

汎用秘匿処理ソフトウェア -ARGUS による集計表の秘匿処理

順次 LP 法と -ARGUS 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

*N S T A C*

---

*Working Paper No.4*

平成 18 年 3 月

独立行政法人 統計センター

本製表技術参考資料は、独立行政法人 統計センター職員の製表技術に関する研究の結果を紹介するものである。

したがって、本資料の内容は当機関としての見解を示すものではないので、利用に際しては注意されたい。

## 目 次

汎用秘匿処理ソフトウェア -ARGUS による集計表の秘匿処理 .....	1
村田 磨理子, 畠山 昌子, 磯部 祥子	
要 旨 .....	1
1. 背景と目的.....	2
2. 集計表の秘匿処理.....	2
3. 実験の対象データ・集計表.....	7
4. 実験結果 .....	8
5. 適切な秘匿処理のためのアプローチ .....	25
6. まとめ.....	25
参考文献 .....	26
付表.....	27
付録 結果表 .....	31
順次 LP 法と -ARGUS 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 .....	48
岡本 政人	
要 旨 .....	48
1 はじめに .....	50
2 集計表の秘匿処理の概要 .....	51
3 実験結果.....	57
4 今後の課題及び最近の集計表開示抑制法の研究動向 .....	62
5 おわりに .....	71
参考文献 .....	72
企業調査における秘匿の免責：公開可能な情報を増やすための系統的アプローチ .....	74
1 序 論 .....	74
2 機密保持の概念.....	74
3 CONFID の方法論.....	76
4 免責の利用.....	77
5 新しいアプローチ：スコア関数.....	77
6 実 例 .....	81
7 結 論 .....	82
追 記 .....	82
参考文献 .....	83



## 汎用秘匿処理ソフトウェア $\tau$ -ARGUS による集計表の秘匿処理

村田 磨理子, 畠山 昌子, 磯部 祥子

### 要 旨

統計ニーズの多様化に対応したデータ提供において、個々の客体が特定される危険性がないかチェックし、必要な場合は適切な秘匿処理を効率的かつ迅速に行う必要がある。本稿は、汎用秘匿処理ソフトウェアの業務適用可能性及び適切な秘匿処理法について検討を行った結果をまとめたものである。

検討では、平成 11 年サービス業基本調査の集計表を対象とし、EU の秘匿処理に関するプロジェクトが開発した汎用秘匿処理ソフトウェア  $\tau$ -ARGUS を用いて、一次秘匿の基準、二次秘匿の方法、複数集計表の一括処理と逐次処理、情報ロスのコスト変数の指定、singleton 秘匿処理について実験を試みた。

日本の統計で一般的な一次秘匿基準である度数 3 の最小度数条件のみの適用は、少数の客体による占有度に基づく  $n-k$  占有ルールや  $p$  %ルールを組み合わせた基準と比較して緩やかな基準である。つまり、客体数に基づく基準のみの適用は、少数の客体による占有度が高いセルを見逃すことがあるという点で、開示リスクがより高いことを示した。開示リスクの適切なレベルの決定や、対応する一次秘匿基準の決定が今後の課題である。

さらに、業務への適用を検討するため、二次秘匿処理の比較や、複数集計表の一括処理、逐次処理などの検証を行った。しかし、 $\tau$ -ARGUS は二次秘匿処理の機能の制約により、単独の 2 次元表の場合を除くと、適切な秘匿処理結果を得ることができなかった。総合的に見て、 $\tau$ -ARGUS は処理機能の制約のために、現行のサービス業基本調査の秘匿処理への実際の適用は難しいと考えられる。そこで、調査や利用者が必要とする条件に従って適切な秘匿結果を得るためのいくつかのアプローチを紹介した。その一つに回答者の了解を得て秘匿を行わない免責アプローチがある。今後は、二次秘匿処理機能の改良のほか、集計事項の分類区分の再編や 2 次元表に限定したクロスの簡略化、免責アプローチの適用も考慮に入れた検討が必要であると考えられる。

## 汎用秘匿処理ソフトウェアτ-ARGUSによる集計表の秘匿処理

村田 磨理子, 畠山 昌子, 磯部 祥子

### 1. 背景と目的

研究センターでは、統計ニーズの多様化に対応した製表方法の研究が主要な研究課題の一つとなっている。統計ニーズの多様化に対応する方策としては、①統計表の充実、②個別ニーズに対応するオーダーメイド集計の実現、③マイクロデータあるいはマイクロデータに相当するデータの提供が考えられるが、いずれの方策に関しても、提供するデータから個々の客体が特定される危険性がないかチェックし、必要な場合は適切な秘匿処理を効率的かつ迅速に行う必要がある。しかし、我が国ではこれまで適切な秘匿処理の方法に関して十分な研究が行われてこなかった。

欧米諸国では集計表の秘匿処理法の研究が進み、ほぼ決定版といえる方法が確立されているが、複雑かつ高度であるため、統計センターにおいてこの方法を実装した秘匿処理システムを内部開発することは極めて困難であった。しかし、EUの秘匿処理に関するプロジェクトSDC及びCASCが汎用秘匿処理ソフトウェアτ-ARGUSの開発・提供を行い、意欲的に機能拡充を図ってきたことから、τ-ARGUSを利用することで高度な秘匿処理法を採用できる可能性がでてきた。そこで、τ-ARGUSの業務適用可能性、及び適切な秘匿処理法について研究を行うこととした。

### 2. 集計表の秘匿処理

#### (1) 開示抑制について

集計表データの公表において、セルの値から個々の客体の機密情報が許容できないほど正確に推定できるとき開示 (disclosure) の問題が発生する。集計表の開示抑制法 (disclosure control methods) として、瀧(2003)は、1) 表の再設計 (table redesign)、2) セルの非表示 (cell suppression)、3) 区間開示 (feasibility intervals)、4) 丸め法 (rounding) の4つの方法を挙げている。表の再設計<sup>1</sup>は、表側や表頭をより大きな分類区分に統合したり、新たな分類区分を用いたりして、開示の問題が起きないように表を作りかえる方法である。詳細な分類が失われ、表の記述的、分析的価値を減じる傾向がある。セルの非表示は、セルの値を表中に示さず、“X”などの記号に置き換える方法である。商業統計、工業統計、サービス業基本統計など経済調査で広く利用されている。いかに情報の損失を少なくするかが重要な点である。区間表示は、非表示ではなく、区間で表示する方法である。非表示よりも情報の損失が少ないが、必ずしも十分広い区間を確保できない場合がある。丸め法は、セルの値をあらかじめ決められた値に置き換える方法である。合計が内訳の和と必ずしも一致しないなどの問題点がある。開示抑制に関する用語は、UN/ECE and Eurostat

<sup>1</sup> 本稿では、グローバル・リコーディング (global recoding) の用語を用いる。

(2003)にまとめられている。

本稿では、「セルの非表示」による開示抑制を秘匿処理と呼ぶ。次に、秘匿処理についてさらに説明を追加する。

そのセルの公開により客体の特定が可能となることが予想されるとき、そのセルはセンシティブ (sensitive) であるとみなす。センシティブと判定する場合は、主に2つある。1つは、セル内の客体数 (度数) が少なすぎる場合で、もう1つは、少数の客体の占有度が高い場合である。前者は、例えば、度数が2以下のセルをセンシティブセルとするなどの場合で、日本では一般的な方法である。実際、度数が1のセルでは客体の値が正確に特定でき、度数が2のセルでは、一方の客体は他方の客体の値を正確に特定できる。以下では、この基準を度数3の最小度数 (minimum frequency) 条件と呼ぶ。閾値ルール (threshold rule) とも呼ばれる。

少数の客体の占有度による判断基準としては、 $n-k$  占有ルール (dominance rule) と  $(p,q)$  事前事後ルール (prior-posterior rule) が一般的に用いられている。

$n-k$  占有ルールは、同一セルに含まれる客体のうち最も大きい  $n$  件の客体が、セルの値の  $k\%$  を上回っている場合に、そのセルはセンシティブセルとみなすルールである。例えば、 $n=3$ 、 $k=75$  の場合は、値の大きい3個の客体の合計値が、セルの値の  $75\%$  を上回っているときに、センシティブセルとみなされる。ここで、セル  $X$  に  $N(X)$  客体があり、それぞれの客体の値  $x_1, x_2, \dots, x_{N(X)}$  は降順になっているとする。このとき、 $n-k$  占有ルールによるセルのセンシティブティ尺度 (sensitivity measure) は次の式で定義され、 $S_n(X) > k\%$  であれば、センシティブとする。

$$S_n(X) = \sum_{i=1}^n x_i / \sum_{i=1}^{N(X)} x_i$$

なお、度数3の最小度数条件は、 $n=2$ 、 $k=100$  の占有ルールに相当する。

$(p,q)$  事前事後ルールは、一般に入手可能な情報から個々の客体の値を  $q$  パーセント以内の誤差で推定できると想定した場合に、統計発表後の推定精度が  $p$  パーセント以内に入るか否かでセンシティブセルを判定するルールである。 $p\%$  ルールは、 $(p,q)$  事前事後ルールの特別な場合で、 $q=100$  としたものである。 $(p,q)$  事前事後ルールによるセルのセンシティブティ尺度は次の式で定義され、 $S_q(X) > -p$  であれば、センシティブとする。

$$S_q(X) = -(q/x_1) \sum_{i=3}^{N(X)} x_i$$

集計表の秘匿処理においては、センシティブセルを一次秘匿、すなわち、セルの数値を表中に示さず、秘匿したことを示す“X”などの記号に置き換える。

一次秘匿だけでは、合計のセルの値から他のセルの値を差し引きして、一次秘匿セルの値を求められる場合がある。一次秘匿セルに対して望ましい機密保護を実現するために、センシティブではないセルを追加的に秘匿する必要がある。これを二次秘匿という。追加的に秘匿するセルの選択は、公表する集計表に含まれる情報の量をできる限り多くしつつ、望ましい機密保護のレベルを確保して行わなければならない。情報量の尺度は、秘匿されるセル数、そのセルに含まれる客体数、そのセルの値などが一般に使われている。

二次秘匿は、望ましい機密保護のレベルを確保し、情報の損失（以下、情報ロスと呼ぶ）を最小化する秘匿パターンを求める問題であり、整数計画法の問題として定式化される。このような整数計画法の問題は、集計表が複雑になると最適解を求めることが非常に難しくなる。そのため、実用性を高めたアルゴリズムや近似的な解法が考案され、秘匿処理のソフトウェアの開発が行われてきた。

## (2) $\tau$ -ARGUS の概要

$\tau$ -ARGUS(Anti-Re-identification General Utility System の略)は、EU のプロジェクトが開発した集計表の秘匿処理用ソフトである。1996年2月から1998年7月に実施されたSDC(Statistical Disclosure Control : 統計的開示抑制)プロジェクトと、それに続くCASC(Computational Aspects of Statistical Confidentiality)プロジェクトがある。CASCプロジェクトの詳細は、<http://neon.vb.cbs.nl/casc/> から参照できる。

$\tau$ -ARGUS は、集計表の開示抑制法として、グローバル・リコーディングと秘匿処理が可能である。 $\tau$ -ARGUS の機能の概要を図 1 に示す。詳細は、CASC-project(2004)を参照されたい。

$\tau$ -ARGUS で採用されている一次秘匿の基準、二次秘匿の方法、情報ロスは、以下のとおりである。なお、本研究ではバージョン 3.0.2 を使用し、OR ソルバーは Cplex を用いた。

### (ア) $\tau$ -ARGUS が採用している一次秘匿の基準

- ①  $n-k$  占有ルール
- ②  $p$  %ルール
- ③最小度数

### (イ) $\tau$ -ARGUS が採用している二次秘匿の方法

#### ①Hypercube/GHMITER 法

階層的に処理を行い、6次元表まで処理可能。市販の OR ソルバーは不要。複数集計表をリンクした秘匿処理が可能。Singleton<sup>2</sup> (該当する客体が1つのセル) 秘匿処

<sup>2</sup> 数学用語では、単集合の訳語が一般的だが、本稿では Singleton のまま用いる。

理が可能。

②Optimal 法 (Fischetti-Salazar 法)

Fischetti と Salazar が提案したアルゴリズムを用いて最適解を求める方法。3次元表まで処理可能。市販の OR ソルバーである Cplex か Xpress が必要。

③HiTaS 法 (Modular 法)

階層構造の各レベルの単純な表に Optimal 法を順次適用していく方法。3次元表まで処理可能。市販の OR ソルバーである Cplex か Xpress が必要。

④Network Flows 法

システム最適化問題の一つであるネットワークフロー問題として定式化し、ヒューリスティックに解を探す方法。1次元の階層的な構造を持つ2次元表にのみ適用できる。市販の OR ソルバーは必須ではない。

(ウ)  $\tau$ -ARGUS が採用しているセル秘匿の情報ロス

- ①非表示とするセル数
- ②非表示とするセルに該当する度数
- ③非表示とするセルの値 (集計項目以外の値を指定することもできる)
- ④ユーザーが設定した任意の値

ここで、二次秘匿の各方法を実用性の点で簡単に説明する。Optimal 法は、理論上は最適解を与えるが、計算時間が長く、大規模な表に対しては実質的に適用できない。Hypercube 法は大規模な表でも解が得られるが、情報ロスが多く、最適解への近似の程度が良くない傾向がある。HiTaS 法は Optimal 法より計算時間が短く、Hypercube 法より情報ロスが少ないが、大規模な表に対しては計算が不安定になる傾向がある。

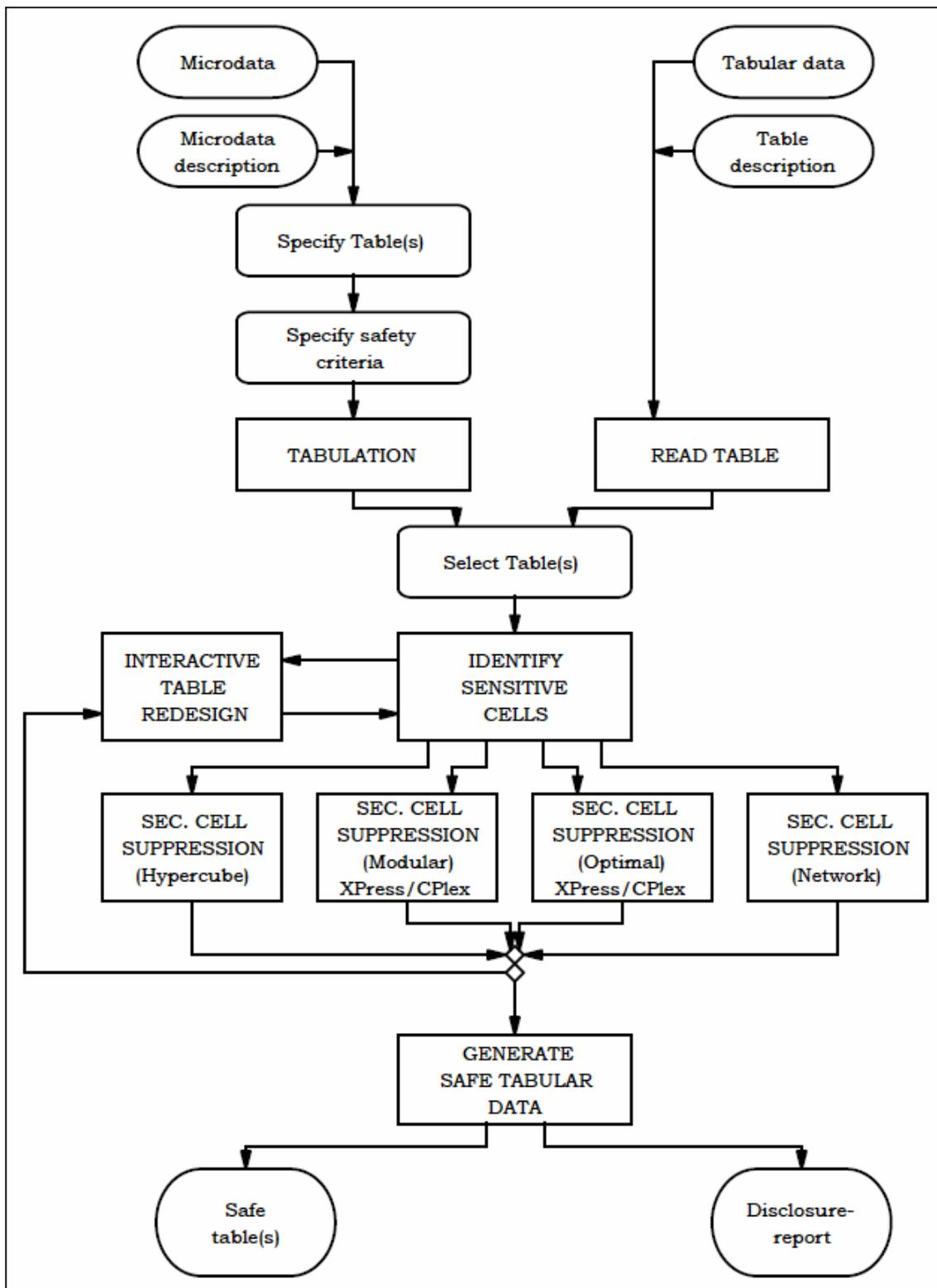


図 1 τ-ARGUS の機能

3. 実験の対象データ・集計表

秘匿処理の重要性が最も高いとみられる統計調査データを研究対象とするのが適当であるため、平成 11 年サービス業基本調査を対象とする。ただし、平成 11 年特定サービス産業実態調査からの取り込み事業所に係るデータは除くほか、取り込み事業所が多い産業区分(791 各種物品賃貸業、792 産業用機械器具賃貸業、793 事務用機械器具賃貸業、821 ソフトウェア業、822 情報処理・提供サービス業、83 広告業、849 その他の専門サービス業、76D ゴルフ練習場、76F テニス場、869 他に分類されない事業サービス業)については対象から除外する。なお、今回の集計・分析に当たって集計に用いる区分として従業者規模が必要となるが、サービス業基本調査では調査をしていないため、平成 11 年事業所・企業統計調査の従業者数を事業所単位でリンクしてデータを作成する。

作成したデータから、表 1 に掲げた A 表から H 表までの統計表を r-ARGUS を用いて集計し、秘匿処理を行う。集計する統計表は計 12 表であるが、平成 11 年サービス業基本調査の統計表の秘匿すべき箇所は、これら 12 表の秘匿処理結果から指定することができる。本研究では、一次秘匿の基準、二次秘匿の方法、複数集計表の一括処理と逐次処理、情報ロスのコスト変数の指定、Singleton 秘匿処理について実験を行う。

なお、サービス業基本調査の公式統計書の集計表に対する秘匿処理は、事業所数そのものは秘匿対象とせず、事業所数の少ないセルの従業者数、収入金額等の値を秘匿対象としている。一次秘匿セルは度数(ここでは、事業所数に相当)3 の最小度数条件により選定し、二次秘匿セルは人手や内部開発したシステムで決定している。松本(1996)は、サービス業基本調査のように複雑な結果表で実行可能な方法として、単一表を計・内訳が階層化していない単純な 2 次元表に分割して処理を行うシステムを紹介している。

表 1 分析の対象とする集計表

統計表		産業区分	属性区分 1	属性区分 2	セル数
<b>【全国編】</b>					
A 表	産業(小分類)、本所・支所(3区分)、経営組織(7区分)別事業所数・収入金額	146	4	11	6424
B 表	産業(小分類)、開設時期(19 区分)別事業所数・収入金額	146	22		3212
C 表	産業(小分類)、従業者規模(6区分)、経営組織(7区分)別事業所数・収入金額	146	9	11	14454
D 表	産業(小分類)、本所・支所(3区分)、資本金階級(11 区分)別事業所数・収入金額 [「外国の会社」を除く会社について]	146	4	17	9928
統計表		地域区分	産業区分	属性区分	セル数
<b>【地域編】</b>					
E-1 表	産業(中間分類)、経営組織(7 区分)別事業所数・収入金額 一都道府県一	48	33	11	17424
E-2 表	産業(中間分類)、経営組織(7 区分)別事業所数・収入金額 一13 大都市一	14	33	11	5082

F-1 表	産業(中間分類)、資本金階級(6区分)別事業所数・収入金額 [「外国の会社」を除く会社について] - 都道府県-	48	33	8	12672
F-2 表	産業(中間分類)、資本金階級(6区分)別事業所数・収入金額 [「外国の会社」を除く会社について] - 13 大都市-	14	33	8	3696
G 表	産業(小分類)別事業所数・収入金額 - 13 大都市・県庁所在市・その他の市町村-	144	146		21024
H-1 表	産業(中間分類)、経営組織(3区分)別事業所数・収入金額 - 13 大都市・ 県庁所在市・人口 30 万以上市・その他の市町村-	175	33	4	23100
H-2 表	産業(中間分類)、経営組織(3区分)別事業所数・収入金額 - 都道府県、 市部・郡部-	48×3	33	4	19008
H-3 表	産業(中間分類)、経営組織(3区分)別事業所数・収入金額 - 県内地域-	259	33	4	34188

#### 4. 実験結果

##### (1) 一次秘匿基準の比較

A 表の各セル (6,424 セル) についてセンシティブリティ尺度を計算した結果を以下で説明する。簡単のため、集計ウェイトは考慮しない。

$n-k$  占有ルール ( $n=3$ ) によるセンシティブリティ尺度のヒストグラムを図 2 に示す。破線は  $k=75$  のしきい値に相当する。あるセルのセンシティブリティ尺度が 75% (0.75) より大きい値をとる場合 (破線より右側の場合)、そのセルは  $n-k$  占有ルール ( $n=3$ 、 $k=75$ ) の基準でセンシティブであると判断される。

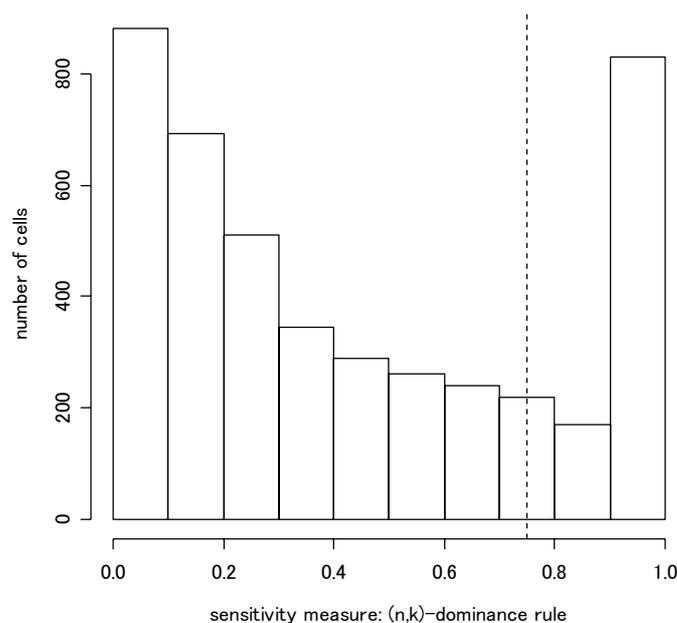


図 2  $n-k$  占有ルール ( $n=3$ ) によるセンシティブリティ尺度

次に、 $p$  %ルールによるセンシティブリティ尺度のヒストグラムを図 3 に示す。値の範囲が広いので、符号反転後に対数変換した値<sup>3</sup>のヒストグラムを図 4 に示す。図 4 の破線は  $p = 25$  のしきい値に相当する。あるセルのセンシティブリティ尺度が  $-25$  より大きい値をとる場合 (符号反転後の図 4 では、破線より左側の場合)、そのセルは  $p$  %ルール ( $p = 25$ ) による基準でセンシティブであると判断される。

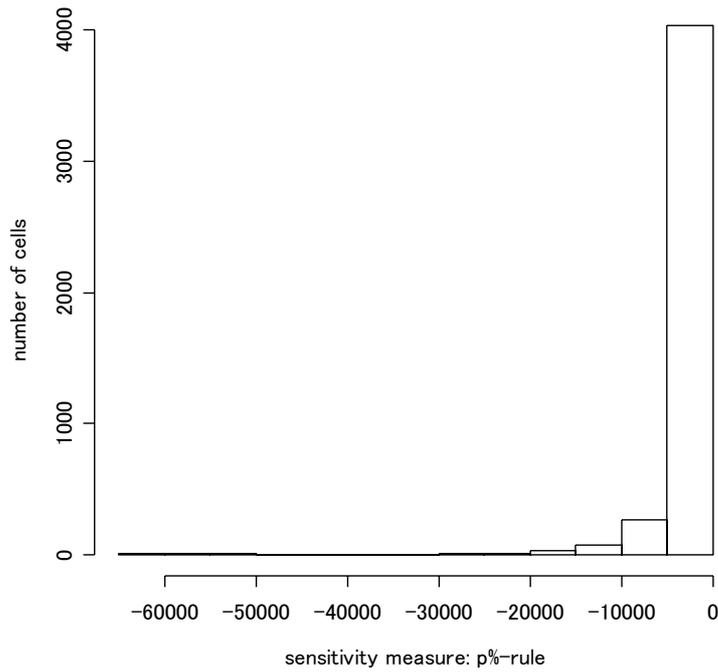


図 3  $p$  %ルールによるセンシティブリティ尺度

<sup>3</sup> 元のセンシティブリティ尺度が 0 の場合は、便宜上-0.01 に置き換えた。

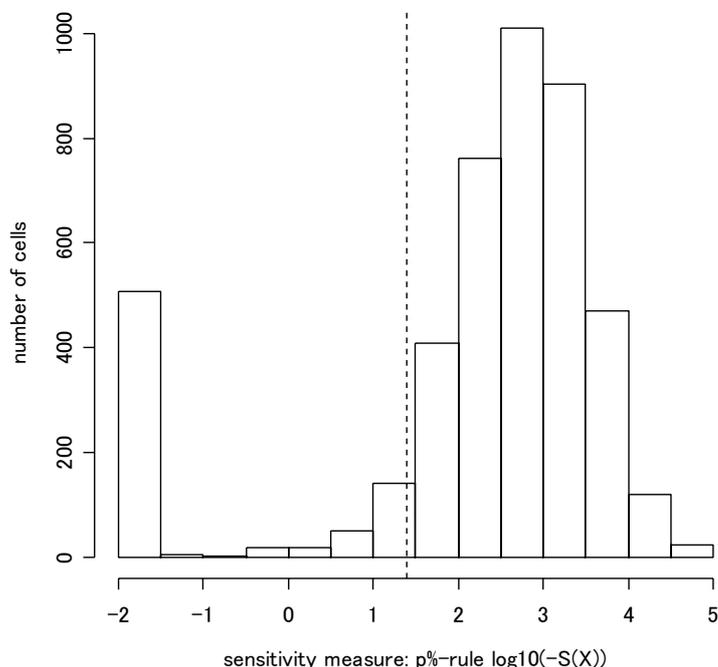


図 4  $p$  %ルールによるセンシティブティ尺度 (符号反転後に対数変換)

図 5 は  $n-k$  占有ルール ( $n=3$ ) によるセンシティブティ尺度と  $p$  %ルールによるセンシティブティ尺度の散布図である。横軸は  $n-k$  占有ルール ( $n=3$ ) によるセンシティブティ尺度を示す。縦の実線は、 $n-k$  占有ルール ( $n=3$ ) の  $k=75$  のしきい値に相当し、実線より右側がセンシティブとなる。縦軸は  $p$  %ルールによるセンシティブティ尺度を示し、対数軸表示のために符号を反転している。横の実線は、 $p$  %ルールの  $p=25$  のしきい値に相当し、実線の下側がセンシティブとなる。この2つのセンシティブティ尺度を比較すると、 $n-k$  占有ルール ( $n=3$ ) によるセンシティブティ尺度の値が小さいときは  $p$  %ルールによるセンシティブティ尺度の値も小さく、逆に、 $n-k$  占有ルール ( $n=3$ ) によるセンシティブティ尺度の値が大きいたときは  $p$  %ルールによるセンシティブティ尺度の値も大きくなる傾向がある。また、 $n-k$  占有ルール ( $n=3$ ) の  $k=75$  のしきい値と  $p$  %ルールの  $p=25$  のしきい値による基準を比較すると、 $p$  %ルールによりセンシティブである場合 (縦の実線より下側の点) は、 $n-k$  占有ルール ( $n=3$ ) によりセンシティブである場合 (横の実線より右側の点) に含まれるため、 $n-k$  占有ルール ( $n=3$ ) の  $k=75$  のしきい値による基準のほうが厳しい<sup>4</sup>ことが分かる。

<sup>4</sup> 本稿では、センシティブセルの数が多い基準を「厳しい」と表現し、少ない場合を「緩い」と表現している。

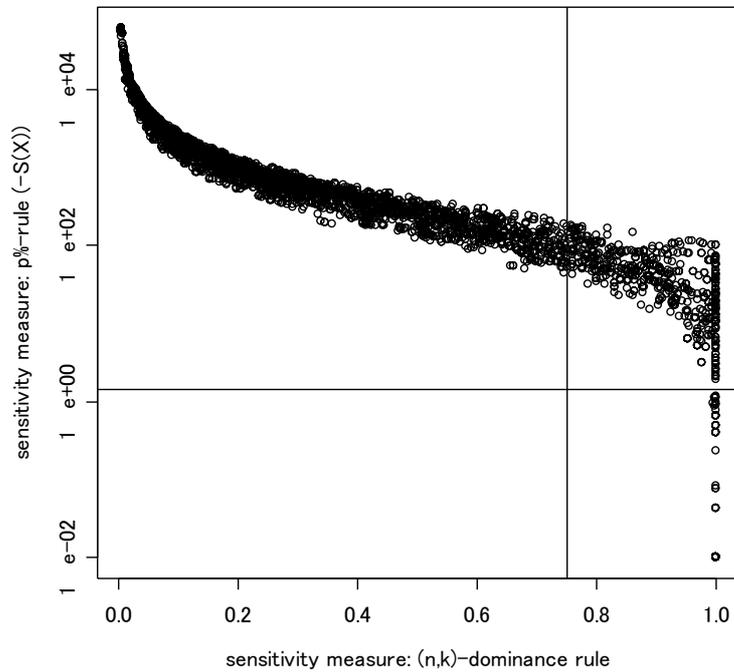


図 5 2つのセンシティブティ尺度

図 6 は  $n-k$  占有ルール ( $n=3$ ) によるセンシティブティ尺度とセル内の客体数の散布図である。実線は  $k=75$  のしきい値に相当し、実線より上側がセンシティブとなる。同様に、図 7 は  $p\%$ ルールによるセンシティブティ尺度とセル内の客体数の散布図である。対数軸表示のためにセンシティブティ尺度の符号を反転している。実線は  $p=25$  のしきい値に相当し、実線より下側がセンシティブとなる。どちらの場合も、セル内の客体数が少ないとセンシティブティ尺度の値が大きく、客体数が多いとセンシティブティ尺度の値が小さくなる傾向が分かる。つまり、セル内の客体数が少ないほどセンシティブとなりやすく、客体数が多いとセンシティブとなりにくいことを意味する。しかし、細かく見ると、客体数が比較的多くてもセンシティブティ尺度の値が小さい場合が見られる。特に、 $n-k$  占有ルール ( $n=3$ ) によるセンシティブティ尺度は、客体数が多いほど分散が大きくなる傾向が著しい。

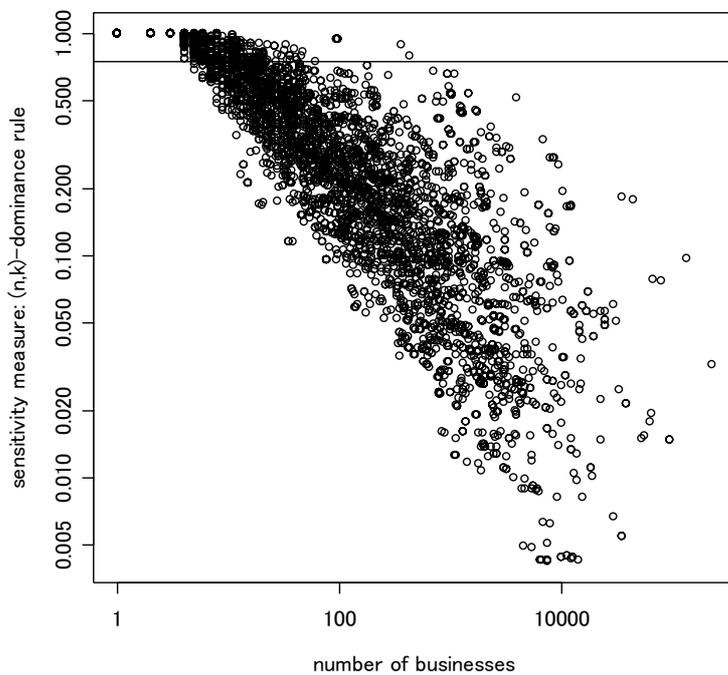


図 6  $n-k$  占有ルール ( $n=3$ ) によるセンシティブティ尺度と客体数

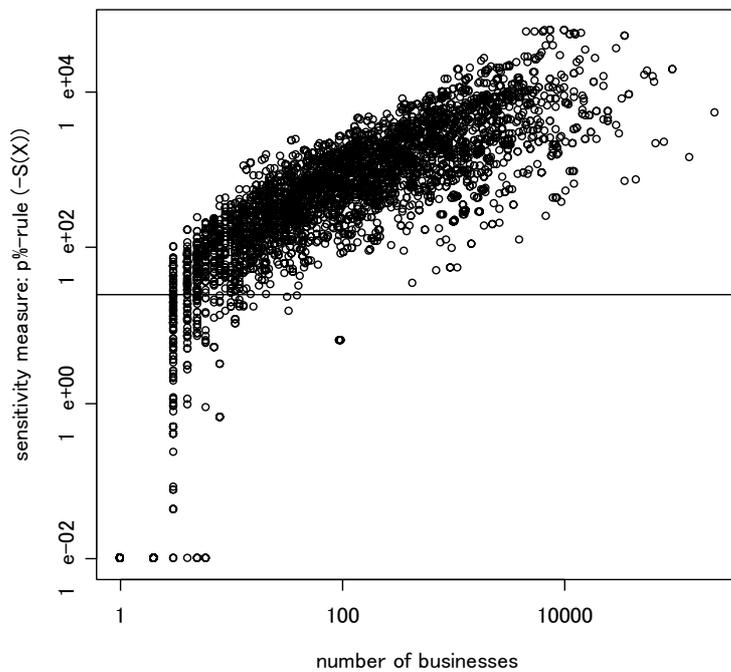


図 7  $p$  %ルールによるセンシティブティ尺度と客体数

$\tau$ -ARGUS では  $n-k$  占有ルール、 $p$ %ルール、最小度数条件を組み合わせてセンシティブセルを特定することができる。A表を用いてさまざまな基準を比較した結果を以下で説明する。

