

## 平成 28 年経済センサス - 活動調査における秘匿処理について

太田 将彰<sup>†</sup>

北原 昌嗣<sup>‡</sup>

The Statistical Disclosure Control on the Economic Census for Business Activity in 2016

OHTA Masaaki  
KITAHARA Masatsugu

### 目 次

はじめに

- I 経済センサスにおける秘匿処理
  - 1 秘匿処理の対象
  - 2 秘匿処理方法
  - 3 平成 28 年経済センサス - 活動調査で実際に行った秘匿処理
  - 4 秘匿処理の審査
- II 秘匿処理の結果
- III 今後の秘匿処理に向けた提案
- IV おわりに

付録

---

<sup>†</sup> 総務省統計局統計調査部経済統計課

<sup>‡</sup> 総務省統計局統計作成支援課

はじめに<sup>1</sup>

経済センサスは、我が国すべての事業所及び企業の経済活動の状態を明らかにし、包括的な産業構造を把握する統計であり、基礎調査と活動調査の2調査から成り立っている。基礎調査は事業所・企業の基本構造（名称、所在地、事業の種類、従業者数等）を調査し、活動調査は経済活動（売上（収入）金額（以下、売上金額）や費用等の経理事項）を調査するというそれぞれの役割を担っている。また、基礎調査は活動調査の名簿を作成するための統計という側面も併せ持っている。

経済センサスのような全数調査では、秘匿処理が公表の際に重要な役割を果たしている。経済センサスの結果表をよく見ると、各所に数値ではなくXの記号を見つけることができる。これは、当該セルに該当する対象数<sup>2</sup>が1又は2であるために、その企業名が公表されていなくとも、地域区分や産業分類などの細分化された結果表情報（メタ情報）や公開している企業情報等により該当する企業が割り出されるおそれがあることから、売上金額等の同業他社に知られると不都合になる可能性のある数値をXで置き換え秘匿しているのである。近年の企業における情報管理に対する意識の高まりから、企業サイドは自社の企業情報の開示には慎重になっており、統計調査に回答した結果、企業情報が漏れてしまうケースを懸念し、調査回答に躊躇する企業も少なからず存在する。統計調査は、個々の企業情報が他者に漏れないことを前提に情報を提供してもらっていることから、もし、秘匿処理が不完全で個々の企業情報が他者に知られるようなことになると統計調査の信頼性が揺らぎ、調査回答拒否につながりかねない。そのため、統計表の秘匿処理は公表の数か月前から準備をし、結果数値の入った統計表が製表された直後から秘匿処理を行い、過少秘匿になっている箇所がないか検証した上で結果を公表している。一方で、せっかく得られた調査情報を秘匿処理により非公表としてしまうのは、公益を損ねることにもつながりかねないため、秘匿処理の過程で過剰な秘匿が生じないように秘匿箇所を調整する作業も併せて行っている。

本稿では、平成28年経済センサス-活動調査で実施した秘匿処理方法を紹介し、その結果を分析することで、秘匿処理の効率化の観点から検討すべき課題を提起する。

## I 経済センサスにおける秘匿処理

### 1 秘匿処理の対象

経済センサスの秘匿処理では、事業所及び企業の売上金額等の経理項目に関連する集計事項を秘匿の対象としており、その処理が講じられるようになったのは経済センサスの前身である事業所統計調査<sup>3</sup>にまで遡る。事業所統計調査では経理項目以外に従業者数も秘匿の対象としていた時期があったが、昭和42年に当時の総理府統計局長から行政管理庁統計基準局長あてに従業者数の秘匿を今後講じないこととしたいと申請があり、それを行政管理庁統計基準局長が承認するという形で従業者数が秘匿対象から外れたのである。その理由として、1. 従業者数の秘匿は、昭和27年に行政管理庁長官が統計審議会に諮問した結果決定されたもので、その後、10余年間の情勢推移をみると、社会的、経済的事情も大きく変わり、統計に対する一般の理解の程度が格段に進

<sup>1</sup> 本稿の作成に当たっては、松本正博氏（独立行政法人統計センター）、関野秀峰氏（総務省統計局）及び経済センサスシステム開発担当の皆様から大変貴重な助言をいただいた。ここに改めて感謝の意を表したい。なお、本稿の内容は個人的見解も含まれており、筆者の所属する機関の公式見解を示すものではない。

<sup>2</sup> 本稿で用いる「対象数」とは、「事業所数（又は企業等数）」のことである。

<sup>3</sup> 「事業所統計調査」は、昭和22年に開始され、平成8年の調査から企業の実態把握を充実させたことに伴い、「事業所・企業統計調査」に名称変更を行い、平成18年の調査を最後に、平成21年から経済センサスに統合された。経済センサス創設時に、「事業所・企業統計調査」、「サービス業基本調査」等は廃止されている。

み、秘匿扱いとする事情が消滅していること、2. 大企業では、従業者数を会社年鑑等の資料で開示しており、それを知るのに不便を感じない実情であり、事業所統計において従業者数を秘匿にする意味がなくなったこと、3. 従業者数を秘匿にすることにより、統計の利用価値を大いに損なうおそれがあること、4. 統計表の細分化に伴い、秘匿箇所が急速に増加し事務量が増大したこと、またその事務量の増大が、集計結果の早期公表の要望の妨げになることをあげている。

経済センサス - 基礎調査<sup>4</sup>（以下、基礎調査）及び経済センサス - 活動調査（以下、活動調査）では、表1に掲げる集計事項を秘匿の対象としている<sup>5</sup>。基礎調査と活動調査を比較すると、活動調査の秘匿対象事項が多いことがわかる。経済センサスでは、売上金額等の経理項目を秘匿対象としているため、必然として詳細な経済活動を調査する活動調査の秘匿対象事項が基礎調査に比べ多くなってしまいうのである。

表1 経済センサスにおいて秘匿対象となる集計事項<sup>6</sup>

	集計事項	
	基礎調査	活動調査
企業等に関する集計	売上（収入）金額、1企業等 当たり売上（収入）金額	売上（収入）金額、費用総額、売上原価、販売費及び一般 管理費、給与総額（個人経営企業では給料賃金）、福利厚 生費、動産・不動産賃借料（個人経営企業では地代家 賃）、減価償却費、租税公課、外注費、支払利息等、付加 価値額、設備投資額、事業活動別売上金額（売上金額の内 訳）、一般消費者と行った電子商取引の額、1企業等当 たり売上（収入）金額、1企業等当たり付加価値額
事業所に関する集計	売上（収入）金額、1事業所 当たり売上（収入）金額	売上（収入）金額、1事業所当たり売上（収入）金額、従 業者1人当たり売上（収入）金額、付加価値額、1事業所 当たり付加価値額、事業従事者1人当たり付加価値額

## 2 秘匿処理方法

秘匿処理の大まかな手順としては、はじめに、統計表において対象数が1又は2であるセルの集計事項（売上金額等の秘匿対象となる集計事項）を秘匿する。これを1次秘匿と呼ぶ。そして、合計（又は上位分類）との差し引きにより、1次秘匿セルの数値が判明してしまう場合、1次秘匿セルの数値が算出できないよう関係するセルを秘匿する。これを2次秘匿と呼ぶ。さらに、2次秘匿したセルの数値が同様の方法で算出できてしまう場合、2次秘匿セルのための秘匿も行い、過少秘匿がなくなるまで秘匿作業を繰り返していく。また、秘匿箇所をもった当該表以外の表にその秘匿された数値の合計値や内訳数値が表章されていることがあるため、併せてそれらなるべく合計値（総数）が秘匿されないように2次秘匿する。前者は表内秘匿で後者は表間秘匿に関する記述である。

具体例をあげると、例えば、表2のような統計表がある場合、1次秘匿箇所は事業所数が1及び2の売上金額のセルとなり、該当するセルをまず始めに秘匿する。（表2—1 1次秘匿は太字Xで表す。以下同じ。）

