

# 評価情報流出の根源的自動防止と相関分析

代表研究者 岩井 淳 群馬大学 社会情報学部 教授

## 1 はじめに

アンケート調査をはじめとする社会調査では、しばしば被調査者に調査者に対する匿名性を保証する。しかし、被調査者自身の回答の特徴により、結果的に匿名性の維持が困難になる場合も多い。例えば、男性19名、女性3名が対象の小規模調査で質問票に性別の質問項目があれば、女性の匿名性は損なわれがちであろう。この結果的な個人情報流出のリスクは、大規模な調査の場合も存在する。性別、年齢、専攻等の個人属性のクロス集計で少人数回答者のセルが生じるためである。本問題は規模と関連なく一般的な社会調査全体に当てはまる。

上述の問題状況では、次の事態が生じがちである：

- ◆1 被調査者が正しいデータを入力することをためらう気持ちになる
- ◆2 調査者によるデータの分析が制限される

授業評価の用語を用いるならば、◆1は、誰が評価を行ったのかを確実に隠蔽する仕組みを導入できなければ、学生が正しい評価データのを入力をためらいかねないということである。◆2は、しばしば事務室などの組織が学生の担当教員に対する匿名性を保証しようと単純集計のみを公開し、「個票」を隠蔽するためである。個票がないため、授業担当者は設問項目を越えての多変量解析などを行うことができない。一方で学生の不安は残る。学校組織の一部である事務室が個票をもつためであり、この点では◆1の問題も解決していない。

この「意図せざる情報流出」の問題を自動的に防ぐ調査システムとして、「動的に質問票構造を変えることで被調査者に匿名性を保証する社会調査システム」の提案がある([3], [5], [6])。すなわち、入力値分析により問題要因を見出し、問題回避のためにデータベース内の質問票設計情報を修正する機構を用いる提案である(「2 先行研究」参照)。調査者が被調査者の入力データを手にする前にデータベース内設定を恒久的に変更するため、将来的にも情報流出の生じ得ない根源的自動防止を実現できる。

先行研究の開発システムには、しかしながら、ごく一部の被調査者の匿名性を保証するために、全被調査者の属性回答を一律に切り離し、これらの設問への回答との相関分析を不能にしてしまう特徴もあった。

本研究は、この課題を克服するための数理的研究と実装作業を行うことを目的とする。すなわち、「動的に質問票構造を変えることで被調査者に匿名性を保証する社会調査システム」の拡張開発を行い、属性設問の回答との相関分析が困難となりやすい従来の課題を除く。

本報告の以下の構成は次の通りである。2では、以降の議論の準備として先行研究を要約する。3では、本研究のシステム設計を説明する。4では、開発システムの利用実験と評価を要約する。

## 2 先行研究

### 2-1 質問票構造の動的変更

先行研究([3], [5], [6])における「動的に質問票構造を変えることで被調査者に匿名性を保証する社会調査システム」の提案を要約する。

先行研究は、任意の質問票の設問は、外的に観察可能な「個人属性に関する設問項目」と、外的に観察不能な「内的評価に関する設問項目」に区分できるという仮定に基づく。本研究においてもこの着想が基本となる。以下、必要に応じて、前者をX項目、後者をY項目と略称する。

先行研究は、X項目とY項目の連結に匿名性が失われるリスクを見出し、リスクの大きな場合にその連結を自動的に除く仕組みを設ける。例えば、前述の授業評価の問題設定では、質問票を「性別の質問項目」だけの質問票1と「それ以外の質問項目」(授業評価等のY項目のみとする)からなる質問票2に分けて調査を行えば、匿名性は損なわれず、各項目について適切な回答が得られるであろう。先行研究の調査システムで

は、この操作を、各被調査者が回答してしまった後で自動的に行う。すなわち、被調査者が入力した回答をシステムが調査者に先行して分析し、匿名性が失われる問題を検出する。この上で、調査者が個票を確認する前に質問票設計を自動変更して問題を回避する。

