経済研究
Vol. 69, No. 2, Apr. 2018

特集：日本の統計――歴史と現在

公的統計の実証分析における秘密計算とその部分計算過程を公開することの安全性の検討

白川清美・千田浩司・田中哲士・高橋慧・菊池亮

公的統計ミクロデータの新たな利用形態として，「オンサイト施設」と「オフサイト利用」がある．日本におけるオンサイト施設の利用は，平成 29 年 1 月から試行運用が開始された．しかしながら，オフサイト利用の環境は未だ整っていない．本論文では，オフサイト利用のセキュリティ課題に着目し，データを暗号化した状態で集計ができる「秘密分散・秘密計算システム」の適用を試行する．公的統計ミクロデータや当該データから得られる集計表や各種統計量の提供においては，統計的開示抑制(Statistical Disclosure Control：SDC)(Hundepool *et al.*(2012))が必要となるが，この SDC に対応した秘密分散・秘密計算システムの実現例はほとんどない．特に，秘密計算は，暗号化していないデータ(平文)を用いた通常処理と比べ計算時間が増加することから，計算時間の増加を抑制しつつ SDC に対応した秘密計算の実現が課題となる．本論文では，公的統計でよく用いられる線形回帰および主成分分析に焦点を当て，計算時間の増加を抑制しつつ SDC に対応した秘密計算の提案を行う．この提案技術の特徴は，SDC に対応した線形回帰や主成分分析において，個人のレコードを用いず統計量のみでも計算可能な処理部分を明らかにし，当該処理部分においては秘密計算ではなく通常の平文の処理とすることで計算時間の増加を抑制することである．この提案手法を秘密分散・秘密計算システムに実装し，すべての処理を平文で計算した結果と比較した．その結果，いずれの計算においても平文と同等の精度で分析ができていることを確認した．
JEL Classification Codes: C44, C61, C88

1. はじめに

「公的統計の整備に関する基本的な計画」(平成 26 年 3 月 25 日閣議決定)において，「オンサイト利用といった新たな利用方法の実現を目指し，実用化に向けた検討を行う．」とのことから，「公的統計ミクロデータ研究コンソーシアム」での検討が進められ(online1: onsite.html)，平成 29 年 1 月からオンサイト施設の試行運用が開始された．この試行では，これまでの「分析に必要な最小限の情報に限り提供される」ことによる探索的・創造的研究が困難な状況から，「フルスペックの情報が利用可能になる」ことにより，探索的・創造的研究を可能とした．それゆえ，研究者が事前に想定しなかった分析が可能となり，これまで以上の研究成果が期待できる．しかしながら，オンサイト施設に設置された PC のハードディスクや USB メモリなどの記憶装置の使用ができないなど，データ利用や成果物の持ち出しに関する制限が多く設定されている(Shirakawa *et al.*(2017), online2: onsite.html)．特に，公的統計ミクロデータなどの研究成果の持ち出しには，統計的開示抑制(Statistical Disclosure Control：SDC)に基づく審査がある．SDC とは個人，企業や他の組織の情報が開示されるリスクを減らすための一連の手法(Hundepool *et al.*(2012))のことであり，本論文では特に，出力される分析結果から個人のデータが開示されるリスクを減らすための手法を指す．この審査結果により，作成したプログラム，集計結果表およびモデル式などの成果物が持ち出しできるか否かを審査し判断する．

一方，利用者の利便性向上のため「オフサイト利用」の実現も検討されている(online3: saishu_honbun.pdf)．オンサイト施設と比べ利用場所を柔軟に対応できることから，公的統計ミクロデータの更なる利活用促進が期待できる．しかしながら，様々な利用場所からネットワークを通じてアクセスができるオフサイト利用においては，不正アクセスなどによる公的統計ミ

クロデータ流出へのセキュリティ対策が大きな課題となる．また，集計結果などの成果物においても，それらの公開に伴う公的統計ミクロデータの情報流出を防ぐための対策が必要となる．

そこで本論文では，公的統計ミクロデータのオフサイト利用の実現を目指し，データを複数の断片(シェア)に分割し，暗号化した状態で集計ができる「秘密分散・秘密計算システム」の適用を試行する．これによりオフサイト利用における公的統計ミクロデータの利用環境のセキュリティリスクの軽減が期待できる．しかしながら，計算速度や集計結果などの安全性に課題が残る．秘密計算は，暗号化していないデータ(平文)を用いた通常処理と比べ計算時間が増加するため，計算時間の増加の抑制が大きな課題である．現在では秘密分散技術に基づく秘密計算技術が，比較的計算時間の増加が少ない手法として知られている．さらに，ある演算に特化した利用が可能な秘密計算の方が，より計算時間の増加を抑えることができる．また，秘密計算であっても集計結果から元のミクロデータの情報が流出する危険性は回避できないため，SDC への対応を図る．SDC に対応した秘密分散・秘密計算システムの実現例はこれまでほとんどない．

本論文では，公的統計でよく用いられる線形回帰および主成分分析に焦点を当て，計算時間の増加を抑制しつつ SDC に基づく持ち出し審査への対応を目指した秘密計算の提案を行う．なお，この持ち出し審査の基準を示したガイドライン((Brandt *et al.*(2010))は存在するが，基準の根拠は明示されていない．それゆえ，Shirakawa *et al.*(2016)において，標準偏差，歪度や尖度などの高次モーメントの組み合わせにおける安全性の検証を行ったが，線形回帰係数などの回帰モデル式における安全性の検証は行っておらず，著者らが知る限り報告されてもいない．そこで SDC の観点による線形回帰係数の安全性の検証を行い，当該検証結果に基づき，SDC に対応した線形回帰や主成分分析の秘密計算の実現手法を提案する．提案手法の特徴は，SDC に対応した線形回帰や主成分分析において個人のレコードを用いず統計量のみでも計算可能な処理部分を明らかにし，当該処理部分については秘密計算ではなく平文を用いた通常処理とすることで計算時間の増加を抑制することである．また，提案手法を秘密分散・秘密計算システムを実装し平文の結果と比較した．その結果，いずれの計算においても平文と同等の精度で分析ができていることを確認した．さらに，秘密計算における SDC の対応により，公的統計の実証分析における秘密計算とその部分計算過程を公開することの安全性を確認した．

以降，第 2 節では，統計量の安全性を評価する基本方式として，Shirakawa *et al.*(2016)による数値パターンの推計のためのデータベースを用いた数量表の安全性の検証方式について解説する．第 3 節では，線形回帰の算出方法とその計算に必要な記述統計量との関係に基づき，線形回帰において SDC を実現する方法について述べている．第 4 節では，計算時間の増加を抑制しつつ SDC に対応可能な秘密計算を用いた線形回帰および主成分分析の手法について述べ，最後の第 5 節においては，本論文における結論を述べている．

2. 関連研究

ある個人レコード(個票)を収集したデータセット(個票データ)をデータセット内の各レコードを 1 つまたは複数の属性の値に基づき，複数の集団(グループ)に分け，そのグループ毎のレコード数(度数)を数え上げるなど，何らかの集計処理を施し表にまとめたものを集計表と呼ぶ．このとき，あるグループに属する個票が極端に少ないなどの要因から，得られた集計表から元の個票に繋がってしまうケースが存在する．この集計表から元の個票に繋がることを防ぐために，集計の前段階でこれらの危険な個票に削除などを施して個票の安全性を担保することを秘匿処理と呼び，秘匿するためのルールを秘匿ルールと呼ぶ．

