

第4章 技術の研究に関する事項

第1 円滑な研究成果の活用のための緊密な連携の推進

研究センターでは、基礎的な製表技術の研究成果を製表実務に活用するため、当該技術を使用する製表担当者との緊密な連携を図っている。具体的には、平成18年度に予定している社会生活基本調査の製表に活用することとして、研究・開発を進めている生活行動分類自動格付システムの実験・検証結果について製表部に説明を行い、業務への適用に際しての留意点等を確認し、これを踏まえたシステムの改良研究を進めているところである。

さらに、次年度の研究テーマを設定するに当たっては、製表部が直面している業務の課題について積極的に対応することとし、製表部から提案された課題の中から研究テーマを設定するなど、一層の連携を推進している。

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16年度に引き続き非常勤職員1人を採用し、集計表の秘匿処理法及び欠測値の補定に関する研究を行った。

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・エディティングの研究を進めるため、平成15、16年度に外部有識者をメンバーとした「データ・エディティング研究会」を開催してきたが、17年度においては、欠測値の補定法の改善に加え、分類の自動格付法、集計データの秘匿処理法など製表技術にかかる事項についても、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

製表業務の中核の一つであるデータ・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17年度においては、欠測値の補定法の改善、多変量外れ値検出法の適用の可能性などについて検討を行った。

研究会の開催実績は、表のとおりである。

表 平成17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. 3.10	・社会生活基本調査の生活行動分類自動格付に関する研究(中間報告) ・製表業務の負担軽減につながる諸外国の研究事例 ・アメリカ経済センサスにおける経理項目の補定法

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「合同統計会議」(8月7日～11日、アメリカ/ミネアポリス)に参加したほか、セレクトティブ・エディティングに関する研究、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についての研究動向などエディティング・補定法に関する文献の収集・分析を行った。

また、原データの品質を高める方策に関するものとして、欧州統計システムにおける品質マネジメント関連文献、統計業務の品質マネジメントに関する欧米の研究動向など主要な文献を収集・翻訳し、製表技術関連資料集3及び4として刊行し、関係部門の参考に供した。さらに、統計データ・エディティングに関するUN/ECE(国連欧州経済委員会)ワークショップで報告されたものの中から、カナダ統計局及びスペイン国家統計院から製表業務の負担軽減につながる研究事例として報告されたインターネット・オプションに関する評価についてのレポートの翻訳を完了した。

参 考

研究の内容

1 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おり、前年度までは全国消費実態調査の年間収入項目を対象に研究を行い、諸外国で用いられている最近隣補定法あるいは最近隣補定法と回帰補定法を組合せた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・企業統計調査では、ホットデック法や最近隣補定法などのドナーを用いる補定法は必ずしも適当ではなく、事業所・企業統計調査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、今年度は、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法を参考にして、現行の補定法である比推定型回帰補定法と、他の回帰モデルや複数の説明変数を用いた補定法との比較検証を行う計画をまとめた。これについては、今後の事業所・企業統計調査の動向を踏まえ、着手時期、詳細な研究内容等を検討することとした。

2 効果的なデータ・エディティングに関する研究

歪んだ分布に対応した外れ値検出法

分布の歪み度合いをロバストに測る調整箱ひげ図が提案されている。しかし、シミュレーションの結果、外れ値の数が多くなると、検出力が大きく低下する恐れがあることが分かった。そこで、外れ値の数が多い場合にも適用可能な方法として、本来の多変量外れ値検出法であるBACONと調整箱ひげ図を組み合わせた外れ値検出法の研究を行った。その結果、この検出法は、裾が比較的重い(ほぼ対象な)分布への適応力に関しても調整箱ひげ図を上回っていることが分かった。

(注) BACON : blocked adaptive computationally efficient outlier nominations

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17年度においては、事業所・企業統計調査の産業分類の自動格付に関する研究について、前年度に改善を行ったシステムを、18年に実施される事業所・企業統計調査の格付符号の検査に

適用した。

また、新たに、格付作業の省力化を目的として、社会生活基本調査の製表業務における生活行動分類の自動格付法の改善に関する研究を行い、格付ルールの作成及び13年社会生活基本調査のデータを用いたテストを行った。

2 産業分類自動格付システムの研究

平成17年度においては、前年度の自動格付法の改善結果に基づいて、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に向けてのデータ作成において、省力化を図るために、自動格付システムを有効に活用することとし、業務への適用に向けた研究を実施した。また、研究成果をまとめて製表技術参考資料3として刊行し関係者へ提供した。

3 生活行動分類自動格付システムの研究

平成17年度においては、社会生活基本調査の生活行動分類自動格付の完全一致方式に用いる格付ルールを確定するための研究を行い、作成したコンスタントを用いて、実際のデータ(平成13年社会生活基本調査の調査票B)に対して自動格付の実験・検証を行った。

4 統計分類自動格付法に関する情報収集及び資料整備

製表事務の質の向上及び効率化を推進する上で、統計分類の自動格付は、手書き文字の自動読み取りとともに最も実現が望まれている製表技術の一つであるが、完全といえる水準に引き上げるには技術的に大きな課題がある。

