

匿名データの作成・提供に係るガイドライン骨子案

第1 ガイドラインの目的

本ガイドラインは、統計法第 35 条及び第 36 条の規定に基づく匿名データの作成及び提供に係る事務処理の明確化、統一化を図ることにより、各行政機関及び独立行政法人等が当該事務処理を円滑に遂行できるようにすることを目的とする。

第2 用語の定義

1 匿名データ

統計法第 2 条第 12 項に規定するもの。

2 調査票情報

統計法第 2 条第 11 項に規定するもの（統計の作成の過程で出来た統計も調査票情報に含まれる。）

3 行政機関及び届出独立行政法人等

統計法第 2 条第 1 項に規定するもの及び同法第 25 条の規定による届出を行った独立行政法人等。

第3 匿名データの作成・提供の実施に当たっての基本原則

1 事務処理要綱の策定

各行政機関及び届出独立行政法人等は、匿名データの作成・提供に係る事務処理の明確化、効率化を図るために、それぞれ事務処理要綱を策定。

2 秘密保護の確保

統計調査に対する調査対象者の信頼を確保する観点から、匿名データの作成を行うため調査票情報を取扱う際に秘密の保護に十分配慮する。

匿名データの作成を外部委託する場合、調査票情報の適正管理義務について契約条項に含める等必要な措置を実施。

匿名データの提供に当たっては、提供を受けた者から匿名データの適正な管理を行う旨の確約を求めるとともに、自己又は第三者の不正な利益を得る目的で提供又は盗用した場合、統計法第 61 条第 3 項の罰則の適用されることを明示。

3 効率的な事務処理の実施

匿名データの作成に当たっては、プログラムの作成やテスト、チェックリストの作成や審査等を行う必要があり、ある程度専門的な知識経験が必要であり、各行政機関及び届出独立行政法人等においては新たに実施すべき業務であることから、必要に応じて外部委託等を検討。

第4 匿名データの作成・提供に関する計画の策定

各行政機関及び届出独立行政法人等は、毎年度当初に、当該年度に提供する予定の匿名データを公表。

第5 匿名データの作成

1 匿名データを作成する統計調査の範囲

行政機関及び届出独立行政法人等が、その実施する統計調査の中から、需要を踏まえて決定。企業や事業所を対象とする統計調査についても秘匿処理の可能性を検討し、匿名化が困難な場合、委託による統計の作成等による対処を検討。

2 匿名データの秘匿処理の方法

(1) 秘匿処理の考え方（別紙1参照）

調査対象である調査単位（世帯や個人、企業）と匿名データの対応関係を特定されないようにすること

(2) 秘匿処理の技法（別紙2参照）

- ・ 識別情報の削除
- ・ 識別情報のトップ・コーディング
- ・ 識別情報のグルーピング
- ・ リサンプリング
- ・ 匿名データのソート（配列順の並べ替え）
- ・ スワッピング
- ・ 誤差の導入 等

(3) 秘匿処理方法の決定

上記の秘匿処理の技法をそれぞれの統計調査の内容等に応じて組み合わせて秘匿処理を実施

秘匿処理が適切であるか等について統計委員会の意見を聴取することを規定。

(4) 秘匿化の目安（別紙3）

審査は各行政機関及び届出独立行政法人等が実施。

第6 匿名データの秘匿処理の実施手順

1 秘匿処理の審査

(1) チェックリストの作成

各行政機関及び届出独立行政法人等内部及び統計委員会における秘匿処理の審査を効率的、効果的に実施するため、統一化された様式（別紙4・5にイメージを提示）で、匿名データを作成しようとする所管の統計調査ごとに、どのような秘匿処理を行うか等を記述したチェックリストを作成。