以下、具体的な設問例を用いて説明する。図1は、一部の国立大学で用いられる授業評価質問票に僅かに修正を加えた説明用の質問票である。(A)の「属性項目」の3項目がX項目に対応する。以下、(B)の△記号の右側の選択肢が選択者を隠蔽すべきケースと前提して議論する(設問2から設問5は設問1と同様)。

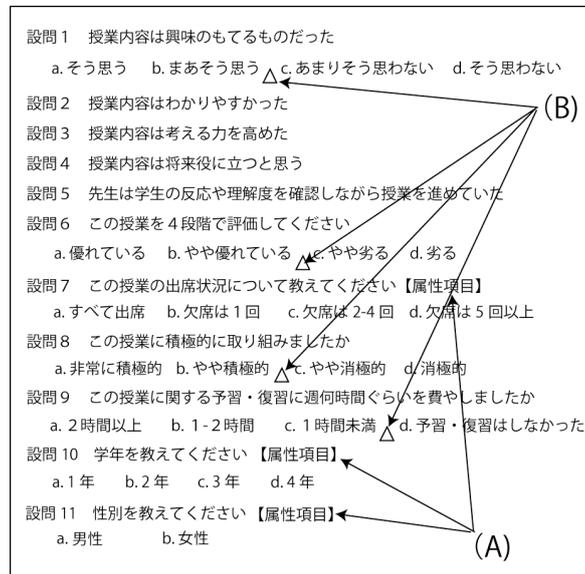


図 1. 授業評価質問票の例 ([3])

図2は、設問番号7, 10, 11の3設問のX項目について、調査者はこの順にY項目との関連に関心を抱いている前提で、システムの動作を例示した図である。調査者は、この優先順位情報をシステムに事前設定する。また、調査者は設問6, 8, 9について多変量解析を行う予定があり、これらをまとめて1ブロックに、他の設問はそれぞれ単一で1ブロックに設定するとする。結果、設問番号として{1}, {2}, {3}, {4}, {5}, {6, 8, 9}の6ブロックを生じる。実際の調査が行われ、被調査者の回答が終了すると、システムは匿名化の自動処理を開始する。

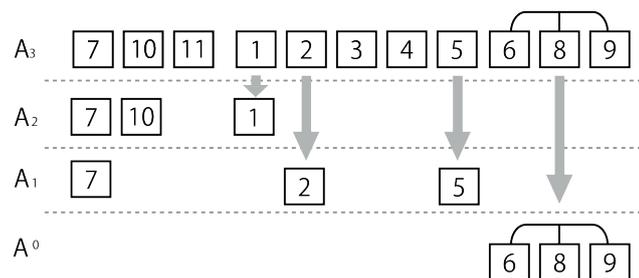


図 2. 質問票構造の動的変更プロセスの例 ([3])

X項目である設問7(4選択)、10(4選択)、11(2選択)の構成から、回答者は理論的に $4 \times 4 \times 2$ の32の属性グループのいずれかに分類される。システムは、各設問ブロック内で、32グループのそれぞれについて匿名性水準の検討を行う。問題があれば、システムは優先順位の低い順に(11, 10, 7の順に)X項目における回答との関連情報を除く処理を施す。ここで問題があるとは、匿名性が失われるリスクが大きい場合であり、全回答者の数Nとその内の否定的回答者(図1における(B)の△記号の右の選択肢の選択者)の数Mの数値から算出される組み合わせ数の対数が一定の閾値を下回る場合として検出する(2.2参照)。ただし、評価に

際して全員が高評価を下した場合（対数の値は0となる）の匿名性水準は問題としないものとする。

この結果、6ブロックは、X項目の回答情報の残り方に0から3までの差が生じる結果となる。X項目の設問の残り方が同様であるようなブロック群は最後にまとめ、A0, A1, A2, A3の4出力を得る。図1の例では、A0は設問6, 8, 9に、A1は設問2, 5, 7に、A2は設問1, 7, 10に、A3は設問3, 4, 7, 10, 11にそれぞれ対応している。（数理的詳細は、[3], [5], [6]を参照。）

