

第4章 技術の研究に関する事項

【中期計画】

より効果的効率的な製表技術の開発に資するための研究を専任で行う組織体制を整備するとともに、国際的な動向等に関する情報収集についても積極的に行いつつ、必要に応じ、国内外の大学や官民の研究所、国際機関や諸外国の統計機関等の外部の機関との間で、技術協力や連携を図りながら、製表業務の高度化や製表結果の品質向上などに重点を置いて研究を実施する。また、調査環境の変化や統計利用者のニーズの複雑多様化に対応すべく、当該研究の成果を的確に活用していくものとする。

第1 研究センターの設置

効率的かつ効果的な製表業務の運営に資するため、製表技術の研究を行う部門として、平成15年4月に研究センターを設置した。

第2 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。さらに、製表技術・実務検討会を開催し、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

外部の研究機関、大学等との人材交流を推進し、統計センター職員の研究能力の向上及び製表技術の高度化・改善を図るため、平成16年度から外部研究者を非常勤職員として採用（16年度及び17年度：各1人、18年度及び19年度：各2人）し、集計表の秘匿処理法、欠測値の補定及びデータアーカイブに関する研究を行うとともに、平成19年度には、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新たに開始した。

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年度以降毎年度、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、欠測値の補定方法の改善、分類の自動格付法、集計データの秘匿処理法などについて検討を進めた。

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及び補定方法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。

各年度の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 データエディティング研究会開催実績

年度 (平成)	回数	開催年月日	議 題
15年度	第1回	15.6.5	<ul style="list-style-type: none"> 統計センターにおけるデータエディティング・補定業務の課題及びデータエディティング研究会の研究方針 諸外国におけるデータエディティング・補定に関する研究の動向 年収データ補定方法に関する当面の研究課題と補定誤差の推定方法
	第2回	15.11.7	<ul style="list-style-type: none"> 年収欠測値の補定方法の比較研究結果 カナダ統計局開発の新補定方法N I Mの概要と平成12年国勢調査世帯類型補定システムとの比較 国連欧州経済委員会UN/ECEによるエディティング及び補定の定義
	第3回	16.3.11	<ul style="list-style-type: none"> 年収欠測値の補定方法の実証研究 欧米諸国における新補定方法N I Mの比較評価結果 最少補定箇所原則に基づく補定法(Fellegi-Holt法)の概要と米国・カナダにおける適用状況
16年度	第1回	16.10.8	<ul style="list-style-type: none"> 年収及びその内訳の欠測値の補定法の研究 多変量外れ値検出法の比較 密度推定法によるイン라이어検出の試み
	第2回	17.3.10	<ul style="list-style-type: none"> 年収及びその内訳の欠測値の補定法の実験結果 密度推定法を利用した二変量外れ値検出法の実験結果 プレプリント方式のデータ品質改善効果に関する欧米諸国の研究動向
17年度	第1回	18.3.10	<ul style="list-style-type: none"> 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
18年度	第1回	18.11.1	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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)
19年度	第1回	19.12.12	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 社会科学的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、表のとおり学会等に参加するとともに、各種研究報告などエディティング及び補定方法に関する文献の収集・分析などを行った。

表 欧米諸国の学会等への参加実績

年度 (平成)	開催地	開催月日	会議名・主催
15年度	アメリカ (サンフランシスコ)	8. 3～ 8. 7	合同統計会議 主催：アメリカ統計学会など
	スペイン (マドリッド)	10.20～ 10.22	統計データエディティングに関するワークショップ 主催：国連欧州経済委員会(ECE)
16年度	カナダ (トロント)	8. 7～ 8.14	合同統計会議 主催：アメリカ・カナダ統計学会など
17年度	オーストラリア (シドニー)	4.5～ 4.12	I S I (国際統計協会)大会 主催：国際統計協会
	アメリカ (ミネアポリス)	8. 7～ 8.11	合同統計会議 主催：アメリカ・カナダ統計学会など
	スイス (ジュネーブ)	11.9～ 11.11	統計データの秘密保護に関するワークショップ 主催：国連欧州経済委員会(ECE)/欧州統計局(Euro Stat)
18年度	ドイツ (ボン)	9.25～ 9.27	統計データエディティングに関するワークショップ 主催：国連欧州経済委員会(ECE)
	アメリカ (シアトル)	8. 6～ 8.10	2006年合同統計会議 主催：アメリカ・カナダ統計学会など
19年度	アメリカ (ソルトレイクシティ)	7.29～ 8. 2	2007年合同統計会議 主催：アメリカ・カナダ統計学会など
	ポルトガル (リスボン)	8.22～ 8.29	I S I (国際統計協会)大会 主催：国際統計協会
	イギリス (マンチェスター)	12.17～ 12.19	統計データの秘密保護に関するワークショップ 主催：国連欧州経済委員会(ECE)/欧州統計局(Euro Stat)

3 経理項目の欠測値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間収入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、前年度に行ったサービス業基本調査の経理項目の回帰式による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的な分析を行い、その研究結果について、本調査の参考に資するため、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、調査票の年齢欄に記入不備がある場合、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、同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毎月公表している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、決められた集計期間中に行うにはスケジュール的に困難であることが予想されること、及び補定されるデータ数が少ないことなどを理由に、現時点においては行わないこととし、今後、集計システム全体の見直しの中で検討していくこととなった。

第2 統計分類の自動格付に関する研究

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、実証研究を進めた。

1 事業所・企業統計調査の産業分類の自動格付の研究

平成15年度から17年度にかけて、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を行い、その成果を平成16年事業所・企業統計調査の格付符号の検査事務に適用し、製表要員の投入量削減を実現するとともに、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適用し、製表業務における審査事務の省力化を図った。