<sup>4</sup> 平成21年基礎調査では売上金額の調査を行わなかったが、平成26年基礎調査では売上金額を調査している。

<sup>5</sup> 活動調査は産業横断的集計に限定した秘匿対象となる集計事項を記載している。

<sup>6</sup> 経済センサスの集計方法は大きく分けて、「企業等に関する集計」と「事業所に関する集計」の2通りがある。本所と支所を含めた企業全体のデータが欲しいときは「企業等に関する集計」、本所と支所を個々に把握したデータが欲しいときは「事業所に関する集計」を利用することになる。企業ではなく「企業等」としているのは、会社企業のほかに独立行政法人などの会社以外の法人や個人経営の事業所も含まれているためである。個人経営では、同一の経営者が複数の事業所を営んでいる場合は、それらをまとめて一つの「企業等」としている。

表2 産業分類、経営組織別売上金額

産業分類	総数		個人		会社		会社以外の法人	
	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)
産業中分類AA	146	368,253	28	28,367	101	297,966	17	41,920
産業小分類AAA	5	53,448	2	4,585	3	48,863	-	-
産業小分類AAB	4	26,647	1	2,212	3	24,435	-	-
産業小分類AAC	97	207,956	17	13,425	68	157,689	12	36,842
産業小分類AAD	2	6,746	-	-	2	6,746	-	-
産業小分類AAE	38	73,456	8	8,145	25	60,233	5	5,078

表2-1 産業分類、経営組織別売上金額

産業分類	総数		個人		会社		会社以外の法人	
	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)
産業中分類AA	146	368,253	28	28,367	101	297,966	17	41,920
産業小分類AAA	5	53,448	2	X	3	48,863	-	-
産業小分類AAB	4	26,647	1	X	3	24,435	-	-
産業小分類AAC	97	207,956	17	13,425	68	157,689	12	36,842
産業小分類AAD	2	X	-	-	2	X	-	-
産業小分類AAE	38	73,456	8	8,145	25	60,233	5	5,078

次に2次秘匿を考える。行方向に1次秘匿されたセルの2次秘匿候補セルを見ていくと、産業小分類AAA及びAABの会社の列の売上金額のセルを2次秘匿する必要がある。(表2-2(2次秘匿は斜体Xで表す。以下同じ。))

表2-2 産業分類、経営組織別売上金額

産業分類	総数		個人		会社		会社以外の法人	
	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)
産業中分類AA	146	368,253	28	28,367	101	297,966	17	41,920
産業小分類AAA	5	53,448	2	X	3	X	-	-
産業小分類AAB	4	26,647	1	X	3	X	-	-
産業小分類AAC	97	207,956	17	13,425	68	157,689	12	36,842
産業小分類AAD	2	X	-	-	2	X	-	-
産業小分類AAE	38	73,456	8	8,145	25	60,233	5	5,078

次に列方向の2次秘匿候補セルを見ていくと、産業小分類AADの総数の売上金額の2次秘匿が必要であるため、産業小分類AADの次に事業所数が少ない産業小分類AABを2次秘匿する。また、このセルを2次秘匿したことにより、産業小分類AABの会社における売上金額の秘匿は必要なくなるため(過剰秘匿の発生)、秘匿の解除をする。(表2-3)

表2-3 産業分類、経営組織別売上金額

産業分類	総数		個人		会社		会社以外の法人	
	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)
産業中分類AA	146	368,253	28	28,367	101	297,966	17	41,920
産業小分類AAA	5	53,448	2	X	3	X	-	-
産業小分類AAB	4	X	1	X	3	24,435	-	-
産業小分類AAC	97	207,956	17	13,425	68	157,689	12	36,842
産業小分類AAD	2	X	-	-	2	X	-	-
産業小分類AAE	38	73,456	8	8,145	25	60,233	5	5,078

これで秘匿処理は完成であるが、もし、この2次元クロス表の秘匿処理完了後に、後続して行った他表の秘匿処理で産業小分類 AAE の総数が秘匿されることが判明した場合、この表においても産業小分類 AAE の総数を秘匿しなくてはならなくなるため、それに対応した2次秘匿（産業小分類 AAE の個人）が発生する。このとき、産業小分類 AAE の行方向を見ると、事業所数が最も少ない会社以外の法人が2次秘匿箇所になるのではないかと思うかもしれないが、ここを2次秘匿してしまうと、秘匿箇所数が増加してしまうため（新たに会社以外の法人の列方向の秘匿が必要になる。）、次に事業所数の少ない個人を2次秘匿する。（表2-4）

表2-4 産業分類、経営組織別売上金額

産業分類	総数		個人		会社		会社以外の法人	
	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)	事業所数	売上金額 (百万円)
産業中分類AA	146	368,253	28	28,367	101	297,966	17	41,920
産業小分類AAA	5	53,448	2	X	3	X	-	-
産業小分類AAB	4	X	1	X	3	24,435	-	-
産業小分類AAC	97	207,956	17	13,425	68	157,689	12	36,842
産業小分類AAD	2	X	-	-	2	X	-	-
産業小分類AAE	38	X	8	X	25	60,233	5	5,078

### 3 平成28年経済センサス-活動調査で実際に行った秘匿処理<sup>7</sup>

#### (1) 前提条件

秘匿処理については、以下を前提条件とした上で行った。

##### ア 秘匿セル数の最小化

結果表単位で秘匿セル数が可能な限り少なくなるように処理を行う。秘匿セル数を最小にする作業の下で、秘匿の優先順位が同位の秘匿セル候補が複数存在する場合は、そのうちの「不詳」など優先して2次秘匿したほうが良いと考えられるセルを秘匿し、そのような区分がないときは対象数のより少ないセルを2次秘匿する。対象数のより少ないセルを優先して2次秘匿する理由は、1. 秘匿による情報の損失を最小限にするため、2. 開示リスクが大きいためである。後者については、例えば、3つの事業所が対象数である売上金額を表章したセルがある場合、そのうちの2つの事業所が懇意にしており、お互いの売上金額を把握していたとすると、残りの1つの事業所の売上金額は当該セルの売上金額合計との差し引きにより、お互い懇意にしている事業所には推計できてしまう。そういったリスクは対象数が少ないほど大きくなるため、優先的に秘匿している。

##### イ 2次元クロス表の切り出し

3次元以上の表になってしまうと、パターン（処理量）が膨大になり、処理時間が大幅にかかってしまう。そのため、多次元表はすべて2次元クロス表に切り出して、秘匿処理を行う。

##### ウ 表間における秘匿

秘匿処理は、「秘匿処理用結果表の構成」（表3-1及び3-2）に記載している処理順で行う。ある結果表の秘匿処理前に他表の秘匿処理で秘匿が決定した集計値については、その秘匿結果を反映した上で結果表の秘匿処理を行い、この処理順で新たな秘匿が発生しなくなるまで反復処理を行う。

<sup>7</sup> ただし、事業所に関する集計の「鉱業、採石業、砂利採取業」及び「卸売業、小売業」の秘匿処理については、基本的な考え方は同じであるが、統計表の作りが異なる等のことから、この限りではない。また、秘匿データを取り扱っている観点から、詳細な点については、記述が省略されている箇所もあることにご留意いただきたい。

エ 企業等に関する集計

事業所数と企業等数の両方を集計する結果表では、企業等数を秘匿処理の対象数とする。

オ 「総数」と内訳の「合計値」

「総数」と内訳の「合計値」が一致しない場合は、2次秘匿処理を行う必要はないものとする。

(2) 秘匿処理の手順

秘匿処理の手順は以下のとおりである。

ア 秘匿処理用結果表の作成

まず始めに、公表する結果表と基本的に同じ様式<sup>8</sup>で、集計事項の結果数値を対象数に限定した表を秘匿処理用結果表として作成する。例として、平成28年活動調査の事業所に関する集計（産業横断的集計）の売上（収入）金額等における4-1表の3次元クロス表の秘匿処理用結果表を下記（例）に示す。なお、当該結果表の数値の「総数」に当たる区分が他の結果表にある場合は、表間秘匿のために、その「総数」の区分を追加した秘匿処理用結果表を作成する。（平成28年活動調査における秘匿処理用結果表の構成については表3-1及び3-2参照）

