

参考資料(Supporting Document)

データ開示における潜在的な露見可能性に関するチェックリスト (仮訳)

バージョン 1.2
2006年3月9日出版

センサス局標準
開示レビュー

Laura Zayatz 著
開示評価委員会 委員長
統計調査部

ドキュメント管理とコントロール

バージョン	出版日	承認	説明
1.0	2004年3月12日	Associate Directors	初版
1.1	2005年6月14日	Configuration Mgr.	Census Bureau Identity Standard and Quality Program Document Management Planに適合させた
1.2	2006年3月9日	Configuration Mgr.	主要標準のハイパーリンクを挿入

このドキュメントの最新バージョンは、センサス局のイントラネットで保守されており、品質管理保管所(the Quality Management Repository)を通じてアクセス可能です。

一旦このフォームに記入されたものは、センサス局の機密事項になります

データ開示における潜在的な露見可能性に関するチェックリスト

あなたは、このフォームに記入する必要がありますか？

あなたの報告義務を削減するために、必ずしも繰り返し実施されるすべての標本調査又はセンサスの公表の度にこのチェックリストを完成させる必要はありません。もし、下記の基準にすべて合致すれば、あなたは、単にDRBの議長あてのメモを用意するだけでよい:地理的情報に変更がない、新規調査事項がまったく導入されていない、最初のデータ提供によって是認された開示回避手段がその後の全ての提供において履行されている、以前の提供においてチェックリストにすべて記入されていた。

たとえチェックリストがこの場合不要だとしても、データが再識別化(reidentification)に使用されかねない情報の公共利用に何らかの変化があることを考慮し、データは引き続き再検討が必要であることに注意して下さい。

標本地域が100,000人未満で地理的区域が識別されない人口統計調査に基づく表形式のデータを除き、どのような形式のデータ提供(例えば、人口統計又は経済統計のマイクロデータ、人口統計又は経済統計の表形式のデータ、オーディオ・テープなど)であっても、このフォームに記入しなければなりません。

このフォームはセンサス局の機密事項であることに注意して下さい。

データ開示における潜在的な露見可能性に関するチェックリスト

センサス/標本調査名: _____ 日付 _____
プロジェクト・マネジャー名: _____ 課. _____ 係. _____ Ph. _____
スポンサーの官庁: _____
提供予定時のデータの経年: _____ (年)

下記のうち、当てはまるものがあればチェックして下さい:

- [] 今回の申請は、一回限りのデータ作成のためのものである。
- [] 今回の申請は、本質的に同じ内容を持つ一連の提供のためのものである。
(次の提供がなされるまでの期間を具体的に記してください。)
- [] 今回の申請は、補足的あるいは以前発表していないデータについてのもので、承認を受けた成果物の再公開のためのものである。
(マークした場合、DRB/MRPに元の成果物を提出した日を記入して下さい)
- (答えが現在と異なるチェックリストの質問があればその質問にのみ、記入する必要があります。)

このチェックリストは3つのセクションに分かれています。該当するセクションについて、すべての質問にお答え下さい。もし、答えるのにスペースが足りない場合は、どうぞ紙を継ぎ足して、質問の番号が分るようにしておいて下さい。

- セクション1(2-10頁)は、マイクロデータに関する質問です。マイクロデータファイルは、回答者レベルのレコードで構成されています。各レコードには、人、世帯、事業所及びその他の構成単位にとって貴重な変数が含まれています。ほとんどのマイクロデータファイルは、人口統計的な情報を含みます。このセクションのいくつかの質問は、事業所ベースのファイルには適用できないかもしれません。
- セクション2(11-12頁)は、表形式(頻度計算)の人口統計データに関するものです。頻度計算データは、1セル内の構成単位(人、家族、など)の数を表します。このチェックリストは、センサスから作表されたもの、あるいは識別される地理的区域が標本数100,000人以下であれば、記入する必要があります。
- セクション3(13-16頁)は、事業所における表形式(規模)のデータに関するものです。規模のデータは、1セル内のあらゆる構成単位の利益の量の合計を表します。

セクション1 ミクロデータ

1.1. ファイルにおける地理的情報

原則: 識別される地理的区域には、標本地域内に最小でも100,000人いなければならない。

1.1.1 ファイル上で表示されるのは地理的にどのレベルですか?