過去の開発システムの匿名性保証の手法では、しかしながら、性別・学年・欠席回数などの属性設問を論理的に切り離すことによって分析上の制約が生じがちであった。すなわち、n個のX項目の設問X1, X2, ..., Xnがあり、調査者の関心が設問番号順に強いとする（匿名性損失のリスクがある場合、Xnから逆順に属性を除く）。また、設問Xiの第j選択肢の回答者群をresG(Xi(j))とする。このとき、例えば、設問X1でresG(X1(j))が例外的に小規模であり、かつ匿名性水準が許容水準を下回れば、X1の他の選択肢への回答を含め他の全体的回答パターンに関連なく全X項目の属性が失われる。結果として、属性設問との相関分析は不能となる。

本研究ではこの課題に対応するため、全被調査者でなく、一部の被調査者の属性回答を選択的に分析不能とすることで、匿名性を保証する仕組みを開発する。最終出力の誤差を最小限に留めることを重視する。

## 2-2 匿名性水準の測定

先行研究および本研究は、匿名性水準の測定手法についても特徴がある。すなわち、報告者が提案した組み合わせ型（Combinatorial Approach）の尺度を用いる（[2], [4]）。同尺度はシャノン理論([8])と整合性があり、k-匿名性([9])や1-多様性([7])の尺度よりも情報量の観点で精緻な特徴がある。例えば、投票の匿名性水準の用語を用いるならば、投票者による各投票内容の告白がもつ情動的価値の総和を投票全体の匿名性水準と位置づけて計算を行う。各投票の情動的価値はシャノンの情報理論の自己情報量に倣い、 $-\log$ （生起確率）と定める（底aは $a>1$ で任意）。単純投票において、全投票者数をN、賛成投票者数（もしくは反対投票者数）をMとするとき、以下の計算式となる。

$$\log \frac{N!}{M!(N-M)!}$$

この計算による匿名性水準の算出は、ネットワークセキュリティの文脈の研究（[1]）で提出された匿名性水準の算出と同形である。

## 3 相関分析支援のためのシステム設計

### 3-1 Method 1（少数者の分離）

本研究の目的上、最もシンプルな手法は、少数の被調査者が対応する属性組み合わせのセルで、その少数者を他の被調査者と単に分離する手法である。前述の例を用いれば、n個のX項目の設問X1, X2, ..., Xnの設問X1でresG(X1(j))が多数の属性削除の原因となる場合、当該のメンバーを最初に分離するして他の属性の削除を回避する手法である。ただし、分離した少数者の集計データを調査者が直接に得ると、当該の少数者の匿名性は損なわれる（属性組み合わせで抽出した少数者は特定が容易になりがちである）。この問題への自然な対応は、他の多数の被調査者から無作為抽出した少数名を同様に分離し、これらを組み合わせた被調査者の集計を調査者が得る手順の実現であろう。以上をMethod 1と略称する。

Method 1は、例えば、50人クラスで、出席回数に関するX項目設問の4選択肢について「すべて出席」40名、「1回欠席」が9名、「欠席は2-4回」が1名、「欠席は5回以上」が0名であり（他のX項目の設問はないとする）、あるY項目の設問で少数の否定的回答があるケースで、「欠席は2-4回」の1名が否定的回答者である場合に、同1名を中心とした分離操作を行う手法である。すなわち、当該の1名および他の49名からの無作為抽出した少数名を分離し、他の多数の被調査者と分けて集計する。従来手法では、全被調査者について出席回数の属性との連結を除くため、出席回数との相関分析が不能となる。しかし、連結を除く被調査者が一部であれば一定の相関分析が可能となり、被調査者のプライバシーも一定水準確保される。

### 3-2 Method 2（属性削除順序の最適化）

Method 1は、しかしながら、データの正確性を損なう上（相関値の信頼性の低下）、属性削除順序に自由度がある場合には、少数者を分離せずにある程度の相関分析を実現できる場合がある。例えば、上述の50