比較した基準は表 2 のとおりである。MF1 は、度数 3 の最小度数条件を表し、他の記号は注のそれぞれの設定を表す。例えば、MF1DR1 は、MF1 と DR1 の両方を組み合わせた基準を表し、度数 3 の最小度数条件または  $n-k$  占有ルール ( $n=3$ 、 $k=75$ ) によってセンシティブとする基準を示す。なお、 $\tau$ -ARGUS では、 $p$ %ルールは他の客体の開示のために複数の客体が結託する場合を想定したルールに拡張しており、結託する客体数  $n$  を指定することができる。また、客体ごとの集計ウェイトをルールに適用することができる。例えば、あるセルに含まれる客体数が 2 であり、客体 1 が収入金額 1 億円でウェイト 4、客体 2 が収入金額 1000 万円でウェイト 7 とすると、セルの値は 4 億 7000 万円となる。以下の実験では、原則として集計ウェイトをルールに適用する。さらに、ルールの中での客体数の数え方にウェイトを適用する場合は、収入金額 1 億円の客体が 4 件、1000 万円の客体が 7 件あるとみなして、上位 2 客体の収入金額は、1 億円と 1 億円となる。表の右側の列は、A表の全 6,424 セルのうち当該基準によってセンシティブと判断されたセルの数を示す。例えば、客体数の数え方にウェイトを適用しない MF1DR1 では、全 6,424 セルのうち 237 セル (3.7%) がセンシティブである。

表 2 一次秘匿基準の比較

組み合わせ	客体数へのウェイトの適用	センシティブセルの数 (全 6,424 セルのうち)	組み合わせ	客体数へのウェイトの適用	センシティブセルの数 (全 6,424 セルのうち)
MF1	なし	175	MF1	あり	175
MF1DR1	なし	237	MF1DR1	あり	358
MF1PR1	なし	200	MF1PR1	あり	234
MF1DR1PR1	なし	237	MF1DR1PR1	あり	358
MF1PR2	なし	179	MF1PR2	あり	198
MF1DR2	なし	190	MF1DR2	あり	267
MF1DR3	なし	217	MF1DR3	あり	271
MF1DR4	なし	179	MF1DR4	あり	208
MF1DR5	なし	258	MF1DR5	あり	495
MF1DR6	なし	194	MF1DR6	あり	331
MF1PR3	なし	241	MF1PR3	あり	264
MF1PR4	なし	209	MF1PR4	あり	274
MF1PR5	なし	183	MF1PR5	あり	235
MF1PR6	なし	253	MF1PR6	あり	311

(注) 組み合わせの略称は以下のとおりである。

記号	最小度数条件の設定	記号	$n-k$ 占有ルールの設定	記号	$p$ %ルールの設定
MF1	最小度数 3	DR1	$n=3, k=75$	PR1	$p=25, n=1$
		DR2	$n=3, k=90$	PR2	$p=10, n=1$
		DR3	$n=2, k=75$	PR3	$p=50, n=1$
		DR4	$n=2, k=90$	PR4	$p=25, n=2$
		DR5	$n=4, k=75$	PR5	$p=10, n=2$
		DR6	$n=4, k=90$	PR6	$p=50, n=2$

表 2 からいくつかの基準を取り上げて説明する。まず、表 3 は、代表的な基準のセンシティブセルの数を示す。度数 3 の最小度数条件 (MF1) のセンシティブセルの数は 175 セルあり、度数 3 の最小度数条件または  $n-k$  占有ルール ( $n=3, k=75$ ) による基準 (MF1DR1) では、237 セルがセンシティブセルとなり、より厳しい基準となる。度数 3 の最小度数条件または  $p$  %ルール ( $p=25$ ) による基準 (MF1PR1) では、200 セルがセンシティブセルとなる。つまり、MF1PR1 は MF1 よりも厳しいが、MF1DR1 よりも緩やかな基準であることが分かる。

以上より、日本の統計で一般的な一次秘匿基準である度数 3 の最小度数条件のみを用いる基準は、 $n-k$  占有ルールや  $p$  %ルールを組み合わせた基準と比較して緩やかな基準であることが分かる。つまり、 $n-k$  占有ルールや  $p$  %ルールの基準でセンシティブとなるセルを見逃すという点で、開示リスクがより高いと言える。外国の統計機関では、 $n-k$  占有ルールや  $p$  %ルールを組み合わせた基準を用いる場合があるが、各パラメータの値は公開されていない。

表 3 一次秘匿基準の組み合わせとセンシティブセルの数

一次秘匿基準の組み合わせ	センシティブセルの数
MF1: 度数 3 の最小度数条件	175
MF1DR1: MF1 と $n-k$ 占有ルール ( $n=3, k=75$ )	237
MF1PR1: MF1 と $p$ %ルール ( $p=25$ )	200

原則として、これ以降の分析では、度数 3 の最小度数条件 (MF1) は必須条件として、 $n-k$  占有ルール、 $p$  %ルールを組み合わせることとする。

度数 3 の最小度数条件は、 $n=2, k=100$  の  $n-k$  占有ルールに相当する。 $n=2$  を固定して、 $k$  を  $k=100, k=90, k=75$  と変えた場合 (表 4) を比較すると、センシティブセルは、それぞれ、175 セル、179 セル、217 セルと順に多くなる。図 2 でも示したように、 $k$  が小さいほど厳しい基準となる。

$n-k$  占有ルールで  $k=75$  に固定して  $n$  を  $n=2, n=3, n=4$  と変えた場合 (表 5) は、センシティブセルは、それぞれ、217 セル、237 セル、258 セルと順に多くなっている

ことがわかる。つまり、 $n$  が大きいほど厳しい基準となる。

表 4  $n-k$  占有ルールにおけるセンシティブセルの数： $n=2$  を固定

$k$	センシティブセルの数
$k=100$	175
$k=90$	179
$k=75$	217

表 5  $n-k$  占有ルールにおけるセンシティブセルの数： $k=75$  を固定

$n$	センシティブセルの数
$n=2$	217
$n=3$	237
$n=4$	258

$p$  %ルールで結託客体数  $n=1$  を固定して、 $p$  を  $p=10$ 、 $p=25$ 、 $p=50$  と変えた場合を比較すると、センシティブセルは、それぞれ、179セル、200セル、241セルとなり、順に多くなっていることがわかる (表 6)。前節の図で示したように、 $p$  が大きいほど ( $-p$  が小さいほど) 厳しい基準となる。また、 $p$  が同じ場合は、結託客体数  $n$  が多いほうが厳しくなる (表 7)。

表 6  $p$  %ルールにおけるセンシティブセルの数：結託客体数  $n=1$  を固定

$p$	センシティブセルの数
$p=10$	179
$p=25$	200
$p=50$	241

表 7  $p$  %ルールにおけるセンシティブセルの数： $p=25$  を固定

結託客体数 $n$	センシティブセルの数
$n=1$	200
$n=2$	209

上で見たように、 $p$  %ルール ( $p=50$ ) ではセンシティブセルが 241セルであり、 $n-k$  占有ルール ( $n=3$ 、 $k=75$ ) の 237セルと同程度の数となっている。表 8 に 2つの基準をセルごとに比較した結果を示す。2つの基準のどちらでもセンシティブとなるセルは 221セルである。 $n-k$  占有ルール ( $n=3$ 、 $k=75$ ) のセンシティブセルのうち 16セルは  $p$  %ルール ( $p=50$ ) ではセンシティブとならない。逆に、 $p$  %ルール ( $p=50$ ) のセンシティブセルのうち 20セルは  $n-k$  占有ルール ( $n=3$ 、 $k=75$ ) ではセンシティブとはならないといった違いがあることが分かる。なお、該当する客体のないセルは「該当なし」として区別した。

表 8  $n-k$  占有ルール ( $n=3$ 、 $k=75$ ) と  $p$  %ルール ( $p=50$ ) の比較

$n-k$ 占有ルール ( $n=3$ 、 $k=75$ )	$p$ %ルール ( $p=50$ )			総計
	センシティブ	非センシティブ	該当なし	
センシティブ	221	16	-	237
非センシティブ	20	4,200	-	4,220
該当なし	-	-	1,967	1,967
総計	241	4,216	1,967	6,424

また、表 2 から、客体数にウェイトを適用した場合は、適用しない場合よりもセンシティブセルが多くなり、厳しい基準となっていることが分かる。特に、 $n-k$  占有ルールは、客体数へのウェイトの適用によるセンシティブセル数の増加が大きい。

まとめると、度数 3 の最小度数条件のみの適用は、 $n-k$  占有ルールや  $p$  %ルールの基準でセンシティブとなるセルを見逃すという点で、開示リスクがより高いと言える。 $n-k$  占有ルールや  $p$  %ルールは、パラメータの値の設定により、センシティブの判断が異なり、開示リスクの適切なレベルの決定や、それに対応する一次秘匿基準の決定が今後の課題である。

## (2) 二次秘匿方法の比較

以下では、 $\tau$ -ARGUS が採用している 4 つの二次秘匿方法のうち、2 次元表のみに適用可能な Network Flows 法を除いた Hypercube 法、HiTaS 法、Optimal 法を比較する。

まず、A 表を使って 3 つの二次秘匿方法を比較する。度数 3 の最小度数条件または  $n-k$  占有ルール ( $n=3$ 、 $k=75$ ) による基準で客体数にウェイトを適用した場合について、3 つの二次秘匿方法による情報ロスを表 9、図 8 に示す。ここで、情報ロスとは、全セルの値の合計に占める秘匿セルの値の合計の割合で定義し、セル数、事業所数、収入金額の 3 通りの値で計算した。

Hypercube 法、HiTaS 法、Optimal 法によるセル数の情報ロスは、それぞれ、34.9%、28.9%、25.5%であり、Hypercube 法では全体の 3 分の 1 以上のセルが秘匿される。収入金額の情報ロスを比較すると、Hypercube 法ではロスが 77.1%であり、収入金額の全体の 8 割近くが秘匿されるが、Optimal 法では 7.0%とロスが大幅に小さくなることが分かる。Optimal 法は情報ロスを最小にする最適解を与えるものであるが、セル数の点では、前述のとおり Optimal 法でも全体の 4 分の 1 のセルが秘匿されることになる。

表 9 3 つの二次秘匿方法による情報ロス

	セル数	事業所数	収入金額
Hypercube 法	34.9%	49.3%	77.1%
HiTaS 法	28.9%	9.7%	20.0%
Optimal 法	25.5%	5.6%	7.0%

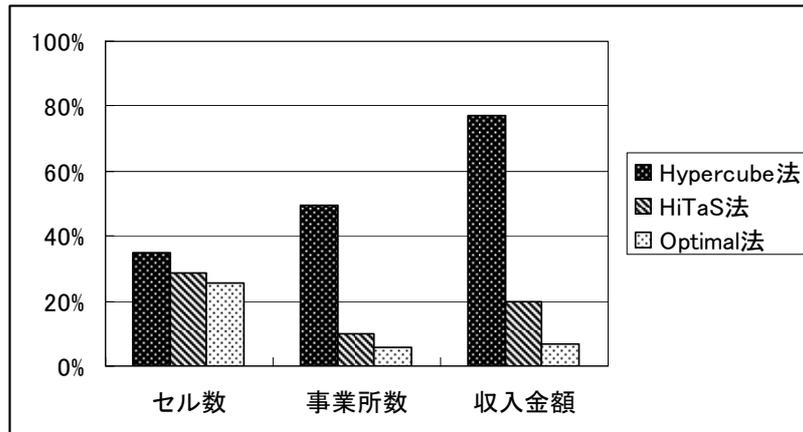


図 8 情報ロスの比較

次に、3つの二次秘匿方法の処理による秘匿箇所の違いを比較する。まず、Hypercube法とOptimal法の秘匿箇所のセル数、収入金額を表10、表11に示す。ARGUSの出力に従って、秘匿しないセルを「Safe」とし、該当する客体のないセルは「該当なし」として区別した。上で示したようにHypercube法はOptimal法よりも情報ロスが大きい。表からHypercube法とOptimal法で共通の二次秘匿セルは784セルであり、Hypercube法での二次秘匿1,882セルのうち、1,098セルはOptimal法では秘匿しないセルであることが分かる。逆に、Optimal法での二次秘匿1,281セルのうち、497セルはHypercube法では秘匿しないセルとなる。収入金額の点から比較すると、Hypercube法の二次秘匿セルの収入金額合計は2767兆2551億円である。一方、Optimal法の二次秘匿セルの収入金額合計は250兆2682億円であり、Hypercube法の約10分の1である。Hypercube法の二次秘匿セルの収入金額合計2767兆2551億円のうち、Optimal法では秘匿しないセルの合計は2554兆8150億円になる。逆に、Optimal法の二次秘匿セルのうちHypercube法では秘匿しないセルの収入金額合計は37兆8281億円であるが、二次秘匿セル全体の収入金額合計に比べて小さい値である。

表 10 Hypercube法とOptimal法の比較：セル数

		Optimal法				計
		Safe	一次秘匿	二次秘匿	該当なし	
Hypercube法	Safe	1,720	-	497	-	2,217
	一次秘匿	-	358	-	-	358
	二次秘匿	1,098	-	784	-	1,882
	該当なし	-	-	-	1,967	1,967
	計	2,818	358	1,281	1,967	6,424

表 11 Hypercube 法と Optimal 法の比較：収入金額合計（億円）

		Optimal 法				計
		Safe	一次秘匿	二次秘匿	該当なし	
Hypercube 法	Safe	7,859,578	-	378,281	-	8,237,859
	一次秘匿	-	29,169	-	-	29,169
	二次秘匿	25,548,150	-	2,124,401	-	27,672,551
	該当なし	-	-	-	-	-
	計	33,407,728	29,169	2,502,682	-	35,939,580

また、HiTaS 法と Optimal 法の秘匿箇所のセル数、収入金額の違いを表 12、表 13 に示す。HiTaS 法は Optimal 法よりも情報ロスが大きいが、Hypercube 法と Optimal 法の処理結果の違いと比べると、HiTaS 法と Optimal 法には共通部分が多いことがわかる。

表 12 HiTaS 法と Optimal 法の比較：セル数

		Optimal 法				計
		Safe	一次秘匿	二次秘匿	該当なし	
HiTaS 法	Safe	2,125	-	473	-	2,598
	一次秘匿	-	358	-	-	358
	二次秘匿	693	-	808	-	1,501
	該当なし	-	-	-	1,967	1,967
	計	2,818	358	1,281	1,967	6,424

表 13 HiTaS 法と Optimal 法の比較：収入金額合計（億円）

		Optimal 法				計
		Safe	一次秘匿	二次秘匿	該当なし	
HiTaS 法	Safe	28,355,398	-	412,626	-	28,768,024
	一次秘匿	-	29,169	-	-	29,169
	二次秘匿	5,052,330	-	2,090,056	-	7,142,386
	該当なし	-	-	-	-	-
	計	33,407,728	29,169	2,502,682	-	35,939,580

Hypercube 法では収入金額の情報ロスが極めて大きいことを示したが、実際、総合計のセルが二次秘匿の対象となっている（付録の結果表を参照）。1次元周辺分布のセルは、158セルのうち53セルが二次秘匿セルとなっている。HiTaS 法では総合計のセルは秘匿されず、1次元周辺分布のセルのうち3セルが二次秘匿セルである。Optimal 法でも総合計のセルは秘匿されず、1次元周辺分布のセルのうち12セルが二次秘匿セルである。

ここまでは、A 表について詳細を見てきたが、同様に、他の表の秘匿処理結果を簡単に説明する。一次秘匿基準は MF1、MF1PR1、MF1DR1 の3種類と客体数へのウェイト適用の有無と組み合わせた6通りとした。この6通りそれぞれについて、Hypercube 法、HiTaS

法、Optimal 法の 3 通りで二次秘匿パターンを決定した。つまり、表ごとに 18 通りで、計 216 通りの処理を行った。

二次秘匿方法別の処理結果を表 14 に示す。処理結果の「成功」、「失敗」、「不可能」は、それぞれ、「処理が（適切かどうかはともかく）完了する」、「処理の途中で異常終了する」、「仕様上の制約で処理できない」ことを意味する。情報ロスの平均は、処理が「成功」の場合の単純平均をとった。Hypercube 法はすべての条件で処理が完了したが、ほとんどの表で情報ロスが極めて大きく、平均が大きな値を示している。HiTaS 法、Optimal 法は情報ロスが比較的小さいが、エラーが発生して処理の途中で異常終了することが多く、業務への適用においては計算の安定性が大きな課題である。さらに、HiTaS 法と Optimal 法は 3 次元表までの制限があり、4 次元表の H-2 表は処理できない。また、セル数の情報ロスは 3 つの方法であまり差がなく、いずれも大きな値を示している。

表 14 二次秘匿方法別処理結果の比較

二次秘匿処理	処理結果				情報ロス(平均)		
	成功	失敗	不可能	計	セル数	事業所数	収入金額
Hypercube 法	72	0	0	72	24.6%	24.7%	34.7%
HiTaS 法	11	55	6	72	19.4%	6.0%	8.1%
Optimal 法	10	56	6	72	27.1%	6.1%	6.8%
計	93	111	12	216	24.3%	20.5%	28.5%

一方で、表 15 の集計表別の処理結果を見ると、2 次元表の B 表や G 表では収入金額や事業所数の情報ロスが十分小さい。特に、比較的単純な構造を持つ B 表では、セル数の情報ロスも 10.4%と小さく、実用的な秘匿処理が行われていると考えられる。τARGUS に付属のサンプルデータを使った試行においても、2 次元表の処理では適切な結果を得られている。

表 15 集計表別処理結果の比較

二次秘匿処理	処理結果				情報ロス(平均)		
	成功	失敗	不可能	計	セル数	事業所数	収入金額
A	10	8	0	18	29.2%	30.9%	48.5%
B	10	8	0	18	10.4%	0.9%	1.6%
C	7	11	0	18	36.0%	40.6%	65.5%
D	7	11	0	18	28.6%	28.8%	37.5%
E-1	8	10	0	18	24.7%	31.9%	41.1%
E-2	7	11	0	18	22.2%	29.4%	45.7%
F-1	7	11	0	18	23.7%	22.3%	25.0%
F-2	7	11	0	18	22.3%	21.1%	23.3%
G	11	7	0	18	21.2%	3.7%	7.2%
H-1	7	11	0	18	27.3%	13.4%	17.5%
H-2	6	0	12	18	25.2%	17.4%	21.6%
H-3	6	12	0	18	26.5%	16.4%	18.8%
計	93	111	12	216	24.3%	20.5%	28.5%

次に、公式統計書の秘匿との比較を行う。比較はA、Bの2表を対象とした。公式統計書には、この2表と同一の表はないため、集計事項の組み合わせが同じ表を使って、セルごと秘匿の有無を転写した。ただし、この研究で用いたデータは特定サービス産業実態調査の対象の事業所や、一部の産業分類の事業所を除外しているため、公式統計書より集計対象の事業所数が少ない。さらに、 $\tau$ -ARGUSの仕様で、集計ウェイトが整数でない場合、度数3の最小度数条件は、3未満をセンシティブと判定するため、結果数値が1または2を秘匿する公式統計書の基準よりも、一次秘匿セルが多くなる。このような違いから、本研究で適用した処理結果は公式統計書よりも秘匿箇所が多い傾向があることに注意する必要がある。

表16にセル数、事業所数、収入金額それぞれの情報ロスを示す。A表については、 $\tau$ -ARGUSによる二次秘匿はHypercube法だけが処理可能であったが、情報ロスが非常に大きい。公式報告書と比較して、セル数の情報ロスが3.7倍、事業所数の情報ロスが45倍、収入金額の情報ロスが40倍となっている。B表はHypercube法とHiTaS法で処理ができたため、情報ロスの少ないHiTaS法の結果を示した。公式報告書と比較すると、セル数の情報ロスが2.3倍、事業所数の情報ロスが4.8倍、収入金額の情報ロスが1.8倍であり、A表ほど大きな差異は見られない。前述のように、B表のような2次元表については、 $\tau$ -ARGUSを用いた秘匿処理が実際に利用可能であると思われる。なお、公式統計書の秘匿箇所を転写する作業の際に、複数表間で秘匿の有無が一致せず、秘匿する必要があると考えられるセルが秘匿されていない箇所が見つかった(付録の結果表を参照)。このような不適切な処理を発見するためには、秘匿結果の適切さや開示リスクの程度を確認する仕組みが必要であると思われる。

表 16 秘匿箇所の比較

		情報ロス		
		セル数	事業所数	収入金額
A表	公式統計書	8.4%	1.1%	2.0%
	ARGUS(MF1, Hypercube)	30.8%	48.9%	79.2%
B表	公式統計書	2.4%	0.0%	0.1%
	ARGUS(MF1, HiTaS)	5.5%	0.2%	0.2%

最後に、二次秘匿における秘匿のレンジの影響について簡単に比較する。 $\tau$ -ARGUSでは、最小度数条件において、レンジ $r$ を指定することによって、一次秘匿セル $x$ の推定値 $\tilde{x}$ に対して望ましい(許容できる)推定の範囲を $x - r \cdot x < \tilde{x} < x + r \cdot x$ に設定することができる。 $r$ が小さいほど正確な推定を許容する、つまり、緩い条件であり、開示リスクが大きくなる。逆に、 $r$ が大きいほど正確な推定を許容しない、つまり、厳しい条件であることを意味する。ここまでの検討では、レンジ $r$ はすべて100%としていた。A表について度数3の最小度数条件で、レンジを30%と100%に設定した場合の秘匿セル数を表17に示す。ここで、二次

秘匿は Hypercube 法を用いた。度数 3 の最小度数条件による一次秘匿のセル数は同じであるが、より厳しい条件である 100% のレンジの方が、二次秘匿セルの数が増えることが確認できる。また、表 18 のとおり、100% のレンジの方が、情報ロスも増える。なお、レンジを 100% にした場合、エラーが発生して処理の途中で異常終了する傾向が顕著に見られるため、実用においては処理の安定性が課題である。

表 17 最小度数条件の秘匿のレンジの比較：セル数

	Safe	一次秘匿	二次秘匿	該当なし	計
レンジ 30%	2554	175	1728	1967	6424
レンジ 100%	2480	175	1802	1967	6424

表 18 最小度数条件の秘匿のレンジの比較：情報ロス

	セル数	事業所数	収入金額
レンジ 30%	29.6%	47.5%	72.6%
レンジ 100%	30.8%	49.3%	79.2%

### (3) 複数集計表の一括（リンク）処理の効果

τ-ARGUS の複数集計表の一括（リンク）処理を検証する。一括処理を行うことで、複数の表間で共通するセルを同時整合的に扱うことができる。ここでは、A、B、C の 3 表に絞って、個々の表を独立に処理する場合と、linked table 機能により 3 表を一括処理する場合を比較する。

A 表は、産業（小分類）、本所・支所（3 区分）、経営組織（7 区分）をクロスした 3 次元表であり、B 表は、産業（小分類）、開設時期（19 区分）をクロスした 2 次元表である。この 2 つの表は、産業（小分類）の 1 次元周辺分布の 146 セルが共通となる。同様に、A 表と C 表の間では産業（小分類）と経営組織（7 区分）をクロスした 2 次元周辺分布の 1,606 セルが共通であり、B 表と C 表の間では産業（小分類）の 1 次元周辺分布の 146 セルが共通である。個々の表を独立に処理する場合は、例えば、A 表では産業（小分類）の 1 次元周辺分布の一部のセルが二次秘匿セルとなるが、B 表ではすべて Safe セルであるといった秘匿処理結果（ステータス）の不一致が発生する。一方、linked table 機能による一括処理では、同時整合的に秘匿処理を行い、不一致は発生しない。

表 19 に 2 表間の整合性を示す。ここで、一次秘匿は度数 3 の最小度数条件または  $n-k$  占有ルール ( $n=3$ ,  $k=75$ ) により、客体数にウェイトを適用する基準を使用した。二次秘匿は、計算可能な方法の中で、情報ロスが最小のものを選択している。個々の表を独立に処理した場合には、共通セルのうち 1 割程度にステータスの不一致が見られる。

同様に、情報ロスを表 20 に示す。個々の表を独立に処理する場合は二次秘匿方法として HiTaS 法や Optimal 法を選択できるが、3 表を一括処理する場合には Hypercube 法のみ適用可能であるため、情報ロスが極めて大きくなる。このため、linked table 機能の適用に

おいては、二次秘匿処理の改善が大きな課題である。

表 19 2表間共通セルの整合性

表間	共通セル数	ステータス不一致の割合	
		個々の表を独立に処理	3表を一括処理
A-B	146	8%	0%
A-C	1,606	12%	0%
B-C	146	13%	0%

表 20 一括処理の情報ロス

表間処理	表番号	二次秘匿方法	セル数	事業所数	収入金額
独立	A	Optimal	25.5%	5.6%	7.0%
	B	Optimal	26.0%	2.4%	2.5%
	C	Optimal	29.9%	9.2%	11.8%
一括	A	Hypercube	38.8%	57.6%	81.0%
	B		16.6%	35.1%	41.6%
	C		40.2%	55.8%	77.5%

#### (4) 逐次処理の効果

$\tau$ -ARGUS の機能の制約のため、A から H 表の全 12 表を同時整合的に秘匿処理することが難しい。そこで、表ごとに順に秘匿処理を行い、前の段階の秘匿処理結果を制約条件として、次の表の処理を行う逐次的な方法を検討する。

処理は、 $\tau$ -ARGUS のヒストリーファイル機能を用い、前段階での秘匿処理結果による秘匿セルを Unsafe とし、表示セルは Safe に指定する。表 21、表 22、表 23 の 3 通りの逐次処理と、個々の表を独立に処理した場合を比較する。一次秘匿は、度数 3 の最小度数条件または  $n-k$  占有ルール ( $n=3$ ,  $k=75$ ) により、客体数にウェイトを適用する基準を使用した。二次秘匿は、実行可能な方法の中で、情報ロスが最小のものを選択している。ただし、逐次処理のほとんどの表では、Hypercube 法以外は実行不可能だったため、二次秘匿を Hypercube 法に限定した独立処理も併せて比較する。

異なる 2 表間の共通セルは延べ 54,994 組あり、独立処理では 12,897 組がステータス不一致、逐次処理 1 では 8,690 組、逐次処理 2 では 9,910 組、逐次処理 3 では 7,365 組がそれぞれステータス不一致である。各表間の共通セル数、不一致数を付表 1、付表 2、付表 3、付表 4、付表 5 に示す。独立処理と比べると、逐次処理は制約条件の効果で不一致が減っていることがわかる。次に、表 24 に収入金額の情報ロスを示し、同様に、付表 6、付表 7 にセル数と事業所数の情報ロスを示す。全体的に Hypercube 法による情報ロスが非常に大きいと逐次処理の効果の比較が難しいが、処理の順序が後になるほど情報ロスが大きくなる傾向がある。

以上のことから、ヒストリーファイルを用いることで、ある程度の同時整合性を確保で

きるが、業務への適用においては、第一に二次秘匿処理の改善が大きな課題であり、第二に適切な処理順を決めることが必要である。

表 21 逐次処理 1

処理順	表番号	制約条件
1	A、B、C 同時	なし
2	D	A
3	E-1	A、C
4	E-2	なし
5	F-1	D、E-1
6	F-2	E-2
7	G	A、B、C、 E-1、E-2
8	H-1	E-1、E-2、G
9	H-2	E-1、G、H-1
10	H-3	E-1、G、H-2

表 22 逐次処理 2

処理順	表番号	制約条件
1	A	なし
2	B	A
3	C	A
4	D	A
5	E-1	A
6	E-2	なし
7	F-1	D、E-1
8	F-2	E-2
9	G	A、E-1、E-2
10	H-1	E-1、E-2、G
11	H-2	E-1、G、H-1
12	H-3	E-1、G、H-2

表 23 逐次処理 3

処理順	表番号	制約条件
1	H-1	なし
2	H-2	H-1
3	H-3	H-2
4	G	H-1、H-3
5	E-1	G、H-3
6	E-2	G、H-1
7	F-1	E-1
8	F-2	E-2
9	D	F-1
10	C	D、E-1、F-1、G
11	B	C
12	A	B、C、D

表 24 逐次処理の比較：収入金額の情報ロス

表番号	情報ロス(収入金額)				
	独立処理	独立処理 (Hypercube)	逐次処理1	逐次処理2	逐次処理3
A	7.0%	77.1%	81.0%	20.0%	85.6%
B	2.5%	3.2%	41.6%	4.0%	56.2%
C	11.8%	74.0%	77.5%	82.7%	86.1%
D	5.5%	55.2%	74.9%	70.6%	72.7%
E-1	9.8%	39.0%	72.0%	66.3%	82.0%
E-2	15.3%	43.7%	43.7%	15.3%	80.2%
F-1	2.3%	26.8%	75.6%	76.9%	75.4%
F-2	2.6%	34.7%	73.8%	62.3%	73.7%
G	12.2%	12.2%	67.6%	58.7%	71.6%
H-1	9.7%	27.2%	81.7%	77.6%	27.2%
H-2	28.9%	28.9%	89.4%	85.9%	68.2%
H-3	28.0%	28.0%	84.1%	82.1%	74.5%

(5) コスト変数の比較

τ-ARGUS では二次秘匿処理において最小化する情報ロスの基準をコスト変数として指定することができる。ここまでの検討では、収入金額の情報ロスを最小にする基準を用いており、コスト変数は収入金額であった。

ここでは、A 表について、セル数、事業所数、収入金額のそれぞれをコスト変数とした場合を比較する。一次秘匿は、度数 3 の最小度数条件または  $n-k$  占有ルール ( $n=3$ 、 $k=75$ )

により、客体数にウェイトを適用する基準を使用した。なお、Hypercube 法による二次秘匿処理はコスト変数によらず同じ結果となり、HiTaS 法は実行不可能だったため、Optimal 法による結果のみを比較する。表 25 にそれぞれのコスト変数を選択した場合の情報ロスを示す。セル数をコスト変数にした場合は、セル数の情報ロスが小さくなり、事業所数をコスト変数にした場合は、事業所数の情報ロスが小さくなり、収入金額をコスト変数にした場合は、収入金額の情報ロスが小さくなることを確認できる。また、セル数をコスト変数とした場合は、セル数の情報ロスは減少するが、収入金額の情報ロスがかなり大きくなることから分かる。事業所数をコスト変数にした場合と収入金額をコスト変数にした場合は、情報ロスは似た傾向を示す。

このように、 $\tau$ -ARGUS では最適化するコスト変数を選択することができるが、二次秘匿処理として HiTaS 法や Optimal 法が実行できない場合が多いため、実際の適用には注意が必要である。

表 25 コスト変数の比較：情報ロス

コスト変数	情報ロス		
	セル数	事業所数	収入金額
セル数	19.2%	38.3%	42.8%
事業所数	26.9%	5.1%	7.6%
収入金額	25.5%	5.6%	7.0%

#### (6) Singleton 秘匿処理の効果

客体数が 1 のセルを Singleton という。それらのセルは一次秘匿の対象であるが、同じ縦列や横列に 2 つの Singleton があると、それぞれの Singleton の客体が互いの値を特定することが可能な場合がある。このため、Singleton セル間の開示リスクが発生する。 $\tau$ -ARGUS では、二次秘匿処理の前に、Hypercube 法を用いた追加的な Singleton 秘匿の処理を行うことができる。例えば、A 表について、一次秘匿は、度数 3 の最小度数条件または  $n-k$  占有ルール ( $n=3$ 、 $k=75$ ) により、客体数にウェイトを適用する基準を使用した場合、1,582 セルが Singleton 秘匿となる。表 26 に Singleton 秘匿処理の有無による結果の情報ロスを示す。どちらも二次秘匿は Optimal 法を用いた。Singleton 秘匿処理をすると、収入金額の情報ロスが 2 倍以上に増加することが分かる。

Singleton 秘匿処理は開示リスクを減らすために有効だが、情報ロスが過大になるため、処理方法の改善が必要である。

表 26 Singleton 秘匿処理の影響：情報ロス

Singleton 秘匿処理	情報ロス		
	セル数	事業所数	収入金額
なし	25.5%	5.6%	7.0%
あり	33.3%	11.7%	16.0%

## 5. 適切な秘匿処理のためのアプローチ

前節までに見たように、 $\tau$ -ARGUSによるサービス業基本調査集計表の秘匿は、二次秘匿処理のアルゴリズムと $\tau$ -ARGUSの処理性能の限界や、集計表の複雑な構造のために、適切な結果を得ることが難しかった。ここでは、調査や利用者が必要とする条件に従って適切な秘匿結果を得るためのいくつかのアプローチを説明する。

より多くのセルを公表する方法の一つである秘匿の免責は、統計機関と回答者の間の協定である。セルに大きな寄与を持つ回答者から開示リスクにもかかわらずセルのデータの公開の許可を得る。つまり、協定の下で、回答者は統計機関が回答者の機密を保持するための義務を免除する。センシティブセルの回答者から免責を得ることによって、当該セルの秘匿を回避でき、それに伴う多くの二次秘匿も不要となる。カナダ統計局の Provençal & Bérard (2003)、Provençal et al. (2004)は、最適な秘匿パターンを導く免責対象を特定する方法を示している。前者の翻訳を本資料に収録した。

オランダ統計局の Van der Meijden (2005)は、詳細な情報をもつ集計表を必要とする利用者の要望に応じて、開示抑制による情報量の損失を抑えるための6つの方法を示している。そのうち4つはマイクロデータの扱い方に関するもので、①一次秘匿や分類階層の高いレベルでの二次秘匿を減らすために、分類の階層構造を区分数の少ないものから多いものへ再構築すること、②占有度をより小さくして一次秘匿を減らすような推定値やその分布を利用するために小地域推定を改良すること、③極端な場合には、多くの秘匿箇所を緩和するために、ある客体が該当する区分を近い区分へ故意に変更するなどの調整を行うこと、④複数の顧客間で必要とする結果表の分類区分が異なる場合は、公表される情報の量を増やすために、顧客の合意の下で分類を統一した表を公表するなど公表義務を変えることである。残りの2つは、 $\tau$ -ARGUSの使用法に関するもので、①ヒストリーファイルを用いて機密保持パターンを割り当てること、②異なる情報ロスの基準を用いて二次秘匿を改善することを挙げている。

## 6. まとめ

$\tau$ -ARGUSによる集計表の秘匿処理の業務適用可能性や適切な秘匿処理法について実験を行った。日本の統計で一般的な一次秘匿基準である度数3の最小度数条件のみの適用は、 $n-k$ 占有ルールや $p\%$ ルールを組み合わせた基準と比較して緩やかな基準である。つまり、度数3の最小度数条件のみの適用は、 $n-k$ 占有ルールや $p\%$ ルールの観点で、開示リスクがより高いことを意味する。開示リスクの適切なレベルの決定や、対応する一次秘匿基準の決定は今後の課題である。

さらに、業務への適用を検討するため、二次秘匿処理の比較や、複数集計表の一括処理、逐次処理などの検証を行った。しかし、実験で適用した集計表に対して、 $\tau$ -ARGUSは二次秘匿処理の際にエラーが発生して異常終了することや秘匿による情報ロスが非常に大きくなることが分かった。これらの処理機能の制約により、単独の2次元表の場合を除くと、

適切な秘匿処理結果を得ることができなかった。なお、検証の過程で公式統計書の秘匿処理結果には複数表間で秘匿の不一致があることが分かった。処理の適切さや開示リスクを確認する仕組みが必要であると考えられる。

総合的に見ると、 $\tau$ -ARGUS は処理機能の制約のために、現行のサービス業基本調査の秘匿処理への実際の適用は難しい。そこで、調査や利用者が必要とする条件に従って適切な秘匿結果を得るためのいくつかのアプローチを紹介した。その一つに回答者の了解を得て秘匿を行わない免責アプローチがある。今後は、二次秘匿処理機能の改良のほか、集計事項の分類区分の再編や2次元表に限定したクロスの簡略化、免責アプローチの適用も考慮に入れた検討が必要であると考えられる。

#### 参 考 文 献

CASC-project (2004).  $\tau$ -ARGUS version 3.0 User's Manual.  
(<http://neon.vb.cbs.nl/casc/Software/taumannualv3.pdf>)

Provençal, J.-S. and Bérard, H. (2003). Waivers in business surveys: A systematic approach to increase the amount of publishable information, *2003 FCSM Conference Papers*, Federal Committee on Statistical Methodology. (Provençal et al. (2004)による改良を追記して本参考資料に翻訳 (村田磨理子訳) を収録)

Provençal, J.-S., Bérard, H., Fillion, J.-M. and Tambay, J.-L. (2004). Approaches to identify the amount of publishable information in business surveys through waivers, in J. Domingo-Ferrer and V. Torra (eds.) *Privacy in Statistical Databases*, Berlin: Springer-Verlag, pp.110-120.