本章では，個票データにおける各属性の平均値や分散値をはじめとした，記述統計量と呼ば

表 1. n 番目の値における最大値と最小値

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
最低値	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
最高値	100	50	33	25	20	16	14	12	11	10	9	8	7	7	6
	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}
最低値	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
最高値	6	5	5	5	5	4	4	4	4	4	3	3	3	3	3

表 2. 度数 2 から 10, 20, 30 の組み合わせパターン数とその標準偏差の最大値と最小値，平均およびパターン数

度数	標準偏差			平均	パターン数
	最大値	最小値	範囲		
2	70.7	0.0	70.7	50.0	51
3	57.7	0.6	57.2	33.3	884
4	50.0	0.0	50.0	25.0	8,037
5	44.7	0.0	44.7	20.0	46,262
6	40.8	0.5	40.3	16.7	189,509
7	37.8	0.5	37.3	14.3	596,763
8	35.4	0.5	34.8	12.5	1,527,675
9	33.3	0.3	33.0	11.1	3,314,203
10	31.6	0.0	31.6	10.0	6,292,069
20	22.4	0.0	22.4	5.0	97,132,873
30	18.3	0.5	17.8	3.3	139,065,026

れるデータを集計した表(数量表)に対する秘匿ルールとして，Shirakawa *et al.*(2016)の数値パターン推計のためのデータベースを用いた数量表の安全性検証方式について解説する．

2.1　数値パターン推計による安全性検証

Shirakawa *et al.*(2016)の安全性検証方式は単純には以下の通りである．

1. 与えられた数量表と同じ記述統計量を持つ個票データのパターン数を数え上げる．
2. パターン数が十分にあれば安全な数量表として公開を許可する．
3. もしパターン数が少ないのであれば，元の個票データを推計しえる危険な数量表であるとして公開しないなどの開示抑制を行う．

ここで，実数を含むデータを取り扱う場合，数量表を満たす組み合わせは無数に存在するため，Shirakawa *et al.*(2016)では以下の制限を加えている．

1. 個票は非負の整数値のみを持つとする．
2. n 個のレコードを持つ(度数 n の)ある属性において，個票の総和が 100 となる．
3. 数値は降順にソートされている．

例として，度数が 30 である場合を考えると各個票 $x_1, \cdots, x_{30}$ は次の式(1)を満たす．

$$\sum_{i=1}^{30} x_i = x_1 + x_2 + \cdots + x_{29} + x_{30} = 100, \quad 0 \le x_{30} \le x_{29} \le \cdots \le x_2 \le x_1. \tag{1}$$

このときの，各 $x_i (1 \le i \le 30)$ がとりえる値の最大値と最小値を表 1 に示す．

式(1)を満たす度数 30 の組み合わせパターン数は，139,065,026 通りである．また，度数 10 の数値の組み合わせパターン数は 6,292,069 通りであり，度数 20 の組み合わせパターン数は 97,132,873 通りである．

Shirakawa *et al.*(2016)の手法では，これらの個票データの総パターン数に対して，記述統計量を用いてどこまでパターン数を絞り込むことができるかで安全性を評価する．例えば，標準偏差について着目し安全性を評価することを考える．表 2 に度数 2 から 10, 20, 30 の個票データの組み合わせにおける標準偏差の最大値と最小値，平均の一覧を示す．一般に，式(1)の制約において標準偏差が最大となるのは，$x_1 = 100$，かつ，$x_2 = \cdots = x_n = 0$ の場合である．したがって，各度数 n における標準偏差の最大値を記録することにより，最大の標準偏差が与えられた場合，個票の数値を推計することが可能となる．

表 3. 度数 100 のテーブルにおけるレコード数，
実数値・ゼロ数値数およびその割合

レコード数	190,569,292		
項目	数	率(%)	容量(GB)
総セル数	19,056,929,200	100	48.3
実数値	4,144,913,179	21.8	33.7
ゼロ値数	14,912,016,021	78.2	14.6

図 1. 記述統計量を記録したデータベースのイメージ

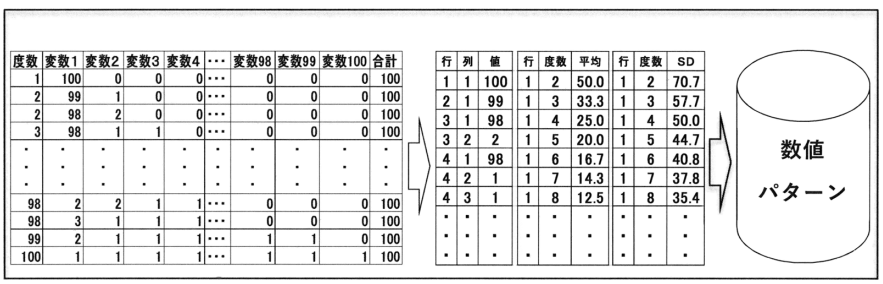


表 4. ターゲットと集計項目

項目	ターゲット	集計	非集計	補助的情報
内容	最大値	度数，平均，標準偏差，歪度および尖度	最大値，最小値，レンジおよびパーセンタイル	侵入者 (第 2 番目に大きい値) 中央値

2.2 数値パターン推計用データベースを用いた安全性検証値

Shirakawa *et al.*(2016)はこれらの記述統計量からどの程度の推計ができるかを評価するために，これらの個票データの組み合わせと記述統計量を記録したデータベースを作成した．具体的には，度数 2 から 100 までの 1 属性の個票データの組み合わせパターンを全て保持し，それぞれの度数，平均，高次モーメント(標準偏差，歪度および尖度)を記録したものである．ただし，例えば度数 2 で $x_1=x_2=50$ のデータは，度数 100 で $x_1=x_2=50, x_3=\cdots=x_{100}=0$ のデータの x_3 から x_{100} を無視したものと同じであるため，実際には，度数 100 の個票データの組み合わせに，取りえる度数 n の各記述統計量を付記している．そのため，実際のデータベース内には多くの 0 のデータを持つセルが存在する．表 3 に度数 100 の場合のデータベースのサイズを，図 1 にデータベースのイメージを示す．

Shirakawa *et al.*(2016)ではこのデータベースを用いた数量表に対する安全性の検証法として，以下のものを検討している．まず，攻撃者として対象となる個票データのうち，2 番目に大きい個票の値 (x_2) を知っている人物(侵入者と呼ぶ)を考える．侵入者は，自身が持つ x_2 と x_2 が属するデータセットに関する数量表から，データベースを用いて最大の個票の値 (x_1) を推定する．このとき，x_1 が取りえる値の範囲が小さい場合，その数量表は安全でない．この安全性の検証方式において，分析に用いる(集計する)記述統計量と分析に使用できない(集計しない)記述統計量をまとめたものを表 4 に示す．表 4 において，最大値などの一部記述統計量が非集計となっているのは，これらの統計量は値そのものが個票であるためである．