人クラスの簡単な例を用いて、出席回数他に X 項目の設問として性別設問があり、「すべて出席」、「1 回欠席」、「欠席は 2-4 回」、「欠席は 5 回以上」の順に男性が 20 人、4 人、0 人、0 人、女性が 20 人、5 人、1 人、0 人とする。性別設問の優先順位が低く、同属性が先に除かれるとすれば問題状況は前述と同様となる。すなわち、「欠席は 2-4 回」の 1 名の問題が残る、Method 1 を用いるか、欠席回数を含めて全属性除く必要が生じる。しかし、もし欠席回数の属性を先に除けば、男性 24 人、女性 26 人の状況となる。この場合は性別属性との相関分析が可能であり、匿名性が損なわれるリスクも小さい。一部データの分離前であるため、データの正確性を損なう問題も回避できる。Method 1 による少数者分離の前に、分離を必要としない順序で属性削除を行えば、データの正確性を保つと共に属性設問との相関分析が可能となる。

この観点より、可能な属性削除順序で最も適切な順序で属性を除く手法を Method 2 と略称する。従来設計では、質問票構造の動的変更が必要となる設問項目の削除順序は、調査者が事前設定する必要があった。しかし、Method 2 では、調査システム側で属性削除順序を自動設定することを前提する。調査者は X 項目の優先順位の設定機会を失うが、この事前設定自体が全 X 項目との相関を不能にする原因となる場合があり、どの結果になるかは実調査を行うまで不明な仕組みであった。Method 2 は、可能な限り調査者側に相関分析の可能性を大きく残す手法であり、調査者側のメリットもより大きな場合が多いと見込まれる。

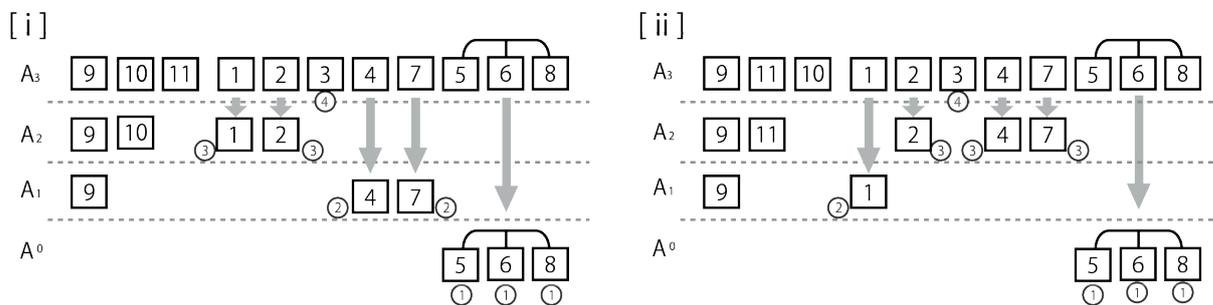


図 3. 属性削除順序の比較例

Step 1:

For all  $zmap \in ZMAP(X)$  do the following:

Sub-step 1:

For each  $AB_j (1 \leq j \leq M)$ , perform the following procedure:

i) If  $Pr_j(AB_j, 2) \cap Z = \emptyset$  or  $Flag(AB_j) = 0$ , then proceed to Step 2.

ii) If  $Pr_j(AB_j, 2) \cap Z \neq \emptyset$  and  $Flag(AB_j) = 1$ ,  $AB_j = Del(AB_j, 2, \{z_{Pr_j(AB_j, 2) \cap Z}\})$  and repeat Step 1.