ファイル上にある明白な地理的識別子だけでなく、データ項目、レコード識別子、あるいはファイル構造が、推測されうる余計な地理情報をもたらしていることもあります。したがって、これらの処置は最小人口100,000人の基準に適合しない地理的区域を不注意により識別されてしまうことを避けるために取られなくてはならないものです。問題が起り得る地域は、以下で吟味します。それぞれの地域ごとに、申請したファイルが提供される前に今まで行ってきた、あるいは、これから行われようとする活動について、どうか指摘願います。

- 1.1.2 第1 標本単位(PSU)あるいは他の地理情報は、内部利用のために設計された管理番号として組み込まれているのが一般的である。

提供されるファイルにおいて、どのようにこの問題が避けられるか:

___ 管理番号を削除するか、地理情報を含めない。

___ 管理番号を暗号化する;具体的に述べて下さい。

___ その他;具体的に述べて下さい。

- 1.1.3 多くのデータベース内のレコードは規則的に連続しているものであり、そのため、最初の事例は小さな番号の付されたPSUあるいはアルファベットの順序で最初になる郡となる。このファイルの記録がそのような地理的な推測が避けるように、どのような順番にならべられているか手短かに説明して下さい。

- 1.1.4 居住についての特定の地理(情報)を含意するデータ項目は、当委員会のために用意した人口統計表上で表示されている明白な識別子以上のものを明らかにし兼ねません。例:スペインの姓が含まれているか否か(5つの南西の州でのみ暗号化されている)という項目はそのグループの州であるという明白な識別子がファイルに存在しない場合であっても明らか; 地下鉄のある領域から地下鉄のない領域までの移動を識別する移動コードは、地理的識別子の一部として地下鉄の有無が含まれていない場合であっても明らか;原子炉または空港のXマイル内の住宅については、識別できる地域がたった1つだけでも明らか;電話局番;地域コード;あるいは、緯度と経度も同様に働きます。

この理由により削除されることになるすべての項目をリストアップして下さい:_____

他の項目において地理的な意味を持つ可能性はあると思うが、削除すべきかどうか決めかねているものがあれば、すべてリストアップして下さい。

-
- 1.1.5 サンプル情報はまた何らかの地理的な指標を提供する可能性があります。例えば、あるウエイトを見ればself-representingとnonself-representingのPSUsが識別される、あるいは意図的に過剰にした標本抽出地域の種類が明らかになる可能性があります。また、「ダービン・タイプ(Durbin Type)」、「ヒットナンバー(Hit Number)」などのためのコードは、地理に関係するかもしれません。

秘密保護の理由やウエイトを識別しにくくするような副標本プランのため、分散が推定されることから削除されるものがあれば、それらを含めてすべてのサンプル情報をリストアップして下さい:

他のサンプル情報において地理的な意味を持つ可能性はあると思うが、削除すべきかどうか決めかねているものがあれば、すべてリストアップして下さい。

- 1.1.6 ファイル上のどのデータ項目も、入手可能な情報を基に、ある特定の種類の施設(例えば刑務所または老人ホーム)の住居であることが明らかになることがありますが、そのようなデータ項目はありませんか。その施設が識別される地域内でたった1つだけのようなケース、あるいは、そのような目的で記録する仕組みが得られているようなケースです。

はい - 施設のタイプを明示して下さい

いいえ

- 1.2. 個人(情報)の開示に係る特段のリスクをもたらすファイルの内容

公開用マイクロデータの開示基準は、申請された内容のいずれもが個人(情報)の開示に係る特段のリスクをもたらすかどうかを決定するため、それぞれのファイルのレビューを必要とします。DRBは、個々の回答(調査対象)者の秘密を保護するために取りうるいくつかの開示回避手段を明らかにしてきました。その手段は下記に述べたとおりであり、また、申請されたファイルに関連した重要な情報は、レビューの際の委員会(審議)の一助となるよう求められます。