Shirakawa *et al.*(2016)では，この方式で推計することにより，度数 20 の場合，表 4 の記述統計量から，20 のすべての数値まで推計できることが明らかにした(図 2 を参照)．

図 2. 度数 20 における元データおよび推計データ（整数と実数）
（Shirakawa *et al.*(2016), Fig. 2 より引用）

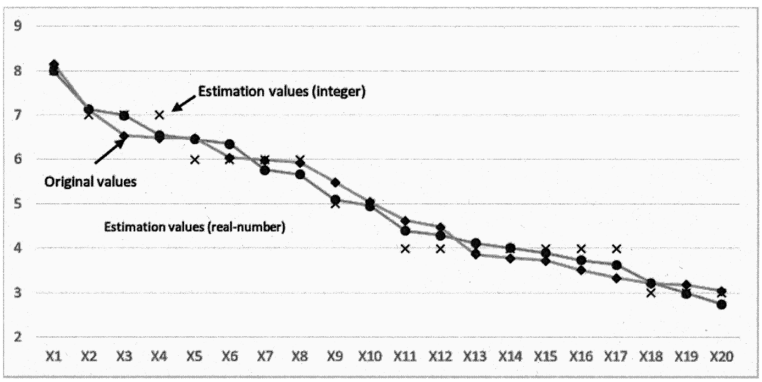


表 5. 個々のデータから算出する統計量

項目	計算に用いる統計量
共分散	(x の残差)×(y の残差)
分散 (標準偏差)	残差平方和，度数
残差平方和	残差平方の和
平方数和	平方数の和
平方数*	数値の二乗
残差平方*	残差の二乗
残差*	数値，平均
平均	総和，度数
総和	すべての数値の和
度数	データの数

注）* はデータの個数と同じ数の値がある．

表 6. 統計量から算出可能な係数など

項目	計算に用いる統計量
相関係数	x と y の分散，共分散
回帰係数	x の分散，共分散
切片	平均，回帰係数

3. 線形回帰における安全性

3.1 線形回帰と記述統計量

2 次元の変数 x と y の場合，この回帰直線のモデル式は以下のとおりである．

$$\hat{y} = ax + b \tag{2}$$

式(2)により，回帰直線によるグラフを描くことができる．しかしながら，このグラフからでは，元データの推計が可能か否かを判断することができない．そこで，この回帰係数 a と切片 b を算出する式を以下に示し，この式を構成する統計量を明確にする．

$$a = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma {x_i}^2 - n\bar{x}^2} \tag{3}$$

$$b = \bar{y} - a\bar{x} \tag{4}$$

さらに，この式に関連する統計量を表 5 と表 6 に示す．表 5 の統計量は，いずれも個々のデータから算出することになる．また，表 6 の係数などは，表 5 の統計量から算出することができる．このことから，本論文の対象である回帰係数の計算には，式(3)または式(4)に示す統計量があればよいことが分かる．したがって，この係数の安全性の検証では，計算に用いる統計量も安全性の検証に含める必要がある．

なお，本論文では，表 5 の * マークがある残差，残差平方和，平方数に加え，表 4 で集計しない統計量としている最大値，最小値，最頻値，レンジ，およびパーセンタイル，などは，個々の数値と一対の統計量であることから，元の数値が容易に推測できる安全でない統計量であることが既知のため，今回の安全性の検証の対象外とした．

なお，多くの統計解析パッケージでは，「記述統計」または「基本統計量」を一括計算する機能を有しているが，この機能を利用した分析では，必然的に，個々の数値を示す中央値(メジアン)，最頻値(モード)，最小値および最大値などの安全でない統計量が含まれることにな

表 7. 度数 10 における各変数の最大値と最小値

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
最大値	100	50	33	25	20	16	14	12	11	10
最小値	10	0	0	0	0	0	0	0	0	0

図 3. x_1 と x_2 を決めた時に取りえる組み合せパターン数

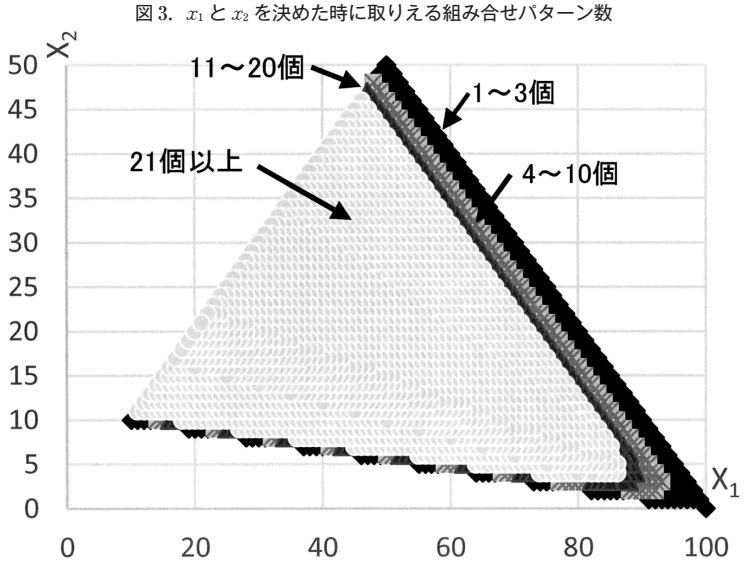


表 8. x_1 と x_2 を決めた時の組合せパターン数による区分

区分	個数	割合
総数	2,051	100.0%
1-3	242	11.8%
4-10	122	5.9%
11-20	112	5.5%
21-100	315	15.4%
100-	1,260	61.4%

る．したがって，この機能を利用することは SDC からみると，安全ではないと言える．

3.2 線形回帰の安全性の検証

本論文では線形回帰係数の安全性を，Shirakawa *et al.*(2016)の先行研究に基づき，個票の総和が 100 となる整数値で構成される個票データの組み合わせパターンを線形回帰係数からどの程度推計できるかで評価する．なお，表 6 の統計量は表 5 の統計量に基づき算出できることから，本論文では表 5 の統計量に着目して安全性の検証を行うことにより，公的統計の実証分析における秘密計算とその部分計算過程を公開することの安全性を検討し，さらに計算速度向上のための議論を行う．

まず，度数が 10 の場合を考える．表 7 より，度数 10，総和 100 の組み合わせパターンは，{100, 0, 0, 0, 0, 0, 0, 0, 0, 0} から {10, 10, 10, 10, 10, 10, 10, 10, 10, 10} まであり，6,292,069 通りの組み合わせがある．

図 3 は，前述の条件下において x_1 と x_2 の値を決めた時に取りえる組み合わせのパターン数を示している．図 3 において，組み合わせパターン数が小さい場合，元の数値パターンが推計できることを意味する．さらに，表 8 に x_1 と x_2 の組み合わせを組み合わせパターン数により区分したものを示す．x_1 と x_2 の組み合わせは 2,051 通りあり，そのうち，「1 から 3」に区分される組み合わせは 242 通りあり，全体の 11.8% となっている．そのため，度数 10 の場合でも，標準偏差の値によっては，元の数値パターンが特定できることになる．

