

個人データの各種利用方法の可否に関する社会調査結果と分析

中川裕志^{1*} 菊池浩明²

概要: 個人データの保護と利活用に関しては、法的には個人情報保護法の匿名加工情報、仮名加工情報などが導入され、実用的には医療情報や移動履歴などに焦点が当たっている。しかし、個人データの利活用や保護は一般人が望むものなのかどうか明確に示されてこなかった。この報告では、この問題に対する具体的答えとして、一般人 1045 人を対象にしたアンケート調査を行い、得られたデータを統計的に分析した結果について報告する。

キーワード: 個人情報保護法, 匿名加工情報, 仮名化, 医療データ, 移動履歴, アンケートによる社会調査

Results and analysis of social surveys on the availability of various uses of personal data

Hiroshi Nakagawa^{1*} Hiroaki Kikuchi²

Abstract: With regard to the protection and utilization of personal data, legally, the Personal Information Protection Law's anonymous processed information and pseudonymized processed information have been introduced, and practically, the focus has been on medical information and movement history. However, it has not been clearly indicated whether the utilization and protection of personal data is what the general public desires. This report describes the answers of a questionnaire survey of 1,045 members of the general public as a concrete answer to this question, and the results of statistical analysis of the answer data.

Keywords: Privacy Act, Anonymous Processed Information, Pseudonymizations, Medical data, Location history, Social survey through questionnaire

1. 問題意識

個人情報とは氏名や個人識別符号だけに限らず、購買記録など個人に係ることが明示されたすべての情報である。情報主体の個人が自分の個人情報を何も加工せずに第三者に渡したり、流通させることは、不要な広告が送られてくる以外により自分への被害が大きい悪用をされることもありえる。そのため、氏名ないし基本 4 情報（氏名、性別、年齢、住所）をハッシュ関数で計算したハッシュ値で置き換える仮名化がよく知られているが、それ以外に個人情報の扱いを規律する改正個人情報保護法[1]（以下では個情法と略記する）では、43 条によって導入された匿名加工情報を用いることもある。個情法 43 条第 1 項に記された個人情報保護委員会が定める匿名加工情報作成の具体的基準であるところの個人情報の保護に関する法律施行規則[2]（以下「規則」と略記する）によれば、匿名加工情報とは、元の個人情報を通常業務で使われる一般的な方法では個人識別できないように加工したものであり、データ主体の同意がなくても第三者に提供できるとしている。ただし、個人情報を匿名加工情報に変換する加工方法は厳密に定義されておらず、特異値除去などの一般的な処理以外は、データに応じて行うように規則には書かれている。しかし、実際

に匿名加工処理を行う事業者は、どのような処理を行えばよいか陽には分からない。

一方で、購買履歴、移動履歴、医療履歴のような個人情報が a) 実名付き元データ、b) 仮名化されたデータ、c) 匿名加工情報のような形態で利用されることに対して、一般人が許容している度合いを知ることは、これらのデータを実利用するにあたって有用である。また、a), b), c) の 3 種類のデータの扱い方に関する法制度的な方向性を見出せるかもしれない。本報告では、一般人 1045 人を対象にして、上記、a), b), c) の 3 種類の個人情報の利用法の許容／否許容についてアンケート調査を行った結果を示し、その分析結果を記す。第 2 節では社会調査の背景、第 3 節では本研究で行った社会調査とその結果を示す。第 4 節では、調査結果を主成分分析、第 5 節では 2 項ロジスティック回帰によって、種々の意見を持つ人々の相関、個々の意見に対する要因分析を行う。第 6 節は性別、年代による意見の差異、第 7 節はまとめである。

2. 社会調査の背景

個人情報の利活用という目的に立ち返ってみると、そもそも個人情報の提供者である一般の人々にとっては、自分の個人情報の利用をどこまで許容できるかが関心事である

1 理化学研究所
RIKEN
2 明治大学

Meiji University

う。個人情報保護に関する社会調査が 2015 年に内閣府によって行われた[3]が、これは 2015 年当時における個人情報保護改正への賛否、個人情報の定義などを尋ねるものであった。2020 年の情報通信白書[4]には、提供した個人情報の安全管理、流出、悪用の懸念を示す日本人が 6 割弱になるという社会調査に結果が記されている。また、[5]の消費者意識調査では、一般の日本人 1000 人を対象に行われた種々の個人情報の利活用に関して不安を感じる項目の調査結果が記載されているので、以下の節において本研究で行った調査結果と比較する。