(例) 事業所に関する集計の産業横断的集計における4-1表を秘匿処理する際の秘匿処理用結果表

地域	産業	総数（単独・本所・支所）	単独事業所	本所・本社・本店	支所・支社・支店
全国	A~R 全産業（S公務を除く）				
	A~B 農林漁業				
	A 農業、林業				
	01 農業				
	02 林業				
	B 漁業				
	03 漁業（水産養殖業を除く）				
	04 水産養殖業				
	C~R 非農林漁業（S公務を除く）				
	∫				
北海道	A~R 全産業（S公務を除く）				
	A~B 農林漁業				
	A 農業、林業				
	01 農業				
	∫				

秘匿処理用結果表では、事業所数のみを集計対象として表章し、秘匿箇所を検出。その秘匿箇所セルに対応する売上金額等を公表用結果表にて秘匿する。

実際の4-1表に全国の表章はないが、全国が「総数」に当たる区分であることから、表間秘匿のために、秘匿処理用結果表では「全国」の区分を追加している。表3-2「秘匿処理用結果表の構成」の△に対応する区分である。

イ 秘匿処理用結果表における処理

秘匿処理は、「秘匿処理用結果表の構成（表3-1及び3-2）」の処理順に行った。処理順によって、前の処理で秘匿されたセルの区分がその後の秘匿処理用結果表のセル区分に表章されている場合は、そのセルを秘匿した上で、当該表の秘匿処理を行った。以下がその処理手順である。

(ア) 2次元クロス表の切り出し

上記（例）をもとに説明をする。「地域<sup>9</sup>」、「産業分類」、「単独・本所・支所の別」の3

<sup>8</sup> 公表する結果表の表頭には、事業所数、従業者数、売上金額などが表章されており、そこから事業所数だけをピックアップし、秘匿処理用結果表を作るイメージである。

<sup>9</sup> 「地域」の「大都市圏」については、都道府県に対する全国のように「総数」に当たる区分がないため、「総数」のない形で切り出している。また、「大都市圏」とそこに属する「都道府県」の数値が一致した場合に限り、「都道府県」と「大都市圏」の形で2次元クロス表を切り出し、秘匿処理を行った。例えば、北海道で、ある産業の数値が秘匿されていた場合、その産業の事業所数が札幌大都市圏で表章されている事業所数と等しいならば、「北海道」＝「札幌大都市圏」と推測できる。この場合、札幌大都市圏においても当該産業の二次秘匿が必要となる。

次元クロス表を2次元クロス表に切り出す場合、組み合わせとしては、以下の3通りがある。

- a. 「産業分類」×「単独・本所・支所の別」(欄外<sup>10</sup>「地域」)
- b. 「地域」×「単独・本所・支所の別」(欄外「産業分類」)
- c. 「地域」×「産業分類」(欄外「単独・本所・支所の別」)

a、b、cのそれぞれの組み合わせの中で、「総数(合計値)」とその内訳に分けて2次元クロス表を切り出す。切り出す順番(秘匿処理の順番)は、「総数」からその内訳の順である。産業分類のような階層がある場合は、上位から下位へ向かう順とする。a、b、cの秘匿処理順は、順列を考えると $3 \times 2 \times 1 = 6$ で6通りあり、秘匿処理順によって秘匿セル数も変わってくるため、6通りすべてにおいて秘匿処理を行う。

a. 「産業分類」×「単独・本所・支所の別」(欄外「地域」)の2次元クロス表の例

欄外(地域(全国・都道府県・大都市圏))

産業	総数	単独事業所	本所・本社・本店	支所・支社・支店
A~R 全産業 (S公務を除く)				
A~B 農林漁業		対象事業所数を表章する。		
C~R 非農林漁業 (S公務を除く)				

→該当表数=全国(1)+都道府県(47)+大都市圏(14)=62

欄外(地域(全国・都道府県・大都市圏))

産業	総数	単独事業所	本所・本社・本店	支所・支社・支店
A~B 農林漁業				
A 農業, 林業		対象事業所数を表章する。		
B 漁業				

→該当表数=(全国(1)+都道府県(47)+大都市圏(14))×(A~B(1)+C~R(1))=124

欄外(地域(全国・都道府県・大都市圏))

産業	総数	単独事業所	本所・本社・本店	支所・支社・支店
G 情報通信業				
G1 情報通信業 (通信業, 放送業, 映像・音声・文字情報制作業)		対象事業所数を表章する。		
G2 情報通信業 (情報サービス業, インターネット附随サービス業)				

→該当表数=(全国(1)+都道府県(47)+大都市圏(14))×1.5桁産業分類を持つ産業大分類(7)=434

欄外(地域(全国・都道府県・大都市圏))

産業	総数	単独事業所	本所・本社・本店	支所・支社・支店
A 農業, 林業				
01 農業		対象事業所数を表章する。		
02 林業				

→該当表数=(全国(1)+都道府県(47)+大都市圏(14))×(産業大分類(18)+1.5桁産業分類(14)-1.5桁産業分類を持つ産業大分類(7))=1550

<sup>10</sup> ここでいう「欄外」とは、統計表の表側でも表頭でもない統計表の外枠に記載されるその統計表が何を示した統計表なのかを表すものである。例えば、aの場合だと、統計表の表側(又は表頭)が「産業分類」、表頭(又は表側)が「単独・本所・支所の別」という統計表の作りで、欄外に「地域」とあるので、その統計表が全国を示した数値なのか、北海道を示しているものなのかが欄外を見るとわかるようになっている。

b. 「地域」×「単独・本所・支所の別」(欄外「産業分類」)の2次元クロス表の例

欄外(産業分類)

地域	総数	単独事業所	本所・本社・本店	支所・支社・支店
全国				
北海道		対象事業所数を表章する。		
青森県				
∫				

→該当表数=総数を含む産業分類数=130

c. 「地域」×「産業分類」(欄外「単独・本所・支所の別」)の2次元クロス表の例

欄外(単独・本所・支所の別)

地域	A~R 全産業 (S公務を除く)	A~B 農林漁業	C~R 非農林漁業 (S公務を除く)
全国			
北海道		対象事業所数を表章する。	
青森県			
∫			

→該当表数=単独・本所・支所の別(4)=4

欄外(単独・本所・支所の別)

地域	A~B 農林漁業	A 農業, 林業	B 漁業
全国			
北海道		対象事業所数を表章する。	
青森県			
∫			

→該当表数=単独・本所・支所の別(4)×(A~B(1)+C~R(1))=8

欄外(単独・本所・支所の別)

地域	G 情報通信業	G1 情報通信業(通信業, 放送業, 映像・音声・文字情報制作業)	G2 情報通信業(情報サービス業, インターネット附随サービス業)
全国			
北海道		対象事業所数を表章する。	
青森県			
∫			

→該当表数=単独・本所・支所の別(4)×1.5桁産業分類を持つ産業大分類(7)=28

欄外(単独・本所・支所の別)

地域	A 農業, 林業	01 農業	02 林業
札幌大都市圏		対象事業所数を表章する。	
北海道			

→該当表数=単独・本所・支所の別(4)×(産業大分類(18)+1.5桁産業分類(14)-1.5桁産業分類を持つ産業大分類(7))×大都市圏(14)=1400

(イ) 2次元クロス表における秘匿処理

切り出した2次元クロス表内で、可能な限り秘匿セルが少なくなるように秘匿処理を行う。2次秘匿を行方向から始める場合と、列方向から始める場合でも秘匿セルの数は変わってくるので、両方試し、秘匿セルの少なくなる方を採用する。また、2次秘匿を行う際に、秘匿対象候補セルが複数上がってくるときがあるが、この場合は、2次秘匿の優先順位※をあらかじめ定めているため、この優先順位の高い順に秘匿を行っている。



