

第5章 技術の研究に関する事項

【中期目標】

- (1) 上記1から3までに掲げる業務に必要な技術について、符号格付業務の自動化の研究及び未回答事項の機械的な補完方法の研究に重点化するとともに、研究成果を業務運営に十分に活用し、調査環境の変化や統計利用者のニーズの多様化に的確に対応できるように努めること。
- (2) できる限り具体的かつ定量的な研究成果の目標等を設定し、その達成度の評価に基づき着実な技術研究の遂行を図ること。

【中期計画】

- (1) 上記1から3までに掲げる業務に必要な技術について、次の①及び②の研究に重点的に取り組む。また、研究成果を業務運営に十分に活用し、調査環境の変化や統計利用者のニーズの多様化に的確に対応する。
 - ① オートコーディングシステムの研究
調査票の記入内容の統計分類符号への格付を自動的に行うオートコーディングシステムの研究を行う。
特に、次に掲げる符号格付業務に研究成果を実際に適用するとともに、その適用に当たっては、格付率等の定量的な目標を年度計画で明らかにしつつ、業務の効率化と品質の維持向上を図る。
 - ・ 平成20年に調査実施が予定されている住宅・土地統計調査の市区町村コード付与
 - ・ 平成21年に調査実施が予定されている経済センサスの産業分類符号格付
 - ・ 平成23年に調査実施が予定されている社会生活基本調査の生活行動分類符号格付また、次に掲げる符号格付業務についても実用化に向けた研究を推進する。
 - ・ 平成21年に調査実施が予定されている全国消費実態調査の収支項目分類符号格付
 - ・ 平成22年に調査実施が予定されている国勢調査の産業分類、職業分類符号格付
 - ・ 平成24年に調査実施が予定されている就業構造基本調査の産業分類、職業分類符号格付
 - ② データエディティングに関する研究
データエディティングにおける調査票の未回答事項に対する機械的な補完方法等の研究、検証を行う。
特に、国勢調査等の製表に研究成果、検証結果を実際に適用することとし、業務の効率化と品質の維持向上を図る。
- (2) 上記(1)の研究に当たっては、国際的な動向等に関する情報収集や、必要に応じて国内外の大学や官民の研究所、国際機関、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も併せて実施する。

【実施結果】

統計センターでは、上記第2章から第4章までに掲げる製表業務に必要な技術として、調査票の記入内容の統計分類符号への格付に係るソフトウェアによる支援としての格付支援（オートコーディング）システムの研究と、データエディティングにおける調査票の未回答事項に対する機械的な補完方法等の研究（データエディティングに関する研究）のほか、統計ニーズの多様化に対応した製表技術に関する研究など、製表実務に適用可能な研究に重点を置いて研究を行った。

この結果、第2期中期目標期間において、格付支援システムの研究では、平成20年住宅・土地統計調査の市区町村コード格付、平成21年経済センサス - 基礎調査の産業小分類符号格付、平成21年全国消費実態調査の収支項目分類符号格付、平成22年国勢調査の市区町村コード格付及び産業・職業の大・小分類符号格付、平成23年社会生活基本調査の生活時間行動分類符号格付、平成24年就業構造基本調査の産業・職業小分類符号格付、平成24年経済センサス - 活動調査の産業小分類符号格付について、それぞれの研究成果を得て実際に適用を行った。また、データエディティングに関する研究では、平成24年経済センサス - 活動調査の経理項目の欠測値補定の研究を終え、24年度には同調査に実際に適用することとしていたが、統計局主催の研究会において、外部有識者から初めての調査で判断材料としてのデータ数も少ないことから適用は難しいとの結論を得たことから、実用化に至っていない。さらに、統計ニーズの多様化に対応した製表技術に関する研究では、匿名デー

タの作成等に係る各種研究を継続的に行い、着実に匿名データの作成に向けた成果を挙げているほか、擬似データ作成に関する研究において統計教育・訓練用データの試行提供を行い、実用化に向けた検証・評価を行っている。

また、上記の研究に当たっては、国際的な動向等に関する情報収集や、必要に応じて国内外の大学や官民の研究所、国際機関、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も併せて実施した。

各研究に関する実施状況は次のとおりである。

第1節 格付支援システムの研究

第1 平成20年住宅・土地統計調査への適用に向けた研究

平成20年度において、平成19年度の「市区町村コードのオートコーディングに関するアルゴリズムの研究」の成果を踏まえ、市区町村符号（コード）格付の格付率75%以上、正解率97%以上を目標として実用化した結果、格付率88%、正解率99%を達成した。

第2 平成21年経済センサス - 基礎調査への適用に向けた研究

平成20年度において、事業所・企業統計調査用に研究・開発した機械学習型²⁴格付支援システムを適用し、平成18年事業所・企業統計調査データで事業所産業小分類符号の格付率65.4%、正解率97.7%、企業産業小分類符号の格付率66.2%、正解率99.2%となった。

平成21年度において、20年度に引き続き産業小分類符号の格付ルールの整備を行い、産業小分類符号の格付率50%以上、正解率97%以上を目標として実用化した結果、格付率57.3%、正解率97.7%を達成した。

第3 平成21年全国消費実態調査への適用に向けた研究

平成20年度において、収支項目分類符号の格付ルールの整備を行い、当初は収支項目分類符号の格付率26.1%、正解率99.0%であったが、平成19年度家計調査データで格付率55.2%、正解率98.7%、平成16年全国消費実態調査データで格付率58.1%、正解率99.3%となった。

平成21年度において、20年度に引き続き収支項目分類符号の格付ルールの整備を行い、収支項目分類符号の一部について試行的導入を行うこととし、目標を二人以上の世帯では格付率50%以上、正解率99%以上、単身世帯では格付率40%以上に設定して実用化した結果、二人以上の世帯の格付率51.4%、正解率99.8%、単身世帯の格付率45.3%を達成した。単身世帯の正解率は、格付対象数が少なく妥当性に乏しいため算出していない。