3. 社会調査の質問と回答結果

本研究では、一般人 1045 人を対象にして以下に記載する質問を用いてアンケート調査を行った。回答者の年齢、性別、職業などの分布は表 1 に示す通りである。調査は 200 万人の回答者候補を有するインテージ社に依頼しておこなった。なお、以下の表では単位は人である。

表 1 アンケート回答者の属性分布

Table 1 Attribute Distribution of Survey Respondents
 (セル内数値の単位は人)

性別	男性	525	女性	520	
年齢	20 代	30 代	40 代	50 代	60 代
平 46.1 歳	170	185	240	243	207
職業	会社員	役員、管理職	公務員	自営業	自由業、専門職
361	56	90	37	22	
学生	専業主婦	派遣、パート	無職	その他	
28	137	212	84	18	

本節では、アンケート調査で用いた複数選択肢からの選択をさせる 14 個の質問を記し、質問ごとに各選択肢を選んだ人数を記載し、その結果を分析する。なお、以下の各質問の回答の表で、各行の右側のセルの値は、左側に回答内容を正しいと答えた回答者数およびその割合%を表す。複数回答を許す質問では%の数字を全部足しても 100 にならないことがある。

質問1: 個人情報の定義として正しいと思うもの(単一選択)。

A) 基本4情報のみ	519 人 (49.7%)
B) 基本4情報+正しい履歴情報	418 人 (40.0%)
C) 基本4情報+正しい履歴情報+正しくない履歴情報	108 人 (10.3%)

2015 年の世論調査[3]で個人情報に含まれる情報を質問しているが、77.6%が基本4情報は当てはまるとしている。逆にいえば、残りの 22.4%は基本4情報のうち氏名などその部分だけを個人情報と考えていたと考えられるが、今回は基本4情報が含

まれることを前提にした質問になっているため、直接の比較はできない。一方、[3]では 34.3%が商品購入履歴、32.1%が移動履歴も個人情報に含まれると答えている。これらは質問 1 の B)、C)を選択した合計値 50.3%に対応する。つまり、[3]に示された 2015 年当時の 32%程度に比べて 20%近く正しい認識を持つ人が増えたといえる。なお、[3]では質問3の C)に対応する質問はしていないので、比較できない。ただし、間違っても個人に紐づいたら個人情報だという解釈は、当時は質問選択肢にすらならなかったという状況だと理解しておくといえよう。

質問2: 匿名化の定義(複数選択)。

A) 基本4情報の削除	238 人(22.8%)
B) 基本4情報のあいまい化	268 人(25.6%)
C) 基本4情報をあいまい化し、同じ情報になった人が $k(\geq 2)$ 人以上	108 人(10.3%)
D) 基本4情報の削除し、さらに履歴情報から一般的な方法では個人が特定されないように履歴情報を作り替える。(これを匿名加工情報と呼ぶ。)	255 人(24.4%)
E) 基本4情報の削除に加えて、履歴情報を加工して同じデータになった人が $k(\geq 2)$ 人以上いる	72 人(6.9%)
F) わからない	355 人(33.98%)

質問3: 個人情報から基本4情報削除し、さらに履歴情報から一般的な方法では個人が特定されないように履歴情報を作り替えた匿名加工情報の第三者提供も含む制度。なお、A)を選ばなかった方は、B)、C)、D)から複数選択可。

A) 賛成できない	285 人(27.3%)
B) 社会に役立つなら賛成	395 人(37.8%)
C) 利用者が特定の企業であっても役立つなら賛成	73 人(7.0%)
D) 自分に被害が及ばないなら賛成	380 人(36.4%)

個人の特定がほとんどできないことを技術的および法的に保証する匿名加工情報の制度自体に賛成しない人が 27.3%にのぼることは、一般人が過剰ともいえるほどに個人情報の保護に傾倒していることがうかがえる。一方で B)、C)、D)が合わせて 72.7%であることは、個人情報の利活用の重要性を認識している人が多数派であることを示している。

質問4: 個人情報を収集した事業者から別の事業者提供

A) 自分の名前が記載された元データでも提供可	27(2.6%)
B) 基本 4 情報を仮名化したなら提供可	159(15.2%)
C) 基本 4 情報を削除したなら提供可	383(36.7%)
D) 拒否	476(45.6%)

B)、C)、D)を合わせると 82.3%を超えるが、[5]で氏名、

住所の提供に不安を感じる人が 81.4%であり、整合性が高い結果となっている。

質問5: 自分自身の身体の状態に関する医療情報の医療機関が医学研究に利用すること(単一選択)

