

第4章 技術の研究に関する事項

第1 研究活動の概要

統計センターでは、製表業務の高度化や製表結果の品質の向上、統計ニーズの多様化への対応などに資するため、製表実務に適用可能な研究に重点を置いて、研究を進めている。この方針の下で、平成19年度は統計分類の自動格付処理、統計データの欠測値の補定処理、個票データの二次的利用の問題などの研究を行い、その成果を製表実務へ効果的に適用させるよう努めてきた。また、これらの研究を進めるため、国際的な研究動向等に関する情報収集を行うとともに、外部からの研究担当職員の採用、外部研究者をメンバーとした研究会の開催など研究体制面での充実を図ってきた。さらに、統計センターに製表技術・実務検討会を新たに設置し、製表方法、製表技術の研究・開発の成果及び製表実務の改善の内容等を共有し、その活用を一体的かつ効果的に推進し、新たな製表技術の適用により効果的な事務改善が得られたもの（あるいは得られそうなもの）について紹介してきたところである。

第1節 技術研究を専任で行う組織の充実

第1 外部研究者の採用及び統計センター内研究会等への外部研究者の参加の推進

平成18年度に引き続き、外部研究者2人を非常勤職員として採用し、18年度に引き続き欠測値の補定に関する研究を行うとともに、新たに調査票情報の秘匿技法の一種であるマイクロアグリゲーションに関する研究を開始した。

また、統計センターの製表業務において中核となる技術の一つであるデータエディティング(欠測又は矛盾したデータ内容を訂正することを目的とした処理)の研究を進めるため、平成15年度以降、外部研究者をメンバーとした「データエディティング研究会」を開催しており、19年度は、1回目は外部講師を招聘し統計分類の自動格付に関する講演会を実施し、2回目は超高次元クロス集計の有効性に関する研究及びサービス産業動向調査の補定方法に関する研究を中心に検討を進めた。

第2節 研究計画

第1 データエディティングに関する研究

1 研究の概要

データエディティングに関する技術の向上及び業務の効率化に資するため、諸外国における研究動向を把握するとともに、実証研究を進めている。

平成19年度は、サービス産業動向調査の売上高等の新たな補定方法について検討を行うとともに、昨年度から研究を進めてきた国勢調査で用いている「世帯類型補定システムの労働力調査への適用の研究」及び「サービス業基本調査の経理項目の欠測値の補定方法の研究」についての報告書を作成した。

データエディティング研究会及び研究会講演会の開催実績は、表のとおりである。

表 平成19年度データエディティング研究会及び研究会講演会開催実績

回数	開催年月日	議 題
第1回	19.12.12	・自動コーディングの利用の評価と問題点 ・産業・職業分類自動コーディングの開発と活用
第2回	20.3.4	・匿名化技法としてのマイクロアグリゲーションの有効性に関する研究 ・諸外国における統計データの秘匿方法
第3回	20.3.11	・社会科学研究のための統計データの公開と機密保護

2 データエディティングに関する情報収集及び資料の整備

データエディティングに関する研究を推進する上で、特に研究の盛んな欧米諸国の研究動向に関する情報収集が重要であることから、アメリカ・カナダ統計学会などが共同開催した「2007年合同統計会議」(7月29日～8月2日、アメリカ/ソルトレイクシティ)に参加したほか、I S I (国際統計協会)大会(8月22日～29日、ポルトガル/リスボン)へ参加し、研究センターが行った「サービス業基本調査の経理項目の補定方法に関する研究」の研究結果を発表するとともに、これらの会議等で発表された研究報告などエディティング・補定方法に関する文献の収集などを行った。

また、原データの品質を高める方策に関するものとして、国連欧州経済委員会が作成・配布した「統計データのエディティングvol.3データ品質への影響」の研究報告を翻訳した。

3 サービス産業動向調査の売上高等の補定方法に関する研究

研究センターでは、主要な調査項目でありながら非回答が比較的多く、補定の精度の影響が特に大きい調査項目について、補定方法の改善の研究を進めてきた。平成15年度、16年度は全国消費実態調査の年収を対象に研究を行い、17年度、18年度は、米国経済センサス（サービス業）における経理項目の欠測値の補定方法を参考にして、平成16年サービス業基本調査のデータを用いて回帰モデルを作成し、その実証研究を行ってきた。

平成19年度は、20年度から始まるサービス産業動向調査の売上等の補定方法の開発を行うため、18年度に行ったサービス業基本調査の経理項目の補定方法を整理した上で、19年度に実施されたサービス産業動向調査試験調査のデータを用いて実証分析を行った。ここでは回帰式の適合性を検証するだけでなく、ホットデック法、コールドデック法の補定方法による実験・検証を行い、研究結果について、本調査の参考に資するため、製表部を經由して総務省統計局へ報告した。

