイト補正とインプューテーションの違い、補助変数を用いる手法は暗黙のうちに無作為な欠測 (Missing at Random, MAR) を仮定していることとその対策、欠測値をそのままにして扱いを分析者に委ねることの是非などについて記述したのち、もっとも安全な方法は欠測値の量を少なく保つことだと結論している。

## 附録2 Andridge and Little (2010) の概要

著者（当時の肩書き）

Rebecca R. Andridge, オハイオ州立大学生物統計学科 (Division of Biostatistics, The Ohio State University)

Roderick J. A. Little, ミシガン大学生物統計学部 (Department of Biostatistics, University of Michigan)

掲載誌

International Statistical Review 78, pp. 40-64, 国際統計協会(ISI)及び傘下7組織の機関誌。

主な内容は以下のとおり。

### 第1節 序

（ホット・デック法は）実地には広範囲に用いられているが、他の手法に比べて背景の理論が開発されていず、この結果研究者とアナリストがこの手法を適用する方法が限られている。

ホット・デック・インピュテーションは、無回答者（レシピエント）の一つ又はそれ以上の変数の欠測した値を、両者で観察された性質について類似した回答者（ドナー）で観察された値で置き換える。

いくつかのバージョンでは、ドナーは我々が「ドナー・プール」と呼ぶ潜在的なドナーの集合から選ばれる。この手法を「ランダム・ホット・デック法」と呼ぶ。

別のバージョンでは、何らかの尺度により「最近隣」となる一つのドナーを特定し、そこから代入する。この手法を、ランダムさを含まないため「確定的ホット・デック法」と呼ぶ。

他の手法で、個別の値ではなく平均値のような集計量を代入するものは、いくつかの共通する性質があるものの、ホット・デック法とは考えない。

ホット・デック法がよく用いられる背景にはいくつかの理由がある。結果が長方形データ（欠測値を含まずそのまま分析できるデータ）になるため、単純な完全データの手法で

二次的な分析が可能であること、モデルに依存せず、パラメトリックなモデルによる手法のようにモデルの誤りの問題がないこと（ただし、マッチングには暗黙の仮定が入っており、決して仮定に依存していないわけではない）、観察された値を代入するため、可能性のある値しか使用されないこと。

などを記述している。

## 第2節 ホット・デック法の例

歴史的に「ホット・デック」の用語はデータ蓄積のためのコンピュータのパンチカードの使用に由来し、無回答者へのドナーとなりうるカードの束を指す。束が「ホット」だというのは現在処理中のものだからであり、ドナーとして前回データや他のデータセットなど事前に処理されたデータを用いる「コールド・デック」と対比される。古典的なホット・デックの処理は、1947年に始まったCPSの所得保障 (Income Supplement) の項目無回答のために開発され、その後発展した。CPSでは、欠測アイテムを補うのに、シーケンシャルなセルによる手法を用いる。

シーケンシャルなホット・デック法は実際に採用されたものとしてはもっとも普及しているが、最近の適用例ではより複雑なマッチング尺度と多変量の欠測を扱うためのより良い手法を用いている。

などを記述している。

## 第3節 ドナー・プールを作成する手法

ドナー・プールは、ドナーとレシピエントの双方で観察される補助変数に基づいて構成されるとして、以下その構成方法を記述している。

3.1 Adjustment Cellによる方法（この方法では、連続変数もグループ化してインピュテーション用分類を構成する。）

3.2 ドナーをレシピエントとマッチングする尺度（3.1の手法ではマッチング関数が0又は1を取るのを一般化して連続変数をカテゴライズしないで住む手法を定義。）

3.3 インピュートする変数の再定義（例えば補助変数  $S$  との関連が強い値  $Y$  が欠測している場合、 $Y$  そのものではなく  $R=Y/S$  をインピュテーションの対象とすることによる改善が可能。）

3.4 インピュテーションとエディット制約（ホット・デック法は、インピュテーション以外にも、Fellegi-Holt法やNIM法など、エディティング・ルールを満たす値を得るのに用いられる。）