A) 利用は全て拒否	156(14.9%)
B) 自分が同意の元で許容	838(80.2%)
C) 無条件に許容	51(4.9%)

質問6: 自分自身の身体の状態に関する仮名化された医療情報を創薬メカが新しい薬の開発への利用(単一選択)

A) 利用は全て拒否	131 人(12.5%)
B) 自分が同意の元で許容	866 人(82.9%)
C) 無条件に許容	48 人(4.6%)

[5]では医療情報の提供にとっても不安を感じる人が 17.8%であり、これは質問 5, 6の A)に対応する人を含む。また、やや不安を感じるが 27.1%であり、あまり不安を感じないが 34.0%、まったく不安を感じない 17.5%で総計 78.6%であったが、これは質問 5, 6 の同意の元に許容という B), C)を合わせた 85.1%, 87.5%を含むと考えれば、類似した割合となる。

質問7: 自分に関する事柄に関して何かを疑われた場合の反論できる機能(単一選択)

A) 必須	326 人(31.2%)
B) 重要	631 人(60.4%)
C) 不要	88 人(8.4%)

質問8: クレジットカード利用履歴をカード会社がサービスレベル変更に利用(単一選択)

A) 許容	473 人(%)
B) 拒否	572 人(%)

質問9: 自分の地理的移動履歴情報を Google, Amazon などのプラットフォーム(以下では PF と略記)事業者が利用することの可否(複数選択)

A) プラットフォーム事業者内部でも不可	397 人(38.0%)
B) プラットフォーム事業者内部なら可	176 人(16.8%)
C) 仮名化されていれば、プラットフォーム事業者以外の事業者に利用可	88 人(8.4%)
D) 匿名加工情報なら、そのプラットフォーム事業者以外の事業者に利用可	166 人(15.9%)
E)元の情報をそのプラットフォーム事業者以外の事業者に利用可	30 人(2.9%)
F) あてはまるものはない	260 人(24.9%)

質問10: 公共交通機関(タクシーを含む)を利用しての地理

的移動履歴の利用(複数選択)

A) 公共交通機関事業者の内部であっても利用不可	352 人(33.7%)
B) 公共交通機関事業者の内部でなら利用可	216 人(20.7%)
C) 仮名化されていれば、外部事業者の利用も可	118 人(11.3%)
D) 匿名加工情報であれば、外部事業者の利用の可	167 人(16.0%)
E) 元情報を外部事業者にも利用可	30 人(2.9%)
F)あてはまるものはない	235 人(22.5%)

[5]では位置情報、行動履歴は 50.6%の人が提供にとっても不安を感じていたが、これは上記に 2 質問の A)と回答した人を含むであろう。また[5]ではやや不安が 33.4%であり、それらの人は上記の 2 質問では B), C), D)という選択肢に吸収されたと考えられる。

質問11: 駅改札口で顔パスで入場/退場できるために自分の顔画像情報と氏名ペアになった情報の利用(単一選択)

A) 顔パス入場/退場だけ可	345 人(33.0%)
B) 顔パス入場/退場に加えて、駅構内、近辺の店での顔パス支払いも可	161 人(15.4%)
C) 不可	539 人(51.6%)

質問12: 公共の場所に顔認識できるビデオカメラを設置の可否(単一選択)

A) 許容	573 人(54.8%)
B) 拒否	472 人(45.2%)

質問13: 政府、自治体、鉄道事業者などが公共の場所や公共交通機関に設置している顔認識できるビデオカメラから得られる顔画像データの利用(単一選択)

A) ビデオカメラの認識結果は認識直後にその場で廃棄され、中央サーバで集中管理しないから可	255 人(24.4%)
B) ビデオカメラの認識結果が中央サーバで集中管理されても、駅内暴力や痴漢行為の抑止効果があるなら可	369 人(35.3%)
C) ビデオカメラの認識結果中央サーバで集中管理されても、駅内暴力や痴漢行為抑止以外の目的、たとえば、人々の通勤経路の調査などで使われるとしても、可	153 人(14.6%)
D) 何があっても許容しない	268 人(25.6%)

[5]では顔画像、指紋の提供にとっても不安が 63.3%だが、これは上記質問の回答 A), D)を合わせた 50%を含むと考えられる。[5]ではパーソナルデータを防犯などの公共目的での提供に