Van der Meijden, R. (2005). Improving confidentiality with  $\tau$ -ARGUS by focusing on clever usage of microdata, *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Working Paper No.42*.

UN/ECE and Eurostat (2003). Glossary on Statistical Disclosure Control (Draft). (統計センター研究センター訳 (2004). 統計データ開示抑制に関する用語集暫定版 (対訳), 「製表関連国際用語集」, No.2, <http://www.nstac.go.jp/services/pdf/skk-yogosyu2.pdf>)

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, pp.337-350.

松本正博 (1996). サービス業基本調査の秘匿処理法について, 「統計局研究彙報」, 第 54 号, pp.57-86.

付表 1 2表間の共通セルの数

	A	B	C	D	E-1	E-2	F-1	F-2	G	H-1	H-2	H-3
A		146	1,606	584	319	0	29	0	146	116	116	116
B			146	0	29	0	0	0	146	29	29	29
C				146	319	0	29	0	146	116	116	116
D					29	0	232	0	0	0	0	0
E-1						0	1,584	0	1,392	6,336	6,336	6,336
E-2							0	462	377	1,716	0	0
F-1								0	0	0	0	0
F-2									0	0	0	0
G										3,828	1,392	1,392
H-1											6,336	6,336
H-2												6,336
H-3												

付表 2 ステータス不一致セルの数：独立処理の場合

	A	B	C	D	E-1	E-2	F-1	F-2	G	H-1	H-2	H-3
A		12	192	92	27	0	10	0	14	9	45	21
B			19	0	3	0	0	0	2	0	12	0
C				40	48	0	13	0	21	12	46	24
D					10	0	43	0	0	0	0	0
E-1						0	167	0	379	1,020	1,874	1,837
E-2							0	51	130	373	0	0
F-1								0	0	0	0	0
F-2									0	0	0	0
G										965	288	352
H-1											1,836	1,787
H-2												1,123
H-3												

付表 3 ステータス不一致セルの数：逐次処理1の場合

	A	B	C	D	E-1	E-2	F-1	F-2	G	H-1	H-2	H-3
A		0	0	39	46	0	6	0	14	21	25	26
B			0	0	6	0	0	0	14	8	8	9
C				13	46	0	6	0	14	21	25	26
D					2	0	18	0	0	0	0	0
E-1						0	68	0	140	1,160	1,504	1,768
E-2							0	35	54	470	0	0
F-1								0	0	0	0	0
F-2									0	0	0	0
G										873	467	542
H-1											344	608
H-2												264
H-3												

付表 4 ステータス不一致セルの数：逐次処理 2 の場合

	A	B	C	D	E-1	E-2	F-1	F-2	G	H-1	H-2	H-3
A		4	399	105	82	0	15	0	27	39	50	51
B			73	0	12	0	0	0	25	17	22	22
C				26	60	0	3	0	56	27	30	29
D					9	0	26	0	0	0	0	0
E-1						0	66	0	167	1,257	1,620	1,880
E-2							0	56	74	565	0	0
F-1								0	0	0	0	0
F-2									0	0	0	0
G										794	454	522
H-1											363	623
H-2												260
H-3												

付表 5 ステータス不一致セルの数：逐次処理 3 の場合

	A	B	C	D	E-1	E-2	F-1	F-2	G	H-1	H-2	H-3
A		30	154	178	69	0	2	0	63	69	38	25
B			2	0	3	0	0	0	33	23	11	6
C				24	42	0	2	0	31	62	31	18
D					4	0	53	0	0	0	0	0
E-1						0	29	0	35	1,428	631	231
E-2							0	15	60	344	0	0
F-1								0	0	0	0	0
F-2									0	0	0	0
G										997	174	53
H-1											797	1,197
H-2												400
H-3												

付表 6 逐次処理の情報ロス (セル数)

表番号	情報ロス(セル数)				
	独立処理	独立処理 (Hypercube)	逐次処理 1	逐次処理 2	逐次処理 3
A	25.5%	34.9%	38.8%	28.9%	42.6%
B	26.0%	15.1%	16.6%	28.5%	19.4%
C	29.9%	40.1%	40.2%	42.7%	41.2%
D	31.0%	32.3%	32.8%	32.1%	32.0%
E-1	24.3%	27.3%	27.7%	27.5%	40.5%
E-2	27.6%	23.3%	23.3%	27.6%	37.5%
F-1	27.5%	28.4%	33.3%	33.2%	35.7%
F-2	26.9%	29.6%	34.1%	31.0%	36.9%
G	33.6%	33.6%	36.4%	37.5%	41.1%
H-1	33.0%	33.0%	41.5%	43.9%	33.0%
H-2	32.6%	32.6%	46.6%	47.2%	41.7%
H-3	33.4%	33.4%	40.1%	40.3%	38.5%

付表 7 逐次処理の情報ロス (事業所数)

表番号	情報ロス(事業所数)				
	独立処理	独立処理 (Hypercube)	逐次処理 1	逐次処理 2	逐次処理 3
A	5.6%	49.3%	57.6%	9.7%	77.6%
B	2.4%	1.6%	35.1%	2.8%	51.3%
C	9.2%	47.8%	55.8%	62.2%	71.7%
D	6.1%	43.5%	65.3%	57.8%	61.8%
E-1	8.0%	28.0%	49.1%	47.8%	74.0%
E-2	12.4%	25.8%	25.8%	12.4%	69.8%
F-1	3.2%	25.4%	67.6%	67.8%	67.2%
F-2	3.1%	29.2%	64.2%	50.6%	68.1%
G	6.8%	6.8%	49.2%	42.4%	60.3%
H-1	10.3%	15.3%	60.5%	60.4%	15.3%
H-2	22.6%	22.6%	75.8%	74.9%	51.5%
H-3	15.6%	15.6%	70.2%	70.0%	58.8%



## 付録 結果表

- A表 産業(小分類), 本所・支所 (3 区分), 経営組織 (7 区分) 別事業所数・収入金額  
一次秘匿: 度数 3 の最小度数条件または  $n-k$  占有ルール ( $n=3$ 、 $k=75$ ) により、  
          客体数にウェイトを適用する基準  
二次秘匿: Hypercube 法

### 凡例

- P 本稿の処理結果で一次秘匿のセル  
S 本稿の処理結果で二次秘匿のセル  
X 公式統計書で秘匿されているが、本稿の処理結果では秘匿されないセル
-  公式統計書で秘匿されているセル  
 公式統計書で秘匿する必要があると考えられるセル



別事業所数・収入金額

総数	収入金額(百万円)										
	個人	会社	国内の会社				外国の会社	その他			
			株式会社	有限会社	合名・合資・相互会社	会社以外の法人		法人以外の団体			
S	9,014,283	S	S	S	17,526,891	337,001	53,032	S	S	1,232,886	L
7,340,270	2,159,617	5,137,431	5,137,431	3,221,867	1,887,822	27,743	-	43,223	39,661	3,561	72
3,273,274	S	2,860,366	2,860,366	1,980,450	859,369	20,546	-	S	S	S	55
2,174,949	S	S	S	S	718,764	13,273	-	S	S	S	55
1,098,325	15,947	S	S	S	140,606	7,273	-	P	P	-	72B
21,053	14,557	6,413	6,413	1,941	4,472	-	-	83	83	-	722
979,960	695,557	283,340	283,340	117,155	165,027	1,158	-	1,063	589	474	723
2,443,752	976,580	1,466,995	1,466,995	760,029	701,999	4,967	-	176	176	-	724
176,772	59,069	113,818	113,818	45,200	67,931	687	-	3,885	3,686	199	725
361,366	S	331,928	331,928	251,222	80,321	384	-	S	S	2,833	726
84,093	9,521	74,572	74,572	65,870	8,703	-	-	-	-	-	729
518,253	94,684	324,545	324,545	252,886	68,936	2,723	-	99,024	90,246	8,778	73
S	323,203	S	S	2,752,508	693,309	S	P	130,448	125,349	5,099	74
1,334,640	204,412	1,129,877	S	S	246,220	X	P	351	351	-	743
129,832	25,702	104,131	104,131	59,235	X	X	-	-	-	-	744
41,402	S	S	S	S	5,551	-	-	S	S	1,312	745
61,530	P	30,999	30,999	29,791	S	P	-	S	S	276	746
S	36,903	S	S	S	353,637	S	-	S	S	2,567	747
S	48,955	S	S	S	87,818	S	P	5,004	4,059	944	749
10,090,738	472,211	8,825,427	8,825,111	7,266,396	1,452,081	106,634	316	793,099	S	S	75
9,182,897	429,867	8,515,312	8,515,042	6,989,985	1,419,460	105,597	271	237,719	232,471	5,248	751
61,999	12,095	46,808	46,808	26,750	20,058	-	-	3,096	2,895	201	752
21,446	16,074	5,373	5,373	1,562	3,811	-	-	-	-	-	753
824,395	14,176	257,935	257,889	248,100	8,752	1,037	46	552,284	S	S	759
580,107	5,239	95,530	95,485	91,300	3,148	1,037	46	479,338	S	S	75A
244,288	8,937	162,404	162,404	156,799	5,605	-	-	72,947	66,855	6,091	75B
S	1,014,186	S	S	S	8,976,716	105,936	P	3,905,261	3,839,013	66,249	76
264,645	2,425	S	S	S	X	461	P	-	P	P	761
182,136	2,141	156,104	156,104	149,911	6,193	-	-	23,891	X	X	762
732,799	3,811	651,252	S	577,898	X	P	P	77,735	72,761	4,974	763
S	-	52,629	52,629	-	-	-	-	S	S	-	764
241,089	58,832	2,276	2,276	1,963	313	-	-	179,981	153,273	26,708	765
S	S	S	S	S	33,337	S	S	S	163,271	S	766
S	5,326	83,948	83,948	77,617	6,331	-	-	S	S	S	76A
48,721	132	P	P	P	-	-	-	S	S	2,917	76B
1,564,957	752	S	S	S	S	S	P	S	S	15,460	76C
200,170	P	S	198,353	188,607	S	S	S	-	-	-	76E
29,669	3,869	25,778	25,778	13,979	11,798	-	-	23	23	-	76G
696,006	320	601,884	601,884	S	X	P	-	93,801	85,948	7,853	767
S	S	S	S	S	8,628,370	S	-	S	S	S	768
168,044	88,375	79,669	79,669	37,671	41,998	-	-	-	-	-	76K
S	X	S	S	S	8,485,097	S	-	P	-	P	76L
797,357	S	762,323	762,323	657,542	101,274	3,506	-	S	S	S	76M
S	95,072	950,125	950,125	724,521	224,153	1,451	-	S	S	3,653	769
27,032	536	23,551	23,551	21,683	1,867	-	-	2,945	2,467	478	76N
S	10,220	7,057	7,057	3,706	3,351	-	-	S	S	S	76P
S	84,316	919,517	919,517	699,132	218,935	1,451	-	S	S	S	76Q
4,075,574	760,325	3,181,029	3,181,029	1,997,334	1,160,677	23,017	-	134,220	133,813	407	77
5,473,482	127,095	5,332,345	S	S	294,043	3,425	P	14,042	S	S	78
5,286,767	73,143	5,201,091	S	S	267,918	3,103	P	12,533	12,533	-	781
35,649	28,413	7,236	7,236	5,764	1,473	-	-	-	-	-	784
151,066	25,540	124,017	124,017	99,043	24,652	321	-	1,509	S	S	789
S	72,607	S	S	S	240,002	S	2,184	S	S	P	79
1,091,109	S	X	X	X	20,600	521	-	P	P	P	794
28,306	2,918	24,192	24,192	14,976	9,215	-	-	1,197	S	P	795
S	S	S	S	S	210,187	S	2,184	S	S	S	799
S	42,662	S	S	S	97,089	-	-	-	-	-	79A
962,476	S	934,687	932,503	S	113,098	S	2,184	S	S	S	79B
S	2,824	S	S	S	91,162	P	P	S	S	696	80, 81
S	S	S	S	S	X	P	P	12,415	12,415	-	80
S	2,566	S	S	S	X	P	P	12,415	12,415	-	801
86,757	P	S	S	S	8,482	-	-	-	-	-	802
S	S	S	S	S	1,155	P	-	S	S	696	81
S	-	-	-	-	-	-	-	S	S	-	811
2,457,483	-	S	S	S	-	-	-	P	P	-	812
349,057	S	S	S	S	1,155	P	-	X	X	696	813
940,108	S	S	S	S	X	X	P	P	P	P	82
872,401	P	S	S	S	2,379	-	P	P	P	P	823
67,707	1,600	66,106	66,106	62,331	3,762	14	-	-	-	-	824
14,086,604	3,324,295	10,043,932	S	S	S	9,593	P	718,377	680,381	37,996	84
834,508	834,508	-	-	-	-	-	-	-	-	-	841
476,885	476,885	-	-	-	-	-	-	-	-	-	84K
357,623	357,623	-	-	-	-	-	-	-	-	-	84L
310,305	310,305	-	-	-	-	-	-	-	-	-	842
1,569,763	1,380,671	-	-	-	-	-	-	189,092	189,092	-	843
367,961	178,869	-	-	-	-	-	-	189,092	189,092	-	84M
1,201,802	1,201,802	-	-	-	-	-	-	-	-	-	84N
238,847	104,622	129,226	129,226	13,500	115,726	-	-	5,000	5,000	-	844
S	S	S	S	S	579,140	S	P	446,266	S	S	845
631,247	52,462	578,405	578,405	S	135,571	P	-	380	-	380	846
10,788	10,788	-	-	-	-	-	-	-	-	-	847
S	S	S	S	S	S	S	-	77,639	S	S	848
S	S	S	S	S	S	S	P	2,429	1,231	1,199	84A
273,939	S	S	S	S	8,392	-	-	S	S	P	84B
330,454	10,243	S	S	S	41,847	S	-	S	S	S	84C
S	13,448	7,736	7,736	7,681	55	-	-	S	S	S	84D
S	27,704	7,037	7,037	1,153	5,884	-	-	S	S	S	84E
182,927	S	141,108	141,108	118,811	S	S	-	P	P	78	84F
38,855	33,091	5,360	5,360	P	S	X	-	404	293	111	84G
713,711	76,450	598,401	598,401	519,135	X	X	-	38,860	31,601	7,260	84J

A表 産業(小分類), 本所・支所(3区分), 経営組織(7区分)

産業小分類	事業所数										
	総数	個人	会社	国内の会社				外国の会社	その他	会社以外の法人	法人以外の団体
				株式会社	有限会社	合名・合資・相互会社	その他				
85 協同組合(他に分類されないもの)	28,434	-	-	-	-	-	-	-	28,434	28,434	-
851 農林水産業協同組合	19,430	-	-	-	-	-	-	-	19,430	19,430	-
852 事業協同組合(他に分類されないもの)	9,004	-	-	-	-	-	-	-	9,004	9,004	-
86 その他の事業サービス業(869を除く)	34,962	3,166	29,349	29,335	21,428	7,822	84	14	2,447	2,089	358
861 速記・筆耕・複写業	3,619	731	2,866	2,866	1,757	1,104	5	-	21	21	-
862 商品検査業	1,657	98	906	894	691	203	-	13	653	644	9
863 計量証明業	873	70	737	737	673	63	-	-	67	67	-
864 建物サービス業	17,828	1,284	16,322	16,321	11,754	4,496	71	1	222	185	37
865 民営職業紹介業	4,489	842	2,194	2,194	1,088	1,103	2	-	1,453	1,142	311
866 警備業	6,496	142	6,324	6,324	5,466	853	5	-	30	29	1
87 廃棄物処理業	12,119	1,468	10,316	10,316	5,237	5,045	34	-	334	306	28
871 一般廃棄物処理業	7,863	1,284	6,319	6,319	2,650	3,646	23	-	259	231	28
872 産業廃棄物処理業	4,166	178	3,914	3,914	2,530	1,373	12	-	73	73	-
879 その他の廃棄物処理業	90	5	83	83	57	26	-	-	2	2	-
88, 89 医療業(病院を除く)	69,637	57,474	8,023	8,021	2,755	5,266	-	1	4,141	3,992	149
885 療術業	68,480	57,382	7,630	7,628	2,421	5,207	-	1	3,468	3,361	107
886 歯科技工所	58,994	53,740	4,900	4,900	1,167	3,733	-	-	354	275	79
887 医療に附帯するサービス業(別掲を除く)	5,110	3,331	1,756	1,756	398	1,358	-	-	24	24	-
889 その他の医療業	1,138	29	875	873	803	71	-	1	235	235	-
89 保健衛生	3,238	283	99	99	54	45	-	-	2,856	2,828	28
892 健康相談施設	1,157	91	393	393	334	59	-	-	673	631	41
899 その他の保健衛生	684	79	79	79	49	31	-	-	526	510	16
90 社会保険, 社会福祉	473	12	313	313	285	28	-	-	147	121	26
90A 保育所	36,832	3,438	1,500	1,500	899	600	-	-	31,894	28,689	3,205
90B その他の児童福祉事業	2,148	-	-	-	-	-	-	-	2,148	2,148	-
904 老人福祉事業	17,261	3,278	833	833	440	393	-	-	13,150	12,050	1,100
905 知的障害・身体障害者福祉事業	14,368	3,167	831	831	440	391	-	-	10,369	9,815	554
909 その他の社会保険, 社会福祉	2,893	110	2	2	-	2	-	-	2,781	2,235	546
91, 92 教育(学校を除く)	8,089	104	657	657	450	208	-	-	7,327	6,948	379
918 社会教育	5,197	38	10	10	10	-	-	-	5,149	3,666	1,484
91C 博物館, 美術館	4,138	18	-	-	-	-	-	-	4,119	3,877	242
91D 動物園, 植物園, 水族館	10,588	493	5,878	5,870	5,061	753	56	8	4,218	3,261	957
91E その他の社会教育	7,866	461	4,200	4,200	3,437	708	56	-	3,205	2,403	802
919 学術研究機関	3,406	224	860	860	735	125	-	-	2,321	1,600	721
92 自然科学研究所	1,634	158	599	599	502	97	-	-	876	732	145
922 人文・社会科学研究所	219	34	92	92	68	24	-	-	93	91	2
93 宗教	1,553	33	169	169	165	4	-	-	1,352	778	574
931 神道系宗教	4,461	236	3,340	3,340	2,701	583	56	-	884	803	81
932 仏教系宗教	2,722	32	1,678	1,670	1,624	45	-	8	1,012	858	154
933 キリスト教系宗教	2,394	26	1,646	1,641	1,604	37	-	6	722	618	104
939 その他の宗教	327	6	31	29	20	9	-	2	290	240	50
94 政治・経済・文化団体	94,747	868	-	-	-	-	-	-	93,878	92,445	1,434
941 経済団体	11,255	158	-	-	-	-	-	-	11,097	10,889	209
942 労働団体	63,765	552	-	-	-	-	-	-	63,213	62,962	287
943 学術・文化団体	6,263	26	-	-	-	-	-	-	6,238	5,641	597
944 政治団体	13,463	133	-	-	-	-	-	-	13,330	12,989	341
949 他に分類されない非営利的団体	37,586	-	-	-	-	-	-	-	37,586	21,849	15,737
95 その他のサービス業	15,610	-	-	-	-	-	-	-	15,610	12,021	3,589
951 集会場	5,116	-	-	-	-	-	-	-	5,116	2,198	2,919
952 と畜場	808	-	-	-	-	-	-	-	808	646	162
959 他に分類されないサービス業	750	-	-	-	-	-	-	-	750	167	583
	15,302	-	-	-	-	-	-	-	15,302	6,816	8,485
	3,850	60	586	586	542	44	-	-	3,205	1,611	1,594
	3,266	52	401	401	365	36	-	-	2,813	1,378	1,435
	89	-	46	46	43	3	-	-	43	38	5
	496	8	139	139	135	4	-	-	349	195	154

別事業所数・収入金額

総数	収入金額(百万円)											
	個人	会社	国内の会社				外国の会社	その他				
			株式会社	有限会社	合名・合資・相互会社	会社以外の法人		法人以外の団体				
S	-	-	-	-	-	-	-	S	S	-	85	
S	-	-	-	-	-	-	-	S	S	-	851	
1,720,766	-	-	-	-	-	-	-	1,720,766	1,720,766	-	852	
8,218,126	52,821	S	S	S	512,159	4,305	3,660	S	S	S	86	
410,467	11,186	394,702	394,702	329,992	64,684	26	-	4,579	4,579	-	861	
322,872	1,711	S	S	S	16,166	-	S	S	S	P	862	
260,153	1,128	224,688	224,688	222,057	2,630	-	-	34,337	34,337	-	863	
S	25,536	S	S	S	308,874	S	P	44,740	X	X	864	
478,627	8,604	240,867	240,867	S	48,536	P	-	229,155	221,116	8,039	865	
S	4,655	S	S	1,797,155	71,269	P	-	13,190	S	P	866	
2,744,275	53,403	S	S	S	592,979	7,267	-	S	S	S	87	
1,364,072	45,925	1,217,109	1,217,109	811,563	399,870	5,676	-	101,038	99,782	1,256	871	
1,368,009	S	S	S	S	190,441	1,591	-	S	S	-	872	
12,194	S	12,136	12,136	9,467	2,669	-	-	P	P	-	879	
S	S	S	S	S	210,724	-	-	P	S	S	9,642	88, 89
S	S	S	S	S	208,215	-	-	P	S	S	S	88
S	S	238,913	238,913	X	X	-	-	11,735	S	S	S	885
136,539	41,071	91,968	91,968	44,490	47,478	-	-	3,500	3,500	-	886	
S	2,110	S	S	S	X	-	-	P	54,827	54,827	-	887
S	1,273	8,397	8,397	6,172	2,224	-	-	S	S	S	S	889
396,754	S	S	S	S	2,509	-	-	S	S	S	S	89
319,572	S	S	S	S	1,717	-	-	296,998	S	S	P	892
77,182	90	S	S	S	792	-	-	S	S	S	812	899
S	58,356	144,182	144,182	117,602	26,580	-	-	S	S	S	66,271	90
S	-	-	-	-	-	-	-	S	S	S	-	901
1,367,502	57,136	S	S	S	18,615	-	-	S	S	S	16,691	903
1,042,446	X	S	S	S	S	-	-	S	S	S	11,333	90A
325,056	X	P	P	-	P	-	-	S	S	S	5,358	90B
1,526,246	726	108,076	108,076	100,111	7,965	-	-	1,417,445	1,407,346	10,099	904	
734,536	243	S	S	S	-	-	-	S	S	S	32,705	905
653,870	252	-	-	-	-	-	-	653,618	646,842	6,776	909	
S	11,545	S	S	S	S	5,597	P	S	S	S	88,680	91, 92
S	11,167	S	S	S	93,066	5,597	-	S	S	S	26,300	91
S	4,148	S	S	S	10,673	-	-	S	S	S	21,093	918
S	1,173	55,236	55,236	S	S	-	-	S	S	X	-	91C
S	2,807	S	S	S	5,605	-	-	25,608	S	P	-	91D
S	167	S	S	S	P	-	-	84,611	S	S	-	91E
S	7,019	S	S	S	82,392	5,597	-	143,377	138,171	5,206	919	
S	378	S	S	S	P	-	-	S	S	S	62,380	92
S	340	S	S	S	P	-	-	P	S	S	58,931	921
83,742	39	6,048	S	4,541	P	-	-	P	77,655	74,205	3,449	922
1,874,488	5,502	-	-	-	-	-	-	1,868,987	1,855,488	13,499	93	
258,609	890	-	-	-	-	-	-	257,719	256,365	1,354	931	
1,129,889	2,719	-	-	-	-	-	-	1,127,170	1,122,556	4,614	932	
114,944	124	-	-	-	-	-	-	114,819	110,420	4,399	933	
371,047	1,768	-	-	-	-	-	-	369,279	366,146	3,132	939	
4,957,129	-	-	-	-	-	-	-	4,957,129	4,088,963	868,166	94	
1,853,781	-	-	-	-	-	-	-	1,853,781	1,625,916	227,865	941	
483,296	-	-	-	-	-	-	-	483,296	267,808	215,488	942	
226,198	-	-	-	-	-	-	-	226,198	211,274	14,924	943	
128,370	-	-	-	-	-	-	-	128,370	99,989	28,381	944	
2,265,485	-	-	-	-	-	-	-	2,265,485	1,883,976	381,509	949	
365,037	S	86,713	86,713	S	S	-	-	S	S	S	S	95
295,349	S	51,990	51,990	S	S	-	-	S	S	S	19,323	951
24,049	-	17,538	17,538	S	P	-	-	6,511	S	S	S	952
45,639	126	17,185	17,185	S	P	-	-	28,329	25,082	3,246	959	

A表 産業(小分類), 本所・支所(3区分), 経営組織(7区分)

産業小分類	事業所数										
	総数	個人	会社	国内の会社				外国の会社	その他	会社以外の法人	法人以外の団体
				株式会社	有限会社	合名・合資・相互会社	その他				
L サービス業	1,037,479	731,901	157,091	157,091	54,454	100,519	2,118	-	148,487	128,533	19,954
72 洗濯・理容・浴場業	344,258	320,598	23,097	23,097	4,106	18,546	446	-	562	392	170
721 洗濯業	60,351	53,156	7,070	7,070	1,786	5,049	235	-	125	125	-
72A 普通洗濯業	58,346	52,592	5,633	5,633	1,198	4,214	221	-	121	121	-
72B リネンサプライ業	2,005	565	1,436	1,436	588	835	13	-	4	4	-
722 洗張・染物業	2,403	2,218	186	186	17	168	-	-	-	-	-
723 理容業	118,697	115,658	2,865	2,865	159	2,642	64	-	174	139	35
724 美容業	151,915	142,053	9,851	9,851	1,074	8,668	109	-	10	10	-
725 公衆浴場業	6,363	4,670	1,582	1,582	326	1,228	28	-	110	28	82
726 特殊浴場業	1,911	451	1,317	1,317	556	752	9	-	143	89	53
729 その他の洗濯・理容・浴場業	2,617	2,392	226	226	188	38	-	-	-	-	-
73 駐車場業	32,219	27,411	4,079	4,079	1,721	2,151	206	-	728	489	240
74 その他の生活関連サービス業	40,457	30,679	9,210	9,210	2,614	6,438	157	-	568	330	237
743 写真業	17,141	12,249	4,888	4,888	1,209	3,595	83	-	4	4	-
744 衣服裁縫修理業	7,949	7,551	397	397	118	245	34	-	-	-	-
745 物品預り業	2,557	2,282	157	157	-	157	-	-	118	87	31
746 火葬・墓地管理業	215	10	60	60	41	18	1	-	145	124	21
747 冠婚葬祭業	3,408	688	2,603	2,603	889	1,678	36	-	117	72	46
749 他に分類されない生活関連サービス業	9,188	7,900	1,106	1,106	357	745	3	-	183	43	139
75 旅館, その他の宿泊所	55,276	33,652	19,788	19,788	7,676	11,536	576	-	1,837	1,421	415
751 旅館	48,871	29,336	19,089	19,089	7,401	11,112	576	-	446	375	71
752 簡易宿所	1,135	855	245	245	72	174	-	-	35	27	8
753 下宿業	2,585	2,519	66	66	-	66	-	-	-	-	-
759 その他の宿泊所	2,685	941	388	388	203	185	-	-	1,356	1,019	336
75A 会社・団体の宿泊所	1,056	225	113	113	61	52	-	-	718	630	88
75B 他に分類されない宿泊所	1,629	716	275	275	142	133	-	-	638	390	248
76 娯楽業(映画・ビデオ制作業を除く)	35,742	22,396	12,187	12,187	5,550	6,569	69	-	1,159	703	456
761 映画館	284	73	207	207	109	85	13	-	4	-	4
762 劇場, 興行場(別掲を除く)	256	85	104	104	66	38	-	-	67	44	22
763 興行団	1,211	197	943	943	599	342	3	-	70	27	44
764 競輪・競馬等の競走場	3	-	3	3	3	-	-	-	-	-	-
765 競輪・競馬等の競技団	857	713	5	5	-	5	-	-	139	17	122
766 スポーツ施設提供業(76D, 76Fを除く)	2,685	567	1,599	1,599	1,247	346	5	-	520	404	115
76A スポーツ施設提供業(別掲を除く)	756	281	220	220	85	136	-	-	255	220	34
76B 体育館	203	18	-	-	-	-	-	-	185	135	50
76C ゴルフ場	1,081	33	968	968	936	31	2	-	80	48	31
76E ボウリング場	220	-	220	220	181	36	3	-	-	-	-
76G バンテイング・テニス練習場	425	235	190	190	46	144	-	-	-	-	-
767 公園, 遊園地	328	26	147	147	101	46	-	-	155	112	43
768 遊戯場	20,065	14,171	5,853	5,853	2,237	3,575	41	-	40	12	28
76K マージャンクラブ	11,902	10,897	1,005	1,005	259	747	-	-	-	-	-
76L パチンコホール	4,751	824	3,926	3,926	1,531	2,359	35	-	1	-	1
76M その他の遊戯場	3,412	2,451	922	922	447	469	6	-	39	12	27
769 その他の娯楽業	10,053	6,563	3,325	3,325	1,188	2,130	7	-	165	87	78
76N マリーナ業	207	53	125	125	98	26	-	-	29	19	10
76P 遊漁船業	1,343	1,181	156	156	41	114	-	-	6	-	6
76Q その他の娯楽業	8,504	5,329	3,045	3,045	1,049	1,990	7	-	129	68	62
77 自動車整備業	57,359	36,354	20,490	20,490	4,994	15,195	300	-	516	503	13
78 機械・家具等修理業(別掲を除く)	16,840	11,113	5,692	5,692	1,685	3,952	55	-	35	27	7
781 機械修理業	8,686	3,887	4,793	4,793	1,427	3,322	44	-	6	6	-
784 表具業	3,900	3,682	218	218	69	149	-	-	-	-	-
789 他に分類されない修理業	4,253	3,543	681	681	189	481	11	-	29	21	7
79 物品賃貸業(791, 792, 793を除く)	8,269	4,282	3,937	3,937	1,357	2,563	16	-	50	45	6
794 自動車賃貸業	989	252	737	737	175	555	7	-	-	-	-
795 スポーツ・娯楽用品賃貸業	542	362	149	149	30	120	-	-	31	31	-
799 その他の物品賃貸業	6,738	3,668	3,051	3,051	1,153	1,889	9	-	20	14	6
79A 音楽・映像記録物賃貸業	3,182	2,028	1,154	1,154	458	696	-	-	-	-	-
79B その他の物品賃貸業	3,556	1,639	1,897	1,897	695	1,193	9	-	20	14	6
80, 81	2,293	212	2,038	2,038	1,446	591	1	-	43	22	21
80 映画・ビデオ制作業	1,964	204	1,752	1,752	1,172	578	1	-	8	8	-
801 映画, ビデオ制作・配給業	1,800	200	1,592	1,592	1,092	498	1	-	8	8	-
802 映画・ビデオサービス業	164	5	160	160	80	80	-	-	-	-	-
81 放送業	329	8	287	287	274	12	-	-	35	14	21
811 公共放送業(有線放送業を除く)	-	-	-	-	-	-	-	-	-	-	-
812 民間放送業(有線放送業を除く)	71	-	69	69	69	-	-	-	2	2	-
813 有線放送業	258	8	217	217	205	12	-	-	33	12	21
82 情報サービス・調査業(821, 822を除く)	374	170	202	202	88	107	7	-	2	2	-
823 ニュース供給業	33	8	23	23	12	11	-	-	2	2	-
824 興信所	341	162	179	179	76	96	7	-	-	-	-
84 専門サービス業(他に分類されないもの, 849を除く)	214,983	179,359	34,556	34,556	14,562	19,813	181	-	1,068	515	554
841 法律事務所, 特許事務所	11,453	11,453	-	-	-	-	-	-	-	-	-
84K 法律事務所	9,505	9,505	-	-	-	-	-	-	-	-	-
84L 特許事務所	1,948	1,948	-	-	-	-	-	-	-	-	-
842 公証人役場, 司法書士事務所	12,947	12,947	-	-	-	-	-	-	-	-	-
843 公認会計士事務所, 税理士事務所	32,718	32,616	-	-	-	-	-	-	102	102	-
84M 公認会計士事務所	3,369	3,268	-	-	-	-	-	-	102	102	-
84N 税理士事務所	29,349	29,349	-	-	-	-	-	-	-	-	-
844 獣医業	7,037	5,345	1,692	1,692	115	1,577	-	-	-	-	-
845 土木建築サービス業	38,022	14,908	22,738	22,738	10,583	12,077	77	-	376	291	85
846 デザイン業	8,003	3,576	4,420	4,420	1,892	2,525	3	-	7	-	7
847 著述家・芸術家業	1,250	1,250	-	-	-	-	-	-	-	-	-
848 個人教授所	103,552	97,262	5,707	5,707	1,972	3,633	102	-	583	122	462
84A 学習塾(各種学校でないもの)	30,724	28,264	2,460	2,460	586	1,874	-	-	-	-	-
84B フィットネスクラブ	413	200	204	204	155	50	-	-	9	7	2
84C スポーツ・健康個人教授所	2,533	1,554	827	827	452	359	16	-	153	44	109
84D 生花・茶道個人教授所	10,974	10,908	52	52	52	-	-	-	14	-	14
84E そろばん個人教授所	9,517	9,383	70	70	-	70	-	-	64	-	64
84F 音楽個人教授所	17,329	16,924	352	352	88	265	-	-	53	-	53
84G 書道個人教授所	14,471	14,196	207	207	1	129	77	-	68	-	68
84J その他の個人教授所	17,592	15,832	1,535	1,535	639	887	9	-	224	70	154