第4 平成22年国勢調査への適用に向けた研究

平成20年度において、開発方針として抽出詳細集計の産業小分類符号及び職業小分類符号の格付率40%以上、正解率97%以上を目標とすることとした。平成17年国勢調査の抽出速報集計用の文字入力済データを用いて、ルールベース型²⁵格付支援システムで、テストを行った結果、産業小分類

²⁴ **機械学習型**：人手による格付結果を学習用データとして、形態素解析など機械処理を行い、確率的に高い順に格付を行うルールを自動的に作成する。

²⁵ **ルールベース型**：調査票の記入内容と辞書に登録された語句が一致する場合のみ、分類符号を格付支援システムにより格付する方法。

符号の正解率97%以上で格付率24.7%、職業小分類符号の正解率95%以上で格付率17.3%となった。

平成21年度において、平成17年国勢調査の抽出速報集計用の文字入力済データを新産業分類符号に組み替え、前年度に引き続きルールベース型格付支援システムの格付テストを行った。この結果、産業小分類符号格付率40.2%、正解率97%以上、職業小分類符号格付率44.2%、正解率97%以上となった。

平成22年度において、平成17年国勢調査の抽出速報集計用の文字入力済データを新職業分類符号に組み替え、ルールベース型格付支援システムの格付テストを行った。また、外部委託で行った研究成果を分析・検討した結果、機械学習型格付支援システムの有効性を確認できたことから、同システムの開発を外部委託で行った。同システムの基本機能として①ルール作成機能、②自動格付機能、③格付結果分析機能を備えたものとし、格付率40%以上、正解率97%以上を目標とした。さらに、抽出速報集計における産業・職業小分類符号格付の一部について、試行的に格付支援システムを適用し、産業・職業小分類符号の格付率40%以上、正解率97%以上を目標とした結果、産業小分類符号の格付率46.7%、正解率99.3%、職業小分類符号の格付率45.8%、正解率99.4%を達成した。

平成23年度において、産業等基本集計における産業大分類符号の一部について、格付支援システムを適用し、格付率40%以上、正解率97%以上を目標として実用化した結果、格付率59.5%、正解率99.6%を達成した。また、職業等基本集計における職業大分類符号の一部について、格付支援システムを適用し、格付率60%以上、正解率98%以上を目標として実用化した結果、格付率は63.5%、正解率99.4%を達成した。

平成24年度において、職業等基本集計における職業大分類符号格付の一部について、格付支援システムを適用し、格付率60%以上、正解率98%以上を目標として実用化した結果、格付率70.4%、正解率99.4%を達成した。また、抽出詳細集計における産業・職業小分類符号格付について、格付支援システムを適用し、格付率40%以上、正解率97%以上を目標として実用化した結果、産業小分類符号の格付率53.3%、正解率99.7%、職業小分類符号の格付率53.5%、正解率99.7%を達成した。

第5 平成23年社会生活基本調査への適用に向けた研究

平成23年度において、生活時間行動分類符号の全文一致方式に加え、更なる格付率の向上と人手による符号格付事務の負担軽減を図るため、調査票に記入された行動内容から形態素解析によりキーワードを抽出して行うキーワード方式の2つの方式により行うこととし、システムを新たに開発した。

平成24年度において、前年度に開発した2つの方式により、生活時間行動分類符号の格付率60%以上、正解率99%以上を目標として実用化した結果、格付率は80.1%、正解率98.9%となった。次回調査に向け、格付支援システムの技術のさらなる研究を進め、正解率の向上を図る。

第6 平成24年経済センサス - 活動調査への適用に向けた研究

平成21年度において、ハードウェアに適合し、更なる精度を持つ事業所・企業に関する格付支援システムを新たに開発することを決定した。新システムは、21年度に外部委託で開発した国勢調査用の格付支援システムの格付手法を参考に21年度から22年度にかけて研究・開発することとし、システムの構築は22年度末までに完了することとした。

平成23年度において、産業小分類符号格付の格付支援システムによる格付率55%以上（正解率97%以上）となるよう、学習用データとして用いる平成21年経済センサス - 基礎調査のデータ整備を行った。

平成24年度において、産業小分類符号の格付率55%以上、正解率97%以上を目標として実用化した結果、格付率33.1%と、格付率は目標を下回った。目標値を下回った要因としては、平成24年経済センサス - 活動調査第2次試験調査の調査票様式に基づき格付ルールを設定していたが、本調査において調査票様式が急遽変更となり、格付に用いる情報項目が減少したためである。

第7 平成24年就業構造基本調査への適用に向けた研究

平成24年度において、産業小分類符号及び職業小分類符号の格付率40%以上、正解率97%以上を目標として実用化した結果、格付率は産業小分類符号で51.4%、職業小分類符号で54.5%を達成した。正解率は産業小分類符号で99.5%、職業小分類符号で99.6%を達成した。

第8 OCR機により認識されたデータを用いて産業・職業大分類を格付する技術の研究

統計分類の格付業務について、調査票に記入された文字を外部委託により入力した後、格付支援システムによる格付を行う場合、文字入力に係る経費及び処理期間の両面において負担が大きく、格付支援システムによる省力化の特性を十分に発揮できているとは言い難い。そこで、格付支援システムによる更なる省力化の可能性を追求するため、OCR機により国勢調査の調査票に記入された文字（イメージデータ）を認識し、その結果を用い、格付ルールに従った格付支援システムによる格付を可能とする技術の研究を、平成22年度から行っている。