63.5%が賛成しており、これらは上記 B), C)の回答者の計の49.9%に対応すると考えられる。

質問14: 自分のゲノム情報(遺伝子情報)保険会社が利用して生命保険の料率に影響すること(単一選択)

A) 許容	310人(29.7%)
B) 拒否	735人(70.3%)

欧米では保険料率にゲノム情報を利用することを禁止するゲノム差別禁止法があるが[12], 日本にはまだそのような法律はない。

4. 主成分分析

第1主成分において係数の値の正から負への値の降順で列挙した。なお、絶対値の大きなものを中心に表示し、中間的な値の回答は省略している。スペースの関係で回答の意味は短縮した表現にしているため、正確な内容は3節を参照していただきたい。

表2 第1主成分
 Table 2 The 1st Principal Component

質問回答#	回答の意味	係数の値
12A	公共のビデオカメラ設置可	0.625
5B	医療情報, 医療機関が同意して利用可	0.579
6B	医療情報, 創薬メーカー同意して利用可	0.550
14A	ゲノム情報の保険会社利用可	0.547
8A	クレカ利用履歴がサービスの影響可	0.486
3B	匿名加工情報の同意なし流通, 社会で役立つなら可	0.419
13B	ビデオ顔画像, 痴漢抑止なら可	0.395
9B	地理的移動履歴情報をPFが利用可	0.392
8B	クレカ利用履歴がサービス影響不可	-0.486
14B	ゲノム情報保険会社利用不可	-0.547
11C	駅の顔パス画像利用不可	-0.585
5A	医療情報の医療機関の研究利用不可	-0.596
6A	医療情報の創薬メーカー利用不可	-0.600
4D	個人情報を別の事業者提供不可	-0.617
12B	公共場所のビデオカメラ設置不可	-0.625
13D	ビデオカメラ顔画像, 利用不可	-0.675

第1主成分によれば、正の係数の回答は個人データ利用の許容のグループ、負の係数の回答は個人データ利用の拒否のグループへと分割されていることが分かる。データ以外についても、顔データ取得が可能なビデオカメラの公共の場所への設置も同様に許容と拒否に分かれている。

第2主成分も第1主成分と同様に係数の値大きいものから降順に列挙した。

表3 第2主成分

Table 3 The 2nd Principal Component

質問回答#	回答の意味	係数の値
10F	公共交通の移動履歴利用可否不明	0.518
9F	移動履歴情報, 匿名加工情報ならそのPF以外の事業者利用可否分らず	0.496
6C	医療情報を医療機関が無条件研究利用可	0.450
14A	医療情報を創薬メーカーが無条件利用可	0.421
14A	ゲノム情報の保険会社利用可	0.307
.....		
6A	医療情報の創薬メーカー利用不可	0.201
13A	ビデオ顔画像即廃棄なら可	0.199
5A	医療情報, 医療組織で研究利用不可	0.184
.....		
5B	医療情報の医療機関の研究同意利用可	-0.408
6B	医療情報を創薬メーカーが同意利用可	-0.411
10A	公共交通の移動履歴利用不可	-0.652
9A	地理的移動履歴情報をPFが利用不可	-0.662

医療情報やゲノム情報の利用を可とするグループが正の係数の多いところにあるが、その一方で医療情報の医療機関、創薬メーカーでの利用拒否も正の係数の上位に位置する。負の係数の場合は、医療情報利用に同意を必要とする意見、移動情報の利用を不可とするものが入る。医療情報と移動情報で意見が分かれる人が多いことが分かる。医療情報の利用不可(5A, 6A)と移動情報の利用不可(9A, 10A, 11C, 12B, 13D)に意見の相関を計算すると表4のようになる。なお、全ての相関係数へP値<0.001である。

表4 医療情報拒否と移動情報拒否の相関係数

Table 4 Correlation coefficient between medical information refusal and mobile information refusal

	5A	6A	9A	10A	11C	12B	13D
5A	1.000	0.806	0.132	0.134	0.233	0.236	0.322
6A	0.806	1.000	0.115	0.140	0.248	0.255	0.341

5A, 6Aと移動履歴そのものの利用拒否(9A, 10A)の相関係数は小さい。一方、顔パス(11C), ビデオカメラ設置(12B), 鉄道駅でのビデオ顔画像利用のように具体的利用法が明確で社会的利益が明確な場合への拒否の場合は相関値が若干高い。つまり、目的によらず個人情報利用拒否というコアな拒否層が医療と重なると解釈できる。