Sub-step 2:

Perform the following procedure:

i)  $A_{Attributes} = Rand(Del(T_0, 2, Y), 1)$

ii) For  $k (0 \leq k \leq n = |Z|)$ ,

$$A_k = Rand\left(\bigcup_{|Pr_j(AB_j, 2) \cap Z|=k} (AB_j), 1\right)$$

iii) Delete the data except for that of  $A_{Attributes}, A_0, A_1, A_2, \dots, A_n$

iv) Record  $A_{Attributes}, A_0, A_1, A_2, \dots, A_n$

and the score of  $EvalOrder = \sum_{Layer=0}^n (|Select(A_{Layer}, 2, Z) + 1| \times |Select(A_{Layer}, 2, Y)|)$

Step 2:

i) For the maximum value of EvalOrder, output  $A_{Attributes}, A_0, A_1, A_2, \dots, A_n$ .

ii) If there are multiple  $zmap$ s that give the same maximum value, one is selected at random.

図 4. 基本動作

Method 2 を用いる場合、「最適な属性削除順序」を決定するアルゴリズムが必要となる。本研究では、各 Y 設問項目とのもつ X 項目の設問との連結数の合計を属性削除順序の評価値として用いる方式とした。この方式は、X 項目との相関分析の可能性を多く残す観点から自然な選択と判断される。図 3 は、最適な属性削除順序を決定するための 2 つの順序の比較例である。ここでは設問番号 9, 10, 11 の 3 設問が X 項目であり、[i]では 11, 10, 9 の順に、[ii]では 10, 11, 9 の順に Y 項目との関連付けを除き、その結果が図の通りに得られるとする。各 Y 項目の設問近くの円を付した数値は、関連付け得る X 項目の設問数に 1 を加えた数値であり、設計手法では、その合計値を調査者の入手情報量の評価値として比較する。すなわち、調査者が X 項目と Y 項目の設問を関連付け得る水準を比較して最大化する手法を取る。図 3 の例では、[i]が 17, [ii]が 18 の値となり、後者がより高く評価される。

図 4 は、システム実装のための形式的記述における上記動作の対応部分である。この形式的記述は、文献 [3]の形式的記述を基礎に Method 2 の機構を加えたもので、集合 Z と関数 ZMAP() の導入が主な変更点である。ZMAP(X) は、X (X 項目の集合) から Z (=X) への可能な全 1 対 1 対応を返す関数である。ZSET(zmap) は、ZMAP(X) の要素である zmap の入力に対して、集合 Z を返す関数である。Z=(z<sub>1</sub>, z<sub>2</sub>, ..., z<sub>n</sub>,) のインデックス番号は、属性削除の優先順位を反映する (数値の大きいものが先に削除される)。また、属性削除順序の評価を実現するのは、主として下部の枠線部分である。(形式的記述の全体および各動作の詳細は「発表資料 1」を参照。)

以上の設計に基づき社会調査システムの実装を行なった。システムの動作は選択設定でき、従来方式と Method 2 の選択を基本とする。Method 2 は、実装上是 Method 1 の拡張であり、可能な全 1 対 1 対応で問題が解決されない場合に Method 1 に移行する。動作の選択は、被調査者が各設問に回答する手順に影響を与えないため、各実装は被調査者からは相違がない。一方で、Method 2 における属性削除順序からの自動選択は被調査者の回答パターンの反映であるため、この自動選択自体が被調査者の匿名性をやや損なう可能性がある。この側面が被調査者の意向に反することが懸念されるが、以下に説明する学生によるシステム評価では概ね良好な評価が得られた。実装システムの利用実験とともに整理する。