以上のことから，線形回帰などの実証分析では，統計量を度数，平均や分散(標準偏差)などの「個々のデータから算出する統計量」と，相

表 9. 度数 5，総和 100 および平方和 4,600 の組み合わせパターン(20 パターンを抽出)

No.	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	P_{16}	P_{17}	P_{18}	P_{19}	P_{20}
1	67	65	64	63	62	61	58	54	59	56	53	55	49	47	43	41	39	37	36	35
2	10	19	22	25	27	29	35	41	33	38	42	39	46	47	43	41	39	37	36	35
3	3	3	4	2	5	6	3	1	5	4	5	7	9	13	30	35	39	37	36	35
4	1	2	2	1	1	1	1	1	2	2	1	2	1	3	1	3	6	22	26	30
5	1	1	0	1	1	1	1	1	1	0	1	1	1	2	1	2	1	3	6	5
総和	82	90	92	92	96	98	98	98	100	100	102	104	106	112	118	122	124	136	140	140
平均	16.4	18	18.4	18.4	19.2	19.6	19.6	19.6	20	20	20.4	20.8	21.2	22.4	23.6	24.4	24.8	27.2	28	28
平方和	4,600																			

表 10. 度数 5，平方和 4,600 となる組み合わせ(データベースから抽出した 10 パターン)

No.	x_1	x_2	x_3	x_4	x_5	総和	標準偏差	度数	平方和
1	56	38	4	2	0	100	25.495	5	4,600
2	59	33	5	2	1				
3	60	30	10	0	0				
4	61	29	5	3	2				
5	62	26	8	4	0				
6	64	18	12	6	0				
7	64	20	8	6	2				
8	65	13	11	9	2				
9	65	14	11	7	3				
10	65	15	10	5	5				

関係数や回帰係数などの「各種統計量から算出可能な係数など」に区分することができる．特に，分析結果の持ち出し審査では，後者の「各種統計量から算出可能な係数など」が対象となっているが，実際の審査においては，「個々のデータから算出する統計量」の値に着目する必要がある．回帰係数の審査における手順では，はじめに度数と分散に SDC の基準値を設け，平均，分散や相関係数などの除算を含む統計量の計算前に，それぞれの基準値との照合をする．その結果，安全性の範囲内である場合，復号化した総和と度数によって平均を計算する．なお，分散など除算を含む基準値の照合では，除算を含まない各変数の平方和に変換した数値での照合をする．具体的な平方和を用いた方法は，度数 n の n 個分の二乗数を作ることである．例えば，度数 5，総和 100，平均 20 および平方和 4,600(標準偏差 25.495)の場合，最大値 (x_1) を平方和 4,600 の正の平方根である 67(端数を切捨てた整数)とする．このとき，67 の二乗数を 4,600 から引いた余りが 111 となる．さらに，正の平方根を求めると x_2 は 10 となる．この計算を繰り返すと，度数 5 の個票データ (67, 10, 3, 1, 1) が算出できる．ただし，この個票データの総和は 82 となるため，総和が 100 の条件を満たさない．したがって，このパターンの個票データはありえないことが分かる．なお，最大値 (x_1) が一番小さい値は 31 と算出できる．このことから，x_1 のレンジは最大 67 から最小 31 であることが分かる．

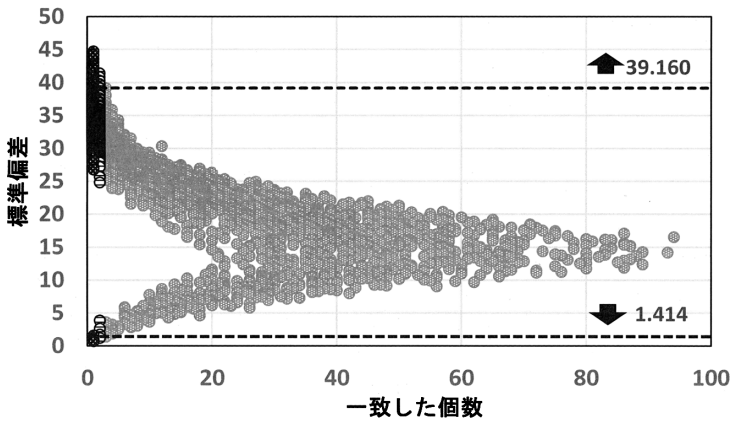
ちなみに，x_1 を 31 とした場合，度数 8(31, 31, 31, 31, 27, 5, 1, 1) となり，度数や総和の条件から外れることになる．

以上，数値パターンの条件(度数 5，総和 100 および平方和 4,600) と x_1 のレンジに基づき抽出したのが表 9 である．

表 9 には，20 パターンがあるが，すべての条件を満たすパターンは P_9 と P_{10} の 2 パターンである．これは，x_1 のレンジと平方和の値に基づき数値パターン抽出した結果であるが，当該数値パターンを導出することは困難なことである．そこで，度数 n (n は 2 から 100) の各種統計量を格納したデータベースに基づき抽出した数値パターンが表 10 である．

表 10 から，度数 5，総和 100 および平方和 4,600 となる数値パターンが 10 パターンであることが分かる．また，x_1 のレンジは先ほどの 67 から 31 までの 37 に対して，65 から 56 までの 10 と狭くなっている．このレンジが狭くな

図 4. 度数 5 における標準偏差が一致した際の各統計量の範囲



ればなる程，数値パターンの特定が容易となる．

次に，図 4 は，度数 5 における標準偏差の閾値設定のため，データベースに格納した数値パターンと標準偏差の値が一致した度数を示している．この図では，標準偏差が 39.160 以上もしくは 1.414 以下となれば，標準偏差の値が一致している個数が 2 以下となる．この場合，平方和が 8,134(x_1 が 90)以上もしくは 2,008(x_1 が 22)以下である．なお，度数 5 の標準偏差の最大値は 44.721(平方和 10,000)であり，最小値は 0(平方和 2,000)である．

この様に，すべての度数 n における標準偏差の閾値を設定し，その標準偏差を平方和に変換することで，除算せずに閾値の計算が可能となる．この結果，表 5 にある統計量を用いて SDC ルールを適用し，暗号化した状態での計算を最小限に留めることにより，計算速度を向上する．

4. 秘密計算を用いた統計分析

秘密計算とは暗号を用いたデータ活用技術であり，データを暗号化して秘匿しつつ，その暗号化されたデータの分析結果のみを得られる技術である．この技術を用いることで，研究者に対して，個票データを暗号化し秘匿しつつ，研究者が所望する分析結果のみを渡すことができる．すなわち，個票データの内容を外部に漏らさずに分析結果のみを得ることができる．

4.1 秘密計算

秘密計算の原理は 1980 年代から知られていたものの(Yao(1986), Goldreich *et al.*(1987), Ben-Or *et al.*(1988)))，当時は処理速度が遅く実用には至らなかった．しかしながら，近年は秘密計算自身の研究の進展やコンピュータの処理能力向上により，実用に耐えうるレベルの処理速度が実現可能となり，急速に秘密計算の実用化が進んできている．実例として，秘密計算システムを用いた医療データの安全な活用などが知られている(Chida *et al.*(2014))．