(2) 行政機関及び届出独立行政法人等の内部における審査

行政機関及び届出独立行政法人等は内部における審査機関を設置し、内部における審査を適切に実施。

(3) 統計委員会への諮問

統計法第35条に基づき、各行政機関及び届出独立行政法人等は、基幹統計に係る匿名データを作成する場合、統計委員会の意見を聴取。

各行政機関は、チェックリストや匿名データの作成対象である統計調査に関する基礎的な情報（調査概要、調査票の様式、標本抽出法等）、匿名データに関する情報（匿名データの作成方針、匿名データのレイアウトフォーム・符号表、匿名化に当

たつての留意事項、実施した匿名化の概要等)を提出。

2 匿名データの作成・確認

統計委員会の答申で認められた秘匿処理の内容で匿名データを作成し、秘匿処理が正しく行われていることを確認。

3 匿名データ提供の周知

匿名データを作成した行政機関及び届出独立行政法人等は、ホームページ等により提供が可能となった匿名データに関して周知。

第7 匿名データ提供の申出

1 匿名データ提供の申出者(以下「申出者」という。)の範囲【P】

統計法第36条では、「学術研究の発展に資すると認める場合その他の総務省令で定める場合」にのみ匿名データの提供ができると規定。

このため、申出者は、自分自身が責任を持って学術研究の発展に資すると認められる研究を行い得る者及び総務省令で定められる目的を実現できる者。

なお、大学における申出者の範囲については次の考え方を参考に判断。

《大学における申出者の範囲の考え方》

教授等の指導の下で、教授とともに大学院生や学部生が研究に携わる場合、教授等共同研究として、教授等と大学院生が研究を行う場合、全ての共同研究者個人として大学院生等が研究を行う場合(学部生の身分を理由とした排除は困難)、大学院生等

2 匿名データの提供申出書(以下「申出書」という。)の記載事項【P】

申出書に記載される申出事項は総務省令で規定(統計法第36条)

(統一の様式を別添で示す予定)

3 申出書の受付窓口

申出書は各行政機関及び届出独立行政法人等が設置する所定の受付窓口へ提出。各行政機関及び独立行政法人等は、事務処理要綱に受付窓口を規定。

4 申出書の受付時期

受付事務の効率化、統計の作成等作業の計画的実施の観点から、各行政機関及び届出独立行政法人等は、年間の受付時期・期間を設定し、受付時期を各年度当初に周知

5 申出書の提出媒体

各行政機関及び届け出独立行政法人等の受付窓口への直接提出、郵送、メールのいずれも可。

6 申出書の添付書類

申出書の提出に当たっては、申出内容の審査に必要な公益性を示す書類や作成する

統計等の様式の添付を要請することとし、事務処理要綱に規定。

《添付書類の例示》

- ・ 公益性を示す書類等

研究計画書，研究費を確保するために作成する既存の書類、過去の著書・論文一覧、発表予定の学会、大会の概要、掲載予定の学術誌等、指導教授や大学、学会からの推薦状等

7 本人確認【P】

申出者の本人確認方法を事務処理要綱に規定。

《本人確認方法の例示》

- ・ 所属機関への電話又はメールによる確認
- ・ その他本人であることを証明する書類の提出
(必要な措置について掲載予定)

8 申出内容の審査

(1) 審査の実施主体

審査は各行政機関等が実施。

(2) 審査基準

統計法第 36 条では、「学術研究の発展に資すると認める場合その他の総務省令で定める場合」にのみ匿名データの提供ができると規定し、利用目的に一定の公益性を要求。また、公益性確保のためには研究成果が公表され、社会に還元されることが必要。

このため、各行政機関等は、申出書の記載内容及び添付書類を基に、利用目的が学術研究の発展に資すると認める場合その他の総務省令で定める場合に合致するか、研究成果が適切に公表され、社会に還元されるかに重点を置いて、審査。

なお、利用目的については次の考え方を参考に判断。

《学術研究の発展に資すると認める場合》

大学・研究機関における場合

大学や研究所などの教授，准教授，講師，ポストドクター，大学院生等が学術を目的として活動する機関が研究活動を行う場合を想定。

大学・研究機関以外における場合

営利企業に属する者の企業活動の一環としての研究も、学術的な研究で、その成果が社会に還元される場合や学術研究に営利目的が一部含まれる場合でも、その成果が社会に還元される場合は当該要件に該当。