## 4 実装システムの利用実験

### 4-1 実装システムによる授業評価

前述のシステム設計に基づき、新たな社会調査システムを実装した。実装システムによる授業評価を紙面調査と既存の電子調査システムと比較する形式で行った。

ある国立大学の「データベース」の授業で受講生 72 名に 3 種類の授業評価を行なった。第 1 授業評価は紙面調査であり、講師により最終講義日 (2019 年 1 月 30 日) に行なった。第 2 授業評価は同大学のオンラインの授業評価システムによる調査であり、入力期間には幅があるが、多くの入力は講師に促されて最終講義日から一週間後の試験日 (2019 年 2 月 6 日) に行われた。第 3 授業評価は、本研究の開発システムによる調査で、やはり試験日に開発システムの説明後に行なった。各調査は図 1 の設問構成を基本とする (第 2 授業評価のみ X 項目が僅かに異なるが Y 項目として同様の 8 項目を含む)。開発システムの説明では、Method 2 を「一部データ方式」、従来方式を「全データ方式」と紹介し、本授業評価では必要な場合に全被調査者の全属性を除く従来の「全データ方式」とすると保証した上で調査を行なった。回答者数は第 1 授業評価が 70 名 (回収率 97.2%)、第 2 授業評価が 47 名 (回収率 65.3%)、第 3 授業評価が 46 名 (回収率 63.9%) だった。

3 調査は、後述の実装システムの評価に先立って行なった。開発システムの実際の利用を通して、より厳密なシステム評価を得ることが目的である。このため特定の授業評価自体はシステム評価と別であるが、システム評価と関連する結果傾向を要約する。各設問について、被調査者学生の回答の選択肢番号の平均値を比較した (選択肢番号の数値は、基本的に被験者の否定的回答傾向の度合いを反映する)。匿名性保証の点から第 3 授業評価の数値が大きいことが予想されたが、統計的に有意に第 3 授業評価の数値が大きな設問は見出されなかった。むしろ、第 3 授業評価の数値は小さい (授業としての評価が高い) 傾向が見られ、設問 1 では (第 1:2.214, 第 2:1.894, 第 3:1.897, Ex. 1-2: p-value = 0.0365\*, Ex. 1-3: p-value = 0.0188\*), 設問 3 では、(第 1:1.629, 第 3:1.397, Ex. 1-3: p-value = 0.0257\*) となった。p-value の値は、T 検定を行なった結果の P 値である。例えば、設問 1 の Ex. 1-2:p-value の値は、第 1 授業評価と第 2 授業評価の設問 1 の回答結果に対する T 検定の P 値に対応する。以上の点以外に、統計的に有意な回答結果は見出されなかった。(詳細は、後述の開発システムに対する評価の点を含めて「発表資料 2」を参照。)

#### 4-2 開発システムに対する評価

前述と同じ被調査者 72 名に対して、開発システムの評価に関する調査を同日に紙面調査として行なった（回収率 100.0%）。以下に設問と回答結果を要約する。基本的に、選択肢 a, b, c, d は「1. と思う, 2. まあと思う, 3. あまりそう思わない, 4. そう思わない」に対応する。ただし、設問 5 の選択肢 a, b, c, d は「1. 優れている, 2. やや優れている, 3. やや劣る, 4. 劣る」に対応する。設問 9 の選択肢 a, b, c, d は「1. 理解できた, 2. まあ理解できた, 3. あまり理解できなかった, 4. 理解できなかった」に対応する。

設問 1 本システムの匿名化処理の仕組み（質問票の分割）は理解できましたか

- |    |             |    |             |
|----|-------------|----|-------------|
| a) | 19 [26.39%] | b) | 37 [51.39%] |
| c) | 15 [20.83%] | d) | 1 [1.39%]   |

設問 2 本システムの匿名化処理により、より正確な授業評価データが得られると思いますか

- |    |             |    |             |
|----|-------------|----|-------------|
| a) | 23 [31.94%] | b) | 36 [50.00%] |
| c) | 13 [18.06%] | d) | 0 [0.00%]   |

設問 3 本システムの匿名化処理の説明は真実であり、教員はあなたをだましていないと思いますか

- |    |             |    |             |
|----|-------------|----|-------------|
| a) | 38 [52.78%] | b) | 27 [37.50%] |
| c) | 4 [5.56%]   | d) | 3 [4.17%]   |

設問 4 本システムの匿名化処理は正確であり、バグによる情報漏洩等はないと思いますか

- |    |             |    |             |
|----|-------------|----|-------------|
| a) | 8 [11.11%]  | b) | 42 [58.33%] |
| c) | 20 [27.78%] | d) | 2 [2.78%]   |

