

オンサイト利用における分析結果等の提供に関する  
標準的なチェック内容の解説と例

令和元年 5 月

独立行政法人統計センター

情報技術センター統計情報提供課



# 目次

はじめに .....	1
1. 分析結果等の提供依頼について .....	1
2. 分析結果等のファイル形式について .....	1
標準的なチェック内容 .....	2
<b>I 統計表 .....</b>	<b>2</b>
1. 度数表、度数の構成比表 .....	2
2. 数量表（総和） .....	4
2.1 数量表（総和）（個人・世帯調査の場合） .....	4
2.2 数量表（総和）（事業所・企業調査の場合） .....	7
3. 数量表（平均、構成比、集中度） .....	18
4. その他の統計表 .....	19
<b>II 統計量 .....</b>	<b>20</b>
5. 最頻値 .....	20
6. 総和、平均、構成比、集中度 .....	21
6.1 総和、平均、構成比、集中度（個人・世帯調査の場合） .....	21
6.2 総和、平均、構成比、集中度（事業所・企業調査の場合） .....	22
7. 線形回帰係数、非線形回帰係数 .....	27
8. 分布の高次モーメント、相関係数、要約統計量及び検定統計量 .....	28
9. 最大値、最小値 .....	29
10. 推定残差 .....	29
11. グラフ .....	29
<b>III その他 .....</b>	<b>30</b>
12. 分析プログラム、操作ログ等 .....	30
統計表における秘匿措置 .....	31
1. 集計区分の変更 .....	31
2. 集計対象の変更 .....	32
3. セルの値を秘匿 .....	33
チェックの具体例 .....	39
【事例 1】 .....	39
【事例 2】 .....	43
参考 オンサイト利用における分析結果等の提供に関する標準的なチェック内容 .....	45

## はじめに

この資料は、調査票情報のオンサイト利用の手引別紙「オンサイト利用における分析結果等の提供に関する標準的なチェック内容（参考参照。以下「標準的なチェック内容」という。）」について、記載内容の解説と例を示すものです。そのため、まずは標準的なチェック内容の【用語の解説・留意事項】をよくお読みください。なお、標準的なチェック内容に記載されている内容については、この資料では斜体にしてあります。

調査票情報には調査客体の秘密が含まれており、分析結果等の公表に当たっては、秘密が漏洩しないよう保護する必要があります。標準的なチェック内容は、広く用いられていると考えられる統計分析の分析結果等について、調査票情報の秘密の保護を担保するため、一般に注意すべきとされている内容となっています。そのため、分析結果等によっては、標準的なチェック内容とは異なるチェック内容となることもあり得ます。チェック内容についてあらかじめ調整しておくことで、提供にかかる時間が短くなる場合もありますので、可能であれば事前に提供を希望される分析結果等のチェック方法等についてご相談ください。

### 1. 分析結果等の提供依頼について

提供を依頼する分析結果等については、申出者が標準的なチェック内容等により、秘匿の措置が行われているかなど安全性をチェックし、提供依頼書を作成します。統計センターは、提供依頼書や補足説明資料により、分析結果等の安全性を確認します。そのため、どのような分析結果等であっても、確認には使用したデータや、利用者が作成した変数、集計・分析方法等の情報が必要ですので、提供依頼書の所定の欄に詳細を記載してください。

申出者が十分なチェックを行っていない場合や、補足説明資料が不十分な場合は、統計センターから申出者に問い合わせるなど、確認に多くの時間がかかります。提供依頼に際しては、十分なチェックを行ってください。

### 2. 分析結果等々のファイル形式について

統計センターにおける確認時には、統計表の行計・列計の確認や、審査資料作成のため分析結果等へのコメントの記入などを行います。そのため、原則として、分析結果等々は csv 形式等 Microsoft Excel で編集が可能な形式で作成してください。ご利用の統計分析用ソフトウェア等が独自形式でしか分析結果等を作成できない等、やむを得ず Microsoft Excel で開けない分析結果等を作成される場合は、あらかじめご相談ください。

また、提供依頼書には各ファイルについて概要を記載しチェックする必要があることから、原則として、一つの分析結果等につき一つのファイルを作成してください。また、原則として、分析結果等と補足説明資料は別のファイルとしてください。

## 標準的なチェック内容

ここでは、標準的なチェック内容における「表1 標準的なチェック内容（以下「チェック内容表 1」という。）」について、「分析結果等の種類」別に記載されている「チェック内容」の解説と「申出者が提示する情報」の例を示します。

標準的なチェック内容では、統計表各セルの値もしくは統計量の値の算出元である各データ値が 0 以上であることを基本として定めています（標準的なチェック内容 2.1 ㉒㉓、2.2 ㉔㉕㉖㉗、6.2 ㉘㉙が該当）。したがって、負の値を含むデータ値の集計結果の提供を依頼する場合は、チェック内容を調整する必要がありますので、事前にご相談ください。