5. 2項ロジスティック回帰

個人情報利用を許容しない回答を目的変数とし、その理由とみなせる説明変数となる回答を2項ロジスティック回帰を用いて計算する。

2項ロジスティック回帰とは、例えば、ある疾病の罹患の有無などの2値の目的変数の値を、既往症の有無などの他の説明変数 a で予測する下記の(1)式を、得られた観測データから統計的最適化で求めるものである。目的変数が1の確率を P_1 、0の確率は P_0 とし、説明変数の値を x_1, \dots, x_n とすると、

$$\log(P_1/P_0) = a_1 \cdot x_1 + \dots + a_n \cdot x_n + b \quad (1)$$

ここで a_1, \dots, a_n は係数、 b は定数である。 $\exp(a_i)$ をのオッズ比(以下では OR と記す)と呼び、 $OR > 1$ なら $\log(P_1/P_0)$ に正に寄与して P_1 が大きくなり、 $OR < 1$ なら負に寄与して P_0 が大きくなる。オッズ比の有意さは p 値として求まる。

気を付けたいのは、(1)式は必ずしも得られた説明変数から目的変数への因果関係を表しているわけではなく、得られたデータにおける依存関係を表していることである。例えば $OR > 1$ が大きい場合は互いに強い正の依存関係がある、言い換えれば相関が強いということが言える。この強い相関を強い予測につながっていると解釈しているだけである。

ここでは、分析のために SPSS を用いた。以下では p 値 < 0.1 以下の説明変数の OR と p 値を、まず正の寄与の $OR > 1$ の場合を OR の降順、次に負の寄与の $OR < 1$ の場合は OR の昇順に示す。また、質問の回答の意味は、紙数の関係から簡素化するので、正確な表現は3節を参照してほしい。また、以下の全て表では、1行目は目的変数である質問番号と回答番号の数英字およびその内容、くわえて回帰式の正解率、2行目以降は説明変数に対応し、1列目が「質問番号と回答英字」2列目が OR の値、3列目が OR の p 値を表す。なお、上記の正解率は機械学習としてみれば、学習データに対する正解率であり、得られた式のデータへの当てはまりの良さを表す b 。

表5 質問5の回答Aの2項ロジスティック回帰
Table 5 Binomial logistic regression for response A to question 5

目的変数 : 5A: 医療情報を医療組織で研究利用不可 正解率=.958		
説明変数に対応する質問への回答内容	OR	p 値
6A: 医療情報創薬メーカ利用不可	435.173	<0.001
2C: 基本4情報をあいまい化し k 匿名	4.563	0.007
.....		
8A: クレカ利用履歴がサービス影響	0.276	0.002

a 説明変数は2値が望ましい。連続値である場合は通常の線形回帰が有力とされる。

可		
9C: 移動履歴、仮名化ならその PF 以外の事業者に利用可	0.035	0.033
9F: 移動履歴情報をその PF が利用可否判断できず	0.016	0.020
9D: 移動履歴情報、匿名加工情報ならその PF 以外の事業者利用可	0.005	0.005

$OR > 1$ の場合、類似内容の目的変数である 6A は表4に示したように 5A との相関係数が大きく、上記の結果でも OR は非常に大きい。2C は匿名化のかなり正確な定義だが、その定義を知っているからこそ、医療組織での自分の医療情報利用に否定的になるのであろう。

$OR < 1$ だと個人の移動履歴情報の利用を許容する人、さらにはクレカの利用履歴がサービスに影響することを許容する意見もあり、これらが医療情報の利用を拒否と逆方向に作用していることが分かる。

表6 質問6の回答Aの2項ロジスティック回帰
Table 6 Binomial logistic regression for response A to question 6

目的変数 : 6A: 医療情報を創薬メーカ利用不可 正解率=.964		
説明変数に対応する質問への回答内容	OR	p 値
5A: 医療情報、医療組織で研究利用不可	217.051	<0.001
10A: 公共交通の移動履歴利用不可	49.132	0.003
10F: 公共交通の移動履歴事業者内部利用可否判断できず	37.992	0.012
10C: 公共交通の移動履歴、仮名化なら外部事業者利用可	19.370	0.012
10B: 公共交通の移動履歴事業者の内部でなら利用可	17.175	0.029
.....		
3B: 匿名加工情報の同意なし流通、社会で役立つなら可	0.097	0.073
3D: 匿名加工情報利用、自分に被害が及ばなければ可	0.109	0.087
7A: 反証システム必須	0.169	0.020
7B: 反証システム重要	0.098	0.002