平成22年度において、外部委託により、「事業の内容」欄を対象とした産業大分類の符号格付について①フリー記入欄内の文字枠の設定の検討、②OCR機による文字認識技術の検討、③文字認識から知識処理、符号格付までのアルゴリズムの検討、の3点を主なテーマとして研究を行った。

平成23年度において、「勤め先・業主などの名称」、「事業の内容」及び「本人の仕事の内容」の3欄を対象として産業・職業分類符号の格付支援を行い、実践的なレベルのテストの実施による研究に着手した。

平成24年度において、外部委託した3者に対し、6月に平成27年国勢調査第1次試験調査の調査票へ記入した擬似データについて、各者のOCR機の読み取り状況の検証を行った。さらに、本研究として、平成27年国勢調査第1次試験調査において文字枠が設定された「勤め先・業主などの名称」、「事業の内容」及び「本人の仕事の内容」の3欄について文字認識の研究を行った。入札で決定した研究用OCR機により調査票の読み取りを行い、文字枠内に記入された文字の認識について、①文字枠内の手書き文字の認識精度向上、②知識処理の精度向上の2点を主なテーマとして実践的なレベルでの研究を行った。また、認識結果を用いて格付ルールを更新し、産業・職業大分類の格付支援を行い、格付への影響について検証を行った。今後も平成27年国勢調査に向けて研究を継続する予定である。

第2節 データエディティングに関する研究

第1 データエディティングの精度評価の研究

平成20年度において、セレクトティブエディティング²⁶を中心とした諸外国の情報を収集するとともに、製表におけるデータ処理方法等を調査し、データエディティング方法について精度評価の手

²⁶セレクトティブエディティング:エラーデータについて集計値への影響度やエラーの度合いを所定の算式によって推定(スコア化)し、その大きさが一定値以上のものについて人手による審査を重点化(それ以外はコンピュータにより補定処理)する方法。

法の研究を進めた。

平成21年度において、セレクトティブエディティングに関する諸外国の情報の収集に加え、平成17年国勢調査の速報集計及び第1次基本集計のデータを用いて、データチェックリスト審査の必要度を予測式により算出する、当該エディティングの考え方を導入した手法の有効性等について検討を行った。

平成22年度において、国勢調査等の大規模調査では、データチェックリストの審査に膨大な人員・時間を必要としている当該審査の効率化を図るため、平成17年国勢調査第1次基本集計の大都市を含む県のデータを用いてテストを行い、その結果を「製表技術参考資料」に取りまとめた。

平成23年度において、前年度に引き続きデータチェックリストの審査の効率化を図るため、平成17年国勢調査第1次基本集計の大都市を含む県のデータを用いて検証を行い、次回国勢調査への実用化に向けてプロジェクトにおいて検討を行った。

平成24年度において、平成22年国勢調査人口等基本集計を用いた記入内容と人手訂正の有無について、個人事項の年齢と続柄に着目し、記入内容と人手訂正の有無を分析しているほか、平成22年国勢調査人口等基本集計データを用いて、データチェック審査の効率化に関する研究を行っている。また、審査省略（エラーが付されたデータについて訂正の有無を統計的手法で分析し、訂正される確率の低いデータを審査対象から外し、審査を省略する。）を行った場合の集計結果の精度について、基準とする最終チェック済データでの集計結果との精度評価に関する研究（乖離表の分析）を行っている。なお、23年度に行った平成17年国勢調査第1次基本集計のデータを用いた実験結果を「製表技術参考資料」に取りまとめた。

第2 多変量外れ値の検出方法の研究

平成20年度及び21年度において、カナダ統計局において実用化されたMSD法²⁷及びその改良手法について、統計解析ソフトR²⁸によりプログラムを作成し、シミュレーションデータによる性能評価を行った。

平成22年度及び23年度において、調査票の未回答事項を補定する際、外れ値（特異値）は精度に大きな影響を与えるものであるため、外れ値を数学的理論により検出する方法の研究を行った。22年度では、企業財務データについて、繰り返し最小二乗法（IRLS）を適用する研究を行い、研究結果を統計関連学会連合大会及び平成22年度統計技術研究会（第1回）において報告した。また、同じく企業財務データについて、乗率を考慮した産業別試算結果の詳細な分析を進めた。なお、平成23年度には、シミュレーションにより様々なデータ条件下での繰り返し最小二乗法（IRLS）の挙動を整理した結果について応用統計学会年会で報告し、企業財務データを用いて売上高の従業者数による補定の試算とともに補定により得られるIRLSウェイトにより標本調査データの乗率調整を試みた結果について国際統計協会（ISI）第58回大会及び統計関連学会連合大会において報告した。

平成24年度において、IRLSよりも性能の良いロバスト回帰²⁹補定法について機能比較を行うため、該当する手法を選定し、統計解析ソフトR上で利用可能な様々なロバスト回帰手法にどのようなものがあるかを調べ、乱数シミュレーションにより機能比較を行った。その研究成果について

²⁷ MSD法：共分散行列と平均値ベクトルを用いて外れ値を検出する、Patak(1990)が提案した方法。

²⁸ 統計解析ソフトR：R言語（統計処理やグラフィックに適したプログラミング言語）を実装した、オープンソースフリーのアプリケーションシステム。

²⁹ ロバスト回帰：外れ値に強い回帰モデルに基づく推計方法

統計関連学会連合大会で報告した。また、サービス産業動向調査の売上高補定への実用可能性を確認するため、調査データを用いた試算を行い、25年1月からのサービス産業動向調査の変更に合わせて、ロバストな売上高のレンジチェック法と、比率補定されることになる売上高について刈込比率補定を統計局に提案し、25年1月の速報集計から採用された。

第3 事業所・企業におけるデータエディティング方法の研究

平成22年度において、平成24年経済センサス-活動調査の経理項目が詳細に調査されることを受け、その詳細さのため未回答が多い場合、結果精度に影響を与えることになることから、これを改善するため、経理項目の補定方法（2次試験調査データを用いた回帰分析法等）を検証した。