そして、2次元クロス表内で、新たな秘匿セルが発生しなくなるまで反復処理を行い、その後、過剰秘匿されてしまった2次秘匿の解除を行う。これは、今までの反復処理で、過剰秘匿されてしまった2次秘匿セルを見つけ出し秘匿の解除をする処理である。

#### ※2次秘匿の優先順位

(秘匿対象の結果表が3次元で、2次元クロス表の行方向から秘匿処理を始めた場合)

- 1位 列方向及び欄外方向共に秘匿セルが一つ又は対象数1のセルが二つあるセル
  - 2位 列方向に秘匿セルが一つ若しくは対象数1のセルが二つあり、欄外方向に秘匿セルが複数(対象数1のセルが二つを除く)あるセル又は列方向に秘匿セルが複数(対象数1のセルが二つを除く)あり、欄外方向に一つ若しくは対象数1のセルが二つあるセル
  - 3位 列方向及び欄外方向のうち一方に秘匿セルが一つ又は対象数1のセルが二つあり、他方が以下(i)に該当しないセル。(i)に該当するセルを秘匿すると、0又は「-」のセルの影響により結果として秘匿セルが増えてしまう。  
(i) 当該方向(列方向又は欄外方向)に対象数が「-」のセルがある場合に、その「-」のセルの行と、2次秘匿を必要としている秘匿セル(対象数1のセルが二つの場合はそのいずれか)の当該方向がクロスするセルが秘匿されている。
  - 4位 列方向及び欄外方向共に秘匿セルが複数(対象数1のセルが二つを除く)あるセル
  - 5位 列方向及び欄外方向の一方に秘匿セルが複数(対象数1のセルが二つを除く)あり、他方が(i)に該当しないセル
  - 6位 列方向及び欄外方向共に(i)及び以下(ii)に該当しないセル。(ii)のセルを秘匿すると、当該方向への秘匿の際に、内訳から「総数」が秘匿されてしまう。  
(ii) 当該方向の他の内訳区分の対象数がすべて「-」
  - 7位 列方向及び欄外方向共に(ii)に該当しないセル
  - 8位 2次秘匿対象のすべてのセル
- (注) 列方向から秘匿処理を始める場合も同様の順位である。その場合、上記優先順位の行を列に置き換えて処理を行う。

#### (ウ) 秘匿対象の確定

上記(ア)及び(イ)で行った秘匿処理から、最も秘匿セル数の少なかった処理順の秘匿結果を採用する。秘匿セル数が最小の秘匿結果が複数ある場合は、秘匿セルに該当している事業所数(又は企業等数)の合計が最小の秘匿結果を採用する。それも複数存在する場合は、最初に処理した秘匿結果を採用することとした。

#### ウ 秘匿処理の反復

イの処理をすべての秘匿処理用結果表で終えたら、再度「秘匿処理用結果表の構成」の処理順で、秘匿処理されたセルを反映させた上で、新たな秘匿セルが発生しなくなるまで秘匿処理の反復をし、秘匿セルを確定する。

#### エ 公表用結果表の秘匿対象セルの秘匿処理

ウの処理を終え、すべての秘匿処理用結果表の秘匿セルが確定することで、公表用結果表における秘匿セルが判明するので、公表用結果表の該当セルに秘匿を反映させ、秘匿処理は完成である。

表3-1 「企業等に関する集計」のうち「産業横断的集計」の秘匿処理結果表の構成

地域区分	産業分類	全企業等											うち企業事業所		うち支店		うち営業所		うち会社企業	資本金融級	企業常用用者規模	企業従業員規模	売上(収入)金額階級	電子商取引の有無(3階級区分)	電子商取引の額(4区分)	電子商取引の額(5区分)		
		1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6									7	8
全国	大	○	△		○	△	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	1.5桁																											
	中																											
	小																											
県	大	○	△		○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	1.5桁																											
	中																											
	大																											
大都市圏	大	○	△		○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	1.5桁																											
	中																											
	大																											

＜凡例＞

- ：結果表と同じ区分
- △：結果表にはないが、秘匿処理用結果表には追加する区分
- ◎：集計されたとおりの対象数とする。
- ◆：「個人」の対象数を「-」とする。
- ：「J62 銀行業」及び「J63 協同組織金融業」の対象数を「-」とする。
- ：「J62 銀行業」及び「J63 協同組織金融業」並びに「個人」の対象数を「-」とする。
- ☆：「J 金融業、保険業」並びに「会社以外の法人」及び「個人」の対象数を「-」とする。

(a) 売上(収入)金額、費用総額、給与総額(個人は給料賃金)、動産・不動産賃借料(個人は地代家賃)、減価償却費、租税公課、付加価値額、設備投資額、事業活動別(売上高内訳)、企業等当たり売上(収入)金額、1企業等当たり付加価値額

(b) 福利厚生費、外注費

(c) 支払利息等

(d) 売上原価、販売費及び一般管理費

(e) 一般消費者と行った電子商取引の額

(f) 一般消費者と行った電子商取引の額

【電子商取引の有無に関する秘匿処理】

- ・ 「電子商取引を行った」が秘匿された場合は、「一般消費者と行った」「電子商取引を行った」「他の企業のみと行った」も一括して秘匿
- ・ 「電子消費者のみと行った」「一般消費者及び他の企業と行った」「他の企業のみと行った」も一括して秘匿
- ※ 「電子商取引を行なった」「電子商取引を行なかつた」以外の項目の対象数が「1,2」の場合でも、1次秘匿は行わない
- ・ 「電子商取引を行なった」より「電子商取引を行なかつた」を優先して秘匿

【国内支所の分布範囲に関する秘匿処理】

- ・ 複数事業所企業と国内複数事業所企業を対象が一致する場合は、2次秘匿を実施

表3-2 「事業所に関する集計」のうち「産業横断的集計」の秘匿処理用結果表の構成

地域区分	産業分類	民営事業所 (除く外国の会社、法人でない団体)																単独・ (3 業 区分) 支所の別			従 (7 業 区分 分規) 業規模						外国の会社			法人でない団体																	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1	2	3	4	5	6	7	8	9	10	11	12	13	14
全国	処理順 結果表 番号	1-1	1-2	2-1	2-2	3-1	3-2	4-1	4-2	5-1	5-2	6	7	8-1	8-2	9-1	9-2	1-1	1-2	2-1	2-2	3-1	3-2	4-1	4-2	5-1	5-2	8-1	8-2	9-1	9-2	1-1	2-1	1-1	2-1	3											
	大	○	○	△	△	○	○	△	△	○	○	○	○	○	○	△	△	○	○	△	△	○	○	△	△	○	○	○	○	○	△	△	○	○	△	△	○	○	△								
	1.5桁	○	○	△	△	○	○	△	△	○	○	○	○	○	○	△	△	○	○	△	△	○	○	△	△	○	○	○	○	○	△	△	○	○	△	△	○	○	△								
	中	○	○	△	△	○	○	△	△	○	○	○	○	○	○	△	△	○	○	△	△	○	○	△	△	○	○	○	○	○	△	△	○	○	△	△	○	○	△								
小	○	○	△	△	○	○	△	△	○	○	○	○	○	○	△	△	○	○	△	△	○	○	△	△	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○								
3.5桁分類	○	○	△	△	○	○	△	△	○	○	○	○	○	○	△	△	○	○	△	△	○	○	△	△	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○								
細	○	○	△	△	○	○	△	△	○	○	○	○	○	○	△	△	○	○	△	△	○	○	△	△	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○								
県	大	○	○	○	○	○	○	○	○	△	△	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
1.5桁	○	○	○	○	○	○	○	○	○	△	△	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
中	○	○	○	○	○	○	○	○	○	△	△	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
大都市	大	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
1.5桁	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
行政区・特別区	大	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
1.5桁	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
大都市以外の市町村	大	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
1.5桁	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
大都市圏	大	○	○	△	△	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
1.5桁	○	○	△	△	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
中	○	○	△	△	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○							
秘匿対象	売上(収入)金額*	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇								
	付加価値額*	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎								