別事業所数・収入金額

総数	収入金額(百万円)										
	個人	会社	国内の会社				外国の会社	その他		法人以外の団体	
			株式会社	有限会社	合名・合資・相互会社	会社以外の法人					
S	8,528,817	S	S	S	8,043,473	210,558	-	S	S	727,629	L
3,253,635	2,009,918	1,219,676	1,219,676	535,462	673,061	11,153	-	24,041	20,535	3,506	72
865,126	S	S	S	S	S	7,894	-	S	S	-	721
S	S	S	S	S	S	S	-	S	S	-	72A
S	15,638	S	S	S	78,742	S	-	36	36	-	72B
18,439	14,557	3,882	3,882	122	3,760	-	-	-	-	-	722
758,313	676,628	80,802	80,802	3,923	75,721	1,158	-	883	409	474	723
1,286,763	S	S	S	122,657	S	1,324	-	115	115	-	724
115,025	58,698	53,501	53,501	18,358	34,749	393	-	2,826	2,627	199	725
197,388	S	S	S	S	63,461	384	-	S	S	2,833	726
12,583	9,521	3,062	3,062	2,817	245	-	-	-	-	-	729
272,120	93,514	151,264	151,264	107,219	41,801	2,244	-	27,341	21,151	6,190	73
1,188,373	296,775	S	S	433,955	367,603	S	-	S	S	S	74
436,079	191,294	244,500	244,500	101,920	139,539	3,042	-	284	284	-	743
48,406	23,951	24,455	24,455	8,415	X	X	-	-	-	-	744
11,730	S	S	S	-	S	-	-	S	S	-	745
19,208	P	S	S	S	S	P	-	S	S	S	746
S	26,370	S	S	S	S	P	-	S	S	S	747
S	48,057	S	S	S	S	P	-	1,748	804	944	749
S	464,064	S	S	S	1,029,544	80,787	-	260,599	S	S	75
S	422,435	S	S	S	1,005,163	80,787	-	96,514	S	S	751
34,567	11,621	21,029	21,029	5,110	15,919	-	-	1,916	1,715	201	752
18,369	15,831	2,538	2,538	-	2,538	-	-	-	-	-	753
200,099	14,176	23,755	23,755	17,831	5,923	-	-	162,168	S	S	759
150,899	5,239	10,008	10,008	S	S	-	-	135,652	S	S	75A
49,200	8,937	13,746	13,746	S	S	-	-	26,517	X	X	75B
S	881,579	S	S	S	3,093,278	45,894	-	S	S	56,879	76
23,149	S	S	S	S	X	461	-	P	-	-	761
X	S	S	S	S	X	-	-	S	S	X	762
300,668	S	S	S	S	41,611	P	-	S	S	S	763
P	-	P	P	P	-	-	-	-	-	-	764
X	58,832	313	313	-	313	-	-	X	S	S	765
S	S	S	S	627,620	S	S	-	S	S	S	766
59,600	S	S	S	S	5,849	-	-	S	S	701	76A
S	132	X	X	X	-	-	-	S	S	2,422	76B
S	S	S	S	S	S	P	-	S	S	S	76C
X	-	X	X	X	S	P	-	-	-	-	76E
15,380	3,869	11,511	11,511	3,807	7,704	-	-	-	-	-	76G
104,400	320	79,247	79,247	77,950	1,296	-	-	24,833	S	S	767
S	S	S	S	S	S	S	-	S	S	S	768
110,292	86,336	23,956	23,956	8,816	15,140	-	-	-	-	-	76K
S	X	S	S	S	S	S	-	P	-	P	76L
122,753	S	93,694	93,694	60,720	32,756	217	-	S	S	S	76M
356,832	S	255,082	255,082	S	104,214	P	-	S	S	3,081	769
S	536	7,509	7,509	6,673	836	-	-	S	S	478	76N
S	10,191	7,057	7,057	3,706	3,351	-	-	S	S	S	76P
X	S	240,516	240,516	S	100,027	P	-	S	S	S	76Q
2,377,980	X	1,578,916	1,578,916	680,393	878,175	20,348	-	X	X	407	77
556,768	122,316	S	S	S	S	S	-	S	S	-	78
475,693	71,600	403,887	403,887	S	S	S	-	207	207	-	781
29,764	26,723	3,042	3,042	1,569	1,473	-	-	-	-	-	784
51,310	23,993	S	S	9,879	S	S	-	S	S	S	789
352,392	S	S	S	S	107,688	S	-	S	S	S	79
67,148	S	S	S	52,425	12,202	S	-	-	-	-	794
9,171	2,918	5,679	5,679	1,877	3,802	-	-	574	574	-	795
276,074	S	S	S	S	91,685	223	-	S	S	S	799
99,301	33,576	65,725	65,725	28,668	37,057	-	-	-	-	-	79A
176,772	S	S	S	S	54,627	223	-	S	S	S	79B
499,999	S	S	S	S	63,226	P	-	S	S	696	80, 81
S	S	S	S	S	S	P	-	P	P	-	80
S	S	S	S	S	S	P	-	P	P	-	801
S	P	S	S	S	7,033	-	-	-	-	-	802
S	32	S	S	S	P	-	-	P	P	696	81
-	-	-	-	-	-	-	-	-	-	-	811
39,575	-	S	S	S	-	-	-	P	P	-	812
S	32	S	S	S	P	-	-	X	X	696	813
20,378	S	S	S	S	X	X	-	-	-	-	82
12,765	P	S	S	S	565	-	-	P	P	-	823
7,614	X	6,221	6,221	3,316	2,892	14	-	-	-	-	824
5,605,736	3,212,053	2,150,008	2,150,008	1,385,321	760,424	4,264	-	243,674	211,619	32,055	84
808,728	808,728	-	-	-	-	-	-	-	-	-	841
474,839	474,839	-	-	-	-	-	-	-	-	-	84K
333,888	333,888	-	-	-	-	-	-	-	-	-	84L
309,843	309,843	-	-	-	-	-	-	-	-	-	842
1,391,239	1,365,983	-	-	-	-	-	-	25,256	25,256	-	843
203,782	178,526	-	-	-	-	-	-	25,256	25,256	-	84M
1,187,457	1,187,457	-	-	-	-	-	-	-	-	-	84N
199,938	101,674	98,264	98,264	7,980	90,284	-	-	-	-	-	844
S	S	S	S	S	440,306	S	-	S	S	S	845
313,691	51,324	261,987	261,987	S	99,172	P	-	380	-	380	846
10,788	10,788	-	-	-	-	-	-	-	-	-	847
S	S	S	S	S	X	X	-	S	S	S	848
S	S	91,229	91,229	38,058	53,172	-	-	-	-	-	84A
44,381	1,407	X	X	37,851	X	-	-	X	S	P	84B
100,496	S	S	S	S	S	X	-	S	3,741	S	84C
19,084	12,559	6,477	6,477	6,477	-	-	-	48	-	48	84D
25,005	21,454	S	S	-	S	-	-	-	-	-	84E
51,056	38,251	12,728	12,728	5,602	7,126	-	-	78	-	78	84F
35,118	S	S	S	P	S	X	-	17	-	17	84G
158,942	60,339	S	S	S	39,461	X	-	S	S	S	84J

A表 産業(小分類), 本所・支所(3区分), 経営組織(7区分)

産業小分類	事業所数										
	総数	個人	会社	国内の会社				外国の会社	その他	会社以外の法人	法人以外の団体
				社	株式会社	有限会社	合名・合資・相互会社				
85 協同組合(他に分類されないもの)	9,822	-	-	-	-	-	-	-	9,822	9,822	-
851 農林水産業協同組合	2,665	-	-	-	-	-	-	-	2,665	2,665	-
852 事業協同組合(他に分類されないもの)	7,157	-	-	-	-	-	-	-	7,157	7,157	-
86 その他の事業サービス業(869を除く)	14,313	3,106	10,043	10,043	4,540	5,438	66	-	1,163	840	323
861 速記・筆耕・複写業	1,975	720	1,253	1,253	521	727	5	-	2	2	-
862 商品検査業	568	96	429	429	313	116	-	-	43	37	6
863 計量証明業	276	67	188	188	137	51	-	-	21	21	-
864 建物サービス業	6,825	1,266	5,463	5,463	2,253	3,154	57	-	96	71	26
865 民営職業紹介業	3,124	834	1,298	1,298	369	927	2	-	992	702	290
866 警備業	1,544	124	1,412	1,412	947	463	1	-	8	7	1
87 廃棄物処理業	6,763	143	5,135	5,135	1,775	3,338	21	-	195	175	21
871 一般廃棄物処理業	4,980	1,257	3,558	3,558	1,009	2,537	12	-	165	144	21
872 産業廃棄物処理業	1,712	171	1,512	1,512	728	775	9	-	29	29	-
879 その他の廃棄物処理業	71	5	64	64	38	26	-	-	2	2	-
88, 89 医療業(病院を除く)	62,916	56,472	4,256	4,256	872	3,384	-	-	2,188	2,083	105
885 療術業	62,446	56,391	4,148	4,148	802	3,345	-	-	1,908	1,827	81
886 歯科技工所	55,673	52,805	2,718	2,718	584	2,134	-	-	150	97	54
887 医療に附帯するサービス業(別掲を除く)	4,531	3,274	1,257	1,257	116	1,142	-	-	-	-	-
889 その他の医療業	216	29	126	126	66	59	-	-	62	62	-
89 保健衛生	2,026	283	47	47	36	11	-	-	1,696	1,668	27
892 健康相談施設	470	81	108	108	70	39	-	-	281	257	24
899 その他の保健衛生	335	69	40	40	25	15	-	-	226	217	9
90 社会保険, 社会福祉	135	12	68	68	44	24	-	-	55	40	15
901 社会保険事業団体	26,837	3,319	584	584	220	364	-	-	22,935	19,980	2,955
903 児童福祉事業	1,404	-	-	-	-	-	-	-	1,404	1,404	-
90A 保育所	13,538	3,158	333	333	100	233	-	-	10,047	9,054	993
90B その他の児童福祉事業	11,727	3,048	331	331	100	231	-	-	8,348	7,811	537
904 老人福祉事業	1,811	110	2	2	-	2	-	-	1,699	1,243	456
905 知的障害・身体障害者福祉事業	5,337	104	243	243	112	131	-	-	4,990	4,644	346
909 その他の社会保険, 社会福祉	3,487	38	8	8	8	-	-	-	3,440	2,037	1,404
91, 92 教育(学校を除く)	3,072	18	-	-	-	-	-	-	3,053	2,841	212
91 教育	4,317	470	1,583	1,583	1,067	499	17	-	2,265	1,458	807
918 社会教育	3,715	441	1,436	1,436	948	470	17	-	1,839	1,109	729
91C 博物館, 美術館	1,889	218	238	238	137	101	-	-	1,433	763	671
91D 動物園, 植物園, 水族館	892	152	127	127	55	73	-	-	612	479	133
91E その他の社会教育	115	34	51	51	27	24	-	-	30	28	2
919 その他の教育施設	883	33	59	59	56	4	-	-	791	255	536
92 学術研究機関	1,826	223	1,198	1,198	811	370	17	-	406	347	59
921 自然科学研究所	602	29	147	147	119	28	-	-	426	348	78
922 人文・社会科学研究所	355	22	129	129	105	24	-	-	203	167	36
93 宗教	247	6	18	18	13	4	-	-	223	181	42
931 神道系宗教	77,599	819	-	-	-	-	-	-	76,780	75,914	866
932 仏教系宗教	9,431	144	-	-	-	-	-	-	9,287	9,141	147
933 キリスト教系宗教	58,480	517	-	-	-	-	-	-	57,963	57,790	173
939 その他の宗教	3,266	26	-	-	-	-	-	-	3,241	2,857	384
94 政治・経済・文化団体	6,422	133	-	-	-	-	-	-	6,289	6,126	163
941 経済団体	24,383	-	-	-	-	-	-	-	24,383	13,025	11,358
942 労働団体	11,199	-	-	-	-	-	-	-	11,199	8,352	2,847
943 学術・文化団体	1,878	-	-	-	-	-	-	-	1,878	645	1,232
944 政治団体	443	-	-	-	-	-	-	-	443	310	133
949 他に分類されない非営利的団体	90	-	-	-	-	-	-	-	90	4	86
95 その他のサービス業	10,774	-	-	-	-	-	-	-	10,774	3,714	7,059
951 集会場	2,459	56	216	216	182	34	-	-	2,187	788	1,399
952 と畜場	2,117	52	135	135	104	31	-	-	1,930	651	1,279
959 他に分類されないサービス業	57	-	31	31	31	-	-	-	26	23	3
	286	4	50	50	47	3	-	-	232	114	117

別事業所数・収入金額

総数	収入金額(百万円)										
	個人	会社	国内の会社				外国の会社	その他			
			株式会社	有限会社	合名・合資・相互会社	会社以外の法人		法人以外の団体			
S	-	-	-	-	-	-	S	S	-	85	
S	-	-	-	-	-	-	S	S	-	851	
1,050,539	-	-	-	-	-	-	1,050,539	1,050,539	-	852	
1,404,306	S	S	S	S	319,409	S	S	S	S	86	
121,289	S	S	S	S	40,688	26	P	P	-	861	
51,690	S	S	S	S	X	-	S	2,814	P	862	
47,903	S	29,771	29,771	27,799	1,972	-	S	S	-	863	
750,170	S	S	S	S	197,624	S	-	X	S	864	
213,649	S	66,398	66,398	S	36,954	P	-	S	7,525	865	
219,604	S	214,708	214,708	S	X	P	-	S	P	866	
907,697	51,138	783,200	783,200	459,121	321,473	2,606	-	73,360	S	87	
562,881	S	S	S	233,785	232,100	S	-	S	S	871	
338,850	S	S	S	222,097	86,704	S	-	S	S	872	
5,966	S	5,908	5,908	3,239	2,669	-	-	P	P	879	
S	S	S	S	S	109,669	-	-	S	S	88, 89	
S	S	S	S	S	107,721	-	-	S	S	88	
S	S	S	S	S	63,237	-	-	1,862	X	X	885
85,257	37,544	47,713	47,713	X	X	-	-	-	-	886	
56,894	2,110	37,050	37,050	32,148	4,902	-	-	17,734	17,734	-	887
S	1,273	3,402	3,402	X	X	-	-	S	S	-	889
S	S	S	S	S	1,948	-	-	S	S	S	89
S	S	S	S	S	1,278	-	-	S	S	P	892
21,134	90	7,528	7,528	6,858	670	-	-	13,516	12,786	730	899
S	55,427	S	S	S	15,875	-	-	S	S	58,082	90
S	-	-	-	-	-	-	-	S	S	-	901
S	54,207	S	S	S	11,543	-	-	S	S	15,516	903
824,267	X	S	S	S	S	-	-	S	S	11,161	90A
S	X	P	P	-	P	-	-	S	S	4,356	90B
1,006,117	726	S	S	S	4,332	-	-	S	S	S	904
S	243	S	S	S	-	-	-	S	S	X	905
351,720	252	-	-	-	-	-	-	351,468	S	S	909
S	9,684	S	S	S	S	S	-	S	S	29,592	91, 92
S	S	S	S	S	S	S	-	S	S	24,269	91
S	S	S	S	S	9,600	-	-	S	S	20,930	918
S	S	13,898	13,898	S	S	-	-	S	S	X	91C
S	2,807	S	S	S	5,605	-	-	8,711	S	P	91D
33,404	167	6,511	6,511	S	P	-	-	26,726	S	S	91E
S	S	S	S	S	S	S	-	S	S	3,339	919
276,123	S	70,359	70,359	69,645	715	-	-	S	S	5,323	92
226,597	S	S	S	S	S	-	-	S	S	1,923	921
49,526	39	S	S	S	P	-	-	S	S	3,400	922
1,243,017	5,186	-	-	-	-	-	-	1,237,831	1,230,434	7,398	93
184,581	877	-	-	-	-	-	-	183,705	183,246	459	931
855,726	2,417	-	-	-	-	-	-	853,310	850,080	3,230	932
49,321	124	-	-	-	-	-	-	49,197	46,760	2,437	933
153,388	1,768	-	-	-	-	-	-	151,620	150,348	1,273	939
2,548,249	-	-	-	-	-	-	-	2,548,249	2,064,149	484,100	94
1,124,552	-	-	-	-	-	-	-	1,124,552	969,446	155,106	941
147,564	-	-	-	-	-	-	-	147,564	S	S	942
94,191	-	-	-	-	-	-	-	94,191	S	S	943
7,325	-	-	-	-	-	-	-	7,325	P	S	944
1,174,617	-	-	-	-	-	-	-	1,174,617	S	S	949
S	S	47,669	47,669	S	S	-	-	S	S	S	95
S	S	S	S	S	S	-	-	S	S	13,074	951
S	-	S	S	S	-	-	-	5,023	S	P	952
28,380	S	S	S	S	P	-	-	S	S	S	959

A表 産業(小分類), 本所・支所(3区分), 経営組織(7区分)

産業小分類	総数	事業所数									
		個人	会社	国内の会社				外国の会社	その他	会社以外の法人	法人以外の団体
				株式会社	有限会社	合名・合資・相互会社	その他				
L サービス業	95,996	6,481	71,636	71,636	44,541	26,633	462	-	17,878	16,268	1,610
72 洗濯・理容・浴場業	16,208	2,471	13,709	13,709	5,561	7,983	165	-	29	29	-
721 洗濯業	7,388	846	6,518	6,518	2,971	3,446	101	-	24	24	-
72A 普通洗濯業	6,537	842	5,674	5,674	2,385	3,196	93	-	21	21	-
72B リネンサプライ業	852	4	845	845	586	250	8	-	3	3	-
722 洗剤・染物業	26	-	26	26	2	24	-	-	-	-	-
723 理容業	1,944	396	1,548	1,548	551	997	-	-	-	-	-
724 美容業	6,263	1,229	5,035	5,035	1,855	3,117	63	-	-	-	-
725 公衆浴場業	226	-	226	226	36	189	-	-	-	-	-
726 特殊浴場業	184	-	179	179	98	81	-	-	5	5	-
729 その他の洗濯・理容・浴場業	177	-	177	177	47	130	-	-	-	-	-
73 駐車場業	1,367	156	1,156	1,156	787	349	20	-	54	38	16
74 その他の生活関連サービス業	5,179	326	4,795	4,795	2,854	1,918	22	-	59	52	7
743 写真業	2,230	233	1,997	1,997	1,175	809	13	-	-	-	-
744 衣服縫製修理業	438	-	438	438	239	198	-	-	-	-	-
745 物品預り業	80	-	77	77	58	18	-	-	3	3	-
746 火葬・墓地管理業	89	-	51	51	45	5	-	-	39	36	3
747 冠婚葬祭業	1,799	80	1,702	1,702	1,089	604	9	-	17	13	4
749 他に分類されない生活関連サービス業	544	13	530	530	247	283	-	-	-	-	-
75 旅館, その他の宿泊所	4,913	123	4,625	4,625	2,963	1,541	121	-	165	160	5
751 旅館	4,451	91	4,313	4,313	2,769	1,423	121	-	48	48	-
752 簡易宿所	114	12	95	95	27	68	-	-	7	7	-
753 下宿業	61	20	41	41	-	41	-	-	-	-	-
759 その他の宿泊所	287	-	176	176	168	9	-	-	110	105	5
75A 会社・団体の宿泊所	119	-	50	50	41	9	-	-	69	64	5
75B 他に分類されない宿泊所	168	-	127	127	127	-	-	-	41	41	-
76 娯楽業(映画・ビデオ制作業を除く)	5,811	187	5,408	5,408	3,477	1,911	19	-	216	200	16
761 映画館	99	-	99	99	79	20	-	-	-	-	-
762 劇場, 興行場(別掲を除く)	58	-	56	56	39	16	-	-	2	2	-
763 興行団	339	3	317	317	198	117	2	-	19	13	6
764 競輪・競馬等の競走場	6	-	6	6	6	-	-	-	-	-	-
765 競輪・競馬等の競技団	46	-	17	17	17	-	-	-	29	29	-
766 スポーツ施設提供業(76D, 76Fを除く)	957	8	834	834	807	26	1	-	115	110	5
76A スポーツ施設提供業(別掲を除く)	118	5	65	65	59	6	-	-	48	48	-
76B 体育館	61	-	2	2	2	-	-	-	59	59	-
76C ゴルフ場	646	3	636	636	628	7	-	-	8	3	5
76E ボウリング場	106	-	106	106	92	13	1	-	-	-	-
76G バンティング・テニス練習場	26	-	26	26	26	-	-	-	-	-	-
767 公園, 遊園地	116	-	65	65	52	11	1	-	51	46	5
768 遊戯場	3,356	97	3,258	3,258	1,896	1,352	10	-	-	-	-
76K マージャンクラブ	394	46	348	348	96	252	-	-	-	-	-
76L パチンコホール	2,424	27	2,396	2,396	1,443	944	10	-	-	-	-
76M その他の遊戯場	538	24	514	514	358	157	-	-	-	-	-
769 その他の娯楽業	835	79	755	755	382	368	6	-	-	-	-
76N マリーナ業	37	-	37	37	32	5	-	-	-	-	-
76P 遊漁船業	8	-	8	8	-	-	-	-	-	-	-
76Q その他の娯楽業	790	71	719	719	350	363	6	-	-	-	-
77 自動車整備業	5,241	240	4,899	4,899	2,303	2,550	46	-	101	101	-
78 機械・家具等修理業(別掲を除く)	2,731	79	2,643	2,643	1,863	778	2	-	10	10	-
781 機械修理業	2,441	34	2,400	2,400	1,702	696	2	-	7	7	-
784 表具業	62	34	29	29	29	-	-	-	-	-	-
789 他に分類されない修理業	228	11	215	215	133	82	-	-	2	2	-
79 物品賃貸業(791, 792, 793を除く)	2,560	128	2,426	2,426	1,656	767	2	-	6	6	-
794 自動車賃貸業	532	3	528	528	349	177	2	-	1	1	-
795 スポーツ・娯楽用品賃貸業	110	-	110	110	36	74	-	-	-	-	-
799 その他の物品賃貸業	1,918	126	1,788	1,788	1,272	516	-	-	5	5	-
79A 音楽・映像記録物賃貸業	768	93	675	675	449	226	-	-	-	-	-
79B その他の物品賃貸業	1,150	33	1,113	1,113	823	290	-	-	5	5	-
80, 81 映画・ビデオ制作業	1,598	5	1,585	1,585	1,374	207	4	-	8	8	-
801 映画, ビデオ制作・配給業	1,106	5	1,094	1,094	898	195	-	-	7	7	-
802 映画・ビデオサービスマン	1,052	5	1,039	1,039	868	171	-	-	7	7	-
81 放送業	492	-	491	491	476	12	4	-	1	1	-
811 公共放送業(有線放送業を除く)	1	-	-	-	-	-	-	-	1	1	-
812 民間放送業(有線放送業を除く)	310	-	310	310	310	-	-	-	-	-	-
813 有線放送業	181	-	181	181	166	12	4	-	-	-	-
82 情報サービス・調査業(821, 822を除く)	196	1	194	194	180	14	-	-	1	1	-
823 ニュース供給業	79	1	76	76	73	3	-	-	1	1	-
824 興信所	118	-	118	118	107	11	-	-	-	-	-
84 専門サービス業(他に分類されないもの, 849を除く)	18,705	2,262	16,121	16,121	11,334	4,763	24	-	323	245	78
841 法律事務所, 特許事務所	113	113	-	-	-	-	-	-	-	-	-
84K 法律事務所	29	29	-	-	-	-	-	-	-	-	-
84L 特許事務所	84	84	-	-	-	-	-	-	-	-	-
842 公証人役場, 司法書士事務所	1	1	-	-	-	-	-	-	-	-	-
843 公認会計士事務所, 税理士事務所	206	176	-	-	-	-	-	-	31	31	-
84M 公認会計士事務所	44	13	-	-	-	-	-	-	31	31	-
84N 税理士事務所	163	163	-	-	-	-	-	-	-	-	-
844 獣医業	255	7	248	248	51	197	-	-	-	-	-
845 土木建築サービス業	9,945	66	9,805	9,805	7,637	2,146	23	-	74	68	6
846 デザイン業	1,330	33	1,297	1,297	895	402	-	-	-	-	-
847 著述家・芸術家業	-	-	-	-	-	-	-	-	-	-	-
848 個人教授所	6,854	1,865	4,770	4,770	2,751	2,018	2	-	218	146	72
84A 学習塾(各種学校でないもの)	3,578	645	2,920	2,920	1,486	1,434	-	-	13	-	13
84B フィットネスクラブ	103	-	102	102	91	11	-	-	1	1	-
84C スポーツ・健康個人教授所	447	97	327	327	248	77	2	-	24	16	8
84D 生花・茶道個人教授所	182	83	40	40	-	40	-	-	59	7	52
84E そろばん個人教授所	556	392	78	78	75	3	-	-	86	86	-
84F 音楽個人教授所	343	100	243	243	129	114	-	-	-	-	-
84G 書道個人教授所	164	160	4	4	2	2	-	-	-	-	-
84J その他の個人教授所	1,481	389	1,056	1,056	719	337	-	-	35	35	-

別事業所数・収入金額

総数	収入金額(百万円)										
	個人	会社	国内の会社				外国の会社	その他			
			株式会社	有限会社	合名・合資・相互会社	会社以外の法人		法人以外の団体			
S	214,800	27,310,080	27,310,080	23,468,252	3,781,521	60,307	-	S	S	298,422	L
1,933,234	51,497	1,871,493	1,871,493	1,203,874	657,634	9,985	-	10,243	10,243	-	72
1,290,841	S	S	S	S	S	6,341	-	S	S	-	721
867,981	S	S	S	S	S	S	-	S	S	-	72A
422,860	P	S	S	S	45,374	P	-	P	P	-	72B
899	-	899	899	P	S	-	-	-	-	-	722
81,954	4,534	77,420	77,420	34,080	43,340	-	-	-	-	-	723
489,047	S	S	S	266,353	S	3,644	-	-	-	-	724
6,440	-	6,440	6,440	1,989	4,451	-	-	-	-	-	725
41,599	-	S	S	S	S	-	-	P	P	-	726
22,453	-	22,453	22,453	X	X	-	-	-	-	-	729
94,307	765	87,996	87,996	69,359	18,568	69	-	5,546	5,225	321	73
S	13,954	S	S	S	201,846	S	-	S	S	P	74
S	5,325	S	S	S	S	997	-	-	-	-	743
41,260	-	41,260	41,260	29,557	11,703	-	-	-	-	-	744
11,524	-	S	S	S	S	-	-	P	P	-	745
19,640	-	8,380	8,380	S	P	-	-	11,260	S	P	746
S	8,100	S	S	S	S	S	-	S	P	P	747
S	529	S	S	S	11,937	-	-	-	-	-	749
2,435,496	4,892	2,380,858	2,380,858	2,132,990	228,146	19,721	-	49,746	49,114	632	75
2,354,371	4,675	2,333,094	2,333,094	2,089,059	224,314	19,721	-	16,602	16,602	-	751
7,123	181	6,624	6,624	4,656	1,968	-	-	318	318	-	752
1,310	37	1,273	1,273	-	1,273	-	-	-	-	-	753
72,692	-	39,866	39,866	39,275	591	-	-	32,826	32,193	632	759
35,547	-	8,883	8,883	8,292	591	-	-	26,664	26,032	632	75A
37,144	-	30,983	30,983	30,983	-	-	-	6,161	6,161	-	75B
S	40,823	S	S	S	1,612,333	18,969	-	S	S	2,899	76
41,316	-	41,316	41,316	39,843	1,473	-	-	-	-	-	761
X	-	S	S	S	X	-	-	P	P	-	762
282,637	P	S	S	S	X	P	-	P	P	P	763
S	-	S	S	S	-	-	-	-	-	-	764
X	-	1,963	1,963	1,963	-	-	-	X	X	-	765
S	P	S	S	S	S	P	-	S	S	P	766
46,085	P	S	S	S	111	-	-	S	S	-	76A
S	-	P	P	P	-	-	-	S	S	-	76B
S	P	S	S	S	S	-	-	3,124	P	P	76C
X	-	X	X	X	S	P	-	-	-	-	76E
0	-	0	0	0	-	-	-	-	-	-	76G
S	-	P	P	P	S	P	-	S	S	S	767
4,760,473	39,581	4,720,892	4,720,892	S	1,553,921	S	-	-	-	-	768
30,489	1,342	29,146	29,146	6,120	23,027	-	-	-	-	-	76K
4,628,717	35,436	4,593,281	4,593,281	S	1,511,821	S	-	-	-	-	76L
101,267	2,803	98,465	98,465	79,390	19,074	-	-	-	-	-	76M
266,639	783	265,855	265,855	S	28,731	S	-	-	-	-	769
9,753	-	9,753	9,753	9,428	325	-	-	-	-	-	76N
29	29	-	-	-	-	-	-	-	-	-	76P
256,857	755	256,102	256,102	S	28,406	S	-	-	-	-	76Q
X	X	820,304	820,304	596,438	S	S	-	X	X	-	77
1,289,076	2,489	S	S	S	S	P	-	S	S	-	78
1,258,720	S	S	S	S	S	P	-	S	S	-	781
5,885	1,690	4,195	4,195	4,195	-	-	-	-	-	-	784
24,471	S	24,385	24,385	19,825	4,561	-	-	P	P	-	789
S	S	S	S	S	55,352	P	-	P	P	-	79
S	P	S	S	S	X	P	-	P	P	-	794
9,287	-	9,287	9,287	5,384	3,904	-	-	-	-	-	795
S	S	S	S	S	45,585	-	-	P	P	-	799
85,272	S	S	S	S	15,954	-	-	-	-	-	79A
S	S	S	S	S	29,630	-	-	P	P	-	79B
S	S	S	S	S	14,508	P	-	P	P	-	80, 81
S	S	S	S	S	S	-	-	S	S	-	80
S	S	S	S	S	S	-	-	S	S	-	801
S	S	S	S	S	1,449	-	-	-	-	-	802
S	-	S	S	S	S	P	-	P	P	-	81
P	-	-	-	-	-	-	-	P	P	-	811
S	-	S	S	S	-	-	-	-	-	-	812
S	-	S	S	S	S	P	-	-	-	-	813
S	P	S	S	S	X	-	-	P	P	-	82
S	P	S	S	S	P	-	-	P	P	-	823
17,932	-	17,932	17,932	S	S	-	-	-	-	-	824
3,945,731	76,410	3,690,825	3,690,825	3,420,763	268,344	1,717	-	178,496	176,294	2,202	84
25,781	25,781	-	-	-	-	-	-	-	-	-	841
2,046	2,046	-	-	-	-	-	-	-	-	-	84K
23,735	23,735	-	-	-	-	-	-	-	-	-	84L
P	P	-	-	-	-	-	-	-	-	-	842
106,095	12,443	-	-	-	-	-	-	93,651	93,651	-	843
93,994	343	-	-	-	-	-	-	93,651	93,651	-	84M
12,100	12,100	-	-	-	-	-	-	-	-	-	84N
S	S	21,691	21,691	2,580	19,111	-	-	-	-	-	844
S	1,642	S	S	S	102,613	S	-	65,621	S	P	845
249,955	845	249,110	249,110	222,723	26,387	-	-	-	-	-	846
-	-	-	-	-	-	-	-	-	-	-	847
S	35,170	S	S	S	X	P	-	19,223	S	S	848
S	14,173	S	S	S	S	-	-	374	-	374	84A
33,810	-	S	S	S	X	-	-	P	P	-	84B
58,688	S	S	S	S	S	P	-	S	S	50	84C
S	51	55	55	-	55	-	-	S	S	S	84D
S	2,801	S	S	1,134	P	-	-	S	S	-	84E
11,668	575	11,093	11,093	5,996	5,096	-	-	-	-	-	84F
2,386	S	P	P	P	P	-	-	-	-	-	84G
169,907	14,936	S	S	S	X	-	-	S	S	-	84J

A表 産業(小分類), 本所・支所(3区分), 経営組織(7区分)