平成23年度において、2次試験調査データにより、ホットデック法の一手法として経理項目の補定方法について研究を行い、比率による補定が真値との乖離が少ないとの結果を得た。

平成24年度において、多重代入法を用いた補定方法について調査・研究を行った。その研究成果について統計関連学会大会、「UNECE統計データエディティングに関するワークショップ」、科学研究費シンポジウム及び応用統計学シンポジウムIにおいて報告したほか、「統計研究彙報」に研究論文を投稿した。また、23年度までに行った平成24年経済センサス-活動調査2次試験調査データを用いた検証結果のほか、諸外国が行っているデータエディティング方法について情報を収集した内容を、「製表技術参考資料」に取りまとめた。

第3節 統計ニーズの多様化に対応した製表技術に関する研究

第1 秘匿方法及び秘匿処理ソフトに関する研究

平成20年度において、オランダ統計局が開発した秘匿処理ソフトウェアである匿名データ作成用の μ -ARGUS、結果表の二次秘匿処理用の τ -ARGUSについて、最新バージョンを使用した機能等の検証作業をサンプルデータ等を用いて進めていたが、平成21年4月から提供する統計調査データの匿名化に際して、 μ -ARGUSの機能では実現できない手法があったことから、プログラムを独自開発した。また、結果表の二次秘匿処理に関して、 τ -ARGUSの最新バージョンでも処理機能に制約があることから、その適用を見送った。

今後も、秘匿処理用のソフトウェアについては、開発の動向等の情報収集を進め、必要に応じ適用の可能性について検討することとした。

第2 ミクロアグリゲーションの有効性に関する研究

平成20年度において、平成19年度に行った匿名化技法としてのミクロアグリゲーションの有効性に関する研究結果を踏まえ、平成16年全国消費実態調査の個別データを用いて、国内外の先行研究を基に評価を試みた。また、全国消費実態調査の個別データを用いた研究に引き続き、他調査（平成15年住宅・土地統計調査及び平成18年事業所・企業統計調査）の個別データを用いた研究を行った。

第3 諸外国のデータ提供方法の調査・研究

平成21年度において、匿名データ及びオーダーメイド集計表の適切かつ円滑な作成・提供を行う方法の研究のため、諸外国におけるリモート集計等のオンライン方式による提供方法の調査を行った。また、ミクロアグリゲーションによる匿名化手法の事業所・企業データへの諸外国での適用例

の調査・研究を行った。

第4 各種匿名化手法の研究

平成22年度において、諸外国におけるデータ提供の趨勢に対応するため、匿名化手法等に関する諸外国の先行研究の情報収集及び文献の翻訳等を実施し、「Handbook on Statistical Disclosure Control」について製表技術関連資料集の原稿を作成した。また、23年度から提供を予定している労働力調査の匿名データの作成方法について、統計局との共同研究を実施した。

平成23年度において、諸外国におけるデータ提供の趨勢に対応するため、匿名化手法等に関する諸外国の先行研究の情報収集及び文献の翻訳等を実施し、製表技術関連資料集を刊行した。また、労働力調査の匿名データ作成について、統計局と共同研究を行い、統計委員会の了承を得て、提供を開始した。さらに、平成24年度中に提供予定の国勢調査の匿名データの作成方法について、統計局との共同研究を実施した。

平成24年度において、統計局との共同研究により、25年上期に提供予定の国勢調査匿名データの作成方法を検討した。その後、24年10月の統計委員会に国勢調査匿名データの作成方法の諮問が行われ、国勢調査匿名データの提供は25年度中に行われることとなった。

第5 各種匿名化技法による有用性と秘匿性の評価方法に関する研究

匿名化技法の違いが匿名データの有用性と秘匿性に与える影響の評価方法について研究を行った。

平成21年度において、定量的な分析に基づく客観的な評価のための指標の作成が可能か、平成16年全国消費実態調査について、マイクロアグリゲーション手法で作成したデータを用いた実証的研究を行った。

平成22年度において、定量的な分析に基づく相対的評価方法について、諸外国における先行研究の調査を行うとともに、平成16年全国消費実態調査の個票データから、各種匿名化手法及びマイクロアグリゲーション手法で作成したデータを用いて、確率的リンケージ手法³⁰やR-Uマップ³¹やなどによる実証的研究を行った。

平成23年度において、定量的な分析に基づく相対的評価方法について、諸外国における先行研究の調査を行うとともに、平成21年家計調査を用いて、かく乱的匿名化手法を適用したデータの作成及びこのデータに対する有用性と秘匿性の定量的評価による検証を実施した。その結果、秘匿化されたマイクロデータの秘匿性と有用性の程度の定量的、客観的な評価の可能性を確認した。

平成24年度において、定量的な分析に基づく相対的評価方法について、諸外国における先行研究の調査を行うとともに、平成17年国勢調査のデータを用いて、かく乱的匿名化手法を適用したデータの作成及びこのデータに対する有用性と秘匿性の定量的評価による検証を実施した。この研究成果について「統計データベースにおけるプライバシーに関する会議（PSD2012）」で報告した。

³⁰ 確率的リンケージ手法：秘匿処理済データの秘匿性の評価に使われる手法の一つ。秘匿前の原データと秘匿処理済データのペア（組合せ）がリンクされる（一致）ペアかリンクされない（不一致）ペアのどちらに属すると判定するかについて属性値の一致基準及び確率値に従って分類する方法。

³¹ R-Uマップ(Risk-Utility Confidentiality map)：秘匿処理済データの秘匿性と有用性の関係を表すもの。例えば、縦軸に情報量損失率、横軸に度数1の減少率をプロットすること（散布図）により関係を視覚的に見ることができる。