※：1事業所当たり、従業員1人当たり及び事業従事者1人当たりを指す。なお、売上(収入)金額については事業活動別を指す。

＜凡例＞

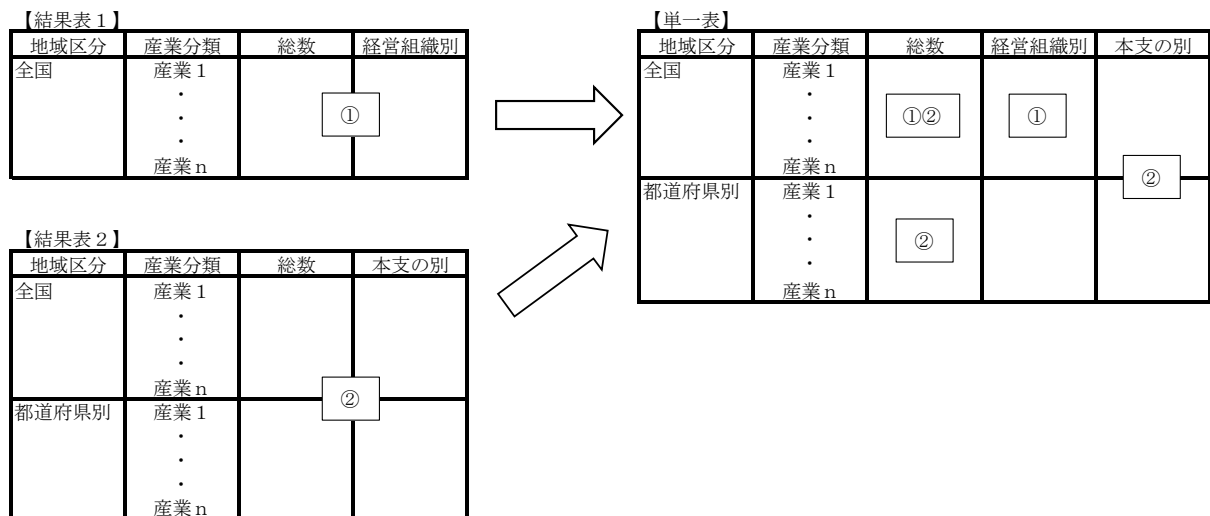
- ：結果と同じ区分
- △：結果表にはないが、秘匿処理用結果表には追加する区分
- ◇：単独事業所を除きネットワーク型産業及びその上位分類の対象数を「1」とする。
- ◎：集計されたおりの対象とする。

#### 4 秘匿処理の審査

平成28年経済センサス - 活動調査では、3で記した秘匿処理のとおり、コンピュータで秘匿を行った後、秘匿箇所が適切か確認を行った。確認には、サービス業基本調査で松本（2003）が提案した個々の結果表を単一表に組み替える方法を実施した。

##### (1) 単一表の構成

複数の結果表を、「秘匿処理用結果表の構成（表3-1及び3-2）」における表章区分間の関係を基に、表側を地域区分・産業分類、表頭をそれ以外の表章区分とした一つの表に編成し直し単一表とした。複数の結果表に存在する表章区分は、単一表においては、一つのセルに集約されている。また、審査しやすさを考慮し、一つの区分に表示するのは、「対象数」と「秘匿対象事項（売上金額）」の2セルに限定した。



##### (2) 単一表のファイル

単一表はExcel形式のファイルとし、Excel機能のフィルタ、グループ化等を活用した。

【単一表】

地域区分	産業分類			総数		経営組織別						本支の別					
	大分類	中分類	小分類	事業所数	売上金額	個人		会社		会社以外の法人		単独事業所		本所・本社・本店		支所・支社・支店	
						事業所数	売上金額	事業所数	売上金額	事業所数	売上金額	事業所数	売上金額	事業所数	売上金額	事業所数	売上金額
地域計	T																
地域計	A																
地域計	A	01															
地域計	A	01	011														
地域計	A	01	012														
地域計	A	01	013														
地域計	A	01	014														
地域計	A	02		1,000	100,000	820	82,000	128	12,800	42	4,200	825	82,500	41	4,100	134	13,400
地域計	A	02	021	350	35,000	300	30,000	50	5,000	-	-	310	31,000	10	1,000	30	3,000
地域計	A	02	022	200	20,000	180	18,000	10	1,000	10	1,000	175	17,500	5	500	20	2,000
地域計	A	02	023	250	25,000	200	20,000	25	2,500	25	2,500	190	19,000	20	2,000	40	4,000
地域計	A	02	024	50	5,000	30	3,000	18	1,800	2	200	40	4,000	5	500	5	500
地域計	A	02	029	150	15,000	120	12,000	25	2,500	5	500	110	11,000	1	100	39	3,900
地域計	B																
地域計	B	03															
地域計	B	03	031														
地域計	B	03	032														
地域計	B	04															
地域計	B	04	041														
地域計	B	04	042														
地域計	.																
地域計	.																
地域計	.																
地域計	.																
地域1	T																
地域1	.																
地域1	.																

ア フィルタ機能

特定の地域区分や産業分類に絞って審査する場合に、フィルタ機能を用いて行を選択できるように、フィルタ用の項目を用意した。

フィルタ用の項目

地域区分	産業分類			総数	
	大分類	中分類	小分類	事業所数	売上金額
地域計	T				
地域計	A				
地域計	A	01			
地域計	A	01	011		
地域計	A	01	012		
地域計	A	01	013		
地域計	A	01	014		
地域計	A	02		1,000	100,000
地域計	A	02	021	350	35,000
地域計	A	02	022	200	20,000
地域計	A	02	023	250	25,000
地域計	A	02	024	50	5,000
地域計	A	02	029	150	15,000
地域計	B				
地域計	B	03			
地域計	B	03	031		
地域計	B	03	032		
地域計	B	04			
地域計	B	04	041		
地域計	B	04	042		

地域区分	産業分類			総数	
	大分類	中分類	小分類	事業所数	売上金額
地域計	A	02		1,000	100,000
地域計	A	02	021	350	35,000
地域計	A	02	022	200	20,000
地域計	A	02	023	250	25,000
地域計	A	02	024	50	5,000
地域計	A	02	029	150	15,000

産業分類「02」を選択

イ グループ化機能

表頭の項目について、列の位置が離れている項目間の状況を審査する場合に、列の表示・非表示を切り替えられるように、グループ化の設定をした。

「総数」欄と「グループ化②」欄との状況を審査する場合は、その間にある「グループ化①」欄を非表示に切り替えて、審査しやすい表示にした。

地域区分	産業分類			総数		経営組織別						本文の別					
	大分類	中分類	小分類	事業所数	売上金額	個人		会社		会社以外の法人		単独事業所		本所・本社・本店		支所・支社・支店	
地域計	A	02		1,000	100,000	830	83,000	128	12,800	42	4,200	825	82,500	41	4,100	134	13,400
地域計	A	02	021	350	35,000	300	30,000	50	5,000	-	-	310	31,000	10	1,000	30	3,000
地域計	A	02	022	200	20,000	180	18,000	10	1,000	10	1,000	175	17,500	5	500	20	2,000
地域計	A	02	023	250	25,000	200	20,000	25	2,500	25	2,500	190	19,000	20	2,000	40	4,000
地域計	A	02	024	50	5,000	30	3,000	18	1,800	2	200	40	4,000	5	500	5	500
地域計	A	02	029	150	15,000	120	12,000	25	2,500	5	500	110	11,000	1	100	39	3,900

