

オンデマンドによる統計作成について —オーストラリア統計局の Table Builder を中心に—

三神 均 ((独) 統計センター)
小島 健一 (野村総合研究所)
伊藤 伸介 (中央大学)

1 はじめに

諸外国の政府統計(公的統計)のマイクロデータにおいては、①匿名化マイクロデータ(個票データに匿名化処理が施されたデータ)の提供、②個票データの提供、③オーダーメイド集計、④オンデマンド型の提供サービス(リモート集計)といった様々な形態による提供が進められてきた。一方、我が国においては、利用者が調査項目(表頭・表側)を選択するだけで集計結果を自動的に出力する、オンデマンド型の集計機能サービスに関する研究が現在進められている。平成 26 年 3 月 25 日に閣議決定された第Ⅱ期「公的統計の整備に関する基本的な計画」では、統計データの有効活用の推進を図る観点から、オンデマンドによる統計作成に関する技術的検証等について実用化に向けた検討を行うことが言及されており、オンデマンド型の集計機能サービスについても、その実用性に関する議論が展開されている。

2 オーストラリアにおけるオンデマンドによる統計作成システム「Table Builder」

集計結果表の公表にあたっては、調査客体が特定されないような秘匿措置が求められるが、一般には、結果表が詳細になるにつれ、また、関連する結果表が増えるにしたがって、秘匿の度合いや困難さは増大する。ゆえに、オンデマンド集計システムにおいて、利用者の様々な要求に応じて集計表を作成・提供するためには、作成された集計表の秘匿に関する十分な検討が求められる。

諸外国では、結果表の有用性と秘匿性のトレードオフ関係を何らかの方法で克服し、オンデマンド集計の提供を実現させている国が存在する。例えば、オーストラリアでは、「Table Builder」というオンデマンド集計システムが開発されている。Table Builder では、インターネットを介してシステムにアクセスし、Census に関する調査項目を指定することによって、自動集計された集計結果表をリアルタイムに受け取ることができる。

TableBuilder には、無料の TableBuilder Basic と、有料の TableBuilder Pro が存在している。TableBuilder Basic では、集計表に用いる変数は、予め設定された分類区分のみが利用可能であって、それらの変数を用いて作成された統計表がオンデマンドで提供可能であるのに対して、TableBuilder Pro においては、利用者がデータベースの中から希望する変数や分類区分を探索的に選定した上で、統計表を作成することが可能となっている。

3 Table Builder における秘匿処理と運用に関する特徴

オーストラリアの“Census and Statistics Act 1905”では、特定の個人や組織が識別可能となるような公表・提供を行ってはならないとされており、いかなる秘匿情報も公開をさせないように、必要な処理をした上で結果を公表することが求められている。そのため、

TableBuilderは、個票データ（非識別データ、deidentified data）を集計した「生の集計表」に対して、意図的に「ノイズ（攪乱値）を付加する」ことで、データの秘匿性を担保している。したがって、TableBuilderにおける秘匿処理の重要な特徴は、TableBuilderの集計システムによって得られる全ての集計表に対して攪乱的措置が施されていることにある。

わずかとは言えノイズを入れることで、分析の対象となる統計表の中の数値が大幅に変わった場合、利用者にとっての情報損失は小さくないと思われる。特に、表の平均値がずれたり、表のデータの分散・共分散構造が崩れたり、また、「0」という結果が「0でない」結果となってしまったりすることは、いずれも、分析対象となる集計表の価値を著しく減じることが考えられる。したがって、オーストラリア統計局では、

- ・ノイズ付与に用いる数値の平均値は0とする
- ・真の結果が0である場合、その値にはノイズを入れない
- ・標本調査についても、ウェイトやウェイトをかける前の値にノイズを入れるのではなく、あくまで集計表に対してノイズを入れる（悉皆調査と標本調査で、ノイズ(攪乱値)の入れ方を極力変えない)
- ・同じ分析をし、同じ表を生成する場合は、常に同じノイズ(攪乱値)を加える

とする攪乱的手法を適用している。特に、最後の特徴は、Differencing と呼ばれる秘匿情報へのアクセス攻撃に対抗する手段である。

また、オーストラリア統計局では、pTable (Perturbation Table) (Andersson *et al.* (2015)) と呼ばれる表を事前に作成してノイズを付加していることや、個票データの1レコードごとに Record Key が割り振られ、これを基に pTable 内の参照値を決定し、同一統計表に対して同一ノイズを付加することを担保している点も特徴として挙げられる。更には、作成された集計表内のセルの値を積み上げた場合に、合計値と整合的（加算性）とできるかどうか、といった点も考慮されている。

このほか、TableBuilder においては、利用にあたって登録が必要なことや、利用規約の遵守状況の確認、組織規模や購入のタイミングに応じた価格体系の設定など運用面においてもいくつかの特徴を挙げることができる。

現在、オーストラリア統計局では、複数の TableBuilder の統合や、利用可能な統計範囲の拡大、更には複数の統計調査のリンケージなどの可能性についても計画が進められている。こうした諸外国における政府統計の動向を踏まえつつ、我が国においてもオンデマンド集計の今後の在り方を検討することが求められる。

参考文献

Andersson, K., Jansson, I., Kraft, K. (2015) “Protection of frequency tables – current work at Statistics Sweden” Paper Presented at UNECE/Eurostat Work Session on Statistical Data Confidentiality, Helsinki, Finland, pp1-20.

谷道正太郎・伊藤伸介・小島健一(2016)「オーストラリアのオンデマンド集計に関する調査研究」(統計センター『製表技術参考資料』No. 33として刊行予定)