第6 擬似マイクロデータ作成に関する研究

平成21年度において、統計調査の公表済み集計結果表から匿名化データを作成する各種方法論を踏まえ、その派生的な方法として、平成16年全国消費実態調査データの各項目を高次元にクロス集計した集計表を作成した。これをベースとして個別データに近い分布と特性を持つ擬似的なデータを作成し、学生、研究者等に教育・訓練用データとして提供すること及び二次利用におけるサテライト機関でのオーダーメイド集計の際の利用者用サンプルデータとして提供するための研究を行った。

平成22年度において、統計調査の公表済み集計結果表から匿名化データを作成する各種方法論を踏まえ、平成16年全国消費実態調査データの各項目を高次元にクロス集計した集計表をベースとして個別データに近い分布と特性を持つ擬似的なデータを作成し、教育・訓練用データとして提供するための研究を行った。

平成23年度において、公的統計のマイクロデータの利用を推進するため、大学などで統計演習に利用可能とする「擬似マイクロデータ」の作成について、統計委員会等の意見・要望があった。これを受けて、統計調査の公表済み集計結果表から匿名化データを作成する各種方法論を踏まえ、平成16年全国消費実態調査データの各項目を高次元にクロス集計した集計表をベースとして個別データに近い分布と特性を持つ擬似的なデータとして、擬似マイクロデータを作成した。23年8月から、教育用擬似マイクロデータの実用性の検証・評価等を行うため、主に大学、研究機関等の教育関係者からの利用を想定した一般への試行提供を実施した。実施に当たり、統計センターホームページに提供方法について紹介したほか、各種学会等において紹介した結果、45件の利用申出があり、大学等での授業、統計演習等において、363名の院生・学生による利用があった。また、今回の試行提供に当たりアンケートを、利用者に対して実施した。

平成24年度において、前年度に引き続き、施行提供を実施し、その結果、29件（前年度45件）の利用申出と、616名（前年度363名）の利用があった。また、平成16年全国消費実態調査データを用いた研究成果を「製表技術参考資料」に取りまとめるとともに、利用者の利便性の向上等を図るため、擬似マイクロデータ簡易版の作成、提供方法の変更等を行った。さらに、平成19年就業構造基本調査データを用いて、質的変数に関する擬似マイクロデータの作成方法に関する検討を行っている。

第4節 次世代統計利用システムの研究開発

統計センターでは、公的統計の利用可能性の拡充及び統計利用者の利便性向上を図るため、公的統計に係る公表情報その他統計センターが蓄積し、又は独自に作成・収集する各種の統計情報を、インターネット環境下でGIS³²等のアプリケーション、API³³を通じて提供することを目的とした次世代統計利用システム開発計画構想プロジェクトを平成23年4月に立ち上げ、研究開発を進めている。

次世代統計利用システムの開発に当たっては、総務省情報流通行政局と連携して、独立行政法人情報通信研究機構が管理・運用している新世代通信網テストベッド（JGN-X）を活用し、平成23年12月からAPI機能による実証実験を開始した。

平成24年度は、引き続きAPI機能による実証実験を行うとともに、平成24年5月からは、G I

³² GIS : Geographic Information System。地図に関する属性情報をコンピュータにより解析するシステム。

³³ API : Application Programming Interface。OSなどの基本ソフトウェアが開発ツールに対し提供している機能を開発ツール（アプリケーション・ソフトウェア）から呼び出す際の取り決め、仕様のこと。

S機能及びセルデータ集計機能の実証実験を開始した。また、平成25年度からの一般利用者への試行提供に向けて、実証実験を通して得られた意見等を踏まえ、各機能の改修を行った。

第5節 情報収集、外部機関との連携等

第1 外部研究者の採用及び統計センター内研究会への外部研究者の活用

1 外部研究者の採用

外部研究者として、統計学の研究に携わっている博士研究員や大学教育初任段階の若手研究者を非常勤研究員及び任期付研究員（上級研究員）として採用し、データエディティングの効率化の研究、匿名データの有用性と秘匿性の評価方法などの研究を行った。

各年度における外部研究者の採用状況は次のとおりである。

表100 外部研究者の採用状況

年度 (平成)	雇用形態	採用者数	業務内容
20年度	非常勤研究員	2	・ 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
21年度	非常勤研究員	2	・ データエディティングの精度の評価の研究、匿名データの秘匿性の評価方法などの研究
22年度	非常勤研究員	3	・ データエディティングの精度の評価の研究、匿名データの秘匿性の評価方法などの研究
23年度	非常勤研究員	2	・ データエディティングの効率化の研究、匿名データの秘匿性の評価方法などの研究
	任期付研究員(上級研究員)	1	
24年度	非常勤研究員	2	・ データエディティングの効率化の研究、匿名データの有用性と秘匿性の評価方法などの研究

2 統計センター内研究会への外部研究者の活用

大学教授等外部から研究者を招へいした「統計技術研究会」を毎年度開催した。また、平成22年度には、外部有識者を講師に招いた「統計技術研究会講演会」を開催した。

なお、研究会等の開催状況等の詳細は、下記第6節を参照。

第2 データエディティング等の研究動向に関する情報収集

データエディティングに関する研究を推進する上で、特に研究の盛んな国内外の研究動向に関する情報収集が重要であることから、毎年度、諸外国等で開催される会議等に参加した。

各年度におけるデータエディティング等の研究動向に関する情報収集の状況は次のとおりである。

表101 データエディティング等の研究動向に関する情報収集の状況

年度 (平成)	開催地等	参加会議名等
20年度	オーストラリア（ウィーン）	Work Session on Statistical Data Editing（統計データエディティングに関するワークショップ）等3会議
21年度	スウェーデン（ストックホルム）	Modernisation of Statistics Production（統計作成の近代化に関する会議）
	統計数理研究所（東京都立川市）	データ解析環境Rの整備と利用