- 1.2.1 氏名、住所、そのほか社会保障(制度)、高齢者医療保険制度又は低所得者医療扶助制度の番号といったユニークな数字の識別子は、ファイルから除外されなければなりません。

- 1.2.2 高収入は、個人又は世帯の明らかな特徴であり、また情報の中でも繊細な取り扱いが必要な項目と考えられています。したがって、ファイルにあるそれぞれの収入額は、世帯か、個人か、あるいは家族であるか、また、合計の所得か、個々の構成員の所得が含まれているかにかかわらず、トップコーディングされるべきです。母集団(人/世帯)全体に適用される収入変数のトップコーディングは、少なくともすべての事例の0.5パーセントが含まれるようにすべきです。

Subpopulations(特定の人口)に適用される収入変数の場合、トップコーディングは、適当

な事例の3パーセント、又はすべての事例の0.5パーセントのどちらかを含まれるようにすべきであり、どちらであってもかなり高いトップコーディングです。この原則の例外は、ある条件の下で起こりえます;例えば、地理的な詳細情報がほとんどない場合です。このようなトップコーディング原則の例外(的許可)はファイルの提供を承認する最終意見提出前に委員会で十分な議論がなされます。

すべての収入のトップコーディングは、適切な原則に合致していますか:

___ はい

___ いいえ - 何パーセントのトップコーディングとなっているか、トップコーディング(対象)の総数を明らかにし、委員会での審議を手短かに要約して下さい。

- 1.2.3 収入だけではなく、そのほかある特徴によっては、ある人と他人との違いがはっきり見えるようなものがあります。例えば、非常に高齢、自身の資産、賃貸料、住宅ローン総額が高価値あるいは高価格なケースです。ファイル上に示された地理の詳細によっては、このような項目が間隔又は順序を表す変数である場合には、これらにトップコーディング(及び/又は統合(collapsing))を施すよう考慮すべきです。当委員会は、(ウエイトを考慮して)ファイル上に代表される母集団(個人/世帯)全体の少なくとも0.5パーセントがこのようなトップコーディングの範疇として含まれることを提案します。ケースとしては少ないのですが、ごく少数の人口のみに適用される変数のような場合、当委員会は、トップコーディングの範疇として、適当なSubpopulations(特定の人口?)の約3~5パーセントを含めるべきであると考慮する場合があります。承認されたトップコーディングの例は:

年齢 - 90歳(2000年人口・住宅センサスの総人口の約0.5パーセント)

現金払いの居所(Housing Unit)の賃貸料 - 1,700ドル(2000年人口・住宅センサスの居所の約0.5パーセント)

暖房用燃料代 - 2,100ドル(2000年人口・住宅センサスにおいて、暖房用燃料が使われ、かつ、賃貸料や共益費とは別に支払われ、現在人が住んでいる居所の約3.0パーセント)

トップコーディング(又は統合(collapsing))及びトップコーディングに相当するものをすべてリストアップして下さい:

トップコーディングの必要性に関して疑問がある項目があれば、そのすべての項目をリストアップして下さい:

- 1.2.4 トップコーディングするデータ項目として提供されるべきであるとして何らかの提案があった場合は、その情報について詳述して下さい(例えば、トップコーディングした変数の平均値又は中央値)。

-
- 1.2.5 そのほかにも、個人を目に見えるようにするかもしれない特徴があります。それは、地理に影響されるものの、非順序変数として表示されており、それゆえトップコーディングできないものです。例えば、外国語、あるいはインドの部族の言語を示す符号;エスキモー人、アレウト族、グアム人、あるいはサモア人といった詳細な人種の識別子;前住居の場所の符号などです。このような場合、詳細値の総計はより大きいカテゴリーへ統合させる(collapsed)必要があるかもしれません。