しかし、研究成果が企業の業務資料や特定の顧客へのレポートである場合、当該要件に該当しない。

《その他総務省令で定める場合》(P)

(1) 審査内容

申出者の属する機関が外形的にみて研究を行っている機関か
研究の成果が社会に還元されるか
提供したデータの管理体制や管理方法に信頼が置けるか
研究内容と利用される統計調査の関係が適切であるか 等

9 手数料【P】

政令で規定された手数料の算定方法に基づき事前に見積もった上で政省令で定める手続きによる方法で納付。

具体的な手続きについて別途提示予定。

第8 審査結果の通知等

審査結果は、申出の受付から 日以内に申出者に対し文書により通知。

申出に応じる場合、通知文書に手数料を明示。

申出に応じられない場合、公平性や透明性を確保する観点から、通知文書にその理由も明示。

また、総務省は各行政機関等の匿名データの作成状況や提供の受付状況と併せて、申出に応じなかった場合の理由を取りまとめ、統計委員会に報告。

第9 記載事項に変更が生じた場合の手続き

記載事項に変更が生じた場合、直ちに変更の申出を行うよう要請。

第10 匿名データの作成・提供を外部委託する場合

1 匿名データ提供の事務の全部を委託する場合

統計法第 37 条では、匿名データの作成・提供の事務を全部委託する場合、政令で定める独立行政法人等に委託することを規定。

申出の受付及び審査は独立行政法人が実施。なお、審査に当たっては必要に応じて調査実施者と意見交換。

審査結果は、申出者及び委託元の行政機関等に通知。

2 匿名データの作成や提供の一部事務を民間に委託する場合

契約の方法は各府省が委託契約をする際の規約に従うものとするが、調査票情報の取扱い等についての取決めが必要である。

- ・ 秘密の保護に関する誓約書の提出
- ・ 調査票情報の受渡し、搬送、管理方法
- ・ 調査票情報の複写、貸与及び提供の禁止
- ・ 作成した匿名データの複写、貸与、依頼者以外への提供の禁止

等、「統計調査の民間委託に係るガイドライン」に準じて規定。

(調査票情報の適正管理義務及び罰則に係ることを明示)

第 11 匿名データの提供

(1) 提供時期

原則として、審査結果の通知日から 日以内。

(2) 提供窓口

原則として、申出書を受けた行政機関等とし、秘匿データの作成を外部委託した場合、委託先から提供することも可とし、この場合、申出書を受けた行政機関等から申出者にその旨を通知。各行政機関等は、事務処理要綱に提供窓口を規定。

(3) 提供手段

申出者は、電子媒体の書留等による送付、提供窓口における直接の受け渡しから選択し、申出書に記載した方法により提供。

第 12 匿名データの提供後の利用の制限

匿名データの利用終了後、返却させるか、破壊が確認できる仕組みを構築。また、匿名データから作成する統計は申出書に記載した利用目的の範囲内（別の研究に用いるような統計を作成してはならない）とすることを明示。

第 13 匿名データの不正利用に対する対応

統計法第 61 条第 3 号では、匿名データの提供を受けた者、匿名データの取り扱いに関する業務委託を受けた者等が匿名データを自己又は第三者の利益を図る目的で提供、盗用した場合罰則の適用を規定。