年度 (平成)	開催地等	参加会議名等
22年度	お茶の水女子大学（東京都文京区）	日本人口学会第62回大会
	ギリシャ共和国（ケルキラ）	Privacy in Statistical Databases 2010（2010年 統計データベースにおけるプライバシーに関する会議）
	オーストラリア統計局	諸外国における人口センサスの匿名データ作成に係る実状把握
	ニュージーランド統計局	
23年度	京都大学（京都府京都市）	日本人口学会第63回大会
	スペイン王国（タラゴナ）	Joint Work Session in Statistical Data Confidentiality（統計データの機密保護に関する合同ワークセッション）
	英国国家統計局	諸外国における人口センサスの匿名データ作成に係る実状把握
24年度	統計数理研究所（東京都立川市）	応用統計学会2012年度年会
	東京大学（東京都目黒区）	日本人口学会第64回大会
	ノルウェー（オスロ）	UNEC E統計データエディティングに関するワークセッション
	イタリア（パレルモ）	統計データベースにおけるプライバシーに関する会議（PSD2012）
	イタリア統計局	匿名化手法等に関する情報収集

第3 データ秘匿に関する情報収集

データ秘匿に関する研究を推進する上で、特に研究の盛んな国内外の研究動向に関する情報収集が重要であることから、毎年度、諸外国等で開催される会議等に参加した。

各年度におけるデータ秘匿の研究動向に関する情報収集の状況は次のとおりである。

表102 データ秘匿の研究動向に関する情報収集の状況

年度 (平成)	開催地等	参加会議名等
20年度	タイ王国（バンコク）	Workshop on Managing Disclosure Risk and Anonymization Tools for Microdata Dissemination(「マイクロデータ普及のための開示リスク管理と匿名化ツールに関するワークショップ」)等3会議
21年度	スペイン王国（ビルバオ）	Work Session on Statistical Data Confidentiality(「統計データの機密保護に関するワークセッション」)等3会議

第4 学会との連携等

統計センターでは、日本統計学会等、統計技術との関連が強い学会に団体加入し、学会の情報を入手するとともに、定期的で開催される会合において発表を行った。また、匿名データの提供及びオーダーメイド集計などの統計センターの業務を広く知らしめるための広報も実施した。

第6節 研究成果の普及等

1 統計技術研究会

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の研究成果を報告し、外部研究者を始めとした有識者から意見等をいただき、それぞれの研究の方向性や手法を検討するため、統計技術研究会を毎年度開催した。

表103 統計技術研究会開催状況

年度(平成)	回数	開催年月日	議 題
20年度	第1回	H20. 12. 19	<ul style="list-style-type: none"> 多変量外れ値の検出について 複数の公表された集計表からの擬似個別データ作成の試み ヨーロッパにおけるデータエディティングと補定に関する取組み動向 「統計データ・エディティング (vol. 3) データ品質への影響」の内容紹介
	第2回	H21. 3. 16	<ul style="list-style-type: none"> 住宅・土地統計調査によるマイクロアグリゲーションの試み マイクロアグリゲーションにおける開示リスクの評価方法について データエディティング及び補定に関する品質評価指標
21年度	第1回	21. 11. 13	<ul style="list-style-type: none"> 擬似データの作成について(中間報告) 各種匿名化技法における有用性と秘匿性の評価方法について 多変量外れ値の検出法-MSD法とその改良版について-
	第2回	22. 3. 16	<ul style="list-style-type: none"> 擬似データの作成方法について 諸外国における統計データの機密保護
22年度	第1回	22. 12. 22	<ul style="list-style-type: none"> 企業財務データを用いた売上高のロバスト回帰による補定 2010年統計データベースにおけるプライバシーに関する会議(PSD2010)出張報告
	講演会	23. 1. 27	<ul style="list-style-type: none"> 統計調査における欠測データの補正に関する研究動向
23年度	第1回	23. 12. 26	<ul style="list-style-type: none"> 経済センサス-活動調査における経理項目の補定方法について 英国における政府統計マイクロデータの作成状況について
	第2回	24. 3. 13	<ul style="list-style-type: none"> 教育用擬似マイクロデータの作成と試行提供状況 統計データの機密保護に関する合同ワークショップ(出張報告)
24年度	第1回	25. 1. 31	<ul style="list-style-type: none"> 統計調査における経理項目の欠測値補定方法について~EMBアルゴリズムによる多重代入法~ UNECE統計データエディティングに関するワークショップの概要(出張報告)
	第2回	25. 3. 12	<ul style="list-style-type: none"> Privacy in Statistical Databases 2012(統計データベースにおけるプライバシーに関する会議)について 匿名化技法としてのスワッピングの可能性-国勢調査マイクロデータを用いた有用性と秘匿性の実証研究について-

2 統計センター実務検討会

統計センター業務についての研究・開発の成果及び事務改善に関する情報等を共有し、その活用を一体的かつ効果的に推進するとともに、職員の人材育成及び専門性の継承を図るため、統計センター実務検討会を毎年度開催した。

各年度における統計センター実務検討会の開催状況は次のとおりである。

表104 統計センター実務検討会開催状況

年度(平成)	20年度	21年度	22年度	23年度	24年度
開 催 回 数	10	10	10	10	10
発 表 テ ー マ 数	12	13	13	18	17