秘密保護の理由で統合(あるいは削除)される項目があればすべてリストアップして下さい:

詳細を統合(collapse)する必要性に関して疑問がある項目があれば、そのすべてをリストアップして下さい:

- 1.2.6 コンテキストチュアルな(contextual)変数(その地域で人または世帯が住んでいることを文脈上物語っている変数)

どのようなコンテキストチュアルな変数があるか、また、暗号化されるレベルを明らかにして下さい。

秘密性理由で統合(あるいは削除)されるコンテキストチュアルな変数があればすべてリストアップして下さい:

詳細を統合(collapse)する必要性に関して疑問があるコンテキストチュアルな項目があれば、そのすべてをリストアップして下さい:

- 1.3. 外部ファイルとのマッチング能力に関する開示リスク

外部のファイルは通常住所と氏名を含んでおり、したがって調査回答(対象)者を識別するために用いることができることから、当該ファイル上のマイクロデータと外部ファイルのデータが一致する潜在的な可能性を回避するための努力なされなければならない。もし、他の行政機関や組織が整備しているメーリング・リストや行政記録上で見つけれられる極めて具体的な特徴がその調査内容に含まれていれば、このように一致する可能性があります。

例えば、特定の地理的識別子と結合して自動車の製造元、モデル、年式が含まれていれば、これらの項目は氏名と住所を含む自動車の登録リストとのマッチングが可能なので、受け入れ難いものとなります。このような項目は広いカテゴリーとして再コードしなければ、たぶんファイルに残すことはできないでしょう。いくつかの例を挙げると：製造業者の所有する特定の主要耐久財(例えば、飛行機)の買い手のリスト;ある州の有権者登録リスト;連邦、州、あるいは地方自治体の納税記録;刑事裁判制度記録;州の狩猟及び漁労免許記録;ある同業組合の会員名簿など。

調査のための標本フレームがセンサス局以外のソースから来ている場合、マッチングはまた極めて可能です。標本フレームを提供した行政機関が調査データと元の記録を一致させることは、とりわけ調査記録に当該機関に由来するデータが含まれていれば、可能です。例えば、事業(プログラム)給付の受領総額、事業(プログラム)の開始年月日など。

1.3.1 外部ファイル

1.3.1.1 この申請されたファイルにあるデータを含む行政記録、メーリング・リストあるいは他の外部ファイルがあることをご存知ですか。

はい - そのリストを明らかにして下さい _____
 いいえ

1.3.1.2 申請されたファイルに含まれる標本がセンサス局以外のソースによって提供されたりリストから抽出されたものですか。

はい - ソースを明らかにして、どのように、また、誰によって標本がリストから抽出されたか記述して下さい。:

 いいえ

1.3.2 マッチング

外部ファイルが存在する場合、調査データとこのファイルをマッチングさせる可能性を避けるために、いくつかの手段が取られる可能性があります。例えば、選択した項目を、削除又は再コードする、あるいは、「ノイズ(誤差)」(すなわち、少量の乱数による変化)をこれらの項目に導入する可能性があります。当委員会は、マッチングの可能性を回避させるためにどちらの手段を取られなければならないか、あらかじめ、正確に示すことはできません。しかし、外部データベースとのマッチングの可能性が存在する場合、当委員会では、ファイル提供に関連したリスクを決定する際に、(次のような)いくつかの要因を考慮します: 1) マッチングする目的で利用できる変数の数; 2) マッチングを実行するのに必要な資源; 3) データの年齢(経年); 4) 外部データの入手しやすさ、信頼性、完全性; そして、5) データの内容が繊細に扱われるべきものか、あるいは独特なものか。マッチングをより容易にするいくつかの要因については下記にリストされています。また、マッチングの可能性を避けるためファイルが提供される前取るべき手段として、情報が必要になります。(注: たとえあなたがマッチングに用いられるかもしれない外部ファイルとしてどのようなものがあるか知らないとしても、この情報は必要です。)