$OR > 1$ の場合、類似内容の目的変数である 5A と同様に相関係数の大きい 5A が非常に大きい。次に公共交通による移動履歴の利用に関するものが多く、不可ないし可否判断できずという回答の OR が大きい、条件付き利用可も $OR > 1$ で上位だった。10A の利用不可は 5A と同じ傾向の回

b 言い換えれば、未知データに対する予測正解率ではない。

答だが、10B, 10C は条件付き利用可なので、自分の仮名化医療情報の創薬メーカーの利用とは相いれない傾向であり、その理由は分からない。

OR<1 の場合は、匿名加工情報の利用を可とするものである。興味深いのは 3D であり、「自分に被害が及ばなければ可」とは逆にいえば自分に被害があれば不可ということになり、5A の回答の人は創薬メーカーへの提供が自分への何らかの被害に結びつくと考えていると考えられる。7A, 7B の反証システムも OR<1 だが、反証システムはそもそも個人情報を利用される前提であるため、5A の人はそのような前提自体を否定的に捉えていると考えられる。

表 7 質問 9 の回答 A の 2 項ロジスティック回帰
 Table 7 Binomial logistic regression for response A to question 9

目的変数: 9A: 移動履歴情報を PF が利用不可 正解率=.856		
説明変数に対応する質問への回答内容	OR	p 値
10A: 移動履歴, 事業者内部利用不可	38.542	<0.001
10B: 移動履歴事業者の内部でなら利用可	3.387	0.003
.....		
11A: ビデオ顔画像即廃棄なら可	0.715	0.171
2C: 基本4情報をあいまい化し k 匿名	0.426	0.013
10F: 移動履歴事業者内部利用の可否判断できず	0.374	0.057
11B: 顔パス入場/退場以外利用も可	0.305	<0.001

OR>1 では同じく公共交通の利用履歴がバス会社や鉄道会社での利用拒否が OR 値も大きく支配的である。9A と 10A の相関係数は 0.685(p 値<0.001)と大きいことから、9A, 10A は同じ人が支持した意見のペアが多いと考えられる。

OR<1 は強い回答は少ないが、顔パスすなわち顔画像の幅広い利用を許容する回答 11B が最強で、(p 値<0.001)である。この回答は、9A の真逆な回答であり、回答の OR 値としては順当なものと考えられる。

表 8 質問 10 の回答 A の 2 項ロジスティック回帰
 Table 8 Binomial logistic regression for response A to question 10

目的変数: 10A: 移動履歴, 事業者内部利用不可 正解率=.872		
説明変数に対応する質問への回答内容	OR	p 値
9A: 移動履歴情報を PF が利用不可	12.905	<0.001
5B: 医療情報の医療機関の研究同意	3.730	0.088

利用可		
13A: ビデオ顔画像即廃棄なら可	2.502	0.004
.....		
4A: 元個人情報を別業者に提供可	0.268	0.069
4B: 基本 4 情報を仮名化なら提供可	0.428	0.025
14A: ゲノム情報保険会社利用可	0.586	0.039

OR>1 の回答に関しては、10A に対して直前に述べた 9A が高いのは当然である。興味深いのは、5B, 13A という利用許可の回答が OR>1 であることだが、いずれも同意、即廃棄などの条件付きである。実は直前の 9A でも 10B という条件付き許容が OR>1 である。このように移動履歴の利用の可否は利用条件に依存していることがうかがわれる。

OR<1 の場合は利用許可が多いが、OR 値が比較的 1 に近く強い回答ではない。

表 9 質問 11 の回答 C の 2 項ロジスティック回帰
 Table 9 Binomial logistic regression for response C to question 11

目的変数: 11C: 顔パスによる入場不可 正解率=.777		
説明変数に対応する質問への回答内容	OR	p 値
6A: 医療情報の創薬メーカー利用不可	2.888	0.122
6B: 医療情報の創薬メーカー利用可	2.043	0.183
.....		
回答者男性	0.556	<0.001
14A: ゲノム情報保険会社利用可	0.447	<0.001
13A: ビデオ顔画像即廃棄なら可	0.341	<0.001
13B: ビデオ顔画像, 痴漢抑止なら可	0.328	<0.001
12A: 公共場所ビデオカメラ設置可	0.133	<0.001