3 製表技術参考資料等の刊行

研究成果の普及を図るため、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。

各年度における製表技術参考資料等の刊行状況は次のとおりである。

表105 製表技術参考資料の刊行状況

年度(平成)	20年度	21年度	22年度	23年度	24年度
刊 行 冊 数	4	4	3	1	8

4 学会等における研究発表

研究成果については、毎年度、学会等において発表を行った。また、平成24年度においては、学術誌「統計研究彙報」への投稿を行った。

各年度における学会等における研究発表の実施状況は次のとおりである。

表106 学会等における研究発表等の実施状況

年度(平成)	年月日	会議等の名称	発表内容	開催地	開催場所
20年度	20. 9. 7 ～ 9. 10	2008年度統計関連学会連合大会(日本統計学会第76回大会)	全国消費実態調査を用いたマイクロアグリゲーションの試み	神奈川県横浜市港北区	慶応義塾大学 矢上キャンパス
	20. 10. 23 ～ 10. 24	研究集会「官庁統計データの公開における諸問題の研究」	マイクロアグリゲーションに関する研究動向 全国消費実態調査を用いたマイクロアグリゲーションの有効性の検証	東京都港区	統計数理研究所
	20. 11. 8	経済統計学会関東支部定例研究会	マイクロデータにおける秘匿性の評価方法に関する一考察	東京都豊島区	立教大学 太刀川記念館
21年度	21. 7. 4	経済統計学会関東支部定例研究会	マイクロデータにおける秘匿性の評価方法に関する一考察	東京都千代田区	専修大学 神田キャンパス
	21. 9. 6 ～ 9. 9	2009年度統計関連学会連合大会(日本統計学会第77回大会)	マイクロアグリゲーションにおける秘匿性の評価方法について	京都府京都市、京田辺市	同志社大学 今出川キャンパス 京田辺キャンパス
	21. 9. 18	「公的統計の有効な提供方法に関する研究」第1回研究会	マイクロデータにおける秘匿性の評価方法に関する一考察	東京都渋谷区	青山学院大学 青山キャンパス
	21. 10. 22 ～ 10. 23	研究集会「官庁統計データの公開における諸問題の研究」	マイクロデータにおける秘匿性の評価方法に関する一考察	東京都立川市	統計数理研究所
	21. 10. 22 ～ 10. 23	研究集会「官庁統計データの公開における諸問題の研究」	マイクロデータにおける秘匿性の評価方法に関する一考察	東京都立川市	統計数理研究所
22年度	22. 4. 10	経済統計学会関東支部2010年4月定例研究会	教育用データの作成について	東京都千代田区	法政大学 市ヶ谷キャンパス
	22. 7. 3	「調査データベース公有化における個人データ保護の統計理論」合同研究会	マイクロデータにおける有用性と秘匿性の評価方法について	東京都立川市	統計数理研究所
	22. 9. 5 ～ 9. 8	2010年度統計関連学会連合大会	マイクロデータにおける有用性と秘匿性の定量的な評価の試み 教育用マイクロデータの作成方法について 多変量外れ値検出法の実データへの適用について - 企業売上高のロバスト回帰による補定 -	東京都新宿区	早稲田大学 早稲田キャンパス
	22. 9. 16 ～ 9. 17	経済統計学会 2010年度(第54回)全国研究大会	教育用マイクロデータ作成の試み - 政府統計マイクロデータの利用拡大に向けて -	大分県大分市	大分大学 旦野原キャンパス
	22. 9. 22 ～ 9. 25	日本行動計量学会第38回大会	高等教育における公的統計の2次利用の枠組み ~ 経済・社会科学における実証分析力の育成	埼玉県さいたま市	埼玉大学
	22. 10. 29 ～ 10. 30	研究集会「官庁統計データの公開における諸問題の研究と他分野への応用」	教育用マイクロデータの作成 マイクロデータにおける有用性と秘匿性の定量的な評価に関する研究	東京都立川市	統計数理研究所
23年度	23. 6. 4	応用統計学会2011年度大会	ロバスト回帰を用いた外れ値に対する乗率の補正	大阪府吹田市	大阪大学 吹田キャンパス
	23. 8. 21 ～ 8. 26	2011年 国際統計協会第58回大会(ISI 2011)	Multivariate Outlier Detection for Regression - Imputation and Aggregation Weight Calibration by ISLS - A Method to Quantitatively Assess Confidentiality and Potential Usage of Official Microdata in Japan	アイルランド ダブリン	タブリンコンベンションセンター