マッチングがより容易になるのは -

- 1.3.2.1 …あるデータ項目あるいはその項目の組み合わせによって、ある小さな、そして容易に識別可能な人の集団を隔離させるような場合。非常に小さな人口区分を示すコードを含めることは避けるべきです。例えば、極めて特定の地理との組み合わせたインディアンの種族又は詳細な職業。普通、1つのファイル又はリスト上に変数のグループが同時に現れることでもない限り、人は一度に1つ以上の変数を考える必要はありません。例えば、年齢と性別は外部ファイル上に同時に現れることはありそうですが、出生した国と職業はありそうにない;このように、ロシア生まれの建築家のような稀にしか出現しない事象を保護する必要はないとすべきです。

ファイルに含めるために申請したデータ項目のうち、小さく、容易に識別可能な人の集団を隔離させるような項目があれば、すべてリストアップして下さい。

この理由により、変更(すなわち、削除、再コード化、誤差(ノイズ)の付加)されるであろうデータ項目をすべてリストアップして下さい。

-
-
- 1.3.2.2 …当該ファイルが人口の実質的な一部を含んでいる(仮に $p > 0.5$ とする)場合。例:大雇用主、高収入の個人、様々なタイプの医者や科学者、または様々なタイプの施設の入所者。追加的な副標本(subsampling)では、データ提供前のある層内においてしばしば必要となります。

このような人の集団がファイル上にいくつか存在している、また、何らかの方法でこれらが副標本となっている場合、これらの人の集団を明らかにして下さい。

-
-
- 1.3.2.3 …当該ファイルに含まれる何らかの情報が記録または他のソースから得られたものであり、その情報が個々の識別子または詳細な地理情報を持っている外部ファイルへのリンクとして働くかもしれないような場合。その例に含まれるものとしては、公益事業会社から燃料消費又は費用の記録;(10年)人口・住宅センサスから得られた近隣、土地またはRDの集計での特徴;政府機関からの福祉または社会保障データ;警察からの逮捕記録;年金と健康保険として従業員に与えられた給付金。

ファイルに含めるために申請したデータ項目のうち、回答者(調査対象者)からのインタビューからは得られないことのない項目があれば、すべてリストアップして下さい。

この理由で変更または削除したデータ項目をすべてリストアップして下さい。

- 1.3.2.4 … マッチングにしばしば用いられる、正確な誕生日、性別、人種のようなデータ項目がファイルに含まれる場合や、または正確な収入総額、不動産税や他の税金、政府支援の(事業)プログラムの開始日と終了日などの両ファイルを一致させる他の項目が含まれる場合。

もしある場合、これらのデータ項目をすべてリストアップして下さい。

この理由で変更または削除したデータ項目をすべてリストアップして下さい。

- 1.3.2.5 … 長期的な(longitudinal)データが収集されている場合。すなわち、同じ回答者/単位を対象にしたデータがいくつか異なる照会期間に収集されるような場合。第一に懸念すべき事として話すべきことは、時系列のデータ項目は潜在的に外部の記録とマッチングしやすいということである。例えば、所得税あるいは雇用記録。もしデータが一度ならず同じ回答者から収集されるのであれば、インタビューの頻度、一つの標本単位として残存するかもしれない期間、ある時期から次の時期までに標本単位として一致する公算に影響する要因を示して下さい。
-
-

- 1.3.2.6 … 極めて特定の地理情報がファイルに含まれている場合。例えば、州、SMSAsなど。(この地理情報は、人口統計表で示されるべきです。)

- 1.3.2.7 … このファイルを外部のデータとマッチングできないよう保護することに関して、以前に述べられなかった考慮すべき事をすべて記述して下さい。例えば、データにある非信頼性またはもともとのノイズ(誤差)など。
-
-

1.3.3 特徴が唯一である集団を明らかにするクロス表

- 1.3.3.1 唯一の特徴を持つ集団が明らかになるようなクロス表を作成されましたか?