このほか、申出内容以外の利用が発覚した場合の行政的な措置をあらかじめ定め、周知、適用。

第 14 実績報告書の作成・提出

次の各項目に関する報告書を作成し、毎年 月までに総務省に提出。総務省は各行政機関及び届出独立行政法人等の提出した実績報告書を取りまとめ、統計委員会に報告。

- (1) 申出書受付の状況
- (2) 審査結果状況
- (3) 申出への対応困難な事案件数と理由
- (4) 匿名データの提供状況

第 15 ガイドラインの施行時期

本ガイドラインは、平成 21 年 月 日から施行

秘匿処理の考え方

(1) 秘匿処理とは

マイクロデータから世帯や個人の秘密の情報を知るということは、調査対象である調査単位(世帯や個人)とマイクロデータの対応関係を特定し、特定されたマイクロデータから調査単位の秘密に属する事項を知るということを意味する。どの調査事項が、秘密の情報に当たるかは一概には決めることができないし、時代とともに変化し、普遍的ではないと思われるので、秘匿処理とは、基本的には、調査単位とマイクロデータの対応関係を特定されないようにするというものである。

(2) 対応関係

提供するマイクロデータには、氏名、住所などの直接的に世帯や個人が特定できる情報は付与されていないので、調査単位とマイクロデータの対応関係は、性別や年齢などの属性(識別情報)が同じかどうかで判断することになる。

全国の全調査単位のマイクロデータが提供されていて、かつ、全調査単位について識別情報が分かる場合、識別情報が一致する調査単位とマイクロデータがそれぞれ1つしかない場合には同じ世帯や個人と判断でき、それぞれ複数ある場合はそのうちのいずれかと判断できる。実際のマイクロデータの提供の場合、一部の調査単位のマイクロデータが提供されていて、かつ、一部の調査単位の識別情報がわかるに過ぎず、このような状況では、対応関係を特定するのは現実的ではないと考えられる。

(3) 特定の可能性

特定の可能性を考えると、地域範囲が狭い場合には、調査対象が絞り込まれるので、識別情報を収集することが容易になり、マイクロデータの地域情報が詳細であれば、特定の可能性が高くなる。また、調査を受けていることが知られていると、その調査単位のマイクロデータに必ず存在することが分かるため、対応関係を特定される可能性が高まる。しかし、調査対象のリストは厳格に管理されており、外部の者が調査を受けている調査単位を知る可能性は低く、調査時から数年が経過すれば外部の者が知ることは不可能と言える。

しかし、特殊なデータのとくに、特定の可能性は高くなる。例えば、100歳以上の高齢者がいる世帯や世帯員が10人いるというような世帯の数は少ないので、母集団のある個別の世帯に対応するデータ数が少なくなり、そのどれに当たるか決定するのが比較的容易になる。また、複数の属性の特殊な組合せも特定の可能性が高くなる。これに対し、標準的な対象の場合には同じ条件のデータが多数出現することになるので、特定の可能性は比較的低いものととどまる。

(4) 識別情報

調査対象である調査単位とマイクロデータの対応関係を特定しようとするときに用いる識別情報とは、提供するマイクロデータに含まれていて、かつ、統計調査以外からも知ることができる情報である

個人又は世帯を対象とした統計の場合、比較的容易に入手できる識別情報としては、外観からでも把握できるような基本的な属性が考えられ、例えば、県、市町村などの地域情報や、世帯員数、世帯員の性別、住宅の大きさなどが挙げられる。このほか、自宅で営業

している世帯であればその産業・職業を知ることができるし、子供の年齢は通学している学年で分かると思われる。ただし、これらの情報だけでは、一般には対応関係を特定することはできない。また、これらの情報の収集は比較的簡単ではあるが、多数の調査単位について情報を収集しようとするれば大きな作業量を必要とする。

実際の問題としては、時間が経つとともに識別情報を正確に知ることは難しくなる。提供されるマイクロデータは数年前の調査の結果であり、そのときに個々の調査対象がどのような属性を有していたか知ることは、たとえ世帯の基本的な属性であっても難しい。既存のリストのようなものの場合も、そのリストとマイクロデータの時点が一致していないと対応関係の特定には多くの誤りが生じることになる。