産業小分類	総数	事業所数										
		個人	会社	国内の会社				外国の会社	その他	会社以外の法人	法人以外の団体	
				株式会社	有限会社	合名・合資・相互会社	その他					
85 協同組合(他に分類されないもの)	3,499	-	-	-	-	-	-	-	-	3,499	3,499	-
851 農林水産業協同組合	2,296	-	-	-	-	-	-	-	-	2,296	2,296	-
852 事業協同組合(他に分類されないもの)	1,203	-	-	-	-	-	-	-	-	1,203	1,203	-
86 その他の事業サービス業(869を除く)	8,750	33	8,427	8,427	6,659	1,752	16	-	-	289	273	16
861 速記・筆耕・複写業	747	9	738	738	509	229	-	-	-	1	1	-
862 商品検査業	261	-	209	209	154	55	-	-	-	52	51	1
863 計量証明業	289	-	283	283	270	12	-	-	-	6	6	-
864 建物サービス業	4,911	9	4,855	4,855	3,831	1,012	12	-	-	47	44	3
865 民営職業紹介業	654	4	472	472	353	119	-	-	-	179	167	12
866 警備業	1,887	12	1,871	1,871	1,542	325	4	-	-	4	4	-
87 廃棄物処理業	3,185	15	3,096	3,096	1,888	1,200	8	-	-	74	69	5
871 一般廃棄物処理業	1,928	11	1,860	1,860	995	860	5	-	-	57	52	5
872 産業廃棄物処理業	1,244	4	1,223	1,223	881	339	3	-	-	17	17	-
879 その他の廃棄物処理業	13	-	13	13	13	-	-	-	-	-	-	-
88, 89 医療業(病院を除く)	2,186	388	1,460	1,460	746	714	-	-	-	338	335	2
88 療術業	1,924	386	1,304	1,304	603	701	-	-	-	235	235	-
885 歯科技工所	1,225	329	839	839	277	562	-	-	-	58	58	-
886 医療に附帯するサービス業(別掲を除く)	341	57	261	261	165	96	-	-	-	24	24	-
887 医療に附帯するサービス業(別掲を除く)	182	-	168	168	160	8	-	-	-	14	14	-
889 その他の医療業	176	-	36	36	2	35	-	-	-	139	139	-
89 保健衛生	262	2	156	156	143	13	-	-	-	103	101	2
892 健康相談施設	103	2	22	22	14	9	-	-	-	78	76	2
899 その他の保健衛生	159	-	134	134	129	5	-	-	-	25	25	-
90 社会保険, 社会福祉	3,119	17	289	289	195	94	-	-	-	2,813	2,747	66
901 社会保険事業団体	355	-	-	-	-	-	-	-	-	355	355	-
903 児童福祉事業	1,041	17	135	135	73	62	-	-	-	889	867	21
90A 保育所	863	17	135	135	73	62	-	-	-	711	711	-
90B その他の児童福祉事業	177	-	-	-	-	-	-	-	-	177	156	21
904 老人福祉事業	750	-	153	153	121	32	-	-	-	597	588	9
905 知的障害・身体障害者福祉事業	475	-	1	1	1	-	-	-	-	474	452	22
909 その他の社会保険, 社会福祉	499	-	-	-	-	-	-	-	-	499	485	14
91, 92 教育(学校を除く)	1,226	13	744	744	644	87	12	-	-	470	444	26
91 社会教育	798	11	572	572	485	75	12	-	-	215	205	10
918 博物館, 美術館	256	3	90	90	84	6	-	-	-	163	157	5
91C 動物園, 植物園, 水族館	105	3	49	49	43	6	-	-	-	53	48	5
91D その他の社会教育	26	-	16	16	16	-	-	-	-	11	11	-
91E その他の教育施設	125	-	25	25	25	-	-	-	-	99	99	-
919 学術研究機関	543	8	482	482	400	69	12	-	-	52	47	5
92 自然科学研究所	428	2	171	171	159	12	-	-	-	255	239	15
921 人文・社会科学研究所	393	2	166	166	155	10	-	-	-	225	218	7
922 宗教	35	-	6	6	4	2	-	-	-	29	21	8
93 宗教	4,095	32	-	-	-	-	-	-	-	4,063	3,980	83
931 神道系宗教	857	14	-	-	-	-	-	-	-	843	826	17
932 仏教系宗教	2,324	18	-	-	-	-	-	-	-	2,306	2,240	66
933 キリスト教系宗教	552	-	-	-	-	-	-	-	-	552	552	-
939 その他の宗教	362	-	-	-	-	-	-	-	-	362	362	-
94 政治・経済・文化団体	5,083	-	-	-	-	-	-	-	-	5,083	3,822	1,261
941 経済団体	1,982	-	-	-	-	-	-	-	-	1,982	1,591	391
942 労働団体	924	-	-	-	-	-	-	-	-	924	607	317
943 学術・文化団体	313	-	-	-	-	-	-	-	-	313	293	20
944 政治団体	33	-	-	-	-	-	-	-	-	33	16	17
949 他に分類されない非営利的団体	1,831	-	-	-	-	-	-	-	-	1,831	1,315	516
95 その他のサービス業	342	4	61	61	56	5	-	-	-	278	248	29
951 集会場	256	-	20	20	20	-	-	-	-	236	207	29
952 と畜場	15	-	9	9	6	3	-	-	-	5	5	-
959 他に分類されないサービス業	71	4	31	31	30	2	-	-	-	36	36	-

別事業所数・収入金額

総数	収入金額(百万円)											
	個人	会社	国内の会社				外国の会社	その他	会社以外の法人		法人以外の団体	
			株式会社	有限会社	合名・合資・相互会社	社						
S	-	-	-	-	-	-	-	S	S	85		
S	-	-	-	-	-	-	-	S	S	851		
535,079	-	-	-	-	-	-	-	535,079	535,079	852		
3,735,565	1,402	S	S	S	153,931	S	-	S	S	86		
167,740	S	S	S	S	18,060	-	-	P	P	861		
112,420	-	S	S	S	X	-	-	S	S	862		
122,244	-	117,760	117,760	117,102	658	-	-	4,484	4,484	863		
2,469,579	P	S	S	S	89,558	S	-	S	14,995	P	864	
153,271	P	81,275	81,275	71,653	9,622	-	-	S	S	X	865	
710,312	103	S	S	S	X	P	-	S	S	-	866	
1,182,540	1,346	S	S	S	200,360	936	-	S	S	S	87	
546,168	S	S	S	S	376,061	134,143	S	-	S	S	S	871
633,559	P	S	S	S	66,216	P	-	P	P	-	872	
2,813	-	2,813	2,813	2,813	-	-	-	-	-	-	879	
S	7,945	S	S	S	S	-	-	S	S	P	88, 89	
S	S	S	S	S	S	-	-	S	S	-	88	
106,160	S	S	S	S	S	-	-	5,598	5,598	-	885	
33,992	3,527	26,965	26,965	X	X	-	-	3,500	3,500	-	886	
S	-	S	S	S	S	-	-	11,542	11,542	-	887	
S	-	S	S	P	X	-	-	S	S	-	889	
134,976	P	S	S	S	373	-	-	S	S	P	89	
101,772	P	S	S	S	251	-	-	S	S	P	892	
33,204	-	S	S	S	122	-	-	S	S	-	899	
S	219	S	S	S	3,635	-	-	S	S	4,513	90	
S	-	-	-	-	-	-	-	S	S	-	901	
S	219	S	S	S	2,725	-	-	S	S	292	903	
86,856	219	S	S	S	2,725	-	-	S	S	-	90A	
S	-	-	-	-	-	-	-	S	S	292	90B	
218,688	-	S	S	S	910	-	-	S	S	P	904	
S	-	P	P	P	-	-	-	S	S	X	905	
239,282	-	-	-	-	-	-	-	239,282	S	P	909	
S	1,730	S	S	S	S	S	-	S	S	860	91, 92	
S	S	S	S	S	S	S	-	S	S	133	91	
116,741	P	S	S	S	714	-	-	S	S	10	918	
40,539	P	19,326	19,326	18,612	714	-	-	S	S	10	91C	
16,001	-	10,155	10,155	10,155	-	-	-	5,846	5,846	-	91D	
60,201	-	S	S	S	-	-	-	S	S	-	91E	
S	P	S	S	S	S	S	-	S	S	123	919	
S	P	S	S	S	P	-	-	S	S	726	92	
S	P	S	S	S	X	-	-	S	S	677	921	
14,941	-	P	P	P	P	-	-	S	S	49	922	
336,538	31	-	-	-	-	-	-	336,507	334,953	1,554	93	
50,004	14	-	-	-	-	-	-	49,991	49,618	373	931	
148,757	17	-	-	-	-	-	-	148,740	147,559	1,181	932	
41,632	-	-	-	-	-	-	-	41,632	41,632	-	933	
96,144	-	-	-	-	-	-	-	96,144	96,144	-	939	
1,868,093	-	-	-	-	-	-	-	1,868,093	1,589,830	278,262	94	
607,057	-	-	-	-	-	-	-	607,057	546,287	60,770	941	
239,783	-	-	-	-	-	-	-	239,783	S	S	942	
120,652	-	-	-	-	-	-	-	120,652	120,109	543	943	
92,866	-	-	-	-	-	-	-	92,866	S	S	944	
807,734	-	-	-	-	-	-	-	807,734	S	P	949	
S	P	14,000	14,000	S	P	-	-	S	S	S	95	
S	-	P	P	P	-	-	-	S	S	S	951	
S	-	S	S	P	P	-	-	583	583	-	952	
9,974	P	S	S	S	P	-	-	S	S	-	959	

A表 産業(小分類), 本所・支所(3区分), 経営組織(7区分)

支所・支社・支店

産業小分類	総数	事業所数									
		個人	会社	国内の会社				外国の会社	その他	会社以外の法人	法人以外の団体
				株式会社	有限会社	合名・合資・相互会社					
L サービス業	238,889	20,401	161,494	161,410	124,189	36,596	625	84	56,994	51,952	5,043
72 洗濯・理容・浴場業	47,569	11,677	35,531	35,531	19,349	15,873	310	-	361	342	19
721 洗濯業	25,733	7,114	18,457	18,457	10,229	7,953	275	-	162	144	19
72A 普通洗濯業	24,000	7,105	16,734	16,734	8,696	7,780	257	-	161	142	19
72B リネンサプライ業	1,734	9	1,724	1,724	1,533	173	18	-	1	1	-
722 洗張・染物業	42	-	25	25	13	13	-	-	17	17	-
723 理容業	4,808	1,508	3,248	3,248	1,677	1,570	-	-	52	52	-
724 美容業	15,234	3,022	12,194	12,194	6,356	5,839	-	-	18	18	-
725 公衆浴場業	664	33	557	557	174	348	35	-	74	74	-
726 特殊浴場業	598	-	560	560	445	115	-	-	38	38	-
729 その他の洗濯・理容・浴場業	490	-	490	490	455	36	-	-	-	-	-
73 駐車場業	3,245	88	2,293	2,293	1,908	350	35	-	864	819	45
74 その他の生活関連サービス業	13,806	965	11,939	11,937	9,255	2,665	17	2	902	755	147
743 写真業	7,264	517	6,737	6,736	5,259	1,463	14	2	10	10	-
744 衣服縫製修理業	1,711	277	1,435	1,435	913	521	-	-	-	-	-
745 物品預り業	923	18	389	389	371	18	-	-	515	401	114
746 火葬・墓地管理業	243	-	111	111	86	25	-	-	132	126	7
747 冠婚葬祭業	2,322	66	2,151	2,151	1,824	325	3	-	104	101	3
749 他に分類されない生活関連サービス業	1,343	87	1,115	1,115	802	313	-	-	141	117	24
75 旅館, その他の宿泊所	19,339	202	15,323	15,308	13,449	1,802	57	15	3,813	3,593	221
751 旅館	7,055	93	6,625	6,617	4,983	1,610	24	8	337	336	1
752 簡易宿所	273	32	191	191	148	42	-	-	50	50	-
753 下宿業	220	69	151	151	151	-	-	-	-	-	-
759 その他の宿泊所	11,791	9	8,356	8,349	8,166	149	34	7	3,426	3,206	219
75A 会社・団体の宿泊所	4,529	-	2,193	2,185	2,134	18	34	7	2,336	2,198	138
75B 他に分類されない宿泊所	7,262	9	6,164	6,164	6,032	132	-	-	1,089	1,088	81
76 娯楽業(映画・ビデオ制作業を除く)	22,642	364	20,852	20,841	15,853	4,933	54	11	1,426	1,314	112
761 映画館	572	3	568	565	527	38	-	3	2	2	-
762 劇場, 興行場(別掲を除く)	75	3	37	37	25	12	-	-	34	34	-
763 興行団	211	-	190	189	163	25	-	1	21	18	3
764 競輪・競馬等の競走場	21	-	11	11	11	-	-	-	10	10	-
765 競輪・競馬等の競技団	59	-	-	-	-	-	-	-	59	58	1
766 スポーツ施設提供業(76D, 76Fを除く)	3,095	3	2,314	2,308	2,125	179	3	7	778	740	38
76A スポーツ施設提供業(別掲を除く)	673	-	253	253	236	16	-	-	421	396	25
76B 体育館	364	-	32	32	32	-	-	-	332	321	12
76C ゴルフ場	1,136	-	1,121	1,121	1,111	7	2	-	16	15	1
76E ボウリング場	571	3	568	561	529	31	1	7	-	-	-
76G バンテイング・テニス練習場	351	-	342	342	217	125	-	-	9	9	-
767 公園, 遊園地	556	-	186	186	184	1	-	-	370	311	60
768 遊戯場	13,308	223	13,056	13,056	9,795	3,227	34	-	29	24	5
76K マージャンクラブ	439	23	416	416	298	118	-	-	-	-	-
76L パチンコホール	8,258	85	8,172	8,172	5,640	2,505	27	-	-	-	-
76M その他の遊戯場	4,611	114	4,468	4,468	3,856	604	7	-	29	24	5
769 その他の娯楽業	4,744	132	4,489	4,489	3,022	1,451	17	-	123	117	6
76N マリーナ業	49	-	39	39	33	6	-	-	10	10	-
76P 遊漁船業	1	-	-	-	-	-	-	-	1	-	1
76Q その他の娯楽業	4,694	132	4,451	4,451	2,990	1,445	17	-	112	107	5
77 自動車整備業	5,190	185	4,608	4,608	3,652	941	15	-	397	397	-
78 機械・家具等修理業(別掲を除く)	9,722	308	9,339	9,337	8,783	550	3	3	74	72	3
781 機械修理業	8,554	52	8,454	8,452	8,135	317	-	3	48	48	-
784 表具業	-	-	-	-	-	-	-	-	-	-	-
789 他に分類されない修理業	1,168	256	885	885	648	233	3	-	26	24	3
79 物品賃貸業(791, 792, 793を除く)	8,881	353	8,502	8,496	7,139	1,342	15	6	26	25	1
794 自動車賃貸業	3,347	9	3,338	3,338	3,269	67	1	-	-	-	-
795 スポーツ・娯楽用品賃貸業	113	-	94	94	84	10	-	-	19	18	1
799 その他の物品賃貸業	5,421	344	5,071	5,064	3,786	1,265	14	6	7	7	-
79A 音楽・映像記録物賃貸業	3,362	254	3,108	3,108	2,151	957	-	-	-	-	-
79B その他の物品賃貸業	2,060	90	1,963	1,956	1,636	307	14	6	7	7	-
80, 81	1,373	13	1,279	1,276	1,142	134	-	3	81	81	-
80 映画・ビデオ制作業	499	9	487	483	349	134	-	3	3	3	-
801 映画, ビデオ制作・配給業	479	9	467	464	330	134	-	3	3	3	-
802 映画・ビデオサービス業	20	-	20	20	20	-	-	-	-	-	-
81 放送業	874	4	793	793	793	-	-	-	78	78	-
811 公共放送業(有線放送業を除く)	78	-	-	-	-	-	-	-	78	78	-
812 民間放送業(有線放送業を除く)	257	-	257	257	257	-	-	-	-	-	-
813 有線放送業	540	4	536	536	536	-	-	-	-	-	-
82 情報サービス・調査業(821, 822を除く)	2,967	24	2,819	2,802	2,751	51	-	17	125	122	3
823 ニュース供給業	2,672	-	2,547	2,530	2,506	24	-	17	125	122	3
824 興信所	296	24	272	272	245	28	-	-	-	-	-
84 専門サービス業(他に分類されないもの, 849を除く)	36,296	5,435	29,249	29,245	23,828	5,334	83	4	1,612	1,259	353
841 法律事務所, 特許事務所	-	-	-	-	-	-	-	-	-	-	-
84K 法律事務所	-	-	-	-	-	-	-	-	-	-	-
84L 特許事務所	-	-	-	-	-	-	-	-	-	-	-
842 公証人役場, 司法書士事務所	27	27	-	-	-	-	-	-	-	-	-
843 公認会計士事務所, 税理士事務所	191	41	-	-	-	-	-	-	149	149	-
84M 公認会計士事務所	149	-	-	-	-	-	-	-	149	149	-
84N 税理士事務所	41	41	-	-	-	-	-	-	-	-	-
844 獣医業	339	89	191	191	68	123	-	-	58	58	-
845 土木建築サービス業	7,300	238	6,700	6,696	6,105	580	11	4	362	359	3
846 デザイン業	670	15	656	656	323	333	-	-	-	-	-
847 著述家・芸術家業	-	-	-	-	-	-	-	-	-	-	-
848 個人教授所	27,768	5,024	21,702	21,702	17,332	4,298	72	-	1,043	693	350
84A 学習塾(各種学校でないもの)	12,780	1,834	10,777	10,777	8,032	2,742	3	-	169	85	84
84B フィットネスクラブ	746	7	723	723	639	84	-	-	16	16	-
84C スポーツ・健康個人教授所	1,537	150	1,202	1,202	1,017	185	-	-	184	153	31
84D 生花・茶道個人教授所	271	167	40	40	40	-	-	-	64	-	64
84E そろばん個人教授所	1,608	1,375	233	233	21	212	-	-	-	-	-
84F 音楽個人教授所	3,607	348	3,226	3,226	2,841	316	69	-	33	33	-
84G 書道個人教授所	1,077	737	62	62	62	-	-	-	278	184	94
84J その他の個人教授所	6,142	405	5,438	5,438	4,679	759	-	-	299	222	77

別事業所数・収入金額

総数	収入金額(百万円)											
	個人	会社	国内の会社				外国の会社	その他				
			株式会社	有限会社	合名・合資・相互会社	会社以外の法人		法人以外の団体				
S	270,666	S	S	S	5,701,897	66,136	53,032	S	S	206,836	L	
2,153,401	98,201	2,046,262	2,046,262	1,482,530	557,126	6,605	-	8,938	8,883	55	72	
1,117,307	32,726	S	S	S	225,385	6,311	-	S	S	55	721	
S	S	S	S	S	208,896	5,260	-	S	S	55	72A	
S	S	S	S	S	16,490	1,052	-	P	P	-	72B	
1,716	-	1,632	1,632	S	S	-	-	83	83	-	722	
139,693	14,395	125,118	125,118	79,152	45,965	-	-	180	180	-	723	
667,942	50,709	617,171	617,171	371,019	246,152	-	-	61	61	-	724	
55,307	371	53,877	53,877	24,852	28,730	294	-	1,059	1,059	-	725	
122,379	-	S	S	S	S	-	-	-	-	-	726	
49,057	-	49,057	49,057	X	X	-	-	-	-	-	729	
151,826	405	85,285	85,285	76,308	8,567	410	-	66,136	63,869	2,267	73	
1,518,083	12,474	S	S	S	123,860	S	P	S	S	S	74	
S	7,792	S	S	S	S	X	P	67	67	-	743	
40,167	1,751	38,416	38,416	21,262	17,153	-	-	-	-	-	744	
18,149	128	S	S	S	296	-	-	S	S	1,130	745	
22,682	-	S	S	S	S	-	-	S	S	P	746	
S	2,433	S	S	S	50,778	P	-	S	S	P	747	
S	370	S	S	S	8,389	-	-	3,256	3,256	0	749	
S	3,255	S	S	S	194,391	6,126	316	482,755	469,122	13,632	75	
S	2,757	S	S	S	189,982	5,089	271	124,602	S	P	751	
20,309	293	19,154	19,154	16,983	2,171	-	-	862	862	-	752	
1,767	206	1,562	1,562	1,562	-	-	-	-	-	-	753	
551,604	0	194,313	194,268	190,993	2,237	1,037	46	357,290	S	S	759	
393,660	-	76,639	76,593	S	P	1,037	46	317,022	S	S	75A	
157,943	0	117,675	117,675	S	S	-	-	40,269	X	X	75B	
S	91,784	S	S	S	4,271,104	41,073	P	3,460,954	3,454,483	6,471	76	
200,180	P	S	S	S	X	-	-	P	P	-	761	
71,126	P	66,781	66,781	66,034	747	-	-	S	S	-	762	
149,494	-	X	S	S	X	-	-	P	X	S	P	763
S	-	20,680	20,680	20,680	-	-	-	-	-	-	764	
19,948	-	-	-	-	-	-	-	19,948	S	S	P	765
S	P	-	-	-	S	P	S	S	S	S	S	766
S	-	S	S	S	372	-	-	S	S	S	S	76A
S	-	289	289	289	-	-	-	S	S	495	76B	
S	-	S	S	S	P	P	-	7,171	S	P	76C	
129,078	P	S	127,261	123,606	S	P	S	-	-	-	76E	
14,289	-	14,266	14,266	10,172	4,094	-	-	23	23	-	76G	
S	-	S	S	S	P	-	-	S	S	S	767	
S	S	S	S	S	S	S	-	S	S	18	768	
27,263	696	26,567	26,567	22,735	3,832	-	-	-	-	-	76K	
S	82,958	S	S	S	S	S	-	-	-	-	76L	
573,336	S	570,164	570,164	517,432	49,444	3,289	-	S	S	18	76M	
S	S	429,187	429,187	337,312	91,208	667	-	S	S	572	769	
S	-	6,288	6,288	5,581	706	-	-	S	S	-	76N	
P	-	-	-	-	-	-	-	P	-	P	76P	
S	S	422,900	422,900	331,731	90,502	667	-	S	S	S	76Q	
X	X	781,809	781,809	720,504	S	P	-	68,417	68,417	-	77	
3,627,638	2,290	S	S	S	S	P	P	S	S	P	78	
3,552,353	S	S	S	S	S	-	-	P	S	S	781	
-	-	-	-	-	-	-	-	-	-	-	784	
75,285	S	S	S	69,340	S	P	-	S	S	P	789	
S	S	S	S	S	76,962	S	2,184	S	S	P	79	
S	154	S	S	S	X	P	-	-	-	-	794	
9,848	-	9,225	9,225	7,715	1,510	-	-	622	S	P	795	
S	S	S	S	S	72,918	S	2,184	S	S	-	799	
S	S	S	S	S	44,078	-	-	-	-	-	79A	
S	642	S	S	S	28,840	S	2,184	S	S	-	79B	
S	238	S	S	S	13,428	-	P	S	S	-	80, 81	
S	S	S	S	S	13,428	-	P	P	P	-	80	
S	S	S	S	S	13,428	-	P	P	P	-	801	
S	-	S	S	S	-	-	-	-	-	-	802	
S	P	S	S	S	-	-	-	S	S	-	81	
S	-	-	-	-	-	-	-	S	S	-	811	
S	-	S	S	S	-	-	-	-	-	-	812	
S	-	S	S	S	-	-	-	-	-	-	813	
S	208	S	S	S	2,438	-	P	5,239	5,239	0	82	
S	-	S	S	S	S	-	P	5,239	5,239	0	823	
42,162	208	41,953	41,953	S	S	-	-	-	-	-	824	
4,535,137	35,832	4,203,099	S	S	S	3,612	P	296,206	292,468	3,738	84	
-	-	-	-	-	-	-	-	-	-	-	841	
-	-	-	-	-	-	-	-	-	-	-	84K	
-	-	-	-	-	-	-	-	-	-	-	84L	
S	S	-	-	-	-	-	-	-	-	-	842	
72,429	2,245	-	-	-	-	-	-	70,184	70,184	-	843	
70,184	-	-	-	-	-	-	-	70,184	70,184	-	84M	
2,245	2,245	-	-	-	-	-	-	-	-	-	84N	
S	S	9,271	9,271	2,940	6,331	-	-	5,000	5,000	-	844	
S	1,662	S	S	S	36,221	S	P	S	S	P	845	
67,600	293	67,307	67,307	57,295	10,012	-	-	-	-	-	846	
-	-	-	-	-	-	-	-	-	-	-	847	
S	28,752	S	S	S	S	S	-	S	S	S	848	
S	19,760	S	S	S	S	P	-	2,055	1,231	824	84A	
195,749	P	S	S	S	1,986	-	-	S	S	-	84B	
171,270	306	S	S	S	12,783	-	-	S	S	114	84C	
2,083	837	1,204	1,204	1,204	-	-	-	42	-	42	84D	
5,683	3,449	2,235	2,235	19	2,215	-	-	-	-	-	84E	
120,203	S	117,288	117,288	107,213	S	S	-	P	P	-	84F	
1,351	848	116	116	116	-	-	-	387	293	94	84G	
384,862	1,176	S	S	S	26,084	-	-	S	S	S	84J	

A表 産業(小分類), 本所・支所(3区分), 経営組織(7区分)

支所・支社・支店

産業小分類	総数	事業所数										
		個人	会社	国内の会社				外国の会社	その他	会社以外の法人	法人以外の団体	
				株式会社	有限会社	合名・合資・相互会社	支店					
85 協同組合(他に分類されないもの)	15,114	-	-	-	-	-	-	-	-	15,114	15,114	-
851 農林水産業協同組合	14,469	-	-	-	-	-	-	-	-	14,469	14,469	-
852 事業協同組合(他に分類されないもの)	644	-	-	-	-	-	-	-	-	644	644	-
86 その他の事業サービス業(869を除く)	11,899	26	10,878	10,864	10,230	632	3	14	995	976	19	19
861 速記・筆耕・複写業	897	2	876	876	728	148	-	-	19	19	-	-
862 商品検査業	828	2	268	255	223	32	-	13	558	557	1	1
863 計量証明業	308	3	266	266	266	-	-	-	39	39	-	-
864 建物サービス業	6,092	10	6,004	6,003	5,670	330	3	1	78	71	8	8
865 民営職業紹介業	710	4	424	424	366	58	-	-	282	273	10	10
866 警備業	3,064	6	3,041	3,041	2,976	64	-	-	18	18	-	-
87 廃棄物処理業	2,170	19	2,086	2,086	1,574	507	5	-	65	63	2	2
871 一般廃棄物処理業	954	16	901	901	647	249	5	-	37	35	2	2
872 産業廃棄物処理業	1,209	3	1,179	1,179	921	258	-	-	28	28	-	-
879 その他の廃棄物処理業	6	-	6	6	6	-	-	-	-	-	-	-
88, 89 医療業(病院を除く)	4,535	613	2,306	2,305	1,137	1,168	-	1	1,615	1,574	41	41
88 療術業	4,110	606	2,178	2,177	1,016	1,161	-	1	1,326	1,300	26	26
885 歯科工務	2,095	606	1,344	1,344	306	1,037	-	-	146	121	25	25
886 医療に附帯するサービス業(別掲を除く)	238	-	238	238	117	121	-	-	-	-	-	-
887 医療に附帯するサービス業(別掲を除く)	740	-	581	580	577	3	-	1	159	159	-	-
889 その他の医療業	1,037	-	16	16	16	-	-	-	1,021	1,020	1	1
89 保健衛生	425	8	128	128	121	7	-	-	289	274	15	15
892 健康相談施設	246	8	17	17	10	7	-	-	222	217	4	4
899 その他の保健衛生	179	-	111	111	111	-	-	-	67	57	11	11
90 社会保険, 社会福祉	6,876	102	627	627	485	142	-	-	6,146	5,962	184	184
901 社会保険事業団体	390	-	-	-	-	-	-	-	390	390	-	-
903 児童福祉事業	2,682	102	365	365	268	98	-	-	2,215	2,129	86	86
90A 保育所	1,777	102	365	365	268	98	-	-	1,310	1,292	17	17
90B その他の児童福祉事業	905	-	-	-	-	-	-	-	905	836	69	69
904 老人福祉事業	2,002	-	262	262	217	45	-	-	1,740	1,716	24	24
905 知的障害・身体障害者福祉事業	1,235	-	-	-	-	-	-	-	1,235	1,176	58	58
909 その他の社会保険, 社会福祉	567	-	-	-	-	-	-	-	567	551	16	16
91, 92 教育(学校を除く)	5,044	10	3,552	3,544	3,350	167	26	8	1,483	1,359	124	124
91 社会教育	3,353	9	2,192	2,192	2,004	162	26	-	1,151	1,089	62	62
918 博物館, 美術館	1,261	3	532	532	514	19	-	-	725	680	45	45
91C 動物園, 植物園, 水族館	638	3	423	423	404	19	-	-	212	205	7	7
91D その他の社会教育	78	-	26	26	26	-	-	-	52	52	-	-
91E その他の教育施設	545	-	84	84	84	-	-	-	461	423	38	38
919 学術研究機関	2,092	6	1,660	1,660	1,490	144	26	-	426	409	18	18
92 自然科学研究所	1,692	1	1,359	1,351	1,347	4	-	8	331	270	61	61
921 人文・社会科学研究所	1,647	1	1,351	1,346	1,344	2	-	6	294	233	61	61
922 その他の自然科学研究所	45	-	8	5	3	2	-	2	37	37	-	-
93 宗教	13,052	17	-	-	-	-	-	-	13,035	12,550	485	485
931 神道系宗教	967	-	-	-	-	-	-	-	967	922	45	45
932 仏教系宗教	2,961	17	-	-	-	-	-	-	2,943	2,895	49	49
933 キリスト教系宗教	2,445	-	-	-	-	-	-	-	2,445	2,232	213	213
939 その他の宗教	6,679	-	-	-	-	-	-	-	6,679	6,501	178	178
94 政治・経済・文化団体	8,121	-	-	-	-	-	-	-	8,121	5,002	3,119	3,119
941 経済団体	2,429	-	-	-	-	-	-	-	2,429	2,077	352	352
942 労働団体	2,315	-	-	-	-	-	-	-	2,315	946	1,369	1,369
943 学術・文化団体	52	-	-	-	-	-	-	-	52	44	8	8
944 政治団体	627	-	-	-	-	-	-	-	627	148	480	480
949 他に分類されない非営利的団体	2,697	-	-	-	-	-	-	-	2,697	1,787	910	910
95 その他のサービス業	1,049	-	310	310	305	5	-	-	740	575	165	165
951 集会場	893	-	246	246	241	5	-	-	647	520	127	127
952 と畜場	18	-	6	6	6	-	-	-	12	10	2	2
959 他に分類されないサービス業	139	-	58	58	58	-	-	-	82	45	37	37

別事業所数・収入金額

総数	収入金額(百万円)										
	個人	会社	国内の会社				外国の会社	その他	会社以外の法人	法人以外の団体	
			株式会社	有限会社	合名・合資・相互会社	社					
S	-	-	-	-	-	-	-	S	S	-	85
S	-	-	-	-	-	-	-	S	S	-	851
135,148	-	-	-	-	-	-	-	135,148	135,148	-	852
3,078,255	S	S	S	S	38,819	P	3,660	S	S	P	86
121,439	P	S	S	S	5,936	-	-	S	S	-	861
158,761	P	S	S	S	X	-	S	S	S	P	862
90,006	P	77,157	77,157	77,157	-	-	-	S	S	-	863
S	P	S	S	S	21,692	P	P	S	16,424	P	864
111,707	P	93,195	93,195	91,235	1,960	-	-	S	S	0	865
S	P	S	S	S	5,921	-	-	S	S	-	866
654,037	919	S	S	S	71,146	3,725	-	S	S	P	87
255,022	S	239,069	239,069	201,717	33,626	3,725	-	S	S	P	871
395,601	P	S	S	S	37,520	-	-	S	S	-	872
3,415	-	3,415	3,415	3,415	-	-	-	-	-	-	879
S	5,315	S	S	S	S	-	-	S	S	S	88, 89
S	S	S	S	S	S	-	P	S	S	S	88
S	S	S	S	X	S	-	-	4,275	S	S	885
17,291	-	17,291	17,291	12,522	4,768	-	-	-	-	-	886
S	-	S	S	S	P	-	P	25,551	25,551	-	887
S	-	S	S	S	-	-	-	S	S	P	889
S	S	S	S	S	188	-	-	S	S	X	89
S	S	S	S	P	188	-	-	S	S	X	892
22,844	-	S	S	S	-	-	-	S	S	-	82
S	2,711	37,136	37,136	30,066	7,070	-	-	S	S	3,677	90
S	-	-	-	-	-	-	-	S	S	-	901
187,368	2,711	9,483	9,483	5,136	4,347	-	-	175,174	174,292	883	903
131,322	2,711	9,483	9,483	5,136	4,347	-	-	119,128	118,956	172	90A
56,046	-	-	-	-	-	-	-	56,046	55,336	710	90B
301,441	-	27,652	27,652	24,930	2,723	-	-	273,788	272,618	1,170	904
182,279	-	-	-	-	-	-	-	182,279	181,084	1,195	905
62,869	-	-	-	-	-	-	-	62,869	62,440	429	909
S	131	S	S	S	S	P	P	S	S	58,228	91, 92
S	S	S	S	S	S	P	-	S	S	1,897	91
S	P	S	S	S	359	-	-	S	S	154	918
36,435	P	22,012	22,012	21,653	359	-	-	S	S	69	91C
27,910	-	16,858	16,858	16,858	-	-	-	11,052	11,052	-	91D
S	-	S	S	S	-	-	-	S	S	85	91E
350,922	S	S	S	S	S	P	-	S	S	1,743	919
S	P	S	S	S	P	-	P	S	S	56,331	92
S	P	S	S	S	P	-	P	S	S	56,331	921
19,275	-	P	P	X	P	-	P	S	S	-	922
294,934	285	-	-	-	-	-	-	294,649	290,101	4,547	93
24,023	-	-	-	-	-	-	-	24,023	23,501	523	931
125,405	285	-	-	-	-	-	-	125,120	124,917	203	932
23,991	-	-	-	-	-	-	-	23,991	22,029	1,962	933
121,514	-	-	-	-	-	-	-	121,514	119,655	1,860	939
540,788	-	-	-	-	-	-	-	540,788	434,983	105,805	94
122,172	-	-	-	-	-	-	-	122,172	110,183	11,988	941
95,949	-	-	-	-	-	-	-	95,949	S	S	942
11,354	-	-	-	-	-	-	-	11,354	S	P	943
28,179	-	-	-	-	-	-	-	28,179	8,476	19,703	944
283,134	-	-	-	-	-	-	-	283,134	263,501	19,633	949
76,152	-	25,044	25,044	25,009	35	-	-	51,108	S	S	95
S	-	S	S	S	35	-	-	45,935	S	S	951
P	-	P	P	P	-	-	-	905	S	P	952
7,286	-	3,018	3,018	3,018	-	-	-	4,268	S	S	959