もし「いいえ」ならば、1.4に跳んで下さい。

1.3.3.2 結果はどうでしたか。

1.3.3.3 これらの結果が基になった開示リスクがある場合、そのリスクを減少させるための追加の手段として、どのようなものが取られるでしょうか。

1.4. 誤差(ノイズ(Noise))

1.4.1 データに何らかの誤差(ノイズ)を加えられましたか。 _____
もし「いいえ」ならば、1.5に跳んで下さい。

1.4.2 どのような手順を使ってデータに誤差を加えましたか。どうかその手順(すなわち、影響のあるレコードは何パーセントか、誤差の分布など)の詳細を教えてください。

想定されるもの：乱数による誤差(random noise)
レコードの交換 (record swapping)
階級の交換(rank swapping)
ブランク (blank)又は補定(imputation)

1.4.3 元のファイルに戻って誤差を付与したデータとのマッチングを行うような試みはありましたか。

もし「いいえ」ならば、1.5に跳んで下さい。

1.4.4 どのように行いましたが、また、マッチングの成功率はどのくらいでしたか。

1.5. 編集済データ(回答者が提供したデータ値を変更したもの)と補定済データ(無回答のために作られたデータ値)は、それ自身組み込まれた誤差(ノイズ)となります。編集(エディティング)と補定(インピュテーション)のプロセスは、開示に対する保護となります。その値が分る場合は、どうかこのセクションの質問に答えて下さい。

1.5.1 編集したデータ項目を少なくとも1つ含むレコードの割合は何パーセントですか。 _____

1.5.2 全データ項目の何パーセントが編集されたのですか。 _____

1.5.3 補定したデータ項目を少なくとも1つ含むレコードの割合は何パーセントですか。 _____

1.5.4 全データ項目の何パーセントが補定されたのですか。 _____

1.6. 他の問題

- 1.6.1 全標本を含むか、あるいはすべての層の標本を極めて高い比率($p > 0.5$)で含むファイルは単に標本のsubsample(副標本)を含んでいるだけのファイルよりも開示に結びつきやすい傾向にある。例えば、ある人が特定の調査を受けたことが知られているような場合、すべての標本がそのファイルで入手可能と仮定すれば、その人の記録と一致するマイクロデータファイルが見つけれられると推測できることとなります。

このようなファイルを含んでいますか

___ すべての標本

___ 標本のsubsample (副標本)(もしそうであれば、抽出率の範囲を明確に述べて下さい)

-
- 1.6.2 プロジェクト・マネジャーは、公開用ファイル(the public-use file)では省かれた詳細情報が含まれる内部バージョンのファイルから特別集計(special tabulations)を実施する場合、秘密保護の問題が引き起こされる可能性があることを知っておくべきです。例えば、公開用ファイルで含まれず、ファイル上の複数のデータ項目によってクロス表にされるような特定の地理情報が提供されてしまう可能性があります。DRBは、これらの作表を再検討するための手順の概略を記述したガイドラインを用意しました。どうか、このガイドライン(「公開用マイクロデータの利用によってもたらされる調査作表の開示可能性("Disclosure Potential of Survey Tabulations Given the Availability of Public-Use Microdata")」)を参照して下さい。そして、もし公開用ファイルで入手できない詳細を用いて統計表を発表する計画があれば、どうかDRBに相談して下さい。

1.6.3 標本設計について手短かに記述して下さい

- 1) 層化、クラスタリング、階層に関する何らかの記述を含めて下さい。そして標本となるある種の単位がどの段階においても $p > 0.5$ となることが明らかになることを含めて;
- 2) 申請のあった標本単位、調査区単位、研究分析の単位との比較・対照を含めて下さい;
- 3) 公的部門に提出予定あるいは予定がなくても標本設計(標本計画と推定量)の情報を明らかにして下さい。
- 4) ユーザーがどのように標本の分散を推定するか記述してください。申請されたファイルレイアウトで何らかの提供された「変数のネスト」や複製のアプローチのために使われる何らかのウエイトも潜在的に明らかにしながら。

