

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、製表技術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、統計分類の自動格付法の開発、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いる。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いる。

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17年度に引き続き、外部研究者1人を非常勤職員として採用し、欠測値の補定に関する研究を行うとともに、18年度半ばから新たに外部研究者1人を非常勤職員として採用し、データアーカイブ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項目を訂正することを目的とした手法)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、18年度においては、経理項目の欠測値の補定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18年度は、経理項目の欠測値の補定法の改善について検討を行うとともに、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を開始した。

データ・エディティング研究会の開催実績は、表のとおりである。

表 平成18年度データ・エディティング研究会開催実績

回数	開催年月日	議 題
第1回	18. 11.1	・サービス業基本調査における経理項目の補定法(中間報告) ・諸外国のデータ・エディティングに関する研究状況 ・データ・エディティングに関する研究課題
第2回	19.3.19	・サービス業基本調査における経理項目の補定法 ・市区町村自動格付に関するアルゴリズムの研究(中間報告) ・平成19年度研究テーマ(案)

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、国連欧州経済委員会(ECE)の主催で開催された「統計データ・エディティングに関するワークショップ」(9月25日～27日、ドイツ/ボン)に参加したほか、アメリカ・カナダ統計学会などが共同開催した「2006年合同統計会議」(8月6日～10日、アメリカ/シアトル)で行われた研究報告などエディティング・補定法に関する文献の収集・分析などを行った。

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に関する手法と技術」の中から、グラフィカルエディティング、エディティングプロセスの評価などに関する研究報告を翻訳し、製表技術関連資料7として刊行し、関係部門の参考に供した。さらに、前記の「統計データ・エディティングに関するワークショップ」で報告されたものの中から、製表業務の負担軽減につながる研究事例とイタリア、カ

ナダのデータ・エディティングに関する研究報告を「第1回データ・エディティング研究会」において紹介した。

3 経理項目の欠測値の補定法に関する研究

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、最近隣補定法あるいは最近隣補定法と回帰補定法を組み合わせる方法を用いることにより補定精度の向上が期待できることを示した。しかし、事業所・企業を対象とした調査の経理項目の補定においては、最近隣補定法や通常多くの場合に用いられるホットデック法は必ずしも適当ではないため、これらのデータの特性に合った新たな補定法を別途研究する必要がある。

そこで、平成17年度に、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法を参考にして、経理項目の補定法の理論面の研究と適用事例の情報収集を行い、18年度は、その補定法を参考として、平成16年サービス業基本調査のデータを用いた実証研究を行った。

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研究を開始した。平成18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の把握を行うとともに、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を開始した。

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18年度は、17年度に引き続き社会生活基本調査の生活行動分類自動格付法の改善に関する研究を行い、作成した格付ルールは製表業務に適用されることとなった。また、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)の研究を開始し、製表業務への適用可能性等について検証を行った。

2 生活行動分類自動格付に関する研究

平成18年度においては、17年度に引き続き社会生活基本調査の生活行動分類の自動格付の研究を行い、平成13年社会生活基本調査の調査票Bを用いて、主行動、同時行動の完全一致方式による格付ルールを確定した。この研究成果については、平成18年社会生活基本調査で活用することとなった。また、研究成果をまとめて「製表技術参考資料5」として刊行し、関係部門の参考に供した。

3 市区町村コード自動格付に関するアルゴリズムの研究

平成15年住宅・土地統計調査の広島県、大分県及び東京都特別区部の調査票を用い、平成15年住宅・土地統計調査の調査票入力時に用いたOCRによる読み取り結果と、平成17年国勢調査の調査票入力時に用いた現有のOCRによる読み取り結果について、読み取り精度、自動格付精度等の比較検証を行った。

検証の結果、現有のOCRによる自動格付後の認識精度向上フラグを活用することにより、格付に人手審査を要しないデータが一定数確保できることとなり、集計業務への適用の見込みがたち、調査票設計やOCR導入時の仕様に関する提言等をまとめた。

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 データアーカイブに関する研究

これまで、集計表の秘匿処理やオーダーメイド集計など統計ニーズの多様化に対応するために必要な各種手法について研究を行ってきたが、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保管・提供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。

第4 情報処理技術に関する研究

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」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、統計センターの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、機種や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18年度は、政府の統計調査等業務・システムの最適化計画に基づき整備される政府統計共同利用システム（オンライン調査システム）において、データの自動審査、特定帳票の作成に用いる言語として推奨されていることを見据え、「Java」言語によるシステム開発の効率性、品質の向上を可能とする開発環境及び処理時間の短縮等効率的なシステム処理を可能とするソフトウェアについて、情報収集を行った。

また、外部研修を受講する等により、「Java」言語の知識を有する人材の育成も行った。

2 プログラミングの標準化等に関する研究

統計センターにおけるシステム開発業務では、システム開発の効率性、品質を維持向上しつつ、近年の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していくため、開発者の経験と技量に依存する部分をできる限り縮小させることを目的に、プログラミングの標準化等について調査、研究を行ってきた。

平成18年度は、17年度までに情報通信技術関連企業へ委託した標準化への技術支援「開発標準策定関連サービス」の活用により、システム設計における基本方針の企画、プログラム共有部品の企画、開発及び各種システムの開発者に対する設計方針の指導を行うとともに、支援に関する業務を行うアーキテクチャ担当を情報処理課に新たに設置し、家計調査及び平成18年社会生活基本調査のシステム開発への参画、また、プログラムのコーディング規約の改定等による各種システム設計、プログラミングの標準化に向けた活動を行っている。

平成19年度は、引き続き、アーキテクチャ担当の活動を中心として、各種システム開発業務におけるシステム設計、プログラミングの標準化を推進していく。

第5 その他の研究

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18年度の刊行実績は、表のとおりである。

表 平成18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
18.6	製表技術関連資料集 6	統計データ開示抑制に関する用語集改訂版（対訳）
19.2	製表技術参考資料 5	社会生活基本調査の自動格付に関する研究
19.2	製表技術関連資料集 7	統計データのエディティングに関する手法と技術 （国連欧州経済委員会が作成・配布した資料の翻訳）

2 学会における研究発表

平成18年度は、日本統計学会第74回大会（9月5日～8日、仙台市）において、社会生活基本調査における生活行動分類の自動格付法に関する研究発表を行った。