4 「世帯類型補定システム」の労働力調査への適用の研究

労働力調査においては、調査票の年齢欄が記入不備の場合、家族構成等を参考にして職員が総合的判断で補定を行っているが、より客観的かつ統一的な補定方法として国勢調査で導入している「世帯類型補定システム」を適用することについて平成17年度に研究を開始した。平成18年度、19年度は、この手法を労働力調査へ適用する場合の問題点等を把握し、「世帯類型補定システム」を労働力調査に適用できるようプログラムの修正作業等を行い、実験・検証を行った。

研究の結果、「世帯類型補定システム」を労働力調査へ適用することについては、手法としては比較的良好な補定結果が得られたものの、チェック審査時間が増加し、集計期間がタイトな中で行うには難があったこと、補定されるデータ数が少ないことなどを理由に本調査へ導入しないこととし、今後、「世帯類型補定システム」を労働力調査に適用するに当たっては、集計システム全体の見直しの中で検討していくことになった。

第2 統計分類の自動格付に関する研究

1 研究の概要

製表業務の中核の一つである分類格付事務の自動化を図るため、国内外における関連研究の動向を把握するとともに、実証研究を進めている。

平成19年度は、18年度に開始したコンピュータによる市区町村コード自動格付に関するアルゴリズム（解法手順）の研究についてとりまとめ、「製表技術参考資料6」として刊行した。なお、

この研究の成果は、平成20年住宅・土地統計調査の市区町村コード付与事務に適用することとしている。

また、平成21年経済センサス-基礎調査において産業分類自動格付を行うことを目標として、従来からの知識や技術に基づく自動格付技法の改良を図る研究を行った。

2 平成21年経済センサス-基礎調査に適用するための産業分類自動格付に関する研究

平成15年度から17年度にかけて平成18年事業所・企業統計調査に用いる産業分類コードのプレプリント用のデータ作成において自動格付システムを活用するための研究を進め、その研究成果は、製表の審査事務の省力化が図られるなど一定の有効性を示した。平成19年度は、同システムについて、新しいオペレーティングシステムにおいて動作するかの確認作業、格付用ルールの更新、格付率・正解率の向上へ向けた検証を行った。

3 統計分類自動格付導入に向けての検討

統計センターにおいては、統計分類の自動格付の導入に向けて平成19年4月に「統計分類自動格付検討プロジェクト」を設置し、国勢調査を始めとする産業・職業分類及び全国消費実態調査を始めとする収支項目分類の自動格付システム構築に向けた研究を開始することとした。同研究に際しては、専門的知見を有する製表部と研究センターの職員からなる2つのワーキンググループを設けて、これまで研究センターにおいて研究を進めてきた自動格付に関する知識や技術を踏まえ、産業・職業分類と収支項目分類の自動格付に関する具体的な研究に着手した。

平成19年度においては、ワーキンググループで研究を進めつつ、民間事業者にも調査研究を委託し、20年3月末に産業・職業分類と収支項目分類のそれぞれの自動格付研究報告書を受領したところである。同研究に当たっては、自動格付の基となるいわゆる「辞書」を作成し、また、格付精度の評価に利用する必要から、平成17年国勢調査(事業の種類等)と家計調査(収支品目)の調査票の文字入力業務を外部委託し、文字情報のテキストデータ化を行っている。

今後は、引き続きワーキンググループによる研究を行い、民間事業者による研究成果を評価・検討した上で、両者のメリットを生かした本格的なシステム開発を進めることとしている。

第3 統計ニーズの多様化に対応した製表方法に関する研究

1 統計データアーカイブに関する研究

平成18年度は、統計調査の個票データを二次的利用のために提供・保管するデータアーカイブ

に関する国内外の情報収集及び資料整備を行った。平成19年度は、これらの資料に基づき、平成21年4月の新統計法の全面施行に向けて、匿名データの提供準備として、就業構造基本調査のデータを使用した匿名データ作成に係るプロトタイプシステムの構築、実験・検証、データ提供に関する運用上の課題等について検討を行っている。

2 超高次元クロス集計の有効性に関する研究

我が国においてマイクロデータに準じたレベルのデータ提供の可能性の方策を探るため、近年欧米諸国で調査研究が進められている「マイクロアグリゲーション(micro-aggregation)」に着目し、超高次元クロス集計の有効性等について研究を開始した。具体的には、全国消費実態調査の個別データを用いて超高次元クロス集計を試み、マイクロアグリゲート・データによる提供の実現の可能性についての研究を行った。