(5) 特定の試み

秘匿処理の方法を決めるときには、現実にとどのような危険があるかについても考えておく必要がある。最近、個人情報の流出がよく問題となるが、そのような例では、住所（メールのアドレス等も含む。）氏名などが流出しており、それは、商業目的などにそのまま利用できる。しかし、統計情報の場合、住所、氏名が流出することはありません。また、前述のとおり、特殊な対象の場合には特定の可能性が比較的高くなるが、多くの標準的な対象の場合には特定の可能性は比較的低いものにとどまる。一部の対象についてだけ特定できたとしても、商業目的での利用価値は少ないであろう。したがって、対象を特定しようとするような試みが、最近問題になっているような商業目的で行われる可能性は低いものと考えられる。そもそも、数年前の統計情報では利用する価値もないであろう。

しかし、もし対象を特定するような試みが実際に行われたら、それはマイクロデータ提供の危険性、ひいては統計調査の危険性を指摘するものとして利用されてしまうであろう。ところが、絶対的な匿名性を担保しようとする、ドイツでの経験のように提供できる情報が極めて限られてしまう。したがって、この問題は秘匿処理だけで対策を考えるべきものではなく、そのような試みを行うこと自体を制限しておくことが必要となる。このため、データを提供するときには、利用目的を限定し、データの管理を適正に行わせることを義務付けておかななくてはならない。

注：ドイツは、1980年の連邦統計法で「絶対的な匿名化」条項によるマイクロデータの提供を行ってきたが、多くの情報が失われることになり、科学研究の要求に応じられず、ほとんど利用されなかった。そのため、1987年の連邦統計法ではマイクロデータが莫大な時間や経費をかけない限り識別できないという「事実上の匿名性」の概念に法規定を改正している。

秘匿処理の技法

(1) 秘匿処理の技法

対応関係を特定しにくくする秘匿処理の方法としては、下記のような方法がある。

識別情報等の削除

対応関係を特定する危険性の高い識別情報である、世帯や居住地を直接的に特定できるような情報を削除する方法である。

識別情報のトップ・コーディング

対応関係を特定できる可能性が高くなる特殊な属性を、まとめる方法である。例えば、100 歳以上の高齢者がいる世帯や世帯員が 10 人いる世帯の数は少ないので、対応関係を特定しやすくなるので、特に大きい値や小さい値を「 以上」、「 以下」というようにまとめる。海外では、トップ・コーディングされるのが対象全体の 0.5%以上としている例などがある。

識別情報のグルーピング

特定の値をグループ分けして階級区分に変更する方法である。例えば、年齢を例にすると、22 歳ではなく、21～25 歳とする方法である。また、市町村コードなどの地域情報の場合は、外部の者にも把握しやすい情報であること、対応関係を調べなくてはならないデータの範囲を限定できることなどから特に注意が必要となる。海外では、人口 10 万人未満の地域区分は提供しないなどの基準が設けられている例などがある。

リサンプリング

マイクロデータをすべて提供するのではなく、そこから抽出した一部のマイクロデータだけを提供する方法である。この方法によれば、提供するマイクロデータが少なくなるので、対応関係を特定できる可能性を低下させることができる。

また、特定できたとの主張に対し、特定できたと考えることが適当ではないと主張する方法でもある。

マイクロデータのソート

マイクロデータの配列順を並べ替えることでランダムにし、対応関係を探り出すことができないようにする方法である。

別の概念からの秘匿処理の技法としては、マイクロデータから正確な対応関係を知ることができないようにする方法がある。具体的には、マイクロデータを加工して正しくないものにしてしまう方法である。

スワッピング

任意の 2 つの調査単位の間で、一部の調査事項の値を入れ替える方法である。

誤差の導入

マイクロデータの一部の調査事項（識別情報又は秘密の情報自体）に誤差を導入する方法である。

(2) 秘匿処理の方法の決定

上記のような問題があるものの、実際に海外で行われている秘匿処理の方法をみるとかなり詳細なデータをそのまま提供しているのが普通である。秘匿処理は、論理的に可能性だけを考えると極めて厳しく行わなくてはならないことになるが、実際には、秘匿の必要性や利用面も考慮して現実的な判断の下で決定している。