地域区分	産業分類			総数		本文の別					
	大分類	中分類	小分類	事業所数	売上金額	単独事業所		本所・本社・本店		支所・支社・支店	
地域計	A	02		1,000	100,000	825	82,500	41	4,100	134	13,400
地域計	A	02	021	350	35,000	310	31,000	10	1,000	30	3,000
地域計	A	02	022	200	20,000	175	17,500	5	500	20	2,000
地域計	A	02	023	250	25,000	190	19,000	20	2,000	40	4,000
地域計	A	02	024	50	5,000	40	4,000	5	500	5	500
地域計	A	02	029	150	15,000	110	11,000	1	100	39	3,900

非表示

(3) 単一表を用いるメリット・デメリット

詳細な地域別の結果は、セル数が膨大になるため、地域別に結果表を分割することがある。そのため、地域間の秘匿状況は、複数の結果表ファイルを開いて確認する必要があり、また、複数の結果表に存在する表章区分の秘匿結果を見直す場合も、複数のファイルを確認する必要

がある。このような状況において、単一表では、一つのファイルを確認すればよいため、効率的に作業を進められるというメリットがある。

一方で、実際に単一表を利用して作業をしてみると一つのファイルに編成するため、ファイル容量が大きくなってしまふというデメリットがあった。特に、今回の審査ではExcelを用いて行ったため、動作が遅くなってしまふことがあった。単一表を利用して秘匿審査を行う際には、データベースを構築する等、システム面での工夫が必要だと考える。

## II 秘匿処理の結果

平成28年経済センサス-活動調査における産業横断的集計では、秘匿処理の結果、約21万セルが機械的に秘匿された。また、その機械処理のあとに単一表による秘匿処理の審査を行っており、それにより約800セルの見直し箇所が検出された。(表4)

表4 集計区分別秘匿セル数

集計区分	セル総数	うち機械処理による 秘匿セル数	単一表による見直しセル数		
			秘匿解除 セル数	追加秘匿 セル数	
事業所に関する集計	541,006	156,884	818	565	253
企業等に関する集計	172,237	57,791	0	0	0
合計	713,243	214,675	818	565	253

※単一表による見直しセル数は、複数の結果表に表章されている区分における秘匿を1か所にして集計しているため、元の表に復元したときのセル数は表章値より増加する。

秘匿処理の審査では、機械処理において秘匿箇所が過剰になっているセルを中心に秘匿セルの見直しを行っており、以下のような事例があった。

表5 複合サービス事業における経営組織別事業所数

	総数	個人	会社	会社以外の法人
複合サービス事業	501	25 ⑤	299	177 ⑦
郵便局	325 ⑥	25 ③	299	1 ①
協同組合	176 ④	-	-	176 ②

※協同組合の経営組織は「会社以外の法人」と定義されており、「個人」「会社」のセルには対象が存在しないため「-」と表章されている。丸囲み数字は下記で説明する秘匿箇所である。

《表5における秘匿処理の流れ》

- ① 「郵便局×会社以外の法人」のセルを1次秘匿
- ② ①に対して、「複合サービス事業＝郵便局＋協同組合」の関係より、「協同組合×会社以外の法人」のセルを2次秘匿
- ③ ①に対して、「総数＝個人＋会社＋会社以外の法人」の関係より、「郵便局×個人」のセルを2次秘匿
- ④ ②に対して、「総数＝個人＋会社＋会社以外の法人」の関係より、「協同組合×総数」のセルを2次秘匿
- ⑤ ③に対して、「複合サービス事業＝郵便局＋協同組合」の関係より、「複合サービス事業×個

人」のセルを2次秘匿

- ⑥ ④に対して、「複合サービス事業＝郵便局＋協同組合」の関係より、「郵便局×総数」のセルを2次秘匿
- ⑦ ⑤に対して、「総数＝個人＋会社＋会社以外の法人」の関係より、「複合サービス事業×会社以外の法人」のセルを2次秘匿

上記の流れでは、「協同組合」は「個人」「会社」に対象が存在せず（「-」表章）、2次元クロス表では2次秘匿ができないことから、秘匿箇所の過剰が起きてしまっている。そのため、②で「協同組合×会社以外の法人」を秘匿せず、「複合サービス事業×会社以外の法人」を秘匿することで、②による2次秘匿セルである④、④による2次秘匿セルである⑥が不要となり、結果として、秘匿セルを7セルから4セルに低減できた。（表6）

表6 複合サービス事業における経営組織別事業所数（秘匿セルの見直し後）

	総数	個人	会社	会社以外の法人
複合サービス事業	501	X ⑤	299	X ⑦
郵便局	325 ⑥	X ③	299	X ①
協同組合	176 ④	-	-	176 ②

この事例のように、2次元クロス表内に、対象が存在しないセルが存在すると、現状の機械処理では秘匿が過剰になってしまうことがあるため、人手による審査が必要となっている。

### Ⅲ 今後の秘匿処理に向けた提案

現行の経済センサスの機械処理による秘匿処理方法では、上述したように秘匿箇所の過剰が生じてしまう可能性がある。その要因として挙げられるのが、2次秘匿セルの選定条件（優先順位）である。秘匿対象として選定されたセルを含む処理において、内訳に対象数が存在しないセル（表6の「協同組合」における「会社以外の法人」が2次秘匿されていた場合の「個人」「会社」のセル）があり、2次秘匿の候補がほかに「総数」に当たるセルしか存在しない場合に秘匿箇所が過剰になる可能性が増加する。このようなセルは、2次秘匿の優先順位が低くなるように、処理を見直す必要がある。また、2次元クロス表ごとに秘匿処理を行うことで、過剰秘匿が生じてしまう可能性もある。当該表を2次元クロス表に切り出すことで、一つの表が複数の2次元クロス表に分解されるわけだが、秘匿処理の過剰秘匿チェックはこの各2次元クロス表内で行われており、表間での過剰秘匿チェックは行われていない。そのため、元の1表に復元したときに過剰秘匿が判明するのである。さらに、他表由来の秘匿箇所も追加されることで、過剰秘匿が生じる可能性が増加する。現在はこういった要因で発生した過剰秘匿を、秘匿処理の審査で対応し、判明した過剰秘匿の解除を行っているところであるが、これらの問題については今後改善の余地があり、次回の秘匿処理作業に向けて以下の2点を提案する。

#### ① 秘匿箇所の過剰について

当該秘匿セルの行及び列の2次秘匿候補セルを選定するに当たり、当該秘匿セル及び行と列の2次秘匿候補セルを頂点とした四角形の残りの頂点が「-」の場合は、秘匿の優先順位を下げる。

#### ② 過剰秘匿について

全統計表秘匿処理後に過剰秘匿チェックを実施する。

## IV おわりに

海外に目を向けてみると、米国では、統計表の秘匿処理方法に非表示法（cell suppression）及び摂動法（noise infusion）を利用している。米国の経済センサス<sup>11</sup>の一部の表には非表示法が使われており、日本と同様  $n=2, k=100$  の  $(n, k)$  ルールによる1次秘匿（primary suppressions）、そして2次秘匿（complementary suppressions）が利用されている。また、事業所数及び上位レベルの売上金額等の合計を秘匿対象としないのも日本と同様である。報告された情報を秘匿する必要がある事業所については、データセル内で数値を  $D$ （日本では  $X$ ）に置換する処理を行っている。非表示法の代わりに区間開示法が利用されることもある。端数処理法については、小さい数値に適用しても元の数値とほとんど変わらないため、利用には消極的なようである。非表示法は経済センサス以外にも、製造業の年次調査、2007年基準以前の群ビジネスパターン（CBP (County Business Pattern)）、2005年基準以前の非雇用主調査（Nonemployer Statistics (NES)）及び2007年基準以前の事業主調査（Survey of Business Owners (SBO)）で利用されている。