#### 第4節 サンプルング・ウェイトの役割

標本調査のホット・デック法において、乗率を反映する方法を 4.1 シーケンシャル・ホット・デック法及び、4.2 ランダム・ホット・デック法について議論。

#### 第5節 多変量欠測データのホット・デック

多変量データの欠測について、5.1 2 パターンの場合（完全データ及び不完全データの 2 種類のみ考慮すればよい場合）、5.2 単調欠測パターン、5.3 一般的な欠測パターン<sup>17</sup> に分けて記述している。

#### 第6節 ホット・デック推定値の性質

ホット・デック法の性質についての理論的・経験的文献のレビュー。全体を一つのドナー・プールとする単純な方法の場合、データが完全にランダムに欠測 (MCAR) の場合のみ適切な推計値をもたらすこと、そうでない場合の対策（インピュテーション・モデル・アプローチと無回答モデル・アプローチ）、平均値以外の統計量への影響、ホット・デック法に代わる手法（パラメトリック・モデルなど）などについて記述している。

#### 第7節 分散の推定

欠測値をインピュートした値のデータの分析は無回答によってもたらされた不確実性を過小評価している。ホット・デック法によってインピュテーションを行ったデータの適切な分散の推定法には、3つのアプローチがある。(1) 無回答を含んだ明示的な分散の公式、(2) ジャックナイフ法やブートストラップ法のようなリサンプリング法、(3) ホット・デック多重代入法 である。この節では、これらの3つの手法について記述している。

#### 第8節 詳細な例

第3回国民健康栄養調査 (National Health and Nutrition Examination Survey) の実データを用いて、いくつかのインピュテーション手法を比較している。具体的には、Adjustment Cell によるランダム・ホット・デック、予測平均値ランダム・ホット・デック、傾向セル・ランダム・ホット・デック、そしてパラメトリックな回帰インピュテーションである。

#### 第9節 結語

ホット・デック法は現実に項目無回答の対策に広く使われている。その利点は現実の値を代入し、強力なパラメトリック仮定を必要とせず、共分散に関する情報を取り入れることができ、インピュテーションに伴う不確実性を正しく伝えることができれば統計量の良い推計値を与えることであり、欠点は共分散に関する情報を反映するドナーと受けての良

---

<sup>17</sup> 単調欠測パターン及び一般的な欠測パターンについては高橋・渡辺 (2017) を参照のこと。

いマッチングが必要なことであると記述した後、いくつかの課題が示されている。

一つは、Adjustment Cell 法で補助変数の利用が限定されることに対する、共変量情報の利用、二番目は、「スイスのチーズ」(一般的な欠測パターン)への対処、最後は、ホット・デック法によるインプテーションを行った後に、有効な推論を行う方法である。

## 参考文献

- [ 1] 宇都宮浄人・園田桂子 (2001) 「『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。
- [ 2] 川崎茂・小林國夫 (1988) 「ホットデック法の決め付け誤差の評価方法に関する考察」, 統計局研究彙報第 46 号、総務庁統計局・統計センター。
- [ 3] 坂下信之 (2017) 「『諸外国の公的統計における欠測値補完 (インピュテーション) の現状～文献調査～』, リサーチペーパー第 40 号、総務省統計研究研修所。
- [ 4] 高橋将宜・渡辺美智子 (2017) 「欠測データ処理 R による単一代入法と多重代入法」, 統計学 one point 5、共立出版、2017 年 12 月。
- [ 5] 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、総務省統計研修所。
- [ 6] 野村総合研究所 (2013) 『統計データの補完推計に関する調査報告書』(平成 25 年 3 月)
- [ 7] 平川貴大・鳩貝淳一郎 (2012) 「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。
- [ 8] Agency for Healthcare Research and Quality (2007), “Overview of Methodology for Imputing Missing Expenditure Data in the Medical Expenditure Panel Survey”, Methodology Report #19.
- [ 9] Andridge, R. R. and Little, R. J. A. (2010), “A Review of Hot Deck Imputation for Survey Nonresponse”, International Statistical Review 78, pp. 40-64.
- [ 10] Beaumont, J. F., Haziza, D., and Ruiz-Gazen, A. (2013), “A unified approach to robust estimation in finite population sampling”, Biometrika, 100:555–569, 2013.
- [ 11] Bechtel, L., Morris, D. S., and Thompson, K. J. (2015), “Using Classification Trees to Recommend Hot Deck Imputation Methods: A Case Study”, Proceedings of the Federal Committee on Statistical Methodology Meeting, Washington, D.C., December 1-3, 2015.
- [ 12] CSRM (2014), “Annual Report of the Center for Statistical Research and Methodology, Research and Methodology Directorate, Fiscal Year 2014”, U.S. Department of Commerce, Economics and Statistics Administration, U.S. CENSUS BUREAU.
- [ 13] CSRM (2016), “Annual Report of the Center for Statistical Research and Methodology, Research and Methodology Directorate, Fiscal Year 2016”, U.S. Department of Commerce, Economics and Statistics Administration, U.S. CENSUS BUREAU.
- [ 14] CSRM (2017), “Annual Report of the Center for Statistical Research and Methodology, Research and Methodology Directorate, Fiscal Year 2017”, U.S. Department of Commerce, Economics and Statistics Administration, U.S. CENSUS BUREAU.
- [ 15] De Waal, T. (2000), “A Brief Overview of Imputation Methods Applied at Statistics Netherlands”, Report, Statistics Netherlands.