また、平成19年度は、平成21年経済センサス-基礎調査の産業分類の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についての研究を行った。

2 生活時間行動分類の自動格付の研究

平成17年度から18年度において、社会生活基本調査の生活時間行動分類自動格付法の改善に関する研究を行い、平成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上で格付ルールを確定した。この成果は、平成18年社会生活基本調査の製表業務に適用するとともに、製表技術参考資料として刊行し、関係部門の参考に供した。

3 市区町村コードの自動格付の研究

平成18年度から19年度において、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)の研究を行い、その成果をとりまとめ、製表技術参考資料とし刊行するとともに、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た。

4 産業・職業分類及び収支項目分類の自動格付導入に向けての検討

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の構築に向けた研究を開始することとし、専門的知見を有した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、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。

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 集計表の秘匿処理に関する研究

集計表の秘匿処理は、プライバシー保護等の観点から行うもので、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その必要性が高まることから、平成17年度において、欧米の研究グループで開発された秘匿処理のための汎用ソフトウェアを用いて検証等を実施し、その結果をとりまとめ、報告書として刊行し、関係部門の参考に供した。

2 統計データアーカイブに関する研究

統計調査の個票データの二次利用のために、そのデータを保管・提供する統計データアーカイブに関する研究を行った。平成18年度は、国内外の情報収集及び資料整備を行い、19年度は、これらの資料に基づき、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。

3 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

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、匿名化技法としての有効性についての研究を平成19年度から開始した。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。

4 データ秘匿及び統計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」に平成17年度と19年度に参加し、17年度についてはその会議で収集した情報を外部委託により翻訳し、製表技術関連資料として刊行した。また、平成19年度(12月17日~19日)については、同ワークショップに参加したほか、英国の統計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

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.NET」としているが、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 プログラミングの標準化に関する研究

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、研究を行っている。

平成16年度及び17年度は、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、情報通信技術関連企業が行う標準化への技術支援「開発標準策定関連サービス¹」を活用することとし、18年度は、そ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った。また、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など、プログラミングの標準化に向けた活動を行った。

平成19年度は、各種統計調査の集計システム開発で使用可能なシステム共通部品(フレームワーク)の整備を行い、19年7月から順次提供を開始し、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た。

第3節 研究成果の普及等

第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。

各年度の刊行実績は、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、設定方法などの標準化への支援をIT(情報通信技術)関連企業が有料で行うサービスのことをいう。

表 製表技術参考資料等刊行実績

刊行年月	資料等名	内 容
15. 8	製表技術参考資料 1	形態素解析等の言語処理方法を用いた生活行動分類自動格付システムの開発
16. 4	2003年度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の研究成果、エディティングの汎用的手法Fellgi-Holt法、新補定法N I Mの欧米諸国における研究動向の調査分析結果など
16. 5	製表技術関連資料集 1	国連欧州経済委員会刊行の「データ・エディティングに関する用語集」及び「データ・エディティングの効率性評価：一般的枠組み」の翻訳
16. 8	製表技術研究レポート 1	多変量外れ値検出法に関する国内外の研究動向
16. 8	製表技術参考資料 2	平成15年度に行った事業所・企業統計調査産業分類の自動格付法の研究結果及び国内外における統計分類の自動格付法の研究動向
17. 1	製表技術関連資料集 2	統計データ機密保護に関する国連欧州経済委員会 / E U 統計局合同ワークショップ刊行の「統計データ開示抑制に関する用語集暫定版」及びE U 統計局刊行の「統計品質に関する用語集」の翻訳
17. 3	製表技術関連資料集 3	E U における統計品質の定義、品質標準報告書、L E G 勧告とその実施状況等をまとめた文献の翻訳
17. 3	製表技術関連資料集 4	職業・産業分類格付業務における継続的な品質管理、物価指数作成におけるISO9000の適用経験、生活行動分類の格付誤りの影響分析などの米英の文献の翻訳
17. 4	2004年度データエディティング研究報告	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、インライアー検出法の研究
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 - A R G U S による集計表の秘匿処理、順次L P 法と - A R G U S 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 / E U 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	諸外国におけるセレクトティブ・エディティングの概要
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）
19. 5	2005-2006年度データエディティング研究会報告	<ul style="list-style-type: none"> ・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ショップ ・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法に関する研究

第2 学会等における研究発表

製表技術に関して学識研究者との情報交流を推進し、研究の促進を図ることを目的として、平成15年度から日本統計学会（団体会員）及び日本人口学会（法体会員）に加入し、16年度から日本統計学会及びI S I（International Statistical Institute：国際統計協会）大会において研究発表を行っている。

各年度の発表実績は、表のとおりである。

表 学会等における研究発表実績

年度 (平成)	学会名等	開催月日	開催地	発表テーマ
16年度	第72回 日本統計学会	9.3～ 9.6	岩手県花巻市	・ Predictive Mean Matching 法による年収 欠測値の補定 ・ 多変量外れ値検出法の比較
17年度	第73回 日本統計学会	9.12～ 9.15	広島県広島市	・ 集計表秘匿処理プロトタイプ・システム S C S の開発状況 ・ 歪んだ分布に対応した外れ値検出法
18年度	第74回 日本統計学会	9.5～ 9.8	宮城県仙台市	・ 社会生活基本調査における生活行動分類 の自動格付法
19年度	第56回 I S I 大会	8.22～ 8.29	ポルトガル (リスボン)	・ サービス業基本調査の経理項目の欠測値 の補定法