順次 LP 法と  $\tau$ -ARGUS 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

岡本 政人\*

要 旨

本稿は、サービス業基本調査等への適用を想定して、米国センサス局などが採用している順次 LP 法と欧州の SDC 及び CASC プロジェクトで開発された集計表秘匿処理システム  $\tau$ -ARGUS が搭載している Optimal 法 (Fischetti-Salazar 法), Modular 法 (HiTaS 法), HyperCube 法の比較研究を行うとともに、これらのセル秘匿方式共通の問題を解消することに焦点を当てて、セル秘匿方式以外の開示抑制法を含めて最近のこの分野の研究動向を紹介している。

サービス業基本調査の集計表は、複雑な階層構造をもつ三重クロス表が多く、階層的に二次秘匿セルを求めていく Modular 法及び HyperCube 法では二次秘匿セルが著しく多くなる場合や、統合レベルの高いセルを二次秘匿しないとといった実際的な制約条件を課すと処理が不可能になる場合のあることが明らかとなった。Optimal 法は、本来的には最適な秘匿パターンを求めることのできる秘匿処理法であるが、1万セルを超える大きさの三重クロス表の処理が行えず、数百～数千セルの小さな集計表でも実際には殆どの場合最適な秘匿パターンが得られず、順次 LP 法 (+クリーンアップ処理) のほうがより適切な秘匿パターンを与える場合も多いことが明らかとなった。これに対し、順次 LP 法は、複雑な構造をもつ多重クロス表であっても安定して比較的良好な秘匿パターンを求めることができ、特に事後的にクリーンアップ処理によって不要な二次秘匿セルを除去すると、比較的構造が単純な二重クロス表でも Modular 法とほぼ同等若しくはより優る結果が得られることが示された。

このようにどちらかという消去法的に順次 LP 法の優位性が実証的に示されたが、大きな集計表の処理時間の問題や、各セル秘匿方式共通した問題でもある複数の集計表間の整合的な処理がサービス業基本調査の集計表の場合容易でないことも示された。この問題の解決は今後の課題であるが、本稿では、簡単な実験を行い、sliding protection 方式や免責アプローチを部分的に採用することを提案している。

さらに、セル秘匿方式以外の集計表開示抑制法のうち、CTA が、より大きな集計表の処

理やより多くの集計表の一括処理を行える可能性があること、米国四半期労働力指標で採用されている個票データにノイズを付加し集計する方式が、複数集計表間の整合的な処理やオーダーメイド集計などの追加的集計への対応の問題を解消する可能性があることを示唆し、これらの新しい集計表開示抑制法の研究が必要になってきていることを指摘している。

\* 総務省統計研修所 (前 統計センター研究センター)

## 順次 LP 法と $\tau$ -ARGUS 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

岡本 政人

#### 1 はじめに

本論文では、集計表における開示抑制法 (Disclosure Control Method) のうち、経済調査 (business survey) で通常用いられているセル秘匿 (Cell suppression) 方式を主として扱っている。さらに、一次秘匿セル (センシティブセル) の選定基準は件数 3 の閾値ルールとし、二次秘匿セルを決定する方法の比較に限定している。

サービス業基本調査の集計表における二次秘匿セルの決定は、全て人手によって処理を行うか、内部開発したシステム (松本 (1996)) で処理を行う方法を採用してきた。この内部開発システムは、複雑な構造の大規模集計表複数を、一応、一括して処理できることが利点であるが、反面、開示リスク抑制の度合いと二次秘匿をなるべく少なくするという意味での情報ロスの抑制の度合いが明確ではない<sup>1</sup>。

そこで、欧米諸国において開発されている秘匿処理法<sup>2</sup>の導入可能性について比較検討することとした。比較した秘匿処理法は、Sande (1984)が提案し、米国センサス局などが採用<sup>3</sup>している順次 LP 法と、欧州のプロジェクト SDC 及び CASC で開発された秘匿処理ソフト  $\tau$ -ARGUS<sup>4</sup>が採用している 3 種類の秘匿処理法 HyperCube 法、Modular 法 (HiTaS 法)、Optimal 法 (Fischetti-Salazar 法) である。なお、 $\tau$ -ARGUS には、ネットワークフロー法も搭載されているが、適用できるのは構造の単純な二重クロス表に限られており、本論文で適用対象として想定しているサービス業基本調査の集計表には (指定した開示リスク抑制の度合いが保証されることを必須条件とすると) 適用不可であるため、比較対象から除外した。

この比較実験により、少なくともサービス業基本調査への適用を前提とした場合、順次

<sup>1</sup> 処理結果の適切さの度合いが不明確なことが必ずしも他のセル秘匿方式に比べて問題があるとは言い切れない面がある。第 4 章に記すように複数の集計表を統合的に処理する場合、指定した秘匿の度合いを厳格に守って (他の) 秘匿処理法を適用すると二次秘匿による情報ロスが著しく増大する可能性があるため、部分的に基準を緩和して、情報ロスの増大を抑制する必要がある。処理時間の制約から、適切な秘匿処理法を採用できない場合も十分に考えられる。また、第 2 章に記す Singleton セルの問題に完全に対応した処理を行うことも困難である。結局、どのセル秘匿方式を用いる場合でも、処理結果をチェックし、適宜判断を加えることが重要となる。(次の脚注 3 も参照)

<sup>2</sup> 本稿では、“秘匿処理”をセル秘匿方式による開示抑制の意味で用いる。ノイズ付加方式など他の方式を含めて説明する場合は“開示抑制”を用いる。

<sup>3</sup> Zayatz (2005)によると、米国センサス局では、二重クロス表についてはネットワークフロー法、比較的小規模な三重クロス表には順次 LP 法、大規模な三重クロス表にはネットワークフロー法を適用している。主としてネットワークフロー法を用いているのは処理が速いためであるが、三重クロス表や、二重クロス表でも構造が複雑な場合は、ネットワークフロー法では、指定した秘匿の度合いを保証することができないため、処理結果をチェックしなければならない。

<sup>4</sup> <http://neon.vb.cbs.nl/casc/>からダウンロードできる。

LP法が他の秘匿処理法よりも優れていることが明らかにされた。ただし、この比較実験では、秘匿処理を集計表個々に行っており、集計表間の二次秘匿の整合性を考慮しておらず、この結果から直ちに順次LP法が容易かつ問題なく業務に適用できることを示すものではない。複数集計表を整合的かつ適切に秘匿処理することは容易ではなく、現時点では、どの秘匿処理法を採用した場合でも、なるべく多重クロス表は避け、二重クロス表にするなど集計表の設計方針も含めて様々な配慮・工夫が必要である。本稿では、セル秘匿方式で複数集計表の整合的処理をなるべく容易にする方策についての若干の検討を行い、さらにセル秘匿方式以外の新しい開示抑制法がこの問題をある程度解決できる可能性があることを述べる。

以下、第2章で集計表におけるセンシティブなセルの存在と、二次秘匿の必要性、二次秘匿セルを決定する方法などについて説明し、第3章で平成11年サービス業基本調査の集計表などに適用した結果、第4章で他の開示抑制法の採用可能性も含めて今後の課題等について述べる。

## 2 集計表の秘匿処理の概要

### (1) 一次秘匿

統計調査（センサスあるいは標本調査であっても経済調査で大規模企業・事業所については全数調査を行っているような場合を前提としている。）の個票データから集計表を作成した場合、各セルに分布される調査客体の件数は、通常大きく異なる。セルによっては、件数ゼロや件数1～2となる場合がある。件数1の場合、どの調査客体の回答値か特定される可能性が高い。件数2の場合も、該当セルに分布された調査客体は、当該セルの値から自身の回答値を控除して他方の客体の回答値を知ることができる。そこで、分布される件数が1～2の場合、そのセルはセンシティブなセルとみなされ、分布された数値を（一次）秘匿するなど開示リスク抑制のための処理が施される（閾値ルール<sup>5</sup>：Threshold rule）。セルに分布される客体のうち上位  $n$  が  $k\%$ 以上を占める場合にセンシティブセルとみなす( $n, k$ )ルール<sup>5</sup>（占有ルール、集中度ルールともいう。）や、セルの値に占める個々の客体のシェアを事前に誤差  $q\%$ で推定できるという前提で、シェア第2位の客体がセルの値の公表後に誤差  $p\%$ 以内でシェア第1位の客体の回答値を推定できる場合にセンシティブセルとみなす( $p, q$ )ルール<sup>5</sup>（曖昧度ルール、事前事後ルールともいう。）といったより厳しい基準もあるが、今回の比較実験では、従来からサービス業基本調査などで採用されている件数3の閾値ルール（件数1～2ならばセンシティブセルとするルール）を用いる。

### (2) 二次秘匿

<sup>5</sup> UN/ECE & Eurostat (2003)参照。

以下の小さな集計表で、右下のセル (X<sub>16</sub>) が一次秘匿セルであるとする。

このセルの値だけを秘匿した場合、縦計、横計のセル (X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>9</sub>, X<sub>13</sub>) があるため、他のセルの値から一次秘匿セル X<sub>16</sub> の値を求めることができってしまう。そこで、他のセルの幾つかを (二次) 秘匿する。以下の例では、X<sub>10</sub>, X<sub>12</sub>, X<sub>14</sub> を二次秘匿セルに選定している。この場合でも、セル X<sub>16</sub> の値は、0~41 の範囲にあることを求めることができる。以下、これを秘匿レンジと呼ぶことにする。秘匿レンジをセルの値の±100%、つまり、0~64 としたい場合、この二次秘匿では不十分ということになる。指定された基準を満たす秘匿レンジを確実に確保するだけであれば全てのセルを二次秘匿すればよいが、利用者に出来る限り多くの情報を提供すべきであるから、指定した秘匿レンジを確保できる二次秘匿パターンのうち、秘匿されるものが出来る限り少ないものが良いことになる。情報ロス最小化の具体的基準としては、秘匿されるセルの値、客体数、セル数の合計などが考えられる。

	計	地域 A	地域 B	地域 C
計	X <sub>1</sub> 185	X <sub>5</sub> 45	X <sub>9</sub> 50	X <sub>13</sub> 90
産業 1	X <sub>2</sub> 55	X <sub>6</sub> 11	X <sub>10</sub> 21	X <sub>14</sub> 23
産業 2	X <sub>3</sub> 70	X <sub>7</sub> 15	X <sub>11</sub> 20	X <sub>15</sub> 35
産業 3	X <sub>4</sub> 60	X <sub>8</sub> 19	X <sub>12</sub> 9	X <sub>16</sub> 32

$$\begin{aligned}
 44 &= X_{10} + X_{14} \\
 41 &= X_{12} + X_{16} \\
 30 &= X_{10} + X_{12} \\
 55 &= X_{14} + X_{16}
 \end{aligned}
 \Rightarrow 0 < X_{16} < 41$$

### (3) 最適な秘匿パターンを求める問題

指定された秘匿レンジを確保し、情報ロスを最小化する秘匿パターンを求める問題は、以下のような混合整数計画法 (MIP) の問題として定式化できる (Kelley の定式化)。

この式は、秘匿パターン  $\mathbf{x}=\{x_i\}$  が、任意の一次秘匿セル (連番  $k$ ) について確保すべき秘匿レンジの上限値をとり、しかも表内の算術的關係を満たすセルの値  $\mathbf{y}^{(k)}$  が存在し、同様に確保すべき秘匿レンジの下限値をとり、しかも表内の算術的關係を満たすセルの値  $\mathbf{z}^{(k)}$  も存在することを制約条件として、そのような秘匿パターンのうち情報ロスを最小化するものを求める問題となっている。式中の  $T$  は  $+\infty$  でもよいが、この問題を解くのに各セルが取り得る値の上限を設定しておいたほうが扱い易い場合があるため、このように定式化した。

この MIP 問題は、変数の数=セル数 × (2 × 一次秘匿セル数 + 1) となり、例えば、1

万セルの集計表に1%の一次秘匿セルが存在すると、201万変数と非常に大規模になるため、実際に解くのはほとんど不可能である。1万セルあるいはそれ以上の大きさの集計表は多く、一次秘匿セルの割合も1%程度あるいはそれ以上存在することが普通であるため、最適解をこの式から直接的に導くことは実用的ではない。

$$\begin{aligned} \min \sum_i c_i \cdot x_i \\ \text{s.t.} \\ \text{for } \forall k \end{aligned}$$

$$\mathbf{M}\mathbf{y}^{(k)} = 0$$

$$a_i - a_i x_i \leq y_i^{(k)} \leq a_i + a_i x_i T$$

$$y_{i_k}^{(k)} = a_{i_k} + r a_{i_k}$$

$$\mathbf{M}\mathbf{z}^{(k)} = 0$$

$$a_i - a_i x_i \leq z_i^{(k)} \leq a_i + a_i x_i T$$

$$z_{i_k}^{(k)} = a_{i_k} - r a_{i_k}$$

$$\left( \begin{array}{l} k: \text{一次秘匿セルの連番} \\ x_i = 0 \text{ or } 1, x_{i_k} = 1 : \text{セル}i\text{の秘匿の有無を示す変数} \\ a_i: \text{セル}i\text{の値}, c_i: \text{セル}i\text{を秘匿した場合のコスト係数} \\ T: \text{十分大きな値}, r: \text{セルの値}a_i\text{に対する比率で表した確保すべき秘匿レンジ} \\ \mathbf{M}: \text{セル間の算術的關係を表す行列} \end{array} \right)$$

#### (4) 比較する秘匿処理法

実用性を高めるため、いくつかの秘匿処理法が考案されてきた。その一つは、Fischetti & Salazar (2000)が提案した Optimal 法で、(3)の MIP 問題を双対問題に変換し、Branch and Cut 法を用いることにより最適解を求めるものである。数千セルの集計表であれば扱える可能性がある。

近似的な解を求めることを目指して開発された方法もある。そのうち、3つを採り上げる。オランダ統計局が開発した Modular 法 (de Wolf (2002)) は、通常、集計表が小計、中計などをもつ階層構造になっており、各レベルでは小さな単純クロス表に分割されることに着目し、分割した単純クロス表に対して Optimal 法を適用していく方法である。基本的には下位レベルから Optimal 法を適用していくが、上位レベルで処理不可になった場合に特別な処理が必要になる。

同様に階層的に処理を行っていくのが HyperCube 法 (Repsilber (1994)) である。HyperCube 法は数理計画法の手法を用いない。

MIP よりも扱い易い線形計画法 (LP) の問題に近似させ、さらに一次秘匿セルごとに分けて順次解いて行く方法が順次 LP 法 (Sande (1984)) である。

左式は、(任意の順序で付与した) 連番  $k$  の一次秘匿セル  $i_k$  について、秘匿レンジの上限を確保するために必要な二次秘匿セルを求める LP 問題、右式は、秘匿レンジの下限を

確保するために必要な二次秘匿セルを求める LP 問題になっている。 $y_i^+ > 0$  または  $y_i^- > 0$  となったセル  $i$  は、秘匿セルとなる。秘匿セルとされたセルの対応するコスト係数  $c_i$  をゼロにして、次の LP 問題を解いていく。処理する一次秘匿セルの順番によって秘匿パターンは異なるが、どの順番が良いか明確な傾向はないと言われている (Kelley et al. (1992))。今回の実験では、集計表の左上 (総計欄) から機械的に付与したセル番号順と、一次秘匿セルの値の昇順の2種類の順番を用いることとした。

$$\begin{array}{ll}
 \min \sum_i c_i (y_i^+ + y_i^-) & \min \sum_i c_i (y_i^+ + y_i^-) \\
 \text{s.t.} & \text{s.t.} \\
 \mathbf{M}(\mathbf{y}^+ - \mathbf{y}^-) = 0 & \mathbf{M}(\mathbf{y}^+ - \mathbf{y}^-) = 0 \\
 0 \leq y_i^+, 0 \leq y_i^- \leq a_i & \Rightarrow 0 \leq y_i^+, 0 \leq y_i^- \leq a_i \\
 y_{i_k}^+ = r \cdot a_{i_k}, y_{i_k}^- = 0 & y_{i_k}^+ = 0, y_{i_k}^- = r \cdot a_{i_k} \\
 \left( c_i = 0 \left| \begin{array}{l} \text{セル}i\text{が一次秘匿セルまたは} \\ \text{前段階の処理で二次秘匿セル} \\ \text{に選定} \end{array} \right. \right) & \left( c_i = 0 \left| \begin{array}{l} \text{セル}i\text{が一次秘匿セルまたは} \\ \text{前段階の処理で二次秘匿セル} \\ \text{に選定} \end{array} \right. \right)
 \end{array}$$

確保すべき秘匿レンジの上限, 下限別々に LP 問題を解くのではなく、連立して解く方法も考えられる。さらには、複数の一次秘匿セルについて連立して解く方法も考えられる<sup>6</sup>。一括して解くことで情報ロスが少なくなる可能性はあるが、処理時間は大幅に増大するため、適用できるのは比較的小さい集計表に限られるであろう。

順次 LP 法は、どのような集計表であっても比較的情報ロスが少ない秘匿パターンを求めることができるが、不要な二次秘匿セルが存在することが多い。そこで、順次 LP 法の処理結果から不要な二次秘匿セルを見つけ出して除外していくクリーンアップ処理を行うことが提案されている (Kelley et al. (1992))。不要か否かの判定は、以下のように、当該二次秘匿セルを秘匿しなかった場合の各一次秘匿セルが取り得る値の範囲を求め、いずれも確保すべき秘匿レンジを含んでいるかどうかを確認する方法で行う。

後続の二次秘匿セルの要否を判定するとき、既に不要と判定された二次秘匿セルは非秘匿セルとして扱う。要否を判定していく二次秘匿セルの順番については、セルの値の降順がよいとされており (Kelley et al. (1992))、今回の実験でもこれに従った。この LP 問題は、変数の数が秘匿されたセルの数の2倍<sup>7</sup>と少ないため、一つ一つの処理時間は

<sup>6</sup> このような順次 LP 法のバリエーションは、Fagan (2001)に紹介されている。

<sup>7</sup> 変数の数を半分 (秘匿されたセル数) に減らす定式化も可能であるが、その場合、計算結果において多くの秘匿セルの  $y_i^+, y_i^-$  がゼロより大きくなってしまい、後続の二次秘匿セルについて確認すべき一次秘匿セルの数 (解くべき LP 問題の数) が必要以上に多くなってしまいう可能性がある。なお、使用している数理計画法ソルバーによって適切な定式化が異なるかもしれない。

短い、単純な手順では二次秘匿セル数×一次秘匿セル数回解く必要がある。例えば、1万セルの集計表で2%の一次秘匿セルが存在し、順次LP法により求められた二次秘匿セルが一次秘匿セルの5倍であったとすると、LP問題を20万回解かねばならないため、処理時間を短縮する工夫が必要になる。今回の実験では、順次LP法により秘匿パターンを求める過程で、一次秘匿セルごとに“利用された”二次秘匿セル ( $y_i^+ > 0$  または  $y_i^- > 0$ ) の情報を保存しておき、“利用されなかった”二次秘匿セルと一次秘匿セルの組合せについては、クリーンアップ処理における判定が不要なため除外することで、解くべきLP問題の数を減らした。ただし、不要と判定された二次秘匿セルが見つかった場合、“利用された”二次秘匿セルの情報を更新する必要がある。

$\begin{aligned} & \max y_{i_k}^+ \\ & \text{s.t.} \\ & \mathbf{M}(\mathbf{y}^+ - \mathbf{y}^-) = 0 \\ & y_i^+ = y_i^- = 0 \quad \text{if } i: \text{非秘匿セル or } j \\ & 0 \leq y_i^+, 0 \leq y_i^- \leq a_i \quad \text{if } i: \text{秘匿セル } \neq j, i_k \\ & 0 \leq y_{i_k}^+ \leq ra_{i_k}, 0 \leq y_{i_k}^- \\ & j: \text{秘匿の要否を判定しているセル} \end{aligned}$	$\begin{aligned} & \max y_{i_k}^- \\ & \text{s.t.} \\ & \mathbf{M}(\mathbf{y}^+ - \mathbf{y}^-) = 0 \\ & y_i^+ = y_i^- = 0 \quad \text{if } i: \text{非秘匿セル or } j \\ & 0 \leq y_i^+, 0 \leq y_i^- \leq a_i \quad \text{if } i: \text{秘匿セル } \neq j, i_k \\ & 0 \leq y_{i_k}^+, 0 \leq y_{i_k}^- \leq ra_{i_k} \\ & j: \text{秘匿の要否を判定しているセル} \end{aligned}$
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

以上に説明した4方法のうち、Optimal法、Modular法、HyperCube法は、 $\tau$ -ARGUSに搭載されているものを用い、順次LP法及びクリーンアップ処理については内部開発を行った。また、Optimal法についても後述するように実際の処理結果が必ずしも最適解になっていない可能性があったため、確認のために基本的には同等のアルゴリズムを採用したプログラムの内部開発を行った。

なお、HyperCube法以外は、数理計画法の問題を解くという中核部分に市販の数理計画法ソルバーを利用する必要がある。Optimal法及びModular法についてはILOG社のCPlex、順次LP法についてはS社の国産ソルバーを用いた。ただし、クリーンアップ処理については、LP問題の変数の数が少ないためフリーウェアのソルバーでも処理可能であり、処理が速いこともあってlp\_solveを用いた<sup>8</sup>。

#### (5) 秘匿処理法に関する付加的事項

##### i) 空のセル及び値ゼロのセルの扱い、保護セル

集計表には、分布される客体が存在しない空のセル(“-”等で表章される)や、分布される客体は存在するが値がゼロのセルが存在する。これらのセルは二次秘匿セルの候補から除外することが一般的である。ただし、値がゼロでも分布される客体が1

<sup>8</sup> [http://groups.yahoo.com/group/lp\\_solve/files/](http://groups.yahoo.com/group/lp_solve/files/)からダウンロードできる。LP問題の種類によってはS社のソルバーのほうが処理が速い。本稿のクリーンアップ処理の定式化の場合、後者のほうが速い。

～2のセルについては、一次秘匿することもある<sup>9</sup>。

実際に秘匿処理を行う場合、一部のセルを二次秘匿の候補から除外（保護セル）したい場合がある。例えば、総計のセルや一次元の周辺分布に相当するセル（総計の次の統合レベルのセル）については、なるべく二次秘匿されないようにするのが自然であろう。今回の実験では、総計及び一次元周辺分布のセルを二次秘匿の候補から除外した。

#### ii) 値ゼロの一次秘匿セルの処理

確保すべき秘匿レンジは、通常、セルの値に対する比率の形式で指定するが、値がゼロの（または値がゼロに近い）一次秘匿セルについては、比率の形式で指定することができない。そこで、値がゼロの一次秘匿セルについては、確保すべき秘匿レンジの上限値を指定する。今回の実験では上限値を10とした。例えば、前述した順次LP法の式で値がゼロの一次秘匿セルについては、 $ra_i$ を10に置き換える。

#### iii) 推論開示の抑制

前述した順次LP法の定式化では、秘匿セルはゼロ以上であればどんな値をとってもよいとして秘匿パターンを決定している。仮に、統計の利用者がセルの値を一定の範囲で推定できるとするならば、実質的に秘匿セルが取り得る値の範囲は狭められることになり、これを考慮して秘匿パターンを決定する必要がある。例えば、セルの値の±100%の範囲で推定できると考えられる場合、 $y_i^+$ に対する制約条件を $0 \leq y_i^+ \leq a_i$ に変更する必要がある。

制約条件が強くなるため、推論開示を考慮した処理を行うと二次秘匿が多くなるが、推論できる範囲の上限・下限がセルの値からみて等距離に指定した場合、すなわち±1%の形式で指定する場合、順次LP法における確保すべき秘匿レンジの上限・下限に関するLP問題のうち、どちらか一方を解けばよいことになり、処理時間を短縮できるメリットはある。確保すべき秘匿レンジを±r%の形式で指定している場合、推論によって実際の秘匿レンジが極端に狭くなることは実際には考えにくく、今回の実験では、秘匿による情報ロスを抑制することを重視したため、推論開示については考慮していない。しかし、後述するように、確保すべき秘匿レンジの上限・下限を固定せずに、上限値 - 下限値  $\geq 2r\%$ の形式で指定するSliding Protection方式を採った場合、推論開示は重要な問題となる可能性がある。

#### iv) Singleton のペア間の開示抑制

<sup>9</sup> 欠測値がある調査項目の集計表の場合、分布される客体は存在するが“空”のセルが発生する可能性がある。このようなセルを二次秘匿の候補とするか、あるいは、件数が1～2の場合に一次秘匿とするか否かについての一般的慣行は不明である。今回の実験では、収入金額を対象としているため、このような問題はない。

分布される客体が1件だけのセル (Singleton セル) が複数存在する場合、これらのセルは一次秘匿セルとなるため、通常の秘匿処理が施されることにより、秘匿セルの値を正確には知らない統計利用者が Singleton セルの値を求めることはできなくなる。しかし、Singleton セルに分布される客体が、自身の回答を用いて他の Singleton セルに分布される客体の回答を求めることが可能になる場合がある。

以下の集計表で、 $X_{10}$ と $X_{16}$ が Singleton セルであったとする。二次秘匿による情報ロスを最小限に止めるには、これら二つの一次秘匿セルを相互に利用して例えば $X_{12}$ と $X_{14}$ を二次秘匿にすることが考えられる。ただし、確保すべき秘匿レンジは±25%とする。しかし、 $X_{10}$ に分布された客体は、自身の回答値21を用いることにより、セル $X_{16}$ の値32を求めることができる。同様に $X_{16}$ に分布された客体もセル $X_{10}$ の値を求めることができる。したがって、二次秘匿するセルを追加する必要がある。

	計	地域 A	地域 B	地域 C
計	$X_1$ 185	$X_5$ 45	$X_9$ 50	$X_{13}$ 90
産業 1	$X_2$ 55	$X_6$ 11	$X_{10}$ 21	$X_{14}$ 23
産業 2	$X_3$ 70	$X_7$ 15	$X_{11}$ 20	$X_{15}$ 35
産業 3	$X_4$ 60	$X_8$ 19	$X_{12}$ 9	$X_{16}$ 32

このような Singleton セル間の開示リスクを考慮した秘匿パターンを求める問題はかなり厄介になる。今回の実験では、以下の理由により Singleton セル間の開示リスクを考慮せずに処理を行った。

- ① 順次 LP 法では、Singleton セルを“利用しない”という制約条件を付加して二次秘匿セルを求めることは可能であるが、二次秘匿が必要以上に多くなってしまう。
- ②  $\tau$ -ARGUS には、HyperCube 法のオプションとして Singleton セル間の開示リスク抑制処理が搭載されているが、二次秘匿が著しく増大してしまい、実用性がない。
- ③ 実際には、十分な秘匿レンジを指定して秘匿処理を行うと、Singleton セル間の開示リスクが発生する可能性が少なくなる傾向がある。
- ④ 処理結果に対して、Singleton セル間の開示リスクをチェックすることは比較的容易であり、これによって問題があると判定された Singleton セルのペアに関して追加処理を行うほうが実用的である。

### 3 実験結果

#### (1) 小規模な集計表に適用した結果

まず、 $\tau$ -ARGUS に添付されているサンプルデータから作成した小さな二重クロス集計表に対して、4種類の秘匿処理法を適用した結果を示す。表頭の属性は、小計の区分があり (Nr, Os, Ws, Zs, と不詳区分を表す 99), 一次秘匿セルは (分布される客体が同一で実質的に重複しているものを含めて) 6セル存在する。秘匿レンジは $\pm 25\%$ , コスト係数はセルの値 (秘匿するセルの値の総計の最小化) とした。

順次 LP 法を適用すると、秘匿されるセルは一次秘匿セルを含めて 18セル, セルの値の総計は 4,353,093 となる。順次 LP 法によって求められた二次秘匿セルには不要なものがあり, クリーンアップ処理を行うと秘匿されるセルは 17セル, セル値計は 4,350,709 となる。クリーンアップ後の秘匿パターンは, **Optimal** 法と同一になる。**Modular** 法では, 秘匿されるセル数が 16 と最も少ないが, 最小化すべき秘匿セル値計は 9,309,351 と, 順次 LP 法や **Optimal** 法を大幅に上回っている。**HyperCube** 法は, 秘匿されるセル数 20, 秘匿セル値計 15,417,336 といずれも最も多くなっている。以上の結果は, 比較的構造が単純な小規模の二重クロス表であっても, 階層的に処理を行う **Modular** 法及び **HyperCube** 法は, 情報ロスが過大になる可能性があることを示している。また, この事例だけでは一般化はできないものの, 順次 LP 法が, 特にクリーンアップ処理を行うと, **Optimal** 法により得られる最適な秘匿パターンのよい近似になっている。

## (2) 平成 11 年サービス業基本調査集計表 (全国編) に適用した結果

サービス業基本調査で秘匿対象となる集計項目は, 収入金額, 設備投資額, 給与支給額などの経理項目と従業員数などである。事業所数 (抽出率調整を別にすると各セルに分布される客体の件数に相当する) は秘匿されない。各集計項目別々に秘匿パターンを求める方式と, 同一の秘匿パターンを共通に用いる方式が考えられる。なお, 閾値ルールによって一次秘匿セルが決定されるため, 一次秘匿セルは各項目とも同一である。最もセンシティブな集計項目が明確である場合, 処理の効率も考慮すると, この項目を対象とした秘匿処理から得られた秘匿パターンを他の集計項目にも用いる後者の方式の採用が考えられる。ちなみに平成 11 年調査では後者の方式, 平成 16 年調査では項目の種類によって, 異なる秘匿パターンが用いられている。今回の実験では, 後者の方式を想定したため, 従業員数に関する集計表も含めて, 収入金額を秘匿対象とした処理を行った。集計表は 4 表 (表 1 の A~D 表) とした。このほか, **Optimal** 法で最適解が得られるか検証する意味もあつて地域編の集計表を都道府県別に分割したものも用いた。なお, サービス業基本調査は, 他の調査データを部分的に用いており, 公式統計表で秘匿されている (非公表の) セルの値も扱うため, 慎重を期して他調査データを除外し再作成した集計表を用いた。

## 順次 LP 法 (秘匿セル数 : 18, 秘匿セル値計 : 4, 353, 093)

	Total	Nr	1	2	3	Os	4	5	6	7	Ws	8	9	10	Zd	11	12	99
Total	86,700,593	22,605,912	10,362,885	9,301,035	2,941,992	19,076,360	642,440	2,728,336	11,503,550	4,202,034	23,714,113	2,536,835	19,014,602	2,162,676	21,304,208	13,926,263	7,377,945	-
2	2,795	3	3	0	-	2,792	1,613	-	1,179	-	-	-	-	-	-	-	-	-
4	2,384	463	463	-	-	463	-	-	463	-	-	-	-	-	1,458	1,458	-	-
5	13,439,750	3,582,412	1,949,998	1,169,001	463,413	3,191,713	207,143	455,450	1,717,764	811,356	3,226,153	325,057	2,662,670	238,426	3,439,472	2,387,135	1,052,337	-
6	11,984,354	3,404,007	1,781,580	1,161,491	460,936	2,637,484	159,493	485,720	1,269,583	722,688	2,915,249	381,398	2,348,650	185,201	3,027,614	1,970,373	1,057,241	-
7	12,854,335	3,624,523	1,870,186	1,303,176	451,161	2,714,897	127,111	580,058	1,273,782	733,946	3,363,216	441,812	2,604,751	316,653	3,151,699	1,838,257	1,313,442	-
8	14,957,779	4,160,792	2,262,488	1,458,608	439,696	3,291,406	104,857	447,369	1,590,901	1,148,279	4,087,556	344,277	3,426,154	317,125	3,418,025	2,153,073	1,264,952	-
9	33,458,811	7,833,327	2,498,167	4,208,374	1,126,786	7,237,605	42,223	759,739	5,649,878	785,765	10,121,939	1,044,291	7,972,377	1,105,271	8,265,940	5,575,967	2,689,973	-
99	385	385	-	385	-	-	-	-	-	-	-	-	-	-	-	-	-	-

## 順次 LP 法 + クリーンアップ処理, Optimal 法 (秘匿セル数 : 17, 秘匿セル値計 : 4, 350, 709)

	Total	Nr	1	2	3	Os	4	5	6	7	Ws	8	9	10	Zd	11	12	99
Total	86,700,593	22,605,912	10,362,885	9,301,035	2,941,992	19,076,360	642,440	2,728,336	11,503,550	4,202,034	23,714,113	2,536,835	19,014,602	2,162,676	21,304,208	13,926,263	7,377,945	-
2	2,795	3	3	0	-	2,792	1,613	-	1,179	-	-	-	-	-	-	-	-	-
4	2,384	463	463	-	-	463	-	-	463	-	-	-	-	-	1,458	1,458	-	-
5	13,439,750	3,582,412	1,949,998	1,169,001	463,413	3,191,713	207,143	455,450	1,717,764	811,356	3,226,153	325,057	2,662,670	238,426	3,439,472	2,387,135	1,052,337	-
6	11,984,354	3,404,007	1,781,580	1,161,491	460,936	2,637,484	159,493	485,720	1,269,583	722,688	2,915,249	381,398	2,348,650	185,201	3,027,614	1,970,373	1,057,241	-
7	12,854,335	3,624,523	1,870,186	1,303,176	451,161	2,714,897	127,111	580,058	1,273,782	733,946	3,363,216	441,812	2,604,751	316,653	3,151,699	1,838,257	1,313,442	-
8	14,957,779	4,160,792	2,262,488	1,458,608	439,696	3,291,406	104,857	447,369	1,590,901	1,148,279	4,087,556	344,277	3,426,154	317,125	3,418,025	2,153,073	1,264,952	-
9	33,458,811	7,833,327	2,498,167	4,208,374	1,126,786	7,237,605	42,223	759,739	5,649,878	785,765	10,121,939	1,044,291	7,972,377	1,105,271	8,265,940	5,575,967	2,689,973	-
99	385	385	-	385	-	-	-	-	-	-	-	-	-	-	-	-	-	-