一方の摂動法は、非表示法に代わるものであり、より多くの情報を開示することができる手法である。個々の事業所のデータをわずかに調整（摂動）することで、データを偽装するのである。その偽装データを集計した結果の多くは、比較的少量の歪みを持っている。そして、いくつかのセルは追加秘匿のため、もしくは数値の信頼性が公表基準を満たしていないために非表示にされることがある。それらの非表示にされた数値の一部は、合計からの引き算等により算出される可能性があるが、その推計値は公式な数値ではなく、また真値から大きく乖離している可能性があるため2次秘匿処理は行っていない。摂動法は、非表示法の後継として、先に記した調査（製造業の年次調査を除く。）に現在利用されている。米国センサス局では絶えず、秘匿に関する新しい手法や技術、方法論を追求している。

カナダでは、企業調査の結果表に非表示法を利用しているが、秘匿箇所数が多くならないようセンシティブセルに含まれる特定の企業に対して、秘匿の権利放棄に関する法的契約<sup>12</sup>を結ぶことでデータの開示をしている。秘匿の権利放棄を要請する企業は、カナダ統計局が開発した **G-Confid** というセンシティブセルを特定する一般化されたシステムにより、すでに特定されており、プログラム分析官がまだ法的契約を交わしていない企業に対して、カナダ統計局から秘匿の権利放棄を要請すべきかどうかの確認を行っている。**G-Confid** は秘匿の権利放棄を要請する企業を客観的な数値で公平に選定することを可能にしている。

欧州連合統計局（Eurostat）では、統計表の秘匿処理方法に関して、次の考えを示している。統計表の秘匿処理方法において、よく利用される方法として、非表示法、区分合併法、端数処理法、制御された統計表の調整（CTA）及び摂動法があるが、これらの手法の多くは、数値の欠損がない状態の統計表を公表したいという考えに基づき、数値にノイズを加えたものである。非表示法は最も有名な手法だが、数値の欠損が生じてしまうため、安全ではあるがデータの有用性を損なってしまう。統計表に多数の秘匿セルが含まれてしまった場合の秘匿処理改善策としては、まず統計表を再設計し、端数処理法を利用することである。しかし、この方法は一部の大企業にとっては数値の秘匿が不十分になる可能性があるため、その場合は非表示法が唯一の選択肢になる。また、特定の大企業の数値を秘匿することは、情報の有用性を損なう可能性があるため、秘匿の権利放棄を当該大企業に要請することが効果的な解決策となる。欧州では秘匿処理のツールとして **tau Argus** と **sdcTable** が開発されており、多くの欧州連合加盟国で使用されている。

<sup>11</sup> 米国の経済センサスは日本の経済センサスのような全数調査ではなく、全数調査とサンプル調査の複合調査となっており、部門によって全数とサンプルの使い分けをしている。また、2018年に実施された2017年経済センサスでは完全なオンライン調査を実施している。調査されなかった企業については、行政記録情報を利用することで補完している。

<sup>12</sup> カナダでは、秘匿の権利放棄を要請する企業と交わす法的契約を「Waiver」といい、原則3年ごとに更新される。



欧州連合統計局は、欧州全体の統計表を公表しているため、その秘匿作業も行っている。欧州全体の統計表の秘匿処理方法は次のようなものである。欧州全体の統計表は28の国のデータから作成されており、例えば、そのうちの2国（A国、B国）の数値がある集計事項において秘匿であったとする。A国の秘匿処理は閾値が3であり、対象企業数が2、同様にB国の閾値は5であり、対象企業数は4であった。欧州全体の統計表の閾値は5であり、この2国の合計秘匿対象企業数は6であることから該当セルにおける閾値を超えるため、欧州全体の数値としては公表が可能となる。

日本の経済センサスの秘匿処理方法は非表示法及び区分合併法である。一方の海外では、上記のとおり、より多くの情報を公開したいという考えから非表示法よりも摂動法が優勢の状況である。近年では、株式会社の有価証券報告書をEDINET<sup>13</sup>などで一般的に閲覧することができ、企業の売上金額を確認することができるため、経済センサスの結果表上で秘匿をしたとしても、他資料を参照することで情報を得ることができるケースが存在している。経済センサスでは、売上金額の報告が原則として暦年ベースであるのに対し、有価証券報告書では、年度ベースの売上金額が記載されているという違いはあるものの両者の売上金額に大きな差異は生じないと考えられる。そのため、ある秘匿セルに売上金額を公開している企業が含まれていることがわかってしまった場合、そのセルに含まれる残りの企業の売上金額が一定の範囲内で推測可能となる。

情報化社会である現代は、様々な角度から情報を取得することが可能な社会になっており、時代の趨勢は、公開情報までも秘匿してしまう現行の秘匿処理方法とは逆方向に進んでいる。このような現状の打開策として、企業が公開している情報に関しては、統計表における秘匿の権利放棄を企業へ要請することが、今後の重要な秘匿処理方法の一つになってくるのではないだろうか。秘匿処理は企業の情報を守るという観点から必要な処理ではあるが、過度な適用は社会利益の損失につながりかねない。調査で得た重要な情報を統計表上で非表示にするのではなく、有効に社会に還元できるようにすることが秘匿処理の理想である。

## 付録

### 【用語解説】

主に本文中で使用した秘匿処理に関連する用語について解説する。

#### ① セル

本稿で用いているセルとは、結果表を構成する個々の表章値の升目のことである。イメージとしては、表計算ソフトのワークシートを構成する升目がこれに該当する。

#### ② クロス表

表頭と表側の項目のクロスで集計事項が表章された表のことである。一般的な統計表のことである。

#### ③ 分類事項及び集計事項

分類事項は、結果表のあるセルに分布されるか否かを決定する事項であり、集計事項はそのセルの値となる事項をいう。例えば、経営組織を例にすると、個人経営、法人、会社以外の法人などが分類事項である。これに対して、集計事項は、事業所数、従業者数、売上金額などである。分類事項と集計事項の組み合わせが、統計表の各表章値（各セル）のメタ情報を構成する要素になる。

<sup>13</sup> EDINET (Electronic Disclosure for Investors' NETwork)とは、日本の「金融商品取引法に基づく有価証券報告書等の開示書類に関する電子開示システム」のことである。

④ 1次秘匿

事業所数（企業等数）が1又は2であるセルの売上金額等の数値は、調査客体が特定されてしまう可能性があるため、情報保護の観点から秘匿の対象となり、これを1次秘匿という。

⑤ 2次秘匿

1次秘匿により、数値が非表示になっているセルは、その区分の合計と当該セル以外の数値との差し引き又は下位分類を合計することなどにより、非表示した数値が算出できてしまう場合がある。そのため、1次秘匿セル以外にも同行、同列の位置に追加で秘匿セルを設け算出できなくする措置を講じなければならない。このように1次秘匿セルの数値を算出できなくする処理を2次秘匿という。なお、2次秘匿セルの数値を算出できなくするための3次秘匿などこれに続く秘匿も併せて2次秘匿としている。

⑥ センシティブセル（sensitive cell）

当該セルに該当する事業所の売上金額等の数値が不当に正確に推計できてしまうセルのことであり、秘匿処理を必要とするセルのこと。センシティブセルは、閾値ルール、(n, k)ルール及び(p, q)ルールにより決定される。（これらのルールの詳細は後述する【秘匿処理方法の種類】①非表示法を参照）

⑦ 多次元表

統計表における各項目を一つの次元として考える。例えば、都道府県別、経営組織、従業者規模の3項目で構成される結果表は三次元表となる。

⑧ 過剰秘匿

事業所数が1又は2の場合は、1次秘匿として明白に秘匿の必要性が判明するが、2次秘匿の場合は秘匿の優先順位をもとに秘匿セルを決定している。しかし、後続して行った秘匿処理で他表由来の2次秘匿セルが当該表に発生した場合、その新たに発生した2次秘匿セルにより2次秘匿の必要がなくなるセルが生じる。これが過剰秘匿であり、後続して行った秘匿処理により、秘匿の必要性がなくなったセルの秘匿をいう。