- [16] De Waal, T. (2017), “Imputation Methods Satisfying Constraints”, Work Session on Statistical Data Editing, United Nations Economic Commission for Europe, The Hague, April 2017.
- [17] De Waal, T., Pannekoek, J., and Scholtus, S. (2011), “Handbook of Statistical Data Editing and Imputation”, John Wiley & Sons, New York.
- [18] Deroyon, T. and Favre-Martinoz, C. (2017), “A comparison of Kokic and Bell and conditional bias methods for outlier treatment”, Work Session on Statistical Data Editing, United Nations Economic Commission for Europe, The Hague, April 2017.
- [19] Di Zio, M., Fursova, N., Gelsema, T., Gießing, S., Guarnera, U., Petrauskienė, J., Quensel-von Kalben, L., Scanu, M., ten Bosch, K.O., van der Loo, M., and Walsdorfer, K. (2016), “Methodology for Data Validation”, Revised edition. ESSnet ValiDat Foundation.
- [20] Gramaglia, L. and Tronet, V. (2017), “The modernisation of validation in the ESS – a multidimensional approach”, Work Session on Statistical Data Editing, United Nations Economic Commission for Europe, The Hague, April 2017.
- [21] Hilbink, K., Hoogteijling, E., Van de Pol, F., and Schulte Nordholt, E. (1995), “Algoritme voor Imputatie bij het JWL (Algorithm for Imputation of the Annual Survey on Employment and Wages)”, Internal report, Statistics Netherlands.
- [22] Kalton, G. and Kasprzyk, D. (1982), “Imputing for Missing Survey Responses”, Proceedings of the Section on Survey Research Methods, American Statistical Association, 22-31.
- [23] Kalton, G. and Kasprzyk, D. (1986), “The Treatment of Missing Survey Data”, Survey Methodology 12, pp. 1-16.
- [24] Klejman, Y. (2017), “Editing and Imputing Income Data in Integrated Census: lessons learned from the 2008 Census - toward the 2020 Census”, Work Session on Statistical Data Editing, United Nations Economic Commission for Europe, The Hague, April 2017.
- [25] Klein, M., Lineback, J. F., and Schafer, J. (2014), “Evaluating Imputation Techniques in the Monthly Wholesale Trade Survey”, Proceedings of the Joint Statistical Meetings.
- [26] Koenig, S. (2017), “Improving the Census Core Topic Occupation by Using New Administrative Data Sources”, Work Session on Statistical Data Editing, United Nations Economic Commission for Europe, The Hague, April 2017.
- [27] Kokic, P. N. and Bell, P. A. (1994). “Optimal winsorizing cut-offs for a stratified finite population estimation”, Journal of Official Statistics, Vol.10, No.4, 1994. pp. 419–435.
- [28] Kowarik, A. and Meraner, A. (2017), “An automatic procedure for selecting weights in kNN imputation”, Work Session on Statistical Data Editing, United Nations Economic Commission for Europe, The Hague, April 2017.
- [29] Little, R.J.A. (1986), “Missing Data in Census Bureau surveys”, Second Annual Census Research Conference, 442-454.