3 データ秘匿及びデータアーカイブに関する情報収集

国連欧州経済委員会とEU統計局が共同主催した「統計データの秘密保護に関するワークショップ」(平成19年12月17日～19日、イギリス/マンチェスター)に参加したほか、イギリスのデータアーカイブ機関の一つであるCCSR(Cathiemarsh Center for Census and Survey Research)を訪問し、関連情報の収集を行った。

第4 情報処理技術に関する研究

統計センターでは、情報通信技術の進展やそれに伴う開発環境の変化に対応するとともに、システム開発の効率化、システムの信頼性の維持向上等を図るため、情報処理技術に関する研究を行っている。

1 プログラミング言語に関する研究

統計センターにおけるクライアント/サーバシステム用のシステム開発業務では、現在、主に使用するプログラミング言語を「Visual Basic.NET」としているが、同言語は、マイクロソフト社製のOS「Windows」においてのみ動作が可能であることから、統計センターLANシステムにおけるOSの選択肢を狭めていること、OSの更改等によって、現有プログラムの動作への影響の検証作業や、書き換え作業が発生すること等の課題を有しているため、OSに依存しないプログラミング言語である「Java」について調査、研究を行ってきた。

平成19年度は、研究・開発用LAN上に、「Java」言語による効率的なシステム開発、開発シ

システムの品質の確保、データ処理時間の短縮等を可能とするソフトウェアを導入の上、試験的に簡易な集計システムを構築し、検証を行った。

これまでの研究では、製表システム開発への「Java」言語の適用は可能であるものの、国勢調査等の大量データを高速に処理するには、データ処理時間の短縮等を可能とするソフトウェアが必要となること、更なるハードウェアの性能向上が必要となること等が明らかになった。

2 汎用ソフト・ツールに関する研究

平成19年8月に導入した「研究・開発用LANシステム」に多次元集計が可能とされる市販の汎用ソフトウェアを導入し、新汎用サマリーシステムの集計エンジンとしての適性について、運用上の自由度、機能、処理速度といった観点から現行システムの集計エンジンとの比較検証を行った。

今後、この検証結果を踏まえ、「新汎用サマリーシステム」の集計エンジンの在り方について検討を行うこととしている。

3 システム共通部品（フレームワーク）に関する研究

平成19年度は、システム共通部品（フレームワーク）に関する研究として、各種統計調査の集計システム開発で使用可能なフレームワークの整備を行い、7月から順次提供を開始した。このフレームワークは、平成19年就業構造基本調査集計システム、家計調査新製表システム等に組み込まれており、引き続き、プログラミングの標準化推進活動の一環として、各種統計調査のシステム開発者に対する指導、支援等を進めることとしている。

第5 その他の研究

1 製表技術・実務検討会

製表技術・実務検討会の開催実績は、表のとおりである。

表 平成19年度製表技術・実務検討会開催実績

回数	開催年月日	議 題
第1回	19. 8. 2	・市区町村コード自動格付に関するアルゴリズムの研究
第2回	19.10.17	・韓国の人口センサスについて(2007日韓統計協力会議報告)
第3回	19.11. 7	・ホストからの移行に伴い開発・整備したシステムの紹介
第4回	20. 2.13	・匿名化技法としてのマイクロアグリゲーションの有効性に関する研究

第3節 研究成果の普及等

第1 製表技術の普及及び研究の促進

1 製表技術参考資料等の刊行

製表技術の普及及び研究の促進を図るため、統計センターにおける製表技術の研究成果や国内外における製表技術の研究動向の調査分析結果、製表業務のマネジメントを含めた製表技術関連文献の翻訳などの資料を刊行した。

平成19年度の刊行実績は、表のとおりである。

表 平成19年度製表技術参考資料等刊行実績

刊行年月	資料等名	内 容
19.5	2005-2006年度データエディティング研究会報告	・集計段階における無回答処理に関する理論的概観 ・サービス業基本調査における経理項目の補定方法 ・社会生活基本調査の自動格付に関する研究 ・市区町村コード自動格付に関するアルゴリズムの研究 ・製表業務の負担につながる研究事例 ・2006年統計データエディティングに関するワークセッション ・研究センターにおける研究実績及び研究計画
19.8	製表技術参考資料6	市区町村コード自動格付に関するアルゴリズムの研究
19.8	製表技術参考資料7	「世帯類型補定システム」の労働力調査への適用の研究
19.8	製表技術関連資料集8	公的統計と品質管理
20.3	製表技術参考資料8	サービス業基本調査の経理項目の欠測値の補定方法に関する研究

2 学会等における研究発表

平成19年度は、I S I (International Statistical Institute : 国際統計協会)第56回大会(平成19年8月22日~29日、ポルトガル/リスボン)において、サービス業基本調査の経理項目の欠測値の補定方法に関する研究について研究発表を行った。