OR>1 の回答は OR が小さく、p 値は大きいため強い影響力のある回答ではない。

OR<1 の回答はいずれも p 値<0.001 で OR 値の確実性が高い。12A の回答は顔パス不可の逆方向の回答であり当然の結果だろう。13A は即廃棄条件付きであり、13B は痴漢抑止という社会的な意義のある回答である。よって、社会的意義から考えれば、11C の顔パスによる入/退場の拒否は、個人情報保護のコア層の意見であり、社会的意義や技術的改善を無視した感情的なものと窺われる。

表 10 質問 12 の回答 B の 2 項ロジスティック回帰
 Table 5 Binomial logistic regression for response B to question 12

目的変数: 12B: 公共場所のビデオカメラ設置不可 正解率=.797		
説明変数に対応する質問への回答内容	OR	p 値

4A: 元個人情報情報を別業者に提供可	1.889	0.239
2E: 匿名化の定義: 基本 4 情報削除 +k(≥ 2)匿名	1.545	0.214
回答者年齢 30 代:	1.459	0.183
.....		
13B:ビデオ顔画像, 痴漢抑止なら可	0.092	<0.001
13C:ビデオ顔画像, 痴漢抑止+その他目的でも可	0.112	<0.001
11A: 顔パス入場/退場だけ可	0.133	<0.001
11B: 顔パス入場/退場以外利用も可	0.143	<0.001
13A: ビデオ顔画像即廃棄なら可	0.148	<0.001
14A: ゲノムの保険会社利用可	0.443	<0.001

OR>1 のビデオカメラの公共の場所への設置に反対の意見は、数は多いが、 $1 < OR < 2$ の小さな値のものが分散している。

OR<1 すなわち設置に賛成の意見は 13B, 13C の痴漢抑止効果ありなら可, など具体的な意見が p 値<0.001 かつ OR=0.092, OR=0.112 と非常に強い。これを見ると、公共の場所のビデオカメラ設置による顔画像認識は、目的が明確なら賛同されるもとを考えられる。さらに顔画像による顔パスを許容する人、駅のビデオ画像利用許容する人も OR は 0.15 より小さく、かついずれも p 値<0.001 であり強い意見である。つまり、目的が明確であれば、駅のような公共の場所へのビデオカメラ設置は許容する人が多いことが読み取れる。

表 11 質問 13 の回答 D の 2 項ロジスティック回帰
Table 11 Binomial logistic regression for response D to question 13

目的変数 13D: ビデオ顔画像の利用不可 正解率=.878		
説明変数に対応する質問への回答内容	OR	p 値
9F:移動履歴をプラットフォーム事業者の利用可否判断できず	7.398	0.025
9A: 移動履歴情報を PF が利用不可	3.077	0.194
.....		
2D:匿名の定義=匿名加工情報	0.442	0.011
11A: 顔パス入場/退場だけ可	0.374	0.001
11B: 顔パス入場/退場以外利用も可	0.351	0.026
4A:元個人情報情報を別業者提供可	0.198	0.055
12A: 公共場所ビデオカメラ設置可	0.116	<0.001

OR>1 で 9F の可否判断できずが大きな OR であることは説得性ある理由が見つからない。しいていえば、判断できないということは暗黙にはあるが否定的な意見ということかもしれない。9A は利用拒否なので 13D と同じ方向の回答である。

OR<1 の場合は小さな OR 値の回答は個人情報利用許容

の方向の回答なので、利用不可の回答である 13D とは明確に逆方向の回答であるため、当然の結果である。

表 12 質問 14 の回答 B の 2 項ロジスティック回帰
Table 12 Binomial logistic regression for response B to question 12

目的変数 14B: ゲノム情報保険会社利用不可 正解率=.785		
説明変数に対応する質問への回答内容	OR	p 値
3A: 匿名加工情報,社会流通不可	3.203	0.003
3B: 匿名加工情報の同意なし流通,社会で役立つなら可	2.284	0.008
3D: 匿名加工情報利用, 自分に被害が及ばなければ可	2.050	0.020
3C: 匿名加工情報の利用者が企業でも役立つなら賛成	1.731	0.103
.....		
13B: ビデオ顔画像, 痴漢抑止なら可	0.479	0.035
4B: 基本 4 情報を匿名化なら提供可	0.455	0.004
13A: ビデオ顔画像即廃棄なら可	0.404	0.010
13C: ビデオ顔画像, 痴漢抑止+その他目的でも可	0.324	0.002
11B: 顔パス入場/退場以外利用も可	0.267	<0.001
4A: 元個人情報情報を別業者に提供可	0.196	0.002