- [30] Little, R. J. A. (1988), “Missing-data adjustments in large surveys”, *Journal of Business & Economic Statistics*, Vol.6, No.3 (1988), 287-296.
- [31] Little, R. J. A. and Rubin, D. B. (2002), “Statistical Analysis with Missing Data (second edition)”, John Wiley & Sons, New York.
- [32] Longford, N. T. (2005), “Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician”. Springer, New York. MR2171708.
- [33] McMillan, S. (2013), “Comparison of Imputation Methods in the Survey of Income and Program Participation”, *Proceedings of the Joint Statistical Meetings*.
- [34] Miller, D. (2017), “Creating an Initial Donor Pool for New Questions in the Census of Agriculture”, *Work Session on Statistical Data Editing*, United Nations Economic Commission for Europe, The Hague, April 2017.
- [35] Morris, D. S., Keller, A., and Clark, B. (2016), “An Approach for Using Administrative Records to Reduce Contacts in the 2020 Census”, *Statistical Journal of the International Association for Official Statistics*, 32(2): 177-188.
- [36] Morris, D. S. (2017), “A Modeling Approach for Administrative Record Enumeration in the Decennial Census”, *Public Opinion Quarterly: Special Issue on Survey Research, Today and Tomorrow*, 81(S1): 357-384.
- [37] Mulry, M. H. and Keller, A. (2017), “Comparison of 2010 Census Nonresponse Follow-Up Proxy Responses with Administrative Records Using Census Coverage Measurement Results”, *Journal of Official Statistics*, 33(2): 455–475.
- [38] Raghunathan, T. E., Lepkowski, J. M., Hoewyk, J. V., and Solenberger, P. (2001), “A Multivariate Technique for Multiply Imputing Missing Values using a Sequence of Regression Models”. *Survey Methodology*, 27, 85-95.
- [39] Rubin, D. B. (1986), “Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations”, *Journal of Business and Economic Statistics* 4(1):87-94 · January 1986.
- [40] Rubin, D. B. (1987), “Multiple Imputation for Nonresponse in Surveys”. John Wiley & Sons, New York.
- [41] Sande, I. G. (1982), “Imputation in Surveys: Coping with Reality”, *The American Statistician*, 36, 145-152.
- [42] Schulte Nordholt, E. (1997), “Imputation in the New Dutch Structure of Earnings Survey (SES)”, *Work Session on Statistical Data Editing*, Statistical Commission and Economic Commission for Europe, Prague, October 1997.
- [43] Schulte Nordholt, E. (1998), “Imputation: Methods, Simulation Experiments and Practical Examples”, *International Statistical Review*, 66, 157-180.



- [44] Spies, L. (2017), “Possible imputation procedures for the Census 2021”, Work Session on Statistical Data Editing, United Nations Economic Commission for Europe, The Hague, April 2017.
- [45] Thibaudeau, Y., Slud, E., and Gottschalck, A. (2017), “Modeling Log-Linear Conditional Probabilities for Estimation in Surveys”, *Annals of Applied Statistics*, 11(2): 680-697.
- [46] Thomas, S. (2017), “Future Development of Statistics Canada’s Edit and Imputation System Banff”, Work Session on Statistical Data Editing, United Nations Economic Commission for Europe, The Hague, April 2017.
- [47] U. S. Bureau of Labor Statistics and U. S. Census Bureau (2006), “Design and Methodology, Current Population Survey, TP66”, Tech rep. Vol. 66, 2006. Technical paper 66.
- [48] Weichert, V. (2017), “The ESSnet ValiDat Integration”, Work Session on Statistical Data Editing, United Nations Economic Commission for Europe, The Hague, April 2017.
- [49] Xu, M., Kim, A. A., and Terrie, L. (2017), “Automated Data Editing and Imputation for Surveys of Multinational Enterprises, a Banff Implementation”, Work Session on Statistical Data Editing, United Nations Economic Commission for Europe, The Hague, April 2017.