1.6.4 補足

この情報は、別の調査への補足情報として集められたものですか _____

もし「いいえ」ならば、チェックリストのこのセクションは、完成です。

このマイクロデータファイルは、主要調査から作成されたファイルとリンクできますか。 _____

もし「はい」ならば、主要ファイルにはどのような地理情報がありますか。

セクション2 人口統計の表形式のデータ

2.1. データ

2.1.1 この成果物(統計表のデータ)を「標準的でなく」しているものは何ですか(すなわち、標本人口が10万人未満の人口センサスデータあるいは識別可能な地理的区域)

2.1.2 これは標本データですかセンサスデータですか。 _____
もしセンサスであれば、2.1.5への跳んで下さい。

2.1.3 抽出率を含め、標本設計について手短かに述べて下さい。

2.1.4 例えば、あるセルが10と表示されていれば、当該調査においてたった1人が当該セルの特定の特性を持たされていることを意味するというように、ウエイトは共通の知識(あるいは簡単に推量されうるもの)となっていますか。

2.1.5 発表される表の記載内容(すなわち次元、変数とそのカテゴリー)について教えて下さい。

2.1.6 発表される地理的なレベルはどの程度ですか。

2.2. これらの表を作成するために使用された行政記録のデータはありますか。 _____
もし「はい」ならば、どうか記述して下さい。

2.3. 編集済データ(回答者によって提供されたデータの値が変更されたもの)と補定データ(無回答のために作られたデータの値)は、それ自身に組み込まれた「誤差(ノイズ)」を有します。加工と補定のプロセスは、開示に対する保護となります。もし、その値がわかるのであれば、このセクションの質問に答えて下さい。

2.3.1 編集したデータ項目を少なくとも1つ含むレコードの割合は何パーセントですか。 _____

- 2.3.2 全データ項目の何パーセントが編集されたのですか。 _____
- 2.3.3 補定したデータ項目を少なくとも1つ含むレコードの割合は何パーセントですか。 _____
- 2.3.4 全データ項目の何パーセントが補定されたのですか。 _____

2.4. 開示回避

このデータに対してどのような開示回避のテクニックを利用したか、また、なぜ使用したか、(もしあるなら)どうか詳細な情報を提供して下さい。

想定される技術：レコードの交換 (record swapping)
ブランク (blank)又は補定(imputation)
階級の交換(rank swapping)
乱数による誤差(random noise)
セルの秘匿(cell suppression)
数値の丸めによる制御(controlled rounding)

セクション3 事業所統計の表形式のデータ

3.1. データ

3.1.1 これは標本データですかセンサスデータですか。 _____

3.1.2 事業所数は公表されますか。 _____

もし、センサスであれば3.1.5へ跳んで下さい。

3.1.3 あるタイプの事業所は、確信をもって選ばれましたか。 _____

3.1.4 抽出率を含む標本設計について手短かに記述して下さい。

3.1.5 発表されるデータは何ですか、また、形式(すなわち表の次元、変数とその詳細)はどうなっていますか。

3.1.6 発表される地理的なレベルはどの程度ですか。

3.2. これらの表を作成するために使用された行政記録のデータはありますか。 _____
もし「はい」ならば、どうか記述して下さい。

3.3. 編集済データ(回答者によって提供されたデータ項目の値が変更されたもの)と補定データ(無回答のために作られたデータの値)は、それ自身に組み込まれた「誤差(ノイズ)」を有します。加工と補定のプロセスは、開示に対する保護となります。もし、その値がわかるのであれば、このセクションの質問に答えて下さい。