## Modular 法 (秘匿セル数 : 16, 秘匿セル値計 : 9, 309, 351)

	Total	Nr	1	2	3	Os	4	5	6	7	Ws	8	9	10	Zd	11	12	99
Total	86,700,593	22,605,912	10,362,885	9,301,035	2,941,992	19,076,360	642,440	2,728,336	11,503,550	4,202,034	23,714,113	2,536,835	19,014,602	2,162,676	21,304,208	13,926,263	7,377,945	-
2	2,795	3	3	0	-	2,792	1,613	-	1,179	-	-	-	-	-	-	-	-	-
4	2,384	463	463	-	-	463	-	-	463	-	-	-	-	-	1,458	1,458	-	-
5	13,439,750	3,582,412	1,949,998	1,169,001	463,413	3,191,713	207,143	455,450	1,717,764	811,356	3,226,153	325,057	2,662,670	238,426	3,439,472	2,387,135	1,052,337	-
6	11,984,354	3,404,007	1,781,580	1,161,491	460,936	2,637,484	159,493	485,720	1,269,583	722,688	2,915,249	381,398	2,348,650	185,201	3,027,614	1,970,373	1,057,241	-
7	12,854,335	3,624,523	1,870,186	1,303,176	451,161	2,714,897	127,111	580,058	1,273,782	733,946	3,363,216	441,812	2,604,751	316,653	3,151,699	1,838,257	1,313,442	-
8	14,957,779	4,160,792	2,262,488	1,458,608	439,696	3,291,406	104,857	447,369	1,590,901	1,148,279	4,087,556	344,277	3,426,154	317,125	3,418,025	2,153,073	1,264,952	-
9	33,458,811	7,833,327	2,498,167	4,208,374	1,126,786	7,237,605	42,223	759,739	5,649,878	785,765	10,121,939	1,044,291	7,972,377	1,105,271	8,265,940	5,575,967	2,689,973	-
99	385	385	-	385	-	-	-	-	-	-	-	-	-	-	-	-	-	-

## HyperCube 法 (秘匿セル数 : 20, 秘匿セル値計 : 15, 417, 336)

	Total	Nr	1	2	3	Os	4	5	6	7	Ws	8	9	10	Zd	11	12	99
Total	86,700,593	22,605,912	10,362,885	9,301,035	2,941,992	19,076,360	642,440	2,728,336	11,503,550	4,202,034	23,714,113	2,536,835	19,014,602	2,162,676	21,304,208	13,926,263	7,377,945	-
2	2,795	3	3	0	-	2,792	1,613	-	1,179	-	-	-	-	-	-	-	-	-
4	2,384	463	463	-	-	463	-	-	463	-	-	-	-	-	1,458	1,458	-	-
5	13,439,750	3,582,412	1,949,998	1,169,001	463,413	3,191,713	207,143	455,450	1,717,764	811,356	3,226,153	325,057	2,662,670	238,426	3,439,472	2,387,135	1,052,337	-
6	11,984,354	3,404,007	1,781,580	1,161,491	460,936	2,637,484	159,493	485,720	1,269,583	722,688	2,915,249	381,398	2,348,650	185,201	3,027,614	1,970,373	1,057,241	-
7	12,854,335	3,624,523	1,870,186	1,303,176	451,161	2,714,897	127,111	580,058	1,273,782	733,946	3,363,216	441,812	2,604,751	316,653	3,151,699	1,838,257	1,313,442	-
8	14,957,779	4,160,792	2,262,488	1,458,608	439,696	3,291,406	104,857	447,369	1,590,901	1,148,279	4,087,556	344,277	3,426,154	317,125	3,418,025	2,153,073	1,264,952	-
9	33,458,811	7,833,327	2,498,167	4,208,374	1,126,786	7,237,605	42,223	759,739	5,649,878	785,765	10,121,939	1,044,291	7,972,377	1,105,271	8,265,940	5,575,967	2,689,973	-
99	385	385	-	385	-	-	-	-	-	-	-	-	-	-	-	-	-	-

表 1. 各種秘匿処理法により秘匿されたセルの割合(%)

	LP(セル番号順)	LP(昇順)		Modular	Hyper Cube	Optimal**	一次秘匿				
		clean-up 前	clean-up 後					clean-up 前	clean-up 後		
A 表 単独・本所・支所×産業 ×経営組織 7 区分	6424 セル	収入額	2.925	2.477	2.515	2.352	不可	不可		0.015	
		セル数	14.695	12.360	14.244	12.687				2.226	
B 表 産業×開設時期	3212 セル	収入額	0.163	0.128	0.270	0.251	0.168	0.151*	0.132	0.011	
		セル数	5.791	4.701	6.040	4.981	4.950	5.199*	4.763	1.775	
C 表 産業×従業員規模 6 区 分×経営組織 7 区分	14454 セル	収入額	6.897	6.225	7.065	6.141	16.595	不可		0.014	
		セル数	16.231	14.584	15.345	14.052	27.183			3.508	
D 表 単独・本所・支所×産業 ×資本金 11 区分(国内の会社)	9928 セル	収入額	2.835	2.318	2.859	2.292	3.691	4.479*		0.052	
		セル数	15.743	13.507	15.330	13.276	16.227	20.054*		3.515	
H1 表 県庁所在地 ・主要都市 ×他の市町村 ×産業中間分類 ×経営組織 3 区分	02 県	396 セル	収入額	1.542	1.542	1.821	1.597	1.746	12.608*	4.799 (1.680)	0.139
			セル数	9.596	9.596	10.606	9.343	10.606	11.111*	7.828 (8.333)	2.273
	03 県	396 セル	収入額	1.019	0.964	1.087	0.984	0.342	0.359*	0.329 (1.050)	0.101
			セル数	9.091	8.838	10.101	9.091	7.071	7.071*	6.061 (8.333)	2.020
	04 県	396 セル	収入額	0.826	0.727	0.826	0.727	3.208	6.047	0.938 (2.241)	0.102
			セル数	11.111	9.343	11.111	9.343	10.101	8.081	9.343 (8.333)	2.273
	05 県	396 セル	収入額	4.281	4.269	4.294	4.244	4.338	3.636*	2.081 (3.916)	0.133
			セル数	13.384	12.374	13.889	13.131	12.626	11.364*	10.859 (10.101)	3.030
	06 県	396 セル	収入額	4.472	4.044	3.978	3.777	4.084	4.340*	3.938 (3.777)	0.119
			セル数	13.384	10.354	10.101	7.323	8.838	9.343*	9.343 (7.323)	2.020
09 県	396 セル	収入額	5.501	5.154	5.501	5.154	4.790	1.289*	2.224 (5.154)	0.111	
		セル数	16.919	9.596	16.919	9.596	12.121	9.091*	11.869 (9.596)	2.273	
10 県	396 セル	収入額	8.104	6.793	8.245	6.965	9.084	4.977*	6.917 (7.074)	0.186	
		セル数	17.172	14.141	17.677	13.636	17.929	20.455*	14.141 (12.121)	3.788	

A～C 表は、一次元周辺分布のセル、D 表は、総計、H1 表は各地域区分の総計を保護セルに指定した。Modular 法及び HyperCube 法については、この保護セルの指定のために処理不可となった表がある。

\* HyperCube 法による秘匿結果は、秘匿レンジが±25%以上でないセルが存在する。

\*\* Optimal 法の括弧内数値は、自作 PGM による処理結果。A, C, D 表については処理結果が得られなかった。

確保すべき秘匿レンジは±25%、コスト係数は事業所数の 1 / 2 乗とした。コスト係数を事業所数の 1 / 2 乗としたのは、セルの値（収入金額）の情報ロス最小化を目標とすると秘匿されるセル数が多くなってしまい、逆に秘匿されるセル数の最小化を目標とすると値が大きいセルが秘匿されてしまうため、両者のバランスに配慮した結果である。

処理結果を表 1 に示す。同表から各秘匿処理法について、以下のことが言える。

・ HyperCube 法

処理速度は非常に速いが、一次元周辺分布のセルを保護セルに指定した三重クロス表の A, C 表については、秘匿処理が不可となり、また、ほとんどの表について秘匿レンジが確保されていないセルが存在するなど実用性に関して根本的な問題がある。D 表のように情報ロスが他の秘匿処理法に比べてかなり多くなる場合があり、最適解への近似度が“不安定”である。

・ Modular 法

処理速度は HyperCube 法に次いで速いが、HyperCube 法と同様に一次元周辺分布のセルを保護セルに指定した三重クロス表のうち A 表について秘匿処理が不可となった。C 表は処理できたが、情報ロスが著しく大きくなっており、やはり実用性に関して

根本的な問題がある。比較的構造が単純な二重クロス表である B 表については、比較的情報ロスが少ないが、都道府県別の表 (H1 表) の一部や(1)に挙げた事例のように比較的単純な集計表であっても情報ロスが大きくなる場合があり、最適解への近似度は、比較的単純な構造の集計表であっても“不安定”である。

#### ・ Optimal 法

規模の大きい三重クロス表の A, C, D 表については、少なくとも実験に割くことの出来た時間内に処理結果を得ることができなかった。数千セルの集計表が実用の範囲とみられるが、最適解であるはずの B 表の処理結果は、秘匿セル数、秘匿セル値計いずれも情報ロスがクリーンアップ処理後の順次 LP 法 (セル番号順) を上回っている。表 1 には掲載していないが、最小化すべきコスト係数 (秘匿セル内の事業所数の平方根) の総和もクリーンアップ処理後の順次 LP 法 (セル番号順) を上回っている。そこで、Optimal 法と基本的には同等のプログラムを開発し、県別に分割した小さな表に適用してみたところ、表 1 に示すように Optimal 法とはかなり異なる結果が得られた。表 1 には掲載していないが、最小化すべきコスト係数の総和は、04 県を除くと自作プログラムのほうが小さく、中には順次 LP 法 (+クリーンアップ処理) のほうが小さい県もあった。したがって、 $\tau$ -ARGUS に搭載されている Optimal 法は、多くの場合、実際には最適な秘匿パターンを与えないことが判明した。Optimal 法のアルゴリズムは非常に繊細 (sophisticate) なものであるため、実装プログラムによって結果が異なることは十分に考えられる。これは方法論の純粋な比較から外れた問題という見方であろうが、実用性に力点を置くと、方法論のデリケートさも方法論自体の短所とみなすべきであろう。

#### ・ 順次 LP 法

指定した条件の下で、全ての集計表について秘匿処理ができたのは順次 LP 法だけであった。クリーンアップ処理前では、Modular 法よりも情報ロスが大きい場合があるが、クリーンアップ処理後は、殆どの場合、Modular 法よりも情報ロスが少ないか、ほぼ同等となっている。複雑な構造の集計表を含めて最適解への近似度は、少なくとも他の秘匿処理法に比べると、比較的“安定”していると言えるであろう。

以上のように、どちらかという消去法的に順次 LP 法が他の秘匿処理法に比べて実用的であることが示されたが、次章に述べるように順次 LP 法についても処理時間を含めて課題があり、実務に適用するには様々な工夫が必要である。一方、 $\tau$ -ARGUS の機能に関しては、搭載している秘匿処理法の方法論上の問題という本稿で扱った問題以外にも実用性に問題がある。例えば、 $\tau$ -ARGUS は、対話型処理の機能が優れているが、実務ではバッチ処理のほうが利用し易い。 $\tau$ -ARGUS にも Ver.2 からバッチ処理機能が追加されたが、機能が不十分である。これと関連するが、複数の集計表の整合的な処理に関する機能が実質的に無く、他の集計表の処理で決定された秘匿セルを手で外部指定す

るか、外部指定のために専用のプログラムを開発する必要がある。τ-ARGUS は二次秘匿セルについて確保すべきレンジを出力しないため、この専用プログラムには、単にセルの対応関係をとるだけでなく、確保すべきレンジを求める機能を備えるべきであるが、それには数理計画法のソルバーが必要になり、開発にかなりの労力を要する。HyperCube 法の場合、複数の集計表を一括して処理する機能が一応あるが、前述のとおり HyperCube 法の処理結果に問題があることと、大量の集計表を一括処理することができない等々、実用的でない。τ-ARGUS の開発を行った SDC 及び CASC プロジェクトは既に終了しているが、オランダ統計局がシステムのアップデートを継続しており、将来的にはこれらの問題が解消されるかもしれないが、今後の方針が明確に打ち出されている訳ではない<sup>10</sup>。

#### 4 今後の課題及び最近の集計表開示抑制法の研究動向

##### (1) 順次 LP 法に関する課題

順次 LP 法 (+クリーンアップ処理) は、集計表を分割し階層的に処理するのではなく、集計表全体の構造を考慮しながら処理を行っていくため、構造が複雑な多重クロス表であっても比較的良好な秘匿パターンが安定して得られる。この点では、現在のサービス業基本調査の集計表の秘匿処理に適していると言える。しかし、実務に適用するには処理時間の短縮を図る必要がある。

表 2. 順次 LP 法による処理時間

			セル番号順	昇順
A 表	単独・本所・支所×産業 ×経営組織 7 区分	6424 セル	46m	32m
B 表	産業×開設時期	3212 セル	6m	6m
C 表	産業×従業員規模 6 区分 ×経営組織 7 区分	14454 セル	19h 42m	19h 49m
D 表	単独・本所・支所×産業 ×資本金 11 区分(国内の会社)	9928 セル	3h 02m	3h 19m

PentiumIV3.4GHz 搭載の PC を用いた。

平成 11 年サービス業基本調査全国編の集計表 A~D 表の順次 LP 法による処理時間を表 2 に示す。集計表が二重クロスか三重クロスか、分類区分の階層構造、一次秘匿の数などによってセル数が同じ集計表であっても処理時間が異なるが、表 2 から、集計表のセル数が多くなると、処理時間が急激に長くなることが分かる。地域編の集計表は、さらにセル数が多くなる。都道府県別に分割して処理が可能なものもあるが、一括して処理したほうがよい表もあり、処理時間の問題がより大きくなる。二次秘匿の多少の増加

<sup>10</sup> 今回の実験では Ver.3.0.2 を用いているが、上記の問題については、それ以前のバージョンから特に改善は行われていない。

を許容するのであれば、第2章(5)のiii)に記したように推論開示を防止する処理に変更すると、かなりの処理時間の短縮となる。また、地域編の表について、例えば、東日本と西日本に分割して処理することも考えられる。LP問題の解法として単体法ではなく内点法を用いても処理時間を短縮する効果はなく、二次秘匿を増加させずに処理時間を短縮できる明確な方策は今のところないが、数理計画法ソルバーによって処理が“得意”なLP問題の種類が異なる可能性があるため、他のソルバーも試してみるべきかもしれない。

## (2) 各秘匿処理法共通の課題

### i) 情報ロスの減少

表1に示すようにサービス業基本調査の一次秘匿セルは2~3%程度存在する。二次秘匿セルがその3~5倍程度になるため、三重クロス表では秘匿されるセル数が10%以上になる。見た目には、秘匿箇所が非常に多く感じられるであろう。セル秘匿を行う限り、この程度の秘匿の多さは受け入れなければならないが、これだけ秘匿が多くなると集計表として意味がないと判断するのであれば、三重クロス表を出来る限り避け、原則として二重クロス表に限るべきである。この原則は、次項に述べる複数集計表の整合的処理の問題も考慮すると、採用を検討すべきである。以下に述べる方策は、多少でも秘匿による情報ロスを減少させるためのものであり、根本的に秘匿箇所を減らす方策ではない。

#### ・ Sliding Protection

確保すべき秘匿レンジの上限・下限を固定せずに、上限値 - 下限値  $\geq 2r\%$ の形式で指定する Sliding Protection 方式は、秘匿レンジ固定方式よりも情報ロスが少ない。しかし、Sliding Protection 方式では、秘匿レンジの上限、下限ごとに解いていた LP 問題を一括して処理する必要があり、処理時間が大幅に長くなる。そこで、処理時間の増大を抑制しながら二次秘匿を減らす方法として、順次 LP 法により二次秘匿セルを求めるときは秘匿レンジ固定方式、クリーンアップ処理では Sliding Protection 方式を採用することを試みた。適用結果を表3に示す。二重クロス表の B 表では、目立った効果は見られないが、A 表で秘匿されたセル数が2ポイント減少するなど、三重クロス表では全体的に一定の効果が得られた。

しかし、Sliding Protection 方式では、一次秘匿セルの値が比較的大きい場合に図に示すように推論開示のリスクが高くなることが考えられる。実際に Sliding Protection 方式を適用した結果をみると、このような場合が必ずしもレアケースではないことが分かる。推論開示を抑制するため、利用者がセルの値を $\pm 1\%$ の範囲で推定可能という条件を追加しようとする、Sliding Protection 方式による情報ロスの減少効果は全く無くなってしまう。したがって、全面的に Sliding Protection 方式を採用することは適当でないとみられるが、次項に述べるように、複数の集計表を整合的に処理す

ることに伴う二次秘匿の増大を抑制するために、部分的に Sliding Protection の考え方を採る意味はあるとみられる。

表3. Sliding Protection 方式のクリーンアップ処理により秘匿されたセルの割合 (%)

				LP(セル番号順)			LP(昇順)			
				通常の clean-up	sliding Protection 方式による clean-up		通常の clean-up	sliding Protection 方式による clean-up		
					(1)	(2)		(1)	(2)	
A 表	単独・本所・支所×産業×経営組織 7 区分	6424 セル	収入額 セル数	2.477 12.360	1.928 10.290	2.022 10.476	2.352 12.687	1.844 10.009	1.910 10.554	
B 表	産業×開設時期	3212 セル	収入額 セル数	0.128 4.701	0.122 4.514	0.122 4.514	0.251 4.981	0.246 4.888	0.246 4.888	
C 表	産業×従業員規模 6 区分×経営組織 7 区分	14454 セル	収入額 セル数	6.225 14.584	5.259 13.588	5.263 13.574	6.141 14.052	5.135 13.159	5.146 13.249	
D 表	単独・本所・支所×産業×資本金 11 区分(国内の会社)	9928 セル	収入額 セル数	2.318 13.507	1.728 12.853	1.806 12.903	2.292 13.276	2.022 12.571	2.094 12.621	
H1 表	県庁所在市・主要都市・他の市町村×産業中間分類×経営組織 3 区分	01 県	528 セル	収入額 セル数	0.396 6.250	0.361 5.303	0.426 5.871	0.527 6.061	0.358 5.303	0.480 6.061
		02 県	396 セル	収入額 セル数	1.542 9.596	1.515 9.343	1.515 9.343	1.597 9.343	1.567 9.596	1.567 9.596
		03 県	396 セル	収入額 セル数	0.964 8.838	0.697 7.828	0.727 7.071	0.984 9.091	0.906 7.828	0.906 7.828
		04 県	396 セル	収入額 セル数	0.727 9.343	0.685 8.333	0.685 8.333	0.727 9.343	0.685 8.333	0.685 8.333
		05 県	396 セル	収入額 セル数	4.269 12.374	3.939 11.364	3.939 11.364	4.244 13.131	3.924 10.859	3.939 11.364
		06 県	396 セル	収入額 セル数	4.044 10.354	3.775 7.828	3.775 7.828	3.777 7.323	3.781 8.838	3.781 8.838
		07 県	660 セル	収入額 セル数	4.304 12.879	4.231 11.667	4.272 12.273	3.972 13.182	3.796 11.970	3.796 11.970
		08 県	396 セル	収入額 セル数	4.574 12.626	3.996 10.606	3.996 10.606	4.235 12.374	3.996 10.606	3.996 10.606
		09 県	396 セル	収入額 セル数	5.154 9.596	4.847 9.091	4.847 9.091	5.154 9.596	4.847 9.091	4.847 9.091
		10 県	396 セル	収入額 セル数	6.793 14.141	6.738 13.131	6.738 13.131	6.965 13.636	6.500 12.879	6.503 13.131

Sliding Protection の秘匿レンジは、上限－下限 = 25×2%とした。  
列(2)は、実際の秘匿レンジの上限・下限が、セルの値に一致しないという条件を追加した場合である。

表4. Sliding Protection 方式クリーンアップ処理の場合の処理時間

				セル番号順		昇順			
				通常の clean-up	sliding Protection 方式による clean-up		通常の clean-up	sliding Protection 方式による clean-up	
					(1)	(2)		(1)	(2)
A 表	単独・本所・支所×産業×経営組織 7 区分	6424 セル	46m	44m	43m	32m	43m	42m	
B 表	産業×開設時期	3212 セル	6m	6m	7m	6m	7m	7m	
C 表	産業×従業員規模 6 区分×経営組織 7 区分	14454 セル	19h 42m	21h 00m	21h 20m	19h 49m	23h 26m	22h 37m	
D 表	単独・本所・支所×産業×資本金 11 区分(国内の会社)	9928 セル	3h 02m	3h 50m	3h 45m	3h 19m	4h 14m	3h 25m	

PentiumIV3.4GHz 搭載の PC を用いた。順次 LP 法により二次秘匿セルを求める時間を含む。

図. Sliding Protection 方式の推論開示リスク

10,000	1	9,999
↑	↑	↑
合計のセル	二次秘匿セル	一次秘匿セル

- ・ 一次秘匿セルが取り得る値は 0~10,000 となり、秘匿レンジ 100%超確保されている。
- ・ しかし、二次秘匿セルの値が非常に小さいと推測できる場合、例えば 2 以下であると推定される場合、一次秘匿セルが取り得る値は 9,998~10,000 となり、秘匿レンジは 0.2%と非常に狭くなってしまふ。

・ 免責アプローチ (Waiver approach)

免責アプローチは、一次秘匿セルに分布された客体の了解を得て、秘匿を行わないようにする方法である。一次秘匿セルに分布される全ての客体の了解を得ることは数が多くなるため現実的ではないが、カナダ統計局の Provençal & Bérard (2003), Provençal et al. (2004)は、二次秘匿を減少させる効果の大きい、免責アプローチ対象セルの選定方法を提案している。

例えば、特別な研究を行うために、特定の地域や産業小分類区分に限定して免責アプローチを適用するような場合を除くと、カナダ統計局が提案している方法を用いたとしても、免責アプローチを適用する効果は小さい可能性がある。しかし、次項に述べるように複数の集計表を統合的に処理する作業過程で、二次秘匿の増大を抑制するために秘匿レンジの基準を緩和したい一次秘匿セルが自動的に絞られることから、これらの一次秘匿セルを免責アプローチの対象、あるいは、これらの中からカナダ統計局の方法で選定するほうが効果的かもしれない。

ii) 複数の集計表の統合的処理法

実際の業務では、複数の集計表を統合的に秘匿処理しなければならない。全部の集計表のセル数の総計がそれほど大きくない場合は、表間のセルの対応関係を制約条件に加えて (順次 LP 法であれば、第 2 章(4)に示した式中の行列 **M** を然るべく変更すればよい。)、一つの集計表のように処理すればよいが、セル数の総計が多くなると、表ごとにあるいは表をいくつかのグループに分けて処理する必要がある。しかし、二次秘匿の増大を抑制しつつ、統合的に処理するのは困難なことが多い。この問題の対応策として、確たるものはないが、今回若干検討したことを以下に述べる。

表 5-1 に確保すべき秘匿レンジを ±25%に指定して D, A, B, C 表の順に逐次処理した結果を示す。第 1 回目は、まず、他表との整合性を考慮せずに D 表の秘匿処理を行い、次にこの処理で求められた D 表の二次秘匿セルのうち、A 表に対応するセルが存在するものについて一種の一次秘匿セル (既決定二次秘匿セル) として扱い A 表の秘匿処理を行う。既決定二次秘匿セルの確保すべき秘匿レンジは、一次秘匿セルの

秘匿レンジ（今回の実験では基本的には 25%）ではなく，D 表の処理過程で各セルに対応する変数  $y_i^+, y_i^-$  がとる値の最大値を用いる<sup>11</sup>。同様に B 表も D, A 表の処理で求められた二次秘匿セルのうち B 表に対応するセルが存在するものを既決定二次秘匿セルとして扱う。第 2 回目では，D 表についても A, B, C 表の第 1 回目の処理結果から既決定二次秘匿セルを取り出し，整合的な処理を行い，以下，同様に A, B, C 表の処理を順次行う。第 2 回目の処理でいずれの表についても新たな既決定二次秘匿セルが発生しなければ処理は完了する。いずれかの表で新たな既決定二次秘匿セルが発生した場合は，第 3 回目の処理が必要となる。今回の実験では，処理時間の問題もあって，処理を打ち切ったが，実際には 3 回以上処理を繰り返す必要がある。なお，D 表を最初に処理したのは，同表が国内の会社のみを集計対象としており，集計対象が狭い表から処理したほうが，途中で秘匿処理不可になる可能性が小さいとみられるためである。

表 5-1 は，逐次処理を行うと二次秘匿が大幅に増えることを示している。例えば，A 表は，秘匿されるセルの数が各表個別に処理した場合の 12%から 28%に著増しており，許容し難い結果と言える。

表 5-1. 複数集計表の逐次処理により秘匿されたセルの割合 (%)

				秘匿レンジ 25%		秘匿レンジ 12.5% A, D 表統合	
				各表個別 処理	逐次* 処理 2 回	各表個別 処理	逐次** 処理 1 回
D 表	単独・本所・支所×産業 ×資本金 11 区分(国内の 会社)	9928 セル	収入額	2.318	6.210	3.037	3.037
			セル数	13.507	15.199	12.983	12.983
A 表	単独・本所・支所×産業 ×経営組織 7 区分	6424 セル	収入額	2.477	10.418	2.368	2.371
			セル数	12.360	28.347	11.924	12.158
B 表	産業×開設時期	3212 セル	収入額	0.128	0.128	0.128	0.128
			セル数	4.701	4.701	4.701	4.701
C 表	産業×従業員規模 6 区分 ×経営組織 7 区分	14454 セル	収入額	6.225	10.126	5.576	6.731
			セル数	14.584	21.171	13.636	14.888

順次 LP 法（セル順）＋クリーンアップ処理を用いた。一次元周辺分布に対応するセルは保護セルに指定しており，他表との整合性の問題はないため，二重クロス表の C 表は個別処理と逐次処理の結果は同一になる。

\* 逐次処理 2 回でも未収束ため，実際にはさらに秘匿されるセルが多くなる。

\*\* C 表の一次秘匿セル 3 つについては，例外的に秘匿レンジ 5%で処理，そのうち実際のレンジが特に狭い 1 つ（表 6-2 参照）は，二次秘匿セルを追加して秘匿レンジの下限 100%，上限 20%とした。

<sup>11</sup>  $\epsilon$ -ARGUS は，二次秘匿セルが確保すべきレンジを出力しないため，逐次処理を行う場合，別途，既決定二次秘匿セルの確保すべきレンジを求める必要がある。

表 5 - 2. 秘匿レンジ 5%で処理した C 表の一次秘匿セル 3 つの実際の秘匿レンジ

下限	上限	レンジ計
78%	6%	84%
27%	7%	34%
7%	5%	12%

表 6 - 1. 基準緩和の効果 (C 表)

		秘匿レンジ 25%	秘匿レンジ 12.5%	一次秘匿セル	194 セル
秘匿されたセルの割合(%)	収入額セル数	6.225 14.584	5.576 13.636	うち上限・下限いずれかが25%未満	46 セル
処理時間*		19h 42m	15h 35m	うちレンジの合計が50%未満	14 セル

\* PentiumIV3.4GHz 搭載の PC を用いた。

表 6 - 2. 秘匿レンジ 12.5%で処理した場合に実際のレンジ幅が 50%未満の一覧 (C 表)

下限	上限	レンジ計
26%	19%	45%
16%	28%	44%
14%	29%	43%
27%	15%	42%
26%	15%	41%
22%	19%	41%
16%	23%	39%
24%	15%	39%
22%	16%	38%
18%	20%	38%
19%	18%	37%
17%	15%	32%
13%	17%	30%
15%	13%	28%

逐次処理で許容できる結果を得るには、何らかの基準緩和措置が必要とみられる。確保すべき秘匿レンジを±12.5%に狭めて C 表に適用した結果を表 6 - 1 に示す。C 表単独で処理を行った場合、秘匿されるセル数は全体の 14.6%から 13.6%へと若干減少する。C 表の一次秘匿セル 194 のうち、秘匿レンジ 12.5%の処理で実際の秘匿レンジの上限・下限どちらかが 25%未満のセルは 46 あった。このうち、上限と下限を合計したレンジ幅が 50%未満のセルは 14 あった。ここで、Sliding Protection 方式を部分的かつ結果的に取り入れるのであれば、追加の二次秘匿が必要なのは少なくとも 14 セル以下に限定される。表 6 - 2 にこれら 14 セルの実際の秘匿レンジを掲げる。秘匿レンジをある程度柔軟に考えてよいのであれば、二次秘匿の追加が必要なセルはさらに絞り込める。なお、秘匿レンジを狭めることによって、処理時間をある程度短縮できる効果もある (表 6 - 1 参照)。

秘匿レンジを狭めることによる秘匿セルの減少は、各表個別の処理でみた場合、それ程大きくはないが、これによって複数集計表の逐次処理の回数を少なくすることが

できれば、大きな効果が得られたことになる。ここでは逐次処理を基本的には 1 回で終わらせることを目指して、A, D 表を統合し、確保すべき秘匿レンジを±12.5%として逐次処理を行ってみた。前の表の処理で秘匿されなかったセルは保護セルに指定した。つまり、他表に対応するセルがあるものは、先に処理した表の結果を変更しないように指定した。このようにすると秘匿処理不可となる一次秘匿セルが出てくるが、その場合は、秘匿レンジを 5%に引き下げて取りあえず 1 回の逐次処理で整合的な処理を終わらせるようにした。この結果、秘匿レンジを 5%に引き下げたセルが C 表に 3 つあった。これら 3 セルの実際の秘匿レンジの合計は、表 5-2 に掲げるように、それぞれ 84%, 34%, 12%となり、レンジが非常に狭いセルは 1 つだけであった。表 5-1 に示す処理結果は、このセルに対してだけ秘匿レンジ 12.5%で追加的な処理を行った後のものである。秘匿されるセルの割合の増大は抑制されている。

追加的な処理が必要な一次秘匿セルの数を、少数に絞ることができるのであれば、免責アプローチを適用することも有効な方法かもしれない。

以上のように、実際の業務適用では秘匿の基準をある程度柔軟に考える必要があるが、具体的にどのような方法を採用するかは各業務それぞれに判断する必要があり、それ程簡単ではないかもしれない。特に特別集計やオーダーメイド集計のような公式統計表の作成後に行う追加的な集計の場合、公式統計表の秘匿パターンと整合をとりながら秘匿処理を行う必要があり、許容範囲内の基準緩和では秘匿処理不可になる可能性がある。秘匿処理ができたとしても、追加的な集計表では二次秘匿がかなり多くなってしまいう可能性が高い。

複数集計表間の整合的な秘匿処理の問題をなるべく避けるには、秘匿処理法の選択・改良や基準の緩和といったことを考える前に、集計表の設計段階で三重あるいはそれ以上の多重クロス表をなるべく作成せず、原則として二重クロス表に限るといった配慮が重要である<sup>12</sup>。なお、次項に述べるセル秘匿以外の開示抑制法を採った場合でも、程度に差があるとしても何らかの問題が生じるため、同様の原則が意味を持つことに留意する必要がある。

### (3) セル秘匿以外の開示抑制法の採用可能性

集計表における開示抑制法には、セル秘匿以外の方法もある。これらは、センシティブなセル（セル秘匿方式でいう一次秘匿セル）を実際のセルの値と異なるものに置き換え、これに伴って他のセルの値も変更する方法である。その中にはセル秘匿方式の問題をある程度解消できると期待されるものが存在する。

丸め法や、丸め法に確率的技法を取り入れたランダム・ラウンディング（小野（1988）参照）などはセル秘匿以外の開示抑制法に該当する。ランダム・ラウンディングは、カ

<sup>12</sup> Provençal & Bérard (2003)には、経済調査の集計表は通常二重クロス表であるといった記述があり、少なくともカナダ統計局では、なるべく二重クロス表を作成するようにしていることが窺える。

ナダ人口センサスの小地域集計に利用されている (March & Norris (1987), Boudreau et al. (2004)) ように度数表 (frequency table) の秘匿処理に適しているが、経済調査の集計表 (magnitude table) の秘匿処理には適さない。Cox & Dandekar (2002)は、CTA (Controlled Tabular Adjustment) という経済調査の集計表の秘匿処理にも適用可能な方法を提案している。CTA は、各セルの値の変更を MIP 問題あるいは LP 問題によって求める方法であり、順次 LP 法の変型と見ることもできるかもしれない。順次 LP 法は、最適な秘匿パターンを求める問題を、センシティブセル以外に値を変更するセルをなるべく少数に抑えるようにした LP 問題で近似させる方法であったが、CTA は、多くのセルの値を変更する代わりに、各セルの値の変更幅をなるべく小さく抑える MIP 問題あるいは LP 問題により定式化される。

$$\begin{aligned} & \min \sum_i c_i (y_i^+ + y_i^-) \\ & \text{s.t.} \\ & \mathbf{M}(\mathbf{y}^+ - \mathbf{y}^-) = 0 \\ & 0 \leq y_i^+ \leq \delta a_i, 0 \leq y_i^- \leq \delta a_i \quad (i: \text{非センシティブセル}) \\ & r \cdot a_i \leq y_i^+, y_i^- = 0 \text{ or } r \cdot a_i \leq y_i^-, y_i^+ = 0 \quad (i: \text{センシティブセル}) \end{aligned}$$