設問 5 本システムを 4 段階で評価してください

- |    |             |    |             |
|----|-------------|----|-------------|
| a) | 21 [29.17%] | b) | 50 [69.44%] |
| c) | 1 [1.39%]   | d) | 0 [0.00%]   |

設問 6 一般に、授業評価は授業の改善に役立っていると思いますか

- |    |             |    |             |
|----|-------------|----|-------------|
| a) | 7 [9.72%]   | b) | 35 [48.61%] |
| c) | 28 [38.89%] | d) | 2 [2.78%]   |

設問 7 紙ベースの授業評価方法を本システムに置き換える方がよいと思いますか

- |    |             |    |             |
|----|-------------|----|-------------|
| a) | 15 [20.83%] | b) | 33 [45.83%] |
| c) | 20 [27.78%] | d) | 4 [5.56%]   |

設問 8 現在のオンラインの授業評価方法を本システムに置き換える方がよいと思いますか

- |    |             |    |             |
|----|-------------|----|-------------|
| a) | 15 [20.83%] | b) | 38 [52.78%] |
| c) | 17 [23.61%] | d) | 2 [2.78%]   |

設問 9 本システムの「全データ方式」と「一部データ方式」の違いが理解できましたか

- |    |             |    |             |
|----|-------------|----|-------------|
| a) | 4 [5.56%]   | b) | 28 [38.89%] |
| c) | 35 [48.61%] | d) | 5 [6.94%]   |

設問 10 本システムでは「一部データ方式」の方が「全データ方式」より良いと思いますか

- |    |             |    |             |
|----|-------------|----|-------------|
| a) | 8 [11.11%]  | b) | 39 [54.17%] |
| c) | 23 [31.94%] | d) | 2 [2.78%]   |

## 5 議論

プライバシーを考慮した本開発システムの設計は、被調査者より概ね高く評価された。開発システムに対する評価のうち、全体設計に関する1から8の設問への回答傾向から、この点を確認できる。また、Method 2（一部データ方式）と従来方式（全データ方式）に関して、前者が後者に比べ相関分析の可能性を増す一方で、最適な属性削除順序の自動選択自体が被調査者に関する情報を与えることが懸念されたが、設問10の回答からは、むしろ前者の方が好意的に評価された。被調査者には、設問6の回答のように授業評価結果の有効活用に疑問があり、この点の影響があるかもしれない。すなわち、被調査者には、調査者による相関分析の可能性が増すことで回収データの有効利用が進むことに関して期待があるかもしれない。ただし、設問9よりMethod2と従来方式の違いの理解不足の問題があることが挙げられる。完全に肯定的とは言えない他の設問項目についても、さらに検討を重ねる必要がある。

設問4における22名の否定的回答は、特に注目される。開発システムのプライバシー保全性を評価しつつも、受講生はなお情報漏洩の可能性が残ると判断する模様である。本開発システムでは、調査者が被調査者の入力データを手にする前にデータベース内設定を恒久的に変更するため、実際に情報流出が生じることは考えにくい。その処理はシンプルであり、誤りが含まれる可能性も小さい。しかし、設問1、設問9の回答結果からも、開発システムの仕組みが被調査者には十分に理解されなかったと判断される。本開発システムの実用には、技術的なシステム実装とは別に、システム構造に関するより明快な説明法の開発が必要と見込まれた。

被調査者には、紙面での授業評価における匿名性保全に対する楽観的傾向も認められた。すなわち、設問7における24名の否定的回答から、紙面調査は必ずしも低評価でない。また、システム評価に先立つ授業評価では、匿名性の保全性の高い本開発システムでより低い授業評価結果が得られると見込まれたが、統計的に有意な差のある低評価は得られなかった。むしろ僅かに統計的に有意な差のある高評価が得られた。第2授業評価、第3授業評価の回収率は紙面調査の第1授業評価より低く、紙面でない電子的な授業評価は、より授業に積極的な受講生が参加する傾向があると判断される。授業に積極的な受講生は授業を高く評価する傾向があり、統計的に有意な差のある高評価はこの事情が影響したと思われる。本開発システムで匿名性保全の機能を有効に活用するには、紙面調査と同様の水準で被調査者に回答を促す仕組みも重要と判断された。