OR>1 の場合の回答は、OR 値はおおよそ 3 以下で大きくはないので強い要因ではないが、全て匿名加工情報の利用に係るものであるという特色がある。3A は匿名加工情報の利用不可なので、14B と同方向であるが、それ以外の 3B, 3C, 3D は匿名加工条件付きなら許容ということであり、これまで説明してきた他の目的変数と異なる傾向である。つまり、匿名加工情報の条件付き利用を許容する人もゲノム情報利用はしてほしくないという傾向を示していると考えざるを得ない。

OR<1 は OR 値の小さな 4A, 11B が個人情報利用を広く認める回答であり、13B, 4B, 13A, 13C は条件付き利用可であるので、論理的に整合性がある。

6. 性別, 年代による差異

性別に関しては質問 12, 13 に注目した。質問 12 の公共の場所へのビデオカメラ設置の拒否は、女性は 53%, 男性は 47%であった。

質問 13 の駅、電車内暴力や痴漢行為を防ぐ目的も含むように顔認識できるビデオカメラから得られる顔画像データの利用に関して、回答 13A,13B,13C,13D の回答の中から 13B:痴漢抑止に効果があれば許容する回答を選んだのは女性 37%, 男性 34%だった。

女性は質問 12 ではカメラ設置拒否、質問 13 では痴漢抑止なら顔画像利用可としており、矛盾した結果にも見えるが、一般論としてはカメラ設置を拒否しても、自分たちが痴漢に合わないなら許容するという微妙な心情を表した結果になっている。このような総論反対、各論賛成はこの例に限らずアンケート調査ではよく見られるものである。

年代による有意な差がある項目は少なかったが、以下の項目が有意水準 1% で差が見られた。なお、下記の%は、全年代に対する割合である。

質問 3 の匿名加工情報については 20 代で不賛成が 10.5% で少なく、自分に害がなければ賛成が 20.8% で多かった。

質問 8 のクレジットカードの設定への影響は、30 代が許容 21.6% で多く、許容できない 14.5% で少ない。60 代では逆に許容 15.9% で少なく、許容できない 23.1% で多い。

質問 9 のプラットフォームによる移動履歴の利用では、30 代が許容できないが 13.6% で少なく、60 代では逆に許容が 24.4% で多い。また、30 代では仮名化されていればデータ収集したプラットフォーム以外の事業者が使うことを許容するが 29.5% で多い。30 代は、移動履歴の利活用に拒否感が少ないといえる。

7. まとめ

この報告では、個人情報の利用法ごとの可否について一般人の意見の傾向を知るために、一般人 1045 人を対象にした 14 個の質問に対するアンケート調査を行った結果について報告した。主成分分析では、個人情報利用の許容と非許容が第 1 主成分を分割する要因となっていることが分かった。各利用法を否定する意見を目的変数として、それ以外の質問への回答を説明変数とする 2 項ロジスティック回帰を行い、各利用法を否定する回答の要因を分析した。おおよその傾向としては、関連する利用法の否定が正の寄与、許容が負の寄与となったが、個別には利用法ごとに異なる要因が見られた。このような分析が、個人情報を利用するための障害となっている事項を取り除くこと、および個人情報の利用を許容する要因を促進する製品、社会システムなどを計画する基礎資料になることを期待したい。

謝辞 本研究は科学研究費・基盤研究 (A) 「オープンな評価コンテストによる匿名加工アルゴリズムとリスク評価の研究人と情報のエコシステム」(課題番号: 18H04099) の補助を受けて行われた。

参考文献

- [1] 個人情報の保護に関する法律施行規則: 施行日: 令和四年四月一日 (令和三年個人情報保護委員会規則第四号による改正) <https://elaws.e-ov.go.jp/document?lawid=428M60020000003>
- [2] 個人情報の保護に関する法律についてのガイドライン (仮名加工情報・匿名加工情報編) 平成 28 年 11 月 (令和 3 年 10 月一部改正)

- https://www.ppc.go.jp/personalinfo/legal/guidelines_anonymous/#a3
- [3] 内閣府: 「個人情報保護法の改正に関する世論調査」の概要. 2015. <https://survey.gov-online.go.jp/hutai/h27/h27-kojin.pdf>
 - [4] 総務省: 令和 2 年度版 情報通信白書. 2020. <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/pdf/index.html>
 - [5] 総務省: データの流通環境等に関する消費者の意識に関する調査研究. 2020. https://www.soumu.go.jp/johotsusintokei/linkdata/r02_03_fuzoku.xlsx