秘密計算には複数の実現方法が知られている．特に本論文では計算コストが低くデータの格納効率が良い秘密分散を用いた方法に着目し，秘密分散を用いた秘密計算を単に秘密計算と呼ぶこととする．秘密分散では，データを複数の断片(シェアと呼ぶ)に変換(暗号化)し，それぞれのシェアを異なるコンピュータ(以降サーバと呼ぶ)に保存する．このとき，シェアに変換する際には，

1. 単一のシェアからは元のデータの情報を一切得られない．
2. 一定数以上のシェアを集めた場合のみ元のデータを復元できる．

様に変換されている．そのため，各サーバは元のデータの情報を得ることはできない．秘密分散の具体的な方法は Shamir(1979) を参照され

たい．

秘密計算では，上記の様にデータがシェアとして各サーバに保存されている状態から，各サーバが互いに通信や計算を行うことで，データを復元することなく，そのデータの加算や乗算などを計算し，最終的に分析結果のシェアを生成する．その後，研究者は各サーバからシェアを受け取り復元することで，個票データを得ることなく分析結果のみを得ることができる．

4.2　秘密計算の得意な演算・不得意な演算

秘密計算では分析結果を得るために，データを復元することなくデータの加算や乗算などの演算を行う．この演算は，データが復元されていない，すなわち暗号化されたままでの演算であり，通常の暗号化されていない状態での演算とは大きく処理が異なる．そのため，暗号化されていない状態では高速に処理できる演算であっても，秘密計算では多くの処理時間がかかってしまう場合がある．本節ではどのような演算が秘密計算において得意もしくは不得意なのかを記す．

4.2.1　秘密計算の(整数上の)加算，乗算

いま，a, b という2つの整数のデータが秘密分散されているとする．また，i 番目のサーバが持つシェアを $[a]_i, [b]_i$ と書く．このとき，秘密計算の加算とは $[a]_i, [b]_i$ から $[a+b]_i$ を計算することであり，秘密計算の乗算とは $[a]_i, [b]_i$ から $[ab]_i$ を求めることである．

秘密計算の加算は，実は通常の加算と同様に可能であることが知られており，得意な演算であると言える．秘密計算の乗算も，暗号化されていない場合に比べ各サーバ間の通信が1回必要になるものの，比較的高速に処理することが可能であり，得意な演算と言える．具体的な方法は Genarro *et al.*(1988) を参照されたい．

4.2.2　秘密計算の(整数上の)積和

秘密計算の積和とは，$([a_1]_i, [a_2]_i, \cdots, [a_m]_i)$ と $([b_1]_i, [b_2]_i, \cdots, [b_m]_i)$ から $[\sum_{k=1}^{m} a_k b_k]_i$ を求めることである．前述の秘密計算の加算と乗算を m 回繰り返すことで積和が計算できるが，さらに高速な方法として，秘密計算の乗算と同じ各サーバ間の通信を1回行えば積和が計算できること知られており，乗算と同レベルに得意な演算と言える．

4.2.3　秘密計算の除算

秘密計算の除算とは，秘密計算の乗算とは $[a]_i, [b]_i$ から $\left[\frac{a}{b}\right]_i$ を求めることである．これまでの加算・乗算・積和に比べ，秘密計算の除算は実現が難しいことが知られている．その一つの大きな理由は，a, b が整数であっても $\frac{a}{b}$ は整数でない場合があり，整数以外の浮動小数点数などを用いる必要があるからである．Aliasgari *et al.*(2013) や Kamm *et al.*(2015) らにより秘密計算において浮動小数点数を扱う方法は提案されているが，これまでの加算・乗算・積和に比べて計算コストが高く，秘密計算にとって比較的不得意な演算であると言える．

4.3　秘密計算での線形回帰

4.3.1　秘密計算における naïve な線形回帰の実装

秘密計算を用いた線形回帰について考える．いま，n 個のレコードからなる属性 $\mathbf{y} := (y_1, \cdots, y_n)$ を，k 個の属性を持つ行列 $\mathbf{X} := (\mathbf{1}_n^T, \mathbf{x}_1^T, \cdots, \mathbf{x}_k^T)$ を用い，次の線形モデル，

$$\mathbf{y}^T = \mathbf{X}\boldsymbol{\beta}^T + \boldsymbol{\varepsilon},$$

で説明することを考える．ただし，$1 \leq i \leq k$ において，$\mathbf{x}_i := (x_{i,1}, \cdots, x_{i,n})$ であり，$\mathbf{1}_n$ は1を n 個並べたベクトルである．このとき，残差列 $\boldsymbol{\varepsilon} := (\varepsilon_1, \cdots, \varepsilon_n)$ から求める残差平方和，

$$E := \sum_{j=1}^{N} \varepsilon_j^2,$$

を最小にするような偏回帰係数 $\beta := (\beta_0, \cdots, \beta_k)$ は正規方程式，

$$n\begin{pmatrix} 1 & \overline{\mathbf{x}_1} & \cdots & \overline{\mathbf{x}_k} \\ \overline{\mathbf{x}_1} & \overline{\mathbf{x}_1^2} & \cdots & \overline{\mathbf{x}_1\mathbf{x}_k} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\mathbf{x}_k} & \overline{\mathbf{x}_1\mathbf{x}_k} & \cdots & \overline{\mathbf{x}_k^2} \end{pmatrix}\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = n\begin{pmatrix} \bar{\mathbf{y}} \\ \overline{\mathbf{x}_1\mathbf{y}} \\ \vdots \\ \overline{\mathbf{x}_k\mathbf{y}} \end{pmatrix} \tag{5}$$

を解くことで計算される．ただし，$\bar{\mathbf{x}}$ はベクトル $\mathbf{x}$ の平均値である．この方程式の解決をすべて秘密計算で解決する場合，以下の３つのステップを秘密計算で処理する必要がある．

1. 総和 $n\overline{\mathbf{x}_1}, \cdots, n\overline{\mathbf{x}_k}, n\bar{\mathbf{y}}$，および，積和 $n\overline{\mathbf{x}_1^2}$, $n\overline{\mathbf{x}_1\mathbf{x}_2}, \cdots, n\overline{\mathbf{x}_k^2}, n\overline{\mathbf{x}_1\mathbf{y}}, \cdots, n\overline{\mathbf{x}_k\mathbf{y}}$ の計算，
2. 線形代数を解決するアルゴリズムを用いて，式(5)の正規方程式を解く．

この内，ステップ１は加算と積和で実現することができるため，秘密計算でも容易に実装できる．このとき，必要な積和の数は $O(k^2)$ に従う．しかし，ステップ２における正規方程式の解決は単純ではない．線形代数を解決するアルゴリズムは，ガウスの消去法に代表される直接解法とヤコブ法に代表される反復解法の２種類に大別される．

直接解法は安定した繰り返し回数で計算を実現できるため，終了条件を気にする必要はない．しかし，計算過程でピボットの選択が必要となる．ピボットの選択では対角成分が０にならない様に行列の行を入れ替える操作を行うが，どの行と入れ替えたかという情報は，データの推定に繋がる情報であるため，秘匿しなければならない．そのため，j 番目 $(0 \leq j < k)$ のピボットの選択においては，$k-j$ 回の判定と $(k-j)^2+(k-j)$ 回の乗算が必要となる．また，逆行列の計算中に最終的に $O(k^3)$ の除算が発生する．秘密計算において除算は特にコストが高く，直接解法は処理時間が大きいものとなる．