中長期的には、本研究の属性削除順序の自動決定手順の再検討も重要である。本研究では最も妥当な手順の一つを実現していると思われるが、一律でなく個別の調査者がより価値があると判断する相関分析を実現できれば有益である。このための代替的な自動決定手順の検討も今後の課題である。

## 6 結び

本研究では、動的に質問票構造を変えることで被調査者に匿名性を保証する新型社会調査システムの拡張開発を行った。属性設問の回答との相関分析が困難となりやすい従来の課題を除く作業を行なった。属性削除順序の自動決定を行うことで最終出力の誤差を最小限に留めつつ、全被調査者でなく一部の被調査者の属性回答を選択的に分析不能とすることで、匿名性を保証する仕組みを開発した。

システム利用を踏まえての学生調査の結果は、開発システムがプライバシーを改善すると評価されたことを示している。一部の被調査者の属性回答を選択的に分析不能とする点を含めて、システム全体に概ね肯定的な評価が得られた。ただし、被調査者には紙面調査にも楽観的な傾向が認められ、また本開発システムについても情報漏洩の可能性についての危惧が示された。本開発システムの実用には、被調査者に対するシステム構造に関するより明快な説明法の開発も必要と見込まれ、この点が主な課題である。中長期的には、本研究の属性削除順序の自動決定手順の再設計も重要であり、この問題の検討も継続していきたい。

### 【参考文献】

- [1] Edman, M., Sivrikaya, F., and Yener, B. (2007): A combinatorial approach to measuring anonymity. In *Intelligence and Security Informatics, 2007 IEEE*, pp. 356–363.
- [2] 岩井淳, 投票行為における匿名性概念の形式化, 第35回数理社会学会大会研究報告要旨集, pp. 92-93, 2003.

- [3] 岩井淳, 個人情報漏洩を入力値分析を用いて防ぐ調査システムの設計, 社会情報学会(JSIS & JASI)合同研究大会研究発表論文集, pp. 433-436, 2011.
- [4] A. Iwai, "Evaluation of an Anonymity Measure as an Index of Voting Privacy", Journal of Socio-Informatics, Vol.5, No.1, pp11-25, 2012.
- [5] A. Iwai, "Reviewing Privacy-Enhanced Social Survey System that Employs Combinatorial Anonymity Measure", IMECS2016 (International Multi-Conference of Engineering and Computer Scientist 2016) proceedings, pp.311-316, 2016.
- [6] A. Iwai, "A Design of Privacy-Enhanced Survey System that can be Used for Hospital Evaluation by Patients", Proceedings of ICMEMIS2017(1st International Conference on Mechanical, Electrical and Medical Intelligent System), 2017.
- [7] Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M.,  $\ell$ -diversity: Privacy beyond k-anonymity, in Proceedings of IEEE International Conference on Data Engineering (ICDE), pp. 24–35., 2006.
- [8] Shannon, C. E., A Mathematical Theory of Communication, The Bell System Technical Journal, Vol.27, pp379-423, pp. 623-656, 1948.
- [9] Sweeney, L., k-Anonymity: A Model for Protecting Privacy, International Journal of Uncertainty Fuzziness and Knowledge Based Systems, Vol. 10, No. 5, pp. 557-570, 2002.

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
Optimization of Attributes Elimination Order in a Privacy-Enhanced Survey System	Proceedings of ICTSS2018 (International Conference on Technology and Social Science 2018), Kiryu, Japan. [査読付] (Best Paper Award)	2018年4月19日
Evaluation of a Privacy-Enhanced Survey System	Proceedings of ICTCBW2019 (International Conference on Technological Challenges for Better World 2019), Cebu, the Philippines. [査読付] (Best Paper Award)	2019年3月28日