年度 (平成)	年月日	会議等の名称	発表内容	開催地	開催場所
23年度 (続き)	23. 9. 4 ～ 9. 7	2011年度統計関連学会連合大会	<ul style="list-style-type: none"> 教育用擬似マイクロデータの試行提供について 多変量外れ値検出法の実データへの適用について - 企業売上高のロバスト回帰による補定と集計乗率の補正 - マイクロデータにおける攪乱的手法の有効性の検証 	福岡県 福岡市	九州大学 伊都キャンパス
	23. 9. 14 ～ 9. 15	経済統計学会第55回(2011年度)全国研究大会	教育用マイクロデータの作成方法と試行提供について	東京都 八王子市	中央大学 多摩キャンパス
	23. 10. 21 ～ 10. 22	研究集会「官庁統計データの公開における諸問題の研究と他分野への応用」	<ul style="list-style-type: none"> 教育用擬似マイクロデータ(全国消費実態調査)の作成方法と試行提供 公的統計の二次利用(匿名データ、オーダーメード集計)の現状 家計調査マイクロデータを用いた攪乱的手法の有効性に関する研究 	東京都 立川市	統計数理研究所
	23. 11. 11	平成23年度 公的統計のマイクロデータの利用に関する研究集会	教育用擬似マイクロデータの作成と試行提供 - 公的統計のマイクロデータ利用の推進に向けた取組 -	東京都 立川市	統計数理研究所
	24. 3. 2 ～ 3. 3	第8回統計教育の方法論ワークショップ	教育用擬似マイクロデータの開発とその提供 - 統計教育の教材として利用するために -	東京都 国立市	一橋大学 国立キャンパス
24年度	24. 9. 9 ～ 9. 12	2012年度統計関連学会連合大会	<ul style="list-style-type: none"> 企画セッション テーマ「マイクロデータの二次利用その仕組みと研究活用事例」 公的統計の匿名データ及びオーダーメード集計の提供 一般講演 経済センサス-基礎調査-集計結果に基づく「企業グループ」に関する考察 経済調査における売上高の欠測値補定方法について～EDINETデータを用いた検証～ 経済調査における売上高の欠測値補定方法について～多重代入法による精度の評価～ 補定に用いるロバスト回帰手法の性能比較 教育用擬似マイクロデータの提供における現状と課題について - アンケート結果を踏まえて - 	北海道 札幌市	北海道大学 高等教育推進機構
	24. 9. 12 ～ 9. 14	2012年国際官庁統計協会大会(I A O S 2012)	Developing value-added small area statistics with operational data of mobile network	ウクライナ キエフ	ウクライナハウス
	24. 9. 24 ～ 9. 26	UNECE統計データエディティングに関するワークショップ	Multiple Imputation of Turnover in EDINET Data: Toward the Improvement of Imputation for the Economic Census	ノルウェー オスロ	オスロコングレスセンター
	24. 9. 26 ～ 9. 28	統計データベースにおけるプライバシーに関する会議(P S D 2012)	The Potential of Data Swapping as a Disclosure Limitation Method for Official Microdata in Japan: An Empirical Study to Assess Data Utility and Disclosure Risk for Census Microdata.	イタリア パレルモ	ステーリ宮
	24. 10. 24 ～ 10. 26	2012年度科学研究費シンポジウム～統計科学における深化と横断的展開～	経済調査における売上高の欠測値補定方法について～多重代入法による精度の評価～	島根県 松江市	松江テルサ
	24. 11. 21 ～ 11. 22	2012年度科学研究費シンポジウム～統計科学の基礎的理論とその応用～	経済調査における経理項目の欠測値補定方法～EMBアルゴリズムによる多重代入法～	奈良県 奈良市	国際奈良学セミナーハウス

年度(平成)	年月日	会議等の名称	発表内容	開催地	開催場所
24年度(続き)	25. 2. 1	応用統計学シンポジウム I 「ビッグデータイニシアティブ：データ科学の高度化と活用の実際」	大規模データベースにおける欠測値補定のための多重代入アルゴリズム	東京都千代田区	日本大学経済学部
	24. 12	統計研究集報第70号	「経済調査における売上高の欠測値補定方法について～多重代入法による精度の評価～」を投稿		

第7節 技術協力の実施

統計センターでは、これまで培ってきた製表や統計情報の蓄積に係るノウハウ及び技術について、国内外の公的統計の発展に役立てることを目的として、国の行政機関や地方公共団体、統計作成能力向上を目指す開発途上国からの要請に応じ、国勢調査等の製表業務を圧迫しない範囲で、技術協力を行った。

第1 海外への技術協力

海外への技術協力としては、平成15年度から始まったカンボジア統計局への技術支援のため、関係機関（日本政府及び国際協力機構）からの要請に応じ、専門職員の派遣を毎年度行った。なお、国際協力の要請等に対応するため、平成20年度には総務課に国際係を設け、国際窓口業務を一元的に行った。

カンボジアでは、日本政府及び国際協力機構の支援の下、平成23年(2011年)3月に、全国の事業所の所在地、従業員数、売り上げ等を全数調査する経済センサスをカンボジア史上はじめて実施し、24年(2012年)3月に確報結果が公表された。

平成24年度においては、我が国で開催されたカンボジア統計局職員に対する研修の際に、統計集計技術に関する研修を実施した。カンボジアでは、日本政府及び国際協力機構の支援の下、2011年3月に、全国の事業所の所在地、従業員数、売り上げ等を全数調査する経済センサスをカンボジア史上はじめて実施し、2012年3月から、逐次、集計結果が公表されている。

なお、このようなカンボジア政府への協力が評価され、2012年9月にカンボジア政府から統計センターに対し特別表彰が授与された。

各年度におけるカンボジアへの専門職員の派遣状況は次のとおりである。

表107 カンボジアへの専門職員派遣状況

年度(平成)	20年度	21年度	22年度	23年度	24年度
区分					
専門職員の派遣回数	4	4	4	3	3

第2 国の行政機関及び地方公共団体への技術協力

製表業務の技術協力の一環として、統計局が主催する地方別事務打合せ会、合同指導会及び実務研修会等に対して、職員の派遣を毎年度行った。

各年度における国の行政機関及び地方公共団体への職員の派遣状況は次のとおりである。

表108 国の行政機関及び地方公共団体への職員派遣状況

区分 \ 年度(平成)	20年度	21年度	22年度	23年度	24年度
派遣先会議数	5	5	6	4	2
派遣職員数	8	21	35	7	2

第8節 官庁統計とモバイル空間統計³⁴に基づく新たな統計の創出に関する共同研究

本研究は、各種の官庁統計と携帯電話ネットワークのしくみを使用して作成されるモバイル空間統計を組み合わせることにより、新たに有用な統計情報を創出することを目的とする。

平成24年度は、①新たな統計の創出可能性の検討、②新たな統計の作成方法の確立、③技術的課題の解決方法の確立、④新たな統計の精度検証方法の確立について、(株)NTTドコモと共同研究を行った。

研究内容として、インタビューによるニーズの調査、モバイル空間統計を利用した新しい統計に期待される要件の抽出、モバイル空間統計の信頼性の検証、ベンチマーク人口の評価と選定等を行い、研究成果を取りまとめて、研究報告書を(株)NTTドコモと共同で作成した。また、第7回日本統計学会春季集会で「モバイル空間統計の信頼性評価」というテーマで講演した。

³⁴ モバイル空間統計： 携帯電話ネットワークのしくみを使用して作成する人口の統計情報。