そこで、国内外における統計分類の自動格付法あるいはテキストの自動分類に関する研究について情報収集を行うため、情報処理学会自然言語処理研究会及び言語処理学会に参加し、新しいテキスト自動分類法の研究動向の把握に努めたほか、国内外の関連文献の収集・分析を行い、関係者へ提供した。

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 研究の概要

製表業務の効率を維持しつつ統計ニーズの多様化に柔軟に対応できる製表方法について検討するため、国内外の研究動向等を把握するとともに、実証研究を行っている。

平成17年度においては、オーダーメイド集計に適した既存ソフトウェアのリストアップ及び機

能比較、オーダーメイド集計や詳細な集計の増大に伴って開発の必要性が高まる秘匿処理法の検討を行った。

2 集計表の秘匿処理に関する研究

小地域統計などの詳細な集計や企業統計などではプライバシー保護等の観点から集計表の秘匿処理が必要となる場合が多い。統計ニーズの多様化に伴い、詳細な集計が増大し、さらにオーダーメイド集計が実施されると秘匿処理を効率的に行う必要性が高まるが、現在は汎用性があり、適正な秘匿処理法が確立されていない。

そこで、平成16年度に平成11年サービス業基本調査結果を用い、EUの秘匿処理に関するプロジェクトSDC及びCASCが開発した秘匿処理ソフトウェア -ARGUSの利用可能性に関してテストを行うとともに、米国センサス局の方法を採用したプロトタイプ・システムSCSを開発し、 -ARGUSとの機能比較を行った。

平成17年度には、 -ARGUSとSCSの実用面での機能についての実験・検証の比較結果をとりまとめ、製表技術参考資料4として刊行して関係部門の参考に供するとともに、プロトタイプ・システムSCSの開発状況については、日本統計学会で発表を行った。

3 オーダーメイド集計及び集計表の秘匿処理法に関する情報収集及び資料整備

集計表の秘匿処理法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「合同統計会議」(8月7日～11日、アメリカ/ミネアポリス)に参加するとともに、国連欧州経済委員会/ EU統計局合同ワークショップ(11月9日～11日、スイス/ジュネーブ)に参加し、秘匿処理法に関する文献の収集・分析等を行った。

また、前記のワークショップで作成された統計機密及びマイクロデータ提供管理のためのガイドライン(製表技術関連資料集5に所収)及びカナダ統計局の秘匿の免責に関するペーパー(製表技術参考資料4に所収)の翻訳を行い、関係者へ提供した。

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

現在、主に使用しているプログラミング言語であるVisual Basicは、マイクロソフト社が提供しているOSであるWindowsにおいてのみ動作が可能であり、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。

このため、OSの選択肢を広げるとともに、LANシステム切替え時に生じるプログラムの書き換え等の負担を軽減する目的から、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・研究を行っている。

平成17年度は、試験的にWebアプリケーションをオブジェクト指向の設計により作成し、製表システムへのJavaの適用の可否について検討を行った。今後は、政府の統計調査等業務・システムの最適化計画の動向を見据え、引き続きJavaについての研究を進めることとしている。

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については、開発者の経験と技量などによっては、近年の情報通信技術の進展に伴う技術の習得や開発環境変化への対応などが負担となることもある。このような状況は、システム開発効率の低下やシステム品質のバラツキを生じさせる要因ともなっている。

このため、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、プログラミングに係る標準化等について調査・研究を行っている。

平成17年度は、16年度に引き続き、プログラムの標準化等に関する研究を進め、プログラム開発業務、電子計算機等の運用業務に係る業務手順について見直し、コンピュータシステムガイド「プログラム開発基準編」及び「運営管理基準編」(ホスト編及びPC編)を整備した。

また、開発標準策定関連サービス¹については、一部の手法を試行したところ、有用性がみられたため標準化に向けてこの手法の適用を開始した。

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各種資料を刊行した。

平成17年度の刊行実績は、表のとおりである。

¹開発標準策定関連サービス：システム開発における開発プロセス、ドキュメント、設計方法などの標準化への支援をIT関連企業が有料で行う技術サービスのことをいう。

表 平成17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
17.4	データ・エディティング研究会報告	前年度のデータ・エディティング研究会における年収欠測値の補定法及び多変量外れ値検出法の研究成果、イン라이어検出法の研究
17.6	製表技術参考資料 3	係り受け関係の考慮、精度の高い学習用データを用いることなどによる産業分類自動格付法の改善効果の検証結果
18.3	製表技術参考資料 4	汎用秘匿処理ソフトウェア τ -ARGUSによる集計表の秘匿処理、順次LP法と τ -ARGUS搭載の集計表秘匿処理法の比較結果及び最近の集計表開示抑制法の研究動向
18.3	製表技術関連資料集 5	国連欧州経済委員会 / EU統計局合同ワークショップで作成された「統計の機密保護とマイクロデータへのアクセスを管理するためのガイドラインドラフト」
18.3	製表技術レポート 2	諸外国におけるセレクトティブ・エディティングの概要

2 学会における研究発表

平成17年度は、日本統計学会第73回大会（9月12日～15日、広島市）において、次の2件の研究発表を行った。

集計表秘匿処理プロトタイプ・システムSCSの開発状況

歪んだ分布に対応した外れ値検出法