反復解法はピボット交換に必要な計算のコストが不要になる．また，収束条件についても，十分に安定できる回数だけ反復すればよく収束の判定は気にしなくてもよい．しかし，反復解法においても各反復内で直接解法と同様に各要素に対する除算が必要となる．そのため，反復回数を ℓ とすれば，全体で $O(\ell k^2)$ の除算が必要となり，秘密計算で処理する場合，直接解法と同様に計算コストは高いものとなる．

このような naïve な線形代数を解決するアルゴリズムを秘密計算上で実装した例も存在する (Bogdanov *et al.*(2016)，Lu *et al.*(2016)) が，本来秘密計算の除算は計算コストが高いため，除算を使わずに逆行列を，そして回帰分析を実行することが出来れば，計算コストが低くなり，より高速な実装が可能である．

4.3.2 統計量を利用した線形回帰

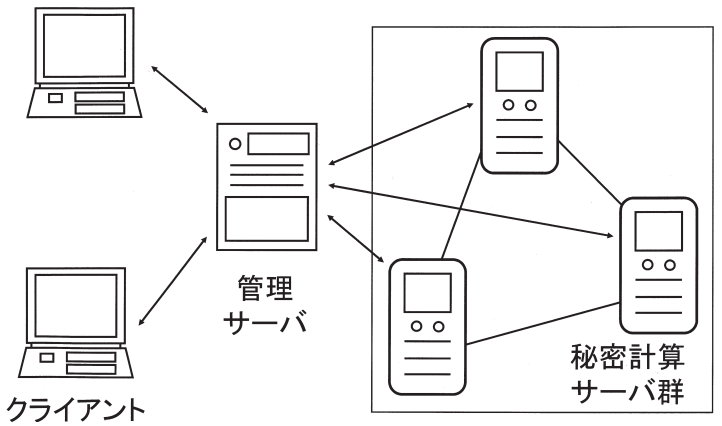
そこで，本論文では 3.2 節で議論した総和・積和を平文に戻し，公開する方式を考える．もし，$\mathbf{X}$ と $\mathbf{y}$ に関する全ての総和・積和を公開しても安全であれば，すなわち，総和・積和から個票データが推定できないのであれば，ステップ２を平文で処理しても安全である．平文の計算ではサーバ間の通信が不要となるため，ステップ２で必要となる $O(k^3)$，もしくは $O(\ell k^2)$ の除算に掛かる通信コストがなくなり，秘密計算よりも高速に処理できる．したがって，実際の秘密計算を用いた線形回帰は，各総和・積和が予め安全であることを先に抽出しておき，ステップ１の総和・積和を求める計算においてのみ秘密計算を用いて安全に求め，総和・積和を平文に戻し，ステップ２を平文で処理することで，高速に計算することができる．

4.3.3 主成分分析への適用

主成分分析についても，線形回帰と同様の手法で実現可能である．いま，レコード数 n，属性数 k のデータ $\mathbf{Z} := (\mathbf{z}_1^T, \cdots, \mathbf{z}_k^T$ に対する主成分分析を考える．このとき，主成分分析は相関係数行列，または，分散・共分散行列に対して固有値・固有ベクトル計算を施すことにより実現される．ここでは相関係数行列を用いるケースについて考える．

秘密計算を用いた線形回帰における逆行列の計算と同様に，秘密計算上で行列の固有・固有ベクトル計算を行うことは難しい．そこで，相

図 5. 秘密計算実験システムのイメージ



関係数行列の各要素に着目する．$\mathbf{Z}$ における相関係数行列の i 行 j 列の要素 $\mathrm{cor}_{i,j}(\mathbf{Z})$ は次式で定義される．

$$\mathrm{cor}_{i,j}(\mathbf{Z}) := \frac{S_{\mathbf{z}_i\mathbf{z}_j}}{S_{\mathbf{z}_i}S_{\mathbf{z}_j}},$$

ただし，$S_{\mathbf{z}_i\mathbf{z}_j}$ は $\mathbf{z}_i$ と $\mathbf{z}_j$ の共分散，$S_{\mathbf{z}_i}$ は $\mathbf{z}_i$ の標準偏差である．標準偏差は分散の平方根であるため，実際には共分散，

$$S_{\mathbf{z}_i\mathbf{z}_j} = \overline{\mathbf{z}_i\mathbf{z}_j} - \overline{\mathbf{z}_i}\,\overline{\mathbf{z}_j},$$

の式で一律に求めることができる．共分散は平均・二乗平均から導出することが可能なため，最終的には秘密計算を用いた線形回帰と同様に，総和と積和，あるいは，共分散を公開しても安全である場合に，総和と積和を公開して相関係数行列を計算し，相関係数行列に対して平文上で固有値・固有ベクトル計算を高速に実施することで安全，かつ，高速に秘密計算を用いた主成分分析が実現できる．

4.4　実験

4.4.1　実験概要

上述の秘密計算を用いた線形回帰・主成分分析の手法について，実際に一橋大学経済研究所の社会科学統計情報研究センター実験作業室に設置されている秘密計算実験システムを用いて実験を行った．秘密計算実験システムは 1 台の管理サーバと 3 台の秘密計算サーバで構成されるサーバ群と，1 台以上のクライアント端末からなるシステムである．この秘密計算実験システムでは，個票データを秘密計算サーバ上に秘密分散で秘匿した状態で格納されており，分析者はクライアント端末を用いることにより，管理サーバを介して分析の命令を秘密計算サーバに送信する．命令を受け取った秘密計算サーバは秘密計算サーバ間で秘密計算の処理を実施し，最終的に分析した結果をクライアント端末に送信することにより，分析者は秘密計算による分析結果を受け取る．また，分析者は秘密計算実験システムから受け取った分析結果を用いてより高度な分析を行うことができる．本実装では，これらの高度な分析の実装に数値計算ソフトである R を利用した．秘密計算システムのイメージを図 5，また，実験作業室に設置されている各機器の構成を表 11 に示す．

本実験では秘密計算実験システムを用いて，秘密計算を用いた線形回帰と主成分分析を実装し，それぞれの分析結果と処理時間について，数値計算ソフトである R が持つ線形回帰の関数(lm)，および，主成分分析の関数(prcomp)と比較した．また，本実験ではデータセットとして，独立行政法人統計センターの教育用擬似ミクロデータを用いた．教育用擬似ミクロデータは 2004 年全国消費実態調査を基に，擬似乱数により生成された擬似データである．このデータは，擬似的に生成された 3 万 2027 世帯の収入と支出に関するデータを 197 の属性にまと

表 11. 秘密計算実験システムの構成

管理／秘密計算サーバ	
OS	Cent-OS 6.4
CPU	Intel Xeon E3-1220 v5 (3.00 GHz, 4 コア)
メモリ	6GB
ホスト OS	Red Hat Enterprise Linux 7.2
ホストメモリ	8GB
クライアント	
OS	Windows 7 (64bit)
CPU	Intel Corei7-6600U (2.80 GHz, 2 コア)
メモリ	16GB