⑨ 秘匿箇所の過剰

⑧の「過剰秘匿」とは意味合いが異なる。それぞれの秘匿セルの秘匿の必要性は存在するが、秘匿箇所を少なくできるパターンが別に存在するときは「秘匿箇所の過剰」という。

⑩ 過少秘匿

過剰秘匿同様、主に2次秘匿を行う際に生じる。他表由来の秘匿セルなどに起因する一目でわかりづらい秘匿処理が存在することによる当該表の2次秘匿が必要なセルの秘匿漏れのことである。

⑪ 表内秘匿

当該表のみで完結する秘匿処理のことである。

⑫ 表間秘匿

当該表とそれ以外の表の間に発生する秘匿処理のことである。

⑬ 開示リスク

秘匿したセルの数値が判明してしまうリスクのことである。

【秘匿処理方法の種類】

秘匿処理方法にはいくつかの方法があり、代表的なものを以下に紹介する。

① 非表示法

個別情報が判明してしまうセルの数値を「X」などに置き換えて非表示とすることで秘匿する手法。セルの1次秘匿基準としては、閾値ルール(threshold rule)、(n, k)ルール及び(p, q)ルールがある。

閾値ルールは、任意のセルに含まれる事業所数が  $n$  未満になると、当該セルの調査客体の情報が一般に推測できてしまうリスクがある場合、 $n$  を閾値として当該セルを秘匿するというものである。日本では、事業所数が1又は2であるセルを秘匿しているため、 $n=3$  の閾値ルールということになる。

( $n, k$ ) ルールは、任意のセルにおいて、そのセルに含まれる事業所の売上金額1位から  $n$  位までの合計がそのセルの数値（セルに含まれる全事業所の売上金額合計）の  $k\%$  以上である場合、そのセルの数値の公開にはリスクが存在し、秘匿対象になるというルールである。例えば、 $n=3, k=70$  の ( $n, k$ ) ルールを基準にした結果表（表 a）があったとする。表 a をみると産業分類 B が秘匿になっている。この秘匿セルの売上金額は 4000（百万円）であったが、当該セルに含まれる全事業所数 17 のうち上位 3 社の売上金額合計が 3200（百万円）であった。これは当該セルの売上金額の 80% に達しており、閾値の 70% 以上となっているため秘匿対象となったのである。日本の経済センサスの場合、事業所数が 2 までのセルを 1 次秘匿対象としているため  $n=2, k=100$  の ( $n, k$ ) ルールで秘匿を行っているということになる。

表 a 産業分類別売上金額

	事業所数	売上金額 (百万円)
産業分類A	40	5,000
産業分類B	17	X
産業分類C	5	3,000

( $p, q$ ) ルールは、事前事後ルールとも呼ばれ、結果表公表前の数値を使って、これから公表する統計表の当該セルに属する事業所の売上金額をどの程度まで推計できるかを秘匿基準としたものである。例えば、任意のセルにおいて、そのセルに含まれるある事業所の売上金額を公的に利用可能な情報を利用して結果表公表前に  $q\%$  以内の誤差で見積もることができたとする。そして、仮に結果表を公表してしまった場合に、セルに含まれる事業所の売上金額の誤差が  $p$  ( $p < q$ ) % 以内で推計可能なとき、このセルは秘匿対象となるのである。(  $p, q$  ) ルールでは、 $p/q$  比を information gain（情報利得）としており、各国の統計局でその許容値を定めている。(  $p, q$  ) ルールを採用している国では、この  $p$  と  $q$  のパラメータを機密情報としている国もある。

② 区分合併法

事業所数が 1 又は 2 の項目区分があるときに、その近接している項目区分を合併させることで区分数を少なくし、該当する事業所数を増やすことで秘匿対象箇所を減らす方法である。例えば、表 b のような場合、常用雇用者規模を 10 区分から 6 区分にすることで、秘匿セルが生じないようにしている。しかし、この手法では、表章区分が少なくなることから、表示できるセル数が少なくなってしまうという問題がある。特に、秘匿対象箇所が近接していない場合、区分合併する項目箇所が一層増えてしまう。また、意味があってその項目区分にしている場合に適用するのは困難である。

表 b 常用雇用者規模別売上金額

常用雇用者規模	事業所数	売上金額 (百万円)
0人	16	858
1~4人	37	2,975
5~9人	47	10,073
10~19人	79	16,890
20~29人	80	18,970
30~49人	26	7,709
50~99人	36	10,799
100~199人	1	X
200~299人	5	13,790
300人以上	2	X

常用雇用者規模	事業所数	売上金額 (百万円)
0人	16	858
1~4人	37	2,975
5~9人	47	10,073
10~19人	79	16,890
20~29人	80	13,508
30人以上	70	43,298

③ 端数処理法（丸め法：rounding）

秘匿したことにより生じる情報の損失を減少させるために、センシティブセルの数値の任意桁数以下を端数処理する手法であるが、この手法では総数とその内訳の合計が一致しないという問題が生じる。この問題を解決するため、制御端数処理法（controlled rounding）と呼ばれる線形プログラミングを活用した総数と内訳の合計を一致させる手法が開発された。また、ランダムラウンディング法（random rounding）は、端数処理をするときに切り捨てるか切り上げるかをランダムに決定することで、偏りが生じないようにする端数処理法である。しかし、端数処理法は、事業所の売上金額を秘匿するような場合、ある範囲で売上金額を推測することができてしまうため好ましくない。

④ 摂動法（perturbation-based method）

秘匿したい数値に乱数などの摂動要素を含ませることで真値を歪ませ秘匿する手法をいう。

⑤ 区間開示法

秘匿したい数値について、区間で表示することにより真値を秘匿する手法である。例えば、秘匿したい数値が30であった場合、区間範囲を20で設定すると、25-45のように表示されることになる。これは25から45の間に真値が存在するという意味である。この手法は、秘匿対象セルの数値が他の秘匿対象セルの数値と比較して著しく小さい（もしくは大きい）場合、区間範囲の設定が結果表全体の精度を低くしてしまう可能性がある。

⑥ 制御された統計表の調整（CTA: Controlled Tabular Adjustment）

セルの数値の選択的調整に基づいて統計表のデータを保護する手法である。センシティブセルの数値は、最も近い安全な値に置換される。また、総数との加法整合性を保つために、他のセルの数値も調整される。

⑦ 秘匿の権利放棄（waiver）

秘匿対象セルに含まれる調査客体の了承を得て、秘匿処理を行わない手法である。特に当該セルに大きな影響を与える企業（1次秘匿の対象となっている企業など）と秘匿の権利放棄に関する法的契約を結ぶことができると秘匿箇所（情報の損失）が著しく減少する。

【参考資料】

- [1] 松本正博（1996）「サービス業基本調査の秘匿処理方法について」、『統計研究彙報』、第54号、1996年5月、pp. 57-86
- [2] 瀧敦弘（2003）「集計表におけるセル秘匿問題とその研究動向」、『統計数理』、第51巻、第2号、2003年、pp. 337-350
- [3] Eurostat (2010) “Handbook on Statistical Disclosure Control Version 1.2”  
[https://ec.europa.eu/eurostat/cros/content/handbook-sdc\\_en](https://ec.europa.eu/eurostat/cros/content/handbook-sdc_en)
- [4] Eurostat (2018) “European business statistics manual”  
[https://ec.europa.eu/eurostat/statistics-explained/index.php?title=European\\_business\\_statistics\\_manual](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=European_business_statistics_manual)
- [5] Peter Wright (2017) “Disclosure control that accounts for survey realities: assessing the risk using G-Confid,” Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Skopje, Macedonia (FYROM), 20-22 September 2017  
[https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2017/7\\_gconfid.pdf](https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2017/7_gconfid.pdf)
- [6] United States Census Bureau (2018) “Economic Census Methodology”  
<https://www.census.gov/programs-surveys/economic-census/technical-documentation/methodology.html>