

## 第5章 技術の研究に関する事項

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。

平成24年度は、統計分類の格付支援システムの研究、データエディティングに関する研究、統計ニーズの多様化に対応した製表技術に関する研究を行った。

### 第1節 格付支援システムの研究

#### 第1 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

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、格付ルールによる格付支援を可能とする技術の研究を平成23年度に行っている。

平成24年度は、外部委託の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機の読み取り状況を検証した。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の向上、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、格付への影響を検証した。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効率化手法及び実用化への研究

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。同審査の効率化を図るため、平成22年国勢調査人口等基本集計のデータを用いて検証を行い、次回国勢調査への実用に向けて「データエディティング手法検討プロジェクト」において検討を行いつつ、研究を進めている。

平成24年度は、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学的手法で分析し、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、平成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

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。

そのため、外れ値を数学的理論により検出する方法を研究している。

平成24年度は、IRLS<sup>16</sup>よりも性能の良いロバスト回帰<sup>17</sup>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<sup>18</sup>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、平成25年1月からのサービス産業動向調査の変更にあわせて、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、平成25年1月の速報集計から採用された。

### **第3 事業所・企業調査におけるデータエディティング方法の研究**

平成23年度に実施した研究を踏まえ、事業所・企業を対象とする調査における経理項目に対するデータエディティングの方法について研究を行っている。

平成24年度は、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連合大会、「UNECE統計データエディティングに関するワークショップ」（オスロ）、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、平成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、諸外国のデータエディティング方法の研究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 各種匿名化手法の研究**

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を行った。

また、統計局との共同研究により、平成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2 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関し、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。そ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。

<sup>16</sup> IRLS：繰返し加重最小二乗法

<sup>17</sup> 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

<sup>18</sup> 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。

### 第3 擬似マイクロデータ作成及び利活用手法に関する研究

匿名データの提供等統計の二次的利用を推進するため、大学などで統計演習等に利用可能な「擬似マイクロデータ<sup>19</sup>」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、統計センターでは、統計調査の公表済み集計結果表から匿名データを作成する各種方法論を踏まえ、平成23年度には平成16年全国消費実態調査データを用いた量的変数の擬似マイクロデータを作成し、実用性の検証・評価等を行うため、8月から試行提供を開始した。

平成24年度は、前年度に引き続き試行提供を実施し、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、質的変数の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GIS<sup>20</sup>等のアプリケーション、API<sup>21</sup>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機能による情報提供の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、24年5月からは、GIS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、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での外部研究者の活用

統計学の研究に携わっている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。また、大学教授等の外部研究者で構成する「統計技術研究会」を2回開催した。

#### 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及びデータ秘匿に関する研究を推進する上で、研究動向に関する情報収

<sup>19</sup> 擬似マイクロデータ：統計調査を特別集計した統計表から乱数を発生させて生成した擬似的なデータ。

<sup>20</sup> 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。

<sup>21</sup> 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。

集が重要であることから、統計数理研究所（東京都立川市）で開催された「応用統計学会2012年度年会」、東京大学（東京都目黒区）で開催された「日本人口学会第64回大会」、ノルウェーのオスロで開催された「UNEC E統計データエディティングに関するワークショップ」及びイタリアのパレルモで開催された「統計データベースにおけるプライバシーに関する会議（PSD 2012）」に参加するとともに、イタリア統計局の担当者から匿名化手法等に関する情報収集を行った。

### 第3 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、定期的開催される会合において発表を行っている。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ている。

## 第6節 研究成果の普及等

### 第1 統計技術及び研究成果の普及等

#### 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2回開催した。

#### 平成24年度 統計技術研究会開催実績

回	開催年月日	議 題
第1回	25. 1. 31	<ul style="list-style-type: none"> <li>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～</li> <li>UNEC E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）</li> </ul>
第2回	25. 3. 12	<ul style="list-style-type: none"> <li>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</li> <li>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－</li> </ul>

#### 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10回開催した。（第50回は平成25年4月に開催）

#### 平成24年度 統計センター実務検討会開催実績

回	開催年月日	発表内容
第41回	24. 7. 25	<ul style="list-style-type: none"> <li>センサデータの統計的開示抑制に関するES Snet（ヨーロッパ統計システムネットワーク）ワークショップについて</li> </ul>
第42回	24. 8. 8	<p>平成24年度 職員提案制度 改善実績部門 二次審査（プレゼンテーション）</p> <ul style="list-style-type: none"> <li>キーワード方式による生活行動分類符号の新自動格付システムの開発</li> <li>結果表審査事務に用いた「審査表」の利活用</li> <li>家計製表業務における事務改善～「家計調査実績報告」の実施による取組及び効果～</li> <li>納品検査のプログラムの共通化・共有化</li> <li>平成23年社会生活基本調査におけるカラーイメージ化によるデータチェックの効率化</li> <li>品質確保のための能動的マトリックスコード可視化審査の導入</li> </ul>