表 12. 実支出額に対する収入総額の線形回帰の結果

有業人員	世帯数	パラメータ		残差平方和	AIC	実行時間(秒)
		収入総額	切片			
全体	32027	0.852	0.981	2313.3	6731.2	1.02
1 人	13913	0.851	1.012	992.9	2762.1	1.07
2 人	13459	0.864	0.831	947.0	2480.7	1.05
3 人	2950	0.854	0.953	224.9	786.1	0.46

めたものである．これらの属性の内，13 属性が世帯の種類を示すカテゴリデータ，33 属性が収入に関する数値データ，150 属性が支出に関する数値データ，そして残る 1 属性がレコードの集計用乗率を示すデータである．

4. 4. 2 実験 1. 線形回帰

本実験では，擬似ミクロデータの世帯人員一人当たりの収入総額と実支出額に対して，対数値を取り，実支出額に対する収入総額の線形回帰(単回帰)の計算を，全レコード，および，有業人員数のグループ毎に行った．ただし，グループ毎の回帰では，有業人員数が 4 人以上のグループを除外している．これは有業人員数が 4 人以上のグループはレコード数が少ないためである．また，収入総額，および，実支出の対数値については事前に暗号化していない状態で計算を行っている．

線形回帰の実行時間，導出されたパラメータと，そのときの残差平方和，および，赤池情報量規準(AIC)の値を表 12 に示す．実行時間においては，いずれの場合においても 1 秒で実行可能であり，実用的な時間で処理ができていると言える．また，各パラメータに対して，秘密計算実験システム上で算出した回帰結果 $\widehat{w}$ と R の lm 関数で計算した回帰結果 w との相対誤差の絶対値，

$$\left|\frac{\widehat{w}-w}{w}\right|,$$

を計算したところ，最大で 1.67×10^{-8} あった．したがって，秘密計算上でも十分に精度の良い回帰計算ができていると言える．

4. 4. 3 実験 2. 主成分分析

本実験では，擬似ミクロデータの収入に関する 33 属性，および，支出に関する 150 属性でそれぞれ，相関係数を用いて主成分分析を行った．収入に関する属性の主成分分析結果について抜粋したものを表 13 に示す．主成分分析の結果では，33 の主成分の内，24 の主成分が有効な値を持ち，残る 9 の主成分が 0 であった．有効な主成分について，秘密計算と R の prcomp 関数による相関係数を用いた主成分分析の結果を比較したところ相対誤差はほぼ 0 であり(絶対値最大で 6.07×10^{-15})，十分な精度で計算できていると言える．一方で，残る 9 の主成分は，相対誤差は大きいものの絶対誤差は絶対値最大でも 2.20×10^{-8} とほぼ 0 である．相対誤差が大きいのは秘密計算実験システムでの小数と R の小数の精度が異なり，0 付近の値が大きくずれたためである．

支出に関する属性の主成分分析結果について抜粋したものを表 14 に示す．主成分分析の結果では 150 の主成分の内，112 の主成分が有効な値を持ち，残りの 38 成分は 0 であった．R による主成分分析の結果と比較したところ，収入に関する属性の結果と同様に，相対誤差はほぼ 0(絶対値最大で 5.56×10^{-15})であり，精度よく計算ができている．また，残りの 38 成分についても，収入の結果と同様に相対誤差は大きいものの，絶対誤差は絶対値最大で 9.83×10^{-8} と値としては誤差がないと言える．

表 13. 収入に関する属性の主成分分析結果(抜粋)

主成分	秘密計算による結果		Rのprcomp関数との差分	
	標準偏差	累積寄与率	絶対誤差	相対誤差
PC1	2.371	0.170	0.000	0.000
PC2	1.769	0.265	0.000	0.000
PC3	1.575	0.340	0.000	0.000
PC4	1.468	0.406	0.000	0.000
PC5	1.399	0.465	0.000	0.000
PC6	1.353	0.521	0.000	0.000
PC7	1.175	0.562	0.000	0.000
PC8	1.067	0.597	0.000	0.000
PC9	1.037	0.629	0.000	0.000
PC10	1.026	0.661	0.000	0.000
⋮				
PC23	0.647	0.999	0.000	0.000
PC24	0.174	1.000	0.000	0.000
PC25	0.000	1.000	0.000	418.0
PC26	0.000	1.000	0.000	466.0
PC27	0.000	1.000	0.000	636.0
PC28	0.000	1.000	0.000	760.0
PC29	0.000	1.000	0.000	1280
PC30	0.000	1.000	0.000	1350
PC31	0.000	1.000	0.000	1390
PC32	0.000	1.000	0.000	1610
PC33	0.000	1.000	0.000	5430000

表 14. 支出に関する属性の主成分分析結果(抜粋)

主成分	秘密計算による結果		Rのprcomp関数との差分	
	標準偏差	累積寄与率	絶対誤差	相対誤差
PC1	4.805	0.1539	0.000	0.000
PC2	3.023	0.2148	0.000	0.000
PC3	2.595	0.2597	0.000	0.000
PC4	2.356	0.2967	0.000	0.000
PC5	2.174	0.3282	0.000	0.000
PC6	2.012	0.3552	0.000	0.000
PC7	1.934	0.3801	0.000	0.000
PC8	1.854	0.4031	0.000	0.000
PC9	1.802	0.4247	0.000	0.000
PC10	1.760	0.4454	0.000	0.000
⋮				
PC111	0.362	0.9994	0.000	0.000
PC112	0.289	1.0000	0.000	0.000
PC113	0.000	1.0000	0.000	33.37
PC114	0.000	1.0000	0.000	25.49
⋮				
PC143	0.000	1.0000	0.000	940.40
PC144	0.000	1.0000	0.000	1051.00
PC145	0.000	1.0000	0.000	1160.00
PC146	0.000	1.0000	0.000	1720.00
PC147	0.000	1.0000	0.000	1932.00
PC148	0.000	1.0000	0.000	2263.00
PC149	0.000	1.0000	0.000	2533.00
PC150	0.000	1.0000	0.000	2802.00

表 15. 主成分分析の実行時間

	秘密計算実験システム	R(prcomp 関数)
収入(33 属性)	175.97 秒	0.28 秒
支出(150 属性)	3506.26 秒	3.24 秒
支出/収入	×19.93	×11.57

また，それぞれの主成分分析の実行時間を表 15 に示す．秘密計算システムの結果を見ると，支出に関する属性の主成分分析と収入に関する属性の主成分分析の実行時間の比が 20 倍となっており，prcomp 関数の実行時間比である 11 倍と大きく異なっている．これは，秘密計算において相関係数を求める際の積和がボトルネックとなっているためである．一般に，行列の相関係数行列の計算では属性数 k に対して，$k(k+1)/2$ 回の積和が必要である．積和の計算に 1 回につき 1 度のサーバ間通信が必要となる．実際に，収入に関する 33 属性と支出に関する 150 属性の積和回数を比較すると，$\frac{150\times151/2}{33\times34/2}\cong20.19$ 倍であり，サーバ間の通信が実行時間に大きく影響を与えていることがわかる．