3.3.1 編集したデータ項目を少なくとも1つ含むレコードの割合は何パーセントですか。 _____

3.3.2 全データ項目の何パーセントが編集されたのですか。 _____

3.3.3 補定したデータ項目を少なくとも1つ含むレコードの割合は何パーセントですか。 _____

3.3.4 全データ項目の何パーセントが補定されたのですか。 _____

3.4. 開示回避

このデータに対してどのような開示回避のテクニックを利用したか、また、なぜ使用したか、(もしあるなら)どうか詳細な情報を提供して下さい。

想定される技術：セルの秘匿(cell suppression)

誤差 (noise)

もし、セルの秘匿が使われなかったら、3.4.2に跳んで下さい。

3.4.1 セルの秘匿

3.4.1.2 一次秘匿の決定のため、どのような基準(また、どのようなパラメータで)を使用しましたか。 _____

もし、センサスのデータであれば、3.4.1.4へ跳んで下さい。

3.4.1.3 当該調査に適合させるため、その基準はどのように採用されたのですか？

3.4.1.4 一次秘匿を施すため、総数に何らかの保護をかけましたか？

3.4.1.5 一次秘匿の決定と各一次秘匿に必要な総数の保護に先立って、企業ごとに事業所を(又は所有者ごとに農場を)結合(名寄せ)しましたか。 _____

3.4.1.6 セルの秘匿を実施する際に、キーとなる項目を選びましたか。もしそうなら、それは何でしたか、又なぜですか。

3.4.1.7 セルの秘匿の実施は、センサス局の秘匿ソフトウェアによるものですか、又は手作業によるものですか。 _____

秘匿ソフトウェアによる場合は、3.4.1.9に跳んで下さい。

3.4.1.8 表の秘匿パターンは検査しましたか？ _____

もし、秘匿が手作業で実施された場合、3.4.2に跳んで下さい。

3.4.1.9 (あるかどうかわかりませんが)何らかのショートカットを使ったか、何らかのパラメータを用いましたか。

3.4.1.10 秘匿を手作業で除去したものはありますか。 _____
もしそうなら、なぜですか。

もしそうなら、それを置き換えるため、他のものをどのように選ばれるのですか。

3.4.1.11 三次元の表もすべて検査しましたか。

3.4.1.12 秘匿される値(すなわち、範囲、中央値、推計値、丸めた値、誤差を持たせた値など)として、何らかの追加情報も公表される予定ですか。

もしそうなら、どうか詳細について教えて下さい。

もし誤差が開示回避の技術として使われなかったら、3.4.3に跳んで下さい。

3.4.2 誤差(ノイズ(Noise))

3.4.2.1 どの項目に誤差を持たせましたか。

3.4.2.2 データへの誤差はどのように加えられましたか。

3.4.2.3 いくつの誤差がデータに加えられましたか。

3.4.3 このデータに何らかの他の開示回避技術が使われましたか。 _____
もしそうなら、どうか詳細について記述して下さい。

3.5. 特別なタイプのデータの取扱い

- 3.5.1 開示回避技術の適用によっては、特別な扱いを必要とするデータもあります。想定されるものとしては:

負の値となるデータ
ネット(正味)値でパーセントに変換したデータ
絶対値データ間の差
加重平均のデータ

特別な扱いを必要としたデータは、何かありましたか。_____

もしそうなら、どのデータで、どのように行いましたか。

3.6. 開示回避の調整

- 3.6.1 これは特別集計(special tabulation)ですか。_____

もし、「いいえ」ならば、3.6.3に跳んで下さい。

- 3.6.2 複数の秘匿パターンが相互に解き明かされないことを保証するため、すべての秘匿は、同じデータ・セットから生み出したすべての表の中で調整されなければなりません。

開示回避技術(例えばセル秘匿パターン)は、以前発表された基本表に用いられたものと調整されましたか。

- 3.6.3 同一(あるいはとてもよく似た)データは、別の課又は係によっても発表されたことがありますか。_____

もし「いいえ」ならば、チェックリストのこのセクションは完成です。

- 3.6.4 開示回避技術(例えばセル秘匿パターン)は、別の課又は係によって使われたものと調整されましたか。