回	開催年月日	発表内容
第43回	24. 9. 12	・平成24年就業構造基本調査の製表についてー働く未来を考えるー
第44回	24. 10. 10	・統計センターにおける管理会計についてー調査別コスト情報等ー
第45回	24. 11. 14	・総務課・福利厚生室の業務についてーもう少し知ってほしい、総務ー
第46回	24. 12. 12	・次期政府統計共同利用システムの概要
第47回	25. 1. 16	・平成23年社会生活基本調査 結果表審査を終えて
第48回	25. 2. 20	・オーダーメイド集計の現状について
第49回	25. 3. 13	・統計センターにおけるLANのサーバ構成について ・職員の健康管理についてーストレス社会で生きるためにー
第50回	25. 4. 23	・国勢調査匿名データの作成方法とその検証について ・平成24年度米国センサス局出張報告

### 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外における製表技術の研究に関連する出版物の翻訳などの資料を刊行した。

#### 平成24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
24. 7	製表技術参考資料16	教育用擬似マイクロデータの開発とその利用ー平成16年全国消費実態調査を例としてー
24. 8	製表技術参考資料17	諸外国のデータエディティング及び混淆正規分布モデルによる多変量外れ値検出法についての研究
24. 8	製表関連国際用語集3 (改定版)	品質関係用語集 (対訳) EU 統計局 “E S S Quality Glossary 2010” Unit B1 “Quality Classifications” (注) 統計センターHPのみに掲載し、印刷物は作成しない。
24. 9	2011年度統計技術研究会報告	・経済センサスー活動調査における経理項目の推定方法について ・英国における政府統計マイクロデータの作成方法 ・統計データの機密保護に関する合同ワークセッション (出張報告) ・教育用擬似マイクロデータ試行提供の状況-アンケート結果及び今後の方向性- ・諸外国のデータエディティング
24. 9	製表技術参考資料18	平成24年経済センサスー活動調査の経理項目補定方法の研究【第1部】売上(収入)金額の補定方法と外れ値除外方法の比較・検討
24. 11	製表技術参考資料19	国勢調査におけるセレクトティブ・エディティングの可能性 精度向上と集計結果への影響評価に関する取り組み
24. 12	製表技術参考資料20	平成24年経済センサスー活動調査の経理項目補定方法の研究【第2部】費用総額及び費用の内訳の補定方法に係る検討並びに経理項目の補定順の提案
25. 3	製表技術参考資料21	収入額の欠測値補定にふさわしい層分けの提案

### 4 学会等における研究発表

#### 平成24年度 学会等における研究発表実績

年月日	会議等の名称	発表内容	開催地	開催場所
24. 9. 9 ～ 9. 12	2012年度統計関連学会 連合大会	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イド集計の提供 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ーEDINETデータを用いた検証ー ・経済調査における売上高の欠測値補定方法についてー多重代入法による精度の評価ー ・補定に用いるロバスト回帰手法の性能比較	北海道 札幌市	北海道大学 高等教育推進機構

年月日	会議等の名称	発表内容	開催地	開催場所
24. 9. 9 ～ 9. 12 (続き)	2012年度統計関連学会 連合大会 (続き)	・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 (続き)	北海道大学 高等教育推 進機構 (続き)
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	UNEC E 統計データ エディティングに關す るワークショップ	・ Multiple Imputation of Turnover in ED I N E T 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關 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
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

#### 平成24年度 学術誌等への投稿実績

投稿年月	学術誌等名	内 容
24. 12	統計研究彙報第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～

## 第7節 技術協力の実施

### 第1 海外への技術協力

平成24年度は、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、3回にわたり延べ5名の専門職員をカンボジア国への派遣を行うと共に、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。

## 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、統計局が主催する都道府県統計担当職員を対象とする実務研修会の講師として職員の派遣を行った。

### 平成24年度の主な派遣実績

業務内容	人数
平成24年度労働力調査 実務研修会	1
平成24年度家計調査 実務研修会	1

## 第8節 官庁統計とモバイル空間統計<sup>22</sup>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している。

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。

<sup>22</sup> 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