そのような現実的な判断を行うために、海外では権威ある委員会などが処理の方法を最終承認する方式をとっている。我が国においても同様の手続きを踏むべきであり、試行的提供では、統計局の「匿名標本データ作成・利用研究会」の承認を得ている。

匿名化の基準（目安）

1 地理的情報について

- (1) 地理的情報としては、地域内に最小でも人口 50 万人以上いなければならない。
- (2) 直接的な地理的情報以外で、地理的情報が明らかになる項目（例えば、サンプリング情報など）についても、上記(1)の最小人口 50 万人の基準に適合させなければならない。
- (3) 地域分析用として、人口 50 万人未満の地理的情報を提供するような匿名データを作成する場合には、他の識別情報などの匿名化の程度を高めなければならない。
- (4) 入手可能な外部情報により、ある特定の種類の施設であることが明らかになるようなことがないようにしなければならない。

2 個人・世帯の識別情報について

- (1) 氏名、住所など個人又は世帯を直接的に識別できる情報は削除されなければならない。
- (2) 間接的に個人又は世帯を識別できる情報、例えば年齢、世帯人員、居住室数などの情報については、年齢の高い個人、世帯員数が多い世帯、居住室数の多い住宅など特定される可能性が高い場合、トップコーディング、グルーピングまたは削除を施す必要がある。トップコーディングにおいては、母集団（個人又は世帯）全体の 0.5% を目安にすることが望ましい。
- (3) 少数の特定の集団を対象とする場合、トップコーディングの基準を 3 ~ 5 % にすることを考慮すべきである。
- (4) トップコーディングするデータ項目については、その情報（平均値や中央値など）を明らかにすることが望ましい。
- (5) 世帯単位のデータを提供する場合、調査単位が特定されないことがないように、必要があれば、匿名化を考慮する必要がある。

3 誤差（ノイズ）

- (1) ミクロデータに誤差を加えることによって、調査データと外部情報との対応関係を特定する可能性を低めることができる。他に適当な匿名化の技法がない場合には、研究・分析上の有用性を損なわない範囲で誤差を付加することを考慮すべきである。
- (2) 誤差を加える方法としては、乱数による誤差の付加（random noise）、調査単位間の調査情報の交換（swapping）、ブランク（blank）への置換え又は補定（imputation）がある。

4 リサンプリング

ミクロデータを全て提供する場合は、その一部を提供する場合に比べて、調査単位の特定の可能性が高くなる。例えば、ある人が調査を受けたことがわかっている場合には、ミクロデータの中に必ずその人のデータがあるはずとの前提で探すことができる。したがって、必要に応じて、ミクロデータの全てではなく、一部のデータだけを提供することを考慮すべきである。

5 外部ファイルとのマッチングの可能性

- (1) ミクロデータと外部の既存ファイルのデータを突き合わせるにより調査単位が識別されるような可能性があれば、それを回避するための措置をとらなければならない。
- (2) 調査のための標本フレームが、国勢調査の母集団情報以外の情報によって提供されている場合には、調査データと標本フレームの元の情報とを一致させることが可能となるおそれがあるので、事前に回避する措置をとらなければならない。

6 その他の問題

- (1) データの一連番号、データの並び順によって、およその地域範囲が推測されるおそれがあるので、削除、付替え又は並べ替えをするべきである。
- (2) サンプリングに関する情報によっては、地理的情報以外に特定の地域や集団であることが明らかになるおそれがあるので、そのような情報は削除すべきである。
- (3) 秘密の情報のうち秘匿の必要性の高い調査項目については、その調査項目自体についてグルーピング、削除等の匿名化を施す必要がある。
- (4) 時間の経過とともに、調査データを外部情報と照合することは困難になる。提供時期は調査時点から最低限2年間以上は離すべきである。