この積和によるボトルネックを解消する手段

としては，複数の計算における通信をまとめる方法が考えられる．ベクトルや行列の要素積の様に，個々の計算を独立して並行に処理できる場合，それぞれの計算データをまとめて通信することにより全体の通信回数を削減できる．これにより，サーバ間の通信に係わる遅延を削減し高速化することが期待できる．

5. 結論

本論文では，データを暗号化した状態で集計ができる「秘密分散・秘密計算システム」に，SDC に基づいた暗号化すべき統計量と復号化後に集計する統計量の基準を設定するため，数量表における安全性の評価基準を示すとともに，線形回帰や主成分分析について秘密計算と平文の計算における計算時間などを比較した．

この結果，SDC に基づいた公的統計ミクロデータの実証分析では，分析に用いる各変数の分散値や共分散が安全性の評価に大きく影響していることが分かった．それゆえ，暗号化計算すべき統計量と復号化後に計算できる統計量の区分を明らかにすることが可能となった．また，この SDC の基準を「秘密分散・秘密計算システム」に導入するため，分散・共分散の値を各変数の平方和に変換することにより，すべてを暗号化した状態により，安全性の検証と計算速度が向上したことが確認できた．このことにより，SDC 機能を付加した「秘密分散・秘密計算システム」を利用することにより，暗号化した状態での公的統計データの実証分析が可能なった．

（一橋大学経済研究所・
NTT セキュアプラットフォーム研究所）

参 照 文 献

［online1］「公的統計ミクロデータ研究コンソーシアム―オンサイトネットワークの形成」，大学共同利用機関法人　情報・システム研究機構ホームページ，url: http: //www. rois. ac. jp/tric/micro/moc/onsite. html，2017 年 11 月 11 日アクセス．

［online2］「オンサイト利用」，独立行政法人　統計センターホームページ，url: http: //www. nstac. go. jp/services/on-site.html，2017 年 11 月 11 日アクセス．

［online3］「統計改革推進会議最終取りまとめ」，首相官邸ホームページ，url:https: //www.kantei.go.jp/jp/singi/toukeikaikaku/pdf/saishu_honbun. pdf，2017 年 11 月 11 日アクセス．

Aliasgari, Mehrdad, Maria Blanton, Yihua Zhang and Aaron Steele (2013) “Secure Computation on Floating Point Numbers,” *20th Annual Network and Distributed System Security Symposium* (*NDSS 2013*).

Ben-Or, Michael, Shafi Goldwasser and Avi Wigderson (1988) “Completeness Theorems for Non-Cryptographic Fault-Tolerant Distributed Computation,” Janos Simon, *Proceedings of the twentieth annual ACM symposium on Theory of computing* (*STOC'88*), pp. 1–10.

Bogdanov, Dan, Sven Laur and Jan Willemson (2008) “Sharemind: A Framework for Fast Privacy-Preserving Computations,” Sushil Jajodia and Javier Lopez, *13th European Symposium on Research in Computer Security* (*Computer Security-ESORICS 2008*), Springer LNCS, Vol. 5283, pp. 192–206.

Bogdanov, Dan, Liina Kamm, Sven Laur and Ville Sokk (2016) “Rmind: a Tool for Cryptographically Secure Statistical Analysis,” *IEEE Transactions on Dependable and Secure Computing*.

Brandt, M., Franconi, L., Guerke, C., Hundepool, A., Lucarelli, M., Mol, J., Ritchie, F., Seri, G. and Welpton, R. Guidelines for the Checking of Output Based on Microdata Research. Final Report of ESSnet Sub-Group on Output SDC, Eurostat, 2010.

Chida, Koji, Gembu Morohashi, Hitoshi Fuji, Akiko Fujimura, Koki Hamada, Dai Ikarashi and Ryuichi Yamamoto (2014) “Implementation and Evaluation of an Efficient Secure Computation System Using ‘R’ for Healthcare Statistics,” *Journal of the American Medical Informatics Association* (*JAMIA*), Vol. 21, Issue e2, pp. e326–e331.

Damgård, Ivan, Matthias Fitzi, Eike Kiltz, Jesper B. Nielsen and Tomas Toft (2006) “Unconditionally Secure Constant-Rounds Multi-party Computation for Equality, Comparison, Bits and Exponentiation,” Shai Halevi and Tal Rabin, *Third Theory of Cryptography Conference* (*TCC 2006*), Springer LNCS, Vol. 3876, pp. 285–304.

Gennaro, Rosario, Michael O. Rabin and Tal Rabin (1998) “Simplified VSS and Fast-track Multiparty Computations with Applications to Threshold Cryptography,” Brian Coan and Yehuda Afek, *Proceedings of the Seventeenth Annual ACM Symposium on Principles of Distributed Computing*, pp. 101–111.

Goldreich, Oded, Silvio Micali and Avi Wigderson (1987) “How to Play any Mental Game,” Alfred V. Aho, *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing* (*STOC'87*),

pp. 218–229.
Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K. and de Wolf, P. P. (2012) *Statistical Disclosure Control*. Wiley, United Kingdom, 2012.
Kamm, Liina and Jan Willemson (2015) "Secure Floating-Point Arithmetic and Private Satellite Collision Analysis," *International Journal of Information Security*, Vol. 14, Issue 6, pp. 531–548.
Lu, Wen-jie, Shohei Kawasaki and Jun Sakuma (2016) "Using Fully Homomorphic Encryption for Statistical Analysis of Categorical, Ordinal and Numerical Data," *IACR Cryptology ePrint Archive 2016*, No. 1163.
Shamir, Adi (1979) "How to Share a Secret," *Communications of the ACM*, Vol. 22, Issue 11, pp. 612–613.
Shirakawa, Kiyomi Yutaka Abe and Shinsuke Ito (2016) Empirical Analysis of Sensitivity Rules: Cells with Frequency Exceeding 10 that Should Be Suppressed Based on Descriptive Statistics, In: Josep Domingo-Ferrer and Mirjana Pejić-Bach eds., Privacy in Statistical Databases: UNESCO Chair in Data Privacy, International Conference, PSD 2016, Dubrovnik, Croatia, September 14–16, 2016, Proceedings, Springer, 2016, pp. 28–40.
Shirakawa Kiyomi and Akinori Sato (2017) "Challenges in Improving the Quality and Amount of Statistical Information for Public Use—New Uses of Official Statistics in Japan—," The 61th World Statistics Congress, 2017, pp. 1–6.
Shirakawa, Kiyomi Yutaka Abe and Shinsuke Ito (2016) Creating an Academic Use File Based on Descriptive Statistics: Synthetic Microdata from the Perspective of Distribution Type, In: Josep Domingo-Ferrer and Mirjana Pejić-Bach eds., *Privacy in Statistical Databases: UNESCO Chair in Data Privacy, International Conference, PSD 2016*, Dubrovnik, Croatia, September 14–16, 2016, Proceedings, Springer, 2016, pp. 149–162.
Yao, Andrew C. (1986) "How to Generate and Exchange Secrets," *27th Annual Symposium on Foundations of Computer Science* (*SFCS 1986*), pp. 162–167.