上式<sup>13</sup>の $\delta$ は、非センシティブセルの値の最大変更率に相当し、なるべく小さい正値を用いたほうが良いが、小さくし過ぎると解が存在しなくなる。センシティブセルの値の変更の方向 (増減) は、コスト (情報ロスに相当) をなるべく小さくする組み合わせを選択したほうが良い。厳密にコストを最小化するには、センシティブセルと同じ数の 0-1 変数を含む混合整数計画法 MIP 問題を解かねばならない。しかし、例えば、センシティブセルの値を降順に並べ、値の変更の方向を+、-交互に指定する方法などで近似することも考えられる。Cox (2005)による最新の研究では、このような単純な近似法では (MIP 問題を解いて得られる) 最適解に比べて情報ロスが 100%以上になるため、メタ・ヒューリスティックな方法でセンシティブセルの値の変更の方向を求めることを提案している。米国運輸統計局 BTS の委託の下で米国の OptTeck 社もメタ・ヒューリスティック法の一つであるタブー探索法<sup>14</sup>を用い近似解を求める研究を行っている (Russell & Kelley (2003), Dulá et al. (2004)参照)。この CTA は、センシティブセルの箇所を開示しないことを前提にしている。サービス業基本調査の場合、各セルに分布される客体の件数 (正確には事業所数) を秘匿しないため、センシティブセルを特定でき、この前提が成り立たないが、例えば、上式の  $r$  を一定にせず、確率的に変動させることにより、開示リスク

<sup>13</sup> この定式化は、各セルの値の最大変更幅をオリジナルのセル値に対する比率で与えており、Giessing & Castro (2005)が restricted CTA と呼んでいるものに相当する。

<sup>14</sup> 例えば、松井知己氏のホームページを参照。遺伝的アルゴリズムもメタ・ヒューリスティック法の一つである。(http://www.simplex.t.u-tokyo.ac.jp/~tomomi/comb-opt/heuristic/ppframe.htm)

を抑制できるとみられる。

簡易化した方法やメタ・ヒューリスティックな方法によって情報ロスが十分少なくできるのであれば、CTA は、順次 LP 法に比べてより大きな集計表の処理あるいはより多くの集計表の一括処理が可能になると期待される。非センシティブセルの値の変更幅は小さいため、推論開示の問題もない。第2章(5)のiv) で述べた Singleton セルの問題も解消されると思われる。CTA を用いる他の利点として、分析的な利用を考慮した処理を行える可能性があることが挙げられる。Cox et al. (2004)は、分析的な利用を考慮したCTA の定式化を試みている。この場合、セルの値の変更幅最小化のコスト関数と他の何らかの関数の合計を最小化する問題を解くことになる。

しかし、このようなCTA の改良が成功したとしても、多重クロス表を用いた分析結果の偏りを十分小さくできるか、あるいは、オーダーメード集計のような追加的集計において、既存の集計表との整合性をとるためにセルの値の変更幅が無視できないほど大きくなれないか等々の問題は残る。

複数集計表の統合的処理やオーダーメード集計、Singleton セルへの対応などを容易にする方法として、集計に用いる個票データに対して何らかの処理を施し、集計表段階での処理をなるべく少なくする方法が考えられる。1990、2000年米国人口センサスで採用されている Confidentiality edit は、これに該当する。個票データの開示抑制処理は、主として一部の世帯を他の地域の世帯と交換（スワッピング）することにより行い、このように処理した個票データを公式集計及びリモート集計用データ、一般公開マイクロデータとしてほぼ共通に使っている（Hawala et al. (2004), Zayatz (2002), Griffin et al. (1989))<sup>15</sup>。

米国人口センサスのスワッピングを主体とする方式は、全国レベルでも開示リスクが大きい経済調査には適さない。経済調査に適用可能な方法としては、Evans et al. (1998) が個票データにノイズを付加する方式を提案している。具体的には、全ての客体に対して値を増加させるか減少させるか無作為に選定し、増加させるものには  $1+s$ 、減少させるものには  $1-s$  を乗じる。 $s$  は平均が  $\mu$  で分散が十分小さい確率分布から発生させた乱数を用いる。平均  $\mu$  は、非センシティブセルの値の誤差の増大を抑制するには小さいほうがよいが、センシティブセルの値を実際の値から十分乖離させるには大きいほうがよい。センシティブセルに分布される客体とそうでない客体で付加するノイズの（セルの値に比べた相対的な）大きさを変えることができればよいが、あらゆる集計表に対応できるようにするにはノイズの平均的な大きさを一律にする必要がある。Evans らは、例示的に  $\mu=10\%$  を用いているが、経済調査の種類によっては十分な大きさではないように思われる。多重クロス表の作成を抑制するなどの原則を設けるならば、センシティブセルに分

<sup>15</sup> Zayatz (2005)によると、人口センサスのロングフォームの代替として継続調査として実施される American Community Survey では、スワッピングの代わりに、当てはめたモデルから推定した値に置き換える部分合成データ方式を用いる可能性が検討されている。これは、同時点で調査を行うサンプルが限られるため、スワッピング方式を採用しにくくなるためと推測される。

布される可能性の高い客体を推定し、それらの客体には大きなノイズを付加するといった工夫の余地はありそうに思われる。少なくとも、予め決められた集計表に限定してよいのであれば、事前に集計を行い、センシティブセルに分布される客体を完全に特定できるため、これらの客体にだけ大きなノイズを付加し、他の客体のノイズをより小さくあるいはゼロにすることができる。客体が2件分布されるセンシティブセルの値が、2件の客体に付加されているノイズが互いに相殺されて本来の値に非常に近くなるのを避けるように調整することも可能になると思われる。主要なセルの値について、ノイズを付加しない本来の値と一致させたいならば、比推定など事後層化の技法を利用して調整することも考えられる。このように実務への適用を考えると様々な工夫が必要になる可能性があるが、セル秘匿方式の場合の処理の困難さに比べると遥かに実用性が高いと思われる。

Evans らの個票データにノイズを付加する方式は、米国の四半期労働力指標 (Quarterly Workforce Indicators) に採用されている (Zayatz (2005), Abowd et al. (2005))。この新しい統計は、就業異動の動向を詳細に捉えることを目的として、事業所と個人の情報をリンクしたパネルデータから作成されている。新たな開示抑制法を採用した理由は不明であるが、売上高などに比べると客体間の格差が小さく、付加するノイズの大きさも比較的小さいもので済むことや、分析的な利用が多く、正確な値を提供することよりも分析に利用し易いデータを提供することのほうが有益であると判断したことなどが推測される。いずれにせよ、新たな試みとして注目される。

## 5. おわりに

本稿では、まず、4種類のセル秘匿方式を実証的に比較検討し、少なくともサービス業基本調査への適用を前提とすると、順次 LP 法が優っていることを明らかにした。しかし、セル秘匿方式共通の問題として、複数集計表を統合的に処理することが容易でないことも示した。これに対する方策として、sliding protection 方式や免責アプローチの部分的な利用を提案するとともに、セル秘匿方式以外の開示抑制法の最近の発展により、セル秘匿方式が有していた問題をある程度解消できる可能性があることを述べた。

全体的に見ると、CTA や個票データにノイズを付加する方式のほうが、セル秘匿方式よりも優っているように思われるが、問題のないセルの値を変更することに躊躇いはある。国際的に見ても、殆どの経済調査では、依然、セル秘匿方式が採用されていることも考慮する必要がある。また、March & Norris (1987)によると、カナダ農業センサスの集計表開示抑制法をノイズ付加方式に変更することを利用者に提案したところ、従来のセル秘匿方式を好む利用者が多く、結局変更されなかった事例があり、利用者が必ずしもセル秘匿方式以外の開示抑制法を支持していないかもしれないことにも留意する必要がある。実際に集計結果がどの利用されているのかを把握しつつ、集計表開示抑制法の研究を進めていく

必要があろう。

#### 参 考 文 献

- Abowd, J.M., Stephens, B. and Vilhuber, L. (2005). Confidentiality protection in the Census Bureau's Quarterly Workforce Indicators, *Proceedings of the Joint Statistical Meeting 2005*, to appear.
- Boudreau, J.-R., Filep, K. And Liu, L. (2004). Iterative rounding for large frequency tables, *Proceedings of the Joint Statistical Meeting 2004*.
- CASC-project (2004).  $\tau$ -ARGUS version 3.0 User's Manual. (<http://neon.vb.cbs.nl/casc/Software/taumannualv3.pdf>)
- Cox, L.H. (2005). Confidentiality protection by controlled tabular adjustment using metaheuristic methods, *Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Working Paper No.32*.
- Cox, L.H. and Dandekar, R.A. (2002). A disclosure limitation method for tabular data that preserves data accuracy and ease-of-use, *Report of FCSM Statistical Policy Seminar, Statistical Policy Working Paper 35*, US Federal Committee on Statistical Methodology, pp.15-30.
- Cox, L.H., Kelley, J.P. and Patil, R. (2004). Balancing quality and confidentiality for multivariate tabular data, in J. Domingo-Ferrer and V. Torra (eds.) *Privacy in Statistical Databases*, Berlin: Springer-Verlag, pp.87-98.
- Dulá, J.H., Fagan, J.T. and Massell, P.B. (2004). Tabular statistical disclosure control: Optimization techniques in suppression and controlled tabular adjustment, *Research Report #2004-04*, US Census Bureau.
- Evans, B.T., Zayatz, L. and Slanta, J. (1998). Using noise for disclosure limitation for establishment tabular data, *Journal of Official Statistics*, Vol.14, pp.537-552.
- Fagan, J.T. (2001). Cell suppression problem formulations – exact solution and heuristics, *SDC Research Paper*, US Census Bureau.
- Fischetti, M. and Salazar-González, J.J. (2000). Models and algorithms for optimizing cell suppression in tabular data with linear constraints, *Journal of the American Statistical Association*, Vol.96, pp.916-928.
- Giessing, S. and Castro, J. (2005). Testing variants of minimum distance controlled tabular adjustment, *Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Working Paper No.40*.
- Griffin, R., Navarro, F. and Florez-Baez, L. (1989). Disclosure avoidance for the 1990 census, *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp.516-521.
- Hawala, S., Zayatz, L. and Rowland, S. (2004). American FactFinder: Disclosure limitation for the Advanced Query System, *Journal of Official Statistics*, Vol.20, pp.115-124.

- Kelley, J.P., Golden, B.L. and Assad, A.A. (1992). Cell suppression: Disclosure protection for sensitive tabular data, *Networks*, 22, pp.397-417.
- March, M.J. and Norris, D.A. (1987). Disclosure avoidance techniques in the Canadian censuses of population and agriculture, *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp.233-238.
- Massell, P.B. (2004). Protecting sensitive cells in a cell suppression program using sliding protection, *Research Report #2005-02*, US Census Bureau.
- Provençal, J.-S. and Bérard, H. (2003). Waivers in business surveys: A systematic approach to increase the amount of publishable information, *2003 FCSM Conference Papers*, Federal Committee on Statistical Methodology. (Provençal et al. (2004)による改良を追記して本参考資料に翻訳 (村田磨理子訳) を収録)
- Provençal, J.-S., Bérard, H., Fillion, J.-M. and Tambay, J.-L. (2004). Approaches to identify the amount of publishable information in business surveys through waivers, in J. Domingo-Ferrer and V. Torra (eds.) *Privacy in Statistical Databases*, Berlin: Springer-Verlag, pp.110-120.
- Repsilber, R.D. (1994). Preservation of confidentiality in aggregated data, *paper presented at the 2<sup>nd</sup> International Seminar on Statistical Confidentiality*, Luxembourg.
- Russell, J.N. and Kelley, J.P. (2003). Bureau of Transportation Statistics' prototype disclosure limitation software for complex tabular data, *Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Working Paper No.39*.
- Sande, G. (1984). Automated cell suppression to preserve confidentiality of business statistics, *Statistical Journal of the UN/ECE*, Vol.2, pp.33-41.
- UN/ECE and Eurostat (2003). Glossary on Statistical Disclosure Control (Draft). (統計センター研究センター訳 (2004). 統計データ開示抑制に関する用語集暫定版 (対訳), 「製表関連国際用語集」, No.2, <http://www.nstac.go.jp/services/pdf/skk-yogosyu2.pdf>)
- Wolf, de P.P. (2002). HiTaS: A heuristic approach to cell suppression in hierarchical tables, *Proceedings of the AMRADS meeting in Luxembourg*.
- Zayatz, L. (2002). SDC in the 2000 US decennial census, in J. Domingo-Ferrer (ed.) *Inference Control in Statistical Databases*, Berlin: Springer-Verlag, pp.183-202.
- Zayatz, L. (2005). Disclosure avoidance practices and research at the US Census Bureau: An update, *Research Report #2005-06*, US Census Bureau.
- 岡本政人 (2005). 集計表秘匿処理プロトタイプ・システム SCS の開発状況, 「2005 年度統計関連学会連合大会報告集」.
- 小野達也 (1988). ランダム・ラウンディングの理論と方法, 「統計局研究彙報」, 第 46 号, pp.97-135.
- 松本正博 (1996). サービス業基本調査の秘匿処理法について, 「統計局研究彙報」, 第 54 号, pp.57-86.

## Waivers in business surveys: A systematic approach to increase the amount of publishable information

企業調査における秘匿の免責：公開可能な情報を増やすための系統的アプローチ

Jean-Sébastien Provençal, Hélène Bérard

Statistics Canada

11-L and 11-K, R.H. Coats Building, Ottawa, Canada K1A06T6

provjea@statcan.ca and berahel@statcan.ca

### 1. 序論

集計表データの公表において、情報（セル）の秘匿は、回答者のデータの機密保持のために使われる一般的な方法である。企業調査では、秘匿のパターンは通常2つの基準によって決まる。すなわち、回答者数の下限と占有ルールである。占有ルールでは、例えば、単独の企業がセル合計の80%より多くを占める場合、そのセルは秘匿される。

少数の大規模企業と多数の小規模企業から構成される非常に偏った企業母集団のために、セルの秘匿は、しばしば、セクター（部門）を支配する単独またはいくつかの企業が存在することから生じる（占有ルール）。あるセルの秘匿は、残存する開示リスクを減らすために、他のセルの秘匿を必要とすることになる。より多くのセルを公表しようとする努力の一つである秘匿の免責（waiver）は、セルに大きな寄与を持つ個体から、開示リスクにもかかわらずセルのデータの公開の許可を得るものである。これらの企業は、しばしば、アドホックな方法で選ばれる。どのセルの回答者から免責を得るかによって、結果として異なる秘匿パターンをもたらす可能性があり、その中のいくつかは、他のものと比べて、調査や利用者の必要とする条件次第で、より適したものになり得る。

この論文では、最適な秘匿パターンを導く潜在的な免責を特定することを意図した方法を示す。最初に、機密保持と開示に関するいくつかの概念について述べる。2番目に、カナダ統計局で秘匿すべき情報を特定するために使われているソフトウェア CONFID について述べる。このソフトウェアで使われている方法論を要約する。3番目に、秘匿される情報の量を最小にするためにどのように免責が利用できるか議論し、調査や利用者の必要とする条件の点でより適した免責対象を特定するために開発した方法を説明する。最後に、カナダの製造業年次調査を使った実例から得られた結果について論じる。

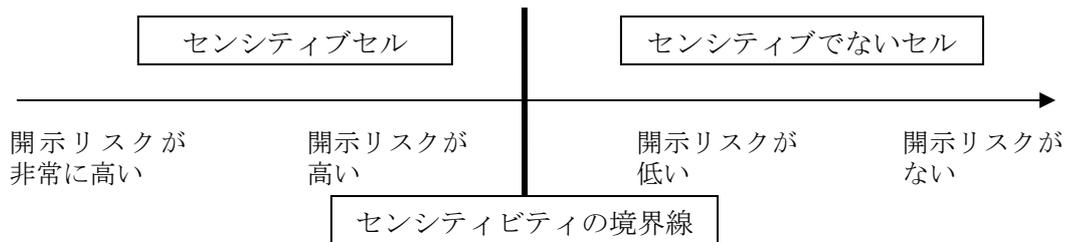
### 2. 機密保持の概念

#### 2. 1 免責

免責は統計機関、この場合はカナダ統計局と回答者の間の協定書である。この協定の下で、回答者は、統計機関が回答者の機密を保持するための一切の義務を免除する。

## 2. 2 センシティブセル

機密保持の分野では、セルの潜在的な開示リスクを評価するためにセンシティブティ (sensitivity) の概念を用いる。セルに開示リスクが予想される時、そのセルはセンシティブ (sensitive) であるとみなす。その境界線は、どの程度の安全性を必要とするかを表す機密保持ルールによって決められる。



このリスクを定量化するために、機密保持ルールを用いる。ルールの1つとして、(n,k)ルールがある。このルールでは、セルのセンシティブティはそのセルの主要な個体 (contributors) によって定義される。上位 n 個体がセルの合計値の k%以上である場合に、そのセルはセンシティブであるとする。このルールは関数として定義することができる。

例えば、(n,k)ルールで n=3, k=yy の場合では、値の大きな方から3つの個体  $X_{(1)}, X_{(2)}, X_{(3)}$  がセルの合計値の yy%以上を占めるときに開示リスクがあると言える。

センシティブティ関数 (sensitivity function)  $S(x)$  は、次のように定式化される。

$$S(x) = \frac{100}{yy} [X_{(1)} + X_{(2)} + X_{(3)}] - t(x)$$

ここで、 $t(x)$  はセル合計値を示す。 $S(x) < 0$  のとき、セルはセンシティブであるとみなす。 $S(x)$  の値によって、セルのセンシティブティの程度を測ることができる。

セルがセンシティブかどうかを決めるとき、ユニット (unit) 数に関するほかの基準を追加することもできる。たとえば、セルのユニット数が3未満の場合に、そのセルはセンシティブであるということが出来る。

## 2. 3 補足的セル<sup>1</sup>

あるセルがセンシティブであると特定された場合は、そのセルには開示のリスクがあるために、公表せずに隠される。公表される表が合計や小計を含む場合、センシティブセルを保護するために、さらに他のセルの除外を必要とする。これらの秘匿によって、隠した値を表の他の値から推測することが不可能になる。これらの補足的セルは残存する機密の保持のために隠される。例えば、次の表を考える。

<sup>1</sup> 二次秘匿セルと同義。

	1	2	3	合計
A	12	8	9	29
B	4	4	6	14
C	5	3	3	11
合計	21	15	18	54

セル C2 はセンシティブセルであるために隠すと仮定する。C2 を保護するために補足的セルも選択しなければならない。したがって、次の結果を得る。(S はセンシティブセルを意味し、C は補足的セルを意味する。)

	1	2	3	合計
A	12	8	9	29
B	C	C	6	14
C	C	S	3	11
合計	21	15	18	54

セル C2 を保護するため、B1、B2、C1 を隠す必要がある。もしこの 3 つのセルのうち 1 つでも隠されないとすると、C2 の値を推測することができる。例えば、B1 が隠されていない場合、 $C1=21-12-B1=5$ 、 $C2=11-3-C1=3$  と計算でき、C2 の機密は維持できなくなる。

機密保持を確実にするために他のセルの組み合わせ、例えば、B2、B3、C3 を除外することもできる。補足的セルの選択は、様々なパラメータや考慮点に依存する。例えば、ひとつの考慮点は、センシティブセルにおける主要な変数のサイズである。補足的セルは、セル C2 の機密保持を守るために十分大きい値をとる必要がある。必要とされる補足的セルのサイズは、前に述べた機密保持ルールに依存する。

### 3. CONFID の方法論

CONFID は、機密保持ルールに従ってセンシティブセルを特定し、原因となる各ユニットの機密性を維持する表を作成するために隠す必要のあるすべての補足的セルを特定するソフトウェアである。

与えられた問題、すなわち、1 組の表の作成に対して、複数の可能な解が存在する。それらの解は、機密保持パターンと呼ばれる。CONFID は、事前に与えられたいくつかの基準に従って、最適解を探す。

目標は、できる限り少ないセルだけを隠し、できる限り少ない情報だけを保留することである（公開できない情報をできる限り少なくすることである）。ときには、2 つの小さなセルの公表か、1 つの大きなセルの公表かを選択しなければならないことがある。その選

扱は調査と公表の目的に依存するだろう。CONFID は金額やセル数の情報量に優先順位をつけることができる。CONFID では、2つの基準の中間の条件での最適な機密保持パターンを探すことも可能である。CONFID は、公表される集計表、データセットそれ自体、機密保持ルールや最適化基準を考慮した連立数値方程式を解くことによって最適解を探すことができる。

#### 4. 免責の利用

CONFID を使って得られた機密保持パターンによって、かなりの量の情報が隠されることがある。これは、少数のユニットの小さなセルからなる非常に詳細な集計表を作成するような調査を行ったときに発生する。免責の利用は、隠される情報の量を減らすために役立つ。次の2つの点で、免責は公開できる情報の増加をもたらす。

- 1) セルのセンシティブ性は、1つまたは2, 3の大きな個体に起因する。これらの企業から免責を得ることでセル全体を公表することができる。
- 2) セルのセンシティブ性を除去したとき、そのセンシティブセルを保護するために当初必要とされた補足的セルは、隠される必要がなくなる。これらのセルを公表することは、金額やセル数に関して相当な情報量の増加をもたらす。

#### 5. 新しいアプローチ：スコア関数

コストや業務上の制約のために、センシティブなユニットそれぞれの免責を得ることは、常に実行できるとは限らない。免責を得ようとするユニットを精選して、サブセットを選ぶ必要がある。我々のアプローチでは、どのユニットを優先すべきかを系統的に特定することを試みている。公開できる情報量の点で最善の結果を与える免責を見つけることを目標とする。言い換えると、どのセンシティブなユニットが、センシティブセルと補足的セルによって秘匿された情報の最大の原因であるかを特定することである。潜在的な免責要求度のランク付けを可能にするために、いくつかの基準を決めた。これらの基準は、センシティブなユニットを含むセルやブロックの特性に基づく。セルやブロックは集計表の2つのレベルである。セルは公表される最も詳細な要素に相当する。企業調査では、1つのセルは、通常、2つの次元から作られる。一般に、4桁の標準産業分類のサブセットのような詳細な産業の次元と、州といった地理的な次元を含む。ブロックは、セル（例えば、州別の4桁産業分類）を結合したもので、1番目の次元のカテゴリーの合計、2番目の次元のカテゴリーの合計や総合計を含む。

	地域A	地域B	地域A + B
産業1	CELL 1A	CELL 1B	TOT 1,A+B
産業2	CELL 2A	CELL 2B	TOT 2,A+B
産業1 + 2	TOT 1+2,A	TOT 1+2,B	TOT 1+2,A+B

## 5. 1 セルのための基準

我々はセルを評価するために使うことのできる2つの特性を定義した。各基準（特性）は比率として表され、パーセンテージで示される。

5. 1. 1) セルが完全に公開可能となるまでに必要な免責の数

5. 1. 2) 主要な変数に基づくセルのサイズ

これらのパーセンテージを組み合わせて単一のスコアを得る。両方の基準は重要度に従って重み付けされる。次に、各基準の詳細を説明する。

### 5. 1. 1 必要な免責の数

必要な免責の数がより少ないセルに取り組むほうがよい。その場合の基準は、あるセルを公開可能とするために必要な免責の数と、すべてのセンシティブセルの中で、必要とする免責の数が最大であるものとの比に基づく。この基準では、より少ない数の免責を必要とするセルが最もよい。基準は次のように定義される。

$$ce_1 = \frac{(\max(x) - x)}{\max(x)} \times 100$$

ここで、 $x$ はセルのセンシティブティを除去するために必要な免責の数で、 $\max(x)$ は、すべてのセンシティブセルの中で $x$ が取りうる最大値である。 $ce_1$ は必要な免責の数が小さいとき、高いパーセンテージとなる。

### 5. 1. 2 主要な変数に基づくセルのサイズ

次に、主要な変数に基づくセルの相対的なサイズを検討する。相対的なサイズを使う理由は、最も大きいセルに高い優先度を与えるためである。この基準の式は、該当の変数に関して、以下に定義される。

$$ce_2 = \frac{y}{\max(y)} \times 100$$

ここで、 $y$ は関心のある変数のセル内の合計値で、 $\max(y)$ はすべてのセンシティブセルの中で $y$ が取り得る最大値である。 $ce_2$ は $y$ が大きいほど、より高いパーセンテージとなる。

### 5. 1. 3 セルのスコア関数

セルの2つの基準を使って、スコア関数を作成し、それぞれのセルをランク付けすることができる。この関数は、以下に定義される。

$$f(ce) = w_1 ce_1 + w_2 ce_2$$

ここで、変数  $ce_i$  は各基準の値を表す。変数  $w_i$  は基準の重要度に応じて適用する重みである。  $w_i$  の和は 1 になる。セルのスコア  $f(ce)$  はパーセンテージで表される。

## 5. 2 ブロックのための基準

センシティブなユニットを含むブロックに対する 4 つの基準を提案する。各基準は比で表され、パーセンテージで示される。

5. 2. 1) セルが完全に公開可能となるまでに必要な免責の数

5. 2. 2) 主要な変数に基づくブロックのサイズ

5. 2. 3) 隠された機密セルと補足的セルの情報の量

5. 2. 4) センシティブなユニットを含んでいるセル数

以下で示すように、これらの基準はセルのための基準と類似している。

それぞれのパーセンテージを組み合わせて、1 つのスコアを作る。それぞれの基準は重要度に応じて重み付けられる。次に、各基準の詳細を説明する。

### 5. 2. 1 必要な免責の数

必要な免責の数がより少ないブロックに取り組むほうがよい。その場合の基準は、あるブロックを公開可能とするために必要な免責の数と、センシティブセルを含むすべてのブロックの中で、必要とする免責の数が最大のものとの比に基づく。この基準では、より少ない数の免責を必要とするセルが最もよい。基準は次のように定義される。

$$bl_1 = \frac{(\max(x) - x)}{\max(x)} \times 100$$

ここで、 $x$  はブロックのセンシティブリティを除去するために必要な免責の数で、 $\max(x)$  は、すべてのブロックの中で  $x$  が取る最大値である。  $bl_1$  は必要な免責の数が小さいとき、高いパーセンテージとなる。

### 5. 2. 2 ブロックのサイズ

最も大きいブロックに、より高い優先度を与えたい。したがって、この基準は主要な変数に関するブロックの相対的なサイズを考慮して定義する。この基準の式は、以下に定義される。

$$bl_2 = \frac{y}{\max(y)} \times 100$$

ここで、 $y$  は該当の変数のブロック内の合計値で、 $\max(y)$  はセンシティブセルを含むすべてのブロックの中で  $y$  が取る最大値である。  $bl_2$  は  $y$  が大きいとき、高いパーセンテージ

となる。

### 5. 2. 3 隠された情報

隠された機密セルと補足的セルの情報の量も評価に入れたい。これを考慮して、次のように定義する。

$$bl_3 = \frac{y_1 + y_2}{\max(y_1 + y_2)} \times 100$$

ここで、 $y_1$ はブロックの中でセンシティブのために隠された該当の変数の合計であり、 $y_2$ は残存する機密のために補足的な目的で隠された変数の合計である。変数 $\max(y_1 + y_2)$ は機密保持のために隠された情報が存在するすべてのブロックの中で、2つの変数の和が取る最大値を示す。 $bl_3$ は大きな情報量が隠されているときに高い値をとる。

### 5. 2. 4 秘匿されたセルの数

多数のセルが秘匿されているブロックを優先したい。目標は隠されたセルの数によってブロックをランク付けすることである。この基準は次の式で定義される。

$$bl_4 = \frac{x}{\max(x)} \times 100$$

ここで、 $x$ はセンシティブセルの数で、 $\max(x)$ はセンシティブセルを含むすべてのブロックの中で $x$ が取り得る最大値である。 $bl_4$ は多くのセルが隠されるときに高いパーセンテージとなる<sup>2</sup>。

### 5. 2. 5 ブロックのスコア関数

ブロックの4つの基準を使ってスコア関数を作成し、それぞれのブロックをランク付けすることができる。この関数は、以下に定義される。

$$f(bl) = w_1 bl_1 + w_2 bl_2 + w_3 bl_3 + w_4 bl_4$$

ここで、変数 $bl_i$ は各基準の値を表す。変数 $w_i$ は基準の重要度に応じて適用する重みである。 $w_i$ の和は1になる。したがって、ブロックのスコア $f(bl)$ はパーセンテージで表される。

<sup>2</sup> 機密セルと補足的セルの数の合計を用いることも考えられる。原文の誤りかもしれない。

セルのスコア関数とブロックのスコア関数の2つを使うことによって、各セル<sup>3</sup>の相対的な優先度を付けることができる。次節では、実例を用いて、このような分類方法の効果を分析する。

## 6 実例

この節では、免責を使うことによる利益を示す。1997年製造業年次調査 (Annual Survey of Manufactures<sup>4</sup> : ASM) から食料品製造業のデータを用いる。この年のASMは、ある一定規模以上のユニットの全数調査を行った。このセクターでは、約3100ユニットで50,468,609千ドルの製造品出荷額がある。データは地域別にさまざまな産業についての合計319セルに渡って集計される(合計や小計を含む)。免責を使わない場合、約55%のセルが公表可能である。

免責ユニットを特定するスコア関数の方法と別の種類のランク付けを比較する。以下の表では、すべてのオプションは、公表されるセル数と製造品出荷額の割合で表されている。この表では4つのオプションを比較している。

- (i) 免責を使わない
- (ii) スコア関数による上位10ユニットの免責を使う
- (iii) 1ユニットあたりの製造品出荷額による上位10ユニットの免責を使う
- (iv) 製造品出荷額のセル合計値による上位10ユニットの免責を使う

産業分類の3つの異なるレベル、すなわち、SIC2 (10:食料品製造業)、SIC3 (例えば、101:肉・家禽肉製品製造業)、SIC4 (例えば、1012:家禽肉製品製造業) 別に結果を示す。

産業分類レベル	セル数 総数	オプション(i)		オプション(ii)		オプション(iii)		オプション(iv)	
		セル数	出荷額割合 (%)	セル数	出荷額割合 (%)	セル数	出荷額割合 (%)	セル数	出荷額割合 (%)
SIC2	13	13	100.0	13	100.0	13	100.0	13	100.0
SIC3	88	61	97.1	65	97.5	61	97.4	62	97.1
SIC4	218	100	87.6	116	91.4	105	90.2	103	89.5

(i)、(ii)、(iii)、(iv)を比較すると、免責の利用によって、公表される情報量を増やせることが分かる。一番細かいレベル(SIC4)では効果がより大きい。また、優先付けされた免責を使う3つのオプションを比較すると、この論文で提案したスコア関数(オプ

<sup>3</sup> 「合計のセル」(広い意味での「セル」)と「内訳のセル」(本論文で定義している「セル」)の集まりであるブロックを同一視していると思われる。

<sup>4</sup> 原文のMiningは誤りと思われる。

ション (ii) <sup>5)</sup> によって、無視できないほど多くの情報量を公表できるようになることが分かる。とくに一番細かいレベルでその傾向が見られる。一番細かいレベルでは、スコア関数による 10 件の免責を利用することによって、製造品出荷額で 5%、セル数で 15% 多く公表できるようになることが分かる。

## 7 結論

要約すると、集計表データを扱わなければならないとき、免責の利用とその優先付けによって、実際に公表する情報をかなり増やすことができるといえる。センシティブセルの特性に基づくセルやブロックのスコア関数は、免責の利用によって得られる利益を最適化することに役立つ。

このスコア関数に関してさらに研究を行うことにより、セルとブロックの特性を組み合わせるより適切な方法を発見できるであろう。また、各関数の重み付けの方法を調べることも可能だろう。重み付けのある側面は最適化できると思われるが、利用者の優先度によって決定すべき部分が常に存在するだろう。

## 追記

Provençal(2004)では、本論文で提案した混合スコアアプローチの研究をさらに進めて、新たにセグメントコストアプローチを提案している。混合スコアアプローチは、1997年ASMでは良好な結果が得られたが、2000年卸売業調査(Wholesale Trade Survey)の3次元表に適用したときに適切な秘匿パターンが得られず、3次元の問題への適用が不十分である。

検討した卸売業調査の表は、産業分類、地域、収入の種類の3次元で、産業分類はトレードグループと呼ばれる区分ごとに小計をとる構造になっている。いくつかのグループには、ごく少数の産業しか含まれない。このような表において、混合スコアアプローチでは、非常に大きなセンシティブセルが見逃される傾向や、表章区分の階層構造がフラットである(あまり階層化されていない)ときには、小さなセンシティブセルが大きなセルを補足的セル(二次秘匿セル)としてしまう傾向が見られた。

そこで、表の構造を反映して、トレードグループごとに表を分割して(分割したものをセグメントと呼ぶ)、セルを保護するために必要な「最小コスト」を定義し、コスト関数によって免責ユニットを特定する方法を提案した。「最小コスト」は、セグメント内で3次元の軸ごとに計算され、セルの値に各軸方向の「最小コスト」を足したものがそのセルのコストとして定義される。このコスト関数は、比較的簡単に計算でき、補足的秘匿セルに依存せず、多くの次元を組み入れることができ、セルのサイズと周囲のセルの追加的情報を取り入れている。欠点は、セグメントの外側の影響を無視することや、免責を得て公表するセンシティブセルが他のセンシティブセルに与える影響を考慮しないことなどである。

卸売業調査の表では、セグメントコストアプローチを使うことで、秘匿されるセルの値

<sup>5)</sup> 原文のオプション (i) は誤りと思われる。

の点で、混合スコアアプローチよりも安定して情報ロスの小さい結果が得られた。一方で、秘匿されるセル数の点では、必ずしもロスが小さくならない。秘匿されるセル数を減少させることや、軸の重要度を考慮したスコア付けなどを課題として示唆している。また、問題とする表の種類に対して適切な処理方法に関する手引きの必要性や、小さなセンシティブセルが大きな補足的セルを取る種類の問題では、丸め法や CTA のような免責に代わるアプローチを試みることについて言及している。

#### 参 考 文 献

Provençal, J.-S., Bérard, H., Fillion, J.-M. and Tambay, J.-L. (2004). Approaches to identify the amount of publishable information in business surveys through waivers, in J. Domingo-Ferrer and V. Torra (eds.) *Privacy in Statistical Databases*, Berlin: Springer-Verlag, pp.110-120.



---

製 表 技 術 参 考 資 料 4

平成 18 年 3 月発行

編集・発行 独立行政法人 統計センター

〒162 - 8668

東京都新宿区若松町 19 - 1

電 話 代 表 03 ( 5273 ) 1200

---

掲載論文を引用する場合は、事前に担当まで連絡してください

担当；研究センター 03 - 5273 - 1286