### I 統計表

提供を依頼する統計表は、1セルに一つの値が入った表にしてください。例えば、一つのセルに平均値と標準偏差が入った統計表を作成したい場合は、平均値の表と標準偏差の表の二つに分割して提供を依頼し、提供後に一つの表に戻してください。

#### 1. 度数表、度数の構成比表

##### チェック内容

##### ① 【度数】各セルが1以上10未満の調査客体から算出した値でないこと（加重なし）

セルの度数が小さい場合、そのセルに含まれる調査客体が特定されるリスクが高くなります。また、集計乗率を用いて（加重ありで）作成された加重度数表であっても、ある調査客体が調査に参加していることを知っている者がいると、この調査客体が含まれるセルの度数が小さい場合、この調査客体が特定されるリスクが高くなります。

そのため、度数表の提供を依頼する場合、各セルの加重なしの度数が10以上である必要があります。ただし、度数が0の場合は、当該セルに含まれる調査客体が存在せず、特定は発生しないことから、度数0をチェック内容から除外しています。

##### 申出者が提示する資料

##### ① 【度数】各セルの度数（加重なし）

提供を依頼する度数表が、チェック内容①を満たすことを示すため、各セルの加重なしの度数を明らかにしてください。

提供を依頼する度数表が、集計用乗率を用いずに（加重なしで）作成されている場合は、その表自体が加重なしの度数の資料ですので、補足説明資料を作成する必要はありません。

提供を依頼する度数表が、表 1 のように集計用乗率を用いて（加重ありで）作成されている場合は、表 2 のように同じ集計区分で加重なしの度数表も作成するなど、各セルの加重なしの度数を示す補足説明資料を作成してください。

**表 1 提供を依頼する加重ありの度数表**

加重あり	区分A-1	区分A-2	区分A-3	計
区分B-1	129	140	125	394
区分B-2	69	87	71	227
区分B-3	137	64	134	335
計	335	291	330	956

**表 2 加重なしの度数表（補足説明資料）**

加重なし	区分A-1	区分A-2	区分A-3	計
区分B-1	14	28	21	63
区分B-2	11	20	14	45
区分B-3	25	12	12	49
計	50	60	47	157

#### チェック内容

- ② 【度数】 行計又は列計の 90%超を占めるセルがないこと（加重なし）
- ③ 【度数】 行計又は列計の 90%超を占めるセルがないこと（加重あり）

度数表を構成する属性から、他の属性の情報が高い確率で推測できてしまうことがあります。例えば、表 3 を見ると、地域 2 において収入が 0～99 万円に入る人は 325 人で、地域 2 の約 91.5%を占めています。

**表 3 行計の 90%超を占めるセルがある度数表**

年収（万円）	0～99	100～199	200～299	300～	計
地域1	150人	170人	240人	320人	880人
地域2	325人	10人	10人	10人	355人
地域3	210人	220人	210人	200人	840人
計	685人	400人	460人	530人	2075人

そのため、ある人が地域 2 に住んでいることが分かれば、その人の年収が高い確率で 100 万円未満であることが推測できてしまいます。このように、提供を依頼する度数表について、行計又は列計の 90%超を占めるセルがある場合は注意が必要です。また、集計乗率を用いて（加重ありで）作成された度数表で 90%超を占めるセルがない場合でも、集計用乗率を

用いずに（加重なしで）作成した場合に 90%超を占めるセルがある場合もあり得ますし、その逆もあり得ますので、注意が必要です。

ただし、90%超を占めるセルがある場合でも、その度数表を公開するリスクが高いかどうかは、集計に用いた属性や、集計区分等によります。そのため、事前に提供を依頼する度数表について、集計内容、集計方法や結果表様式等を相談し、チェックの必要性についてお問合せください。

### 申出者が提示する資料

② 【度数】各セルの構成比（行計及び列計に占める割合）（加重なし）

③ 【度数】各セルの構成比（行計及び列計に占める割合）（加重あり）

事前に提供を依頼する度数表について相談し、確認の必要があると判断された場合は、求められた内容の補足説明資料を作成してください。集計内容、集計方法や結果表様式により、作成をお願いする資料の内容や様式が変化することがありますので、ご注意ください。例えば、三つの変数から作成した表であっても、そのうちの二つの変数に関する構成比表の作成をお願いすることもあり得ます。また、提供を依頼している表とは別の集計区分で構成比表の作成をお願いすることもあり得ます。

## 2. 数量表（総和）

数量表は様々な種類が考えられますが、標準的なチェック内容では、総和の数量表に対するチェック内容を基本として定めています。

公的統計調査には、大きく分けて個人・世帯を対象とした調査と、事業所・企業を対象とした調査の二つがあります。それぞれの調査により、チェック内容及び提示する補足説明資料が異なりますので、ご注意ください。

### 2.1 数量表（総和）（個人・世帯調査の場合）

#### チェック内容

① 【度数】各セルが1以上10未満の調査客体から算出した値でないこと（加重なし）

数量表に含まれるセル値の算出元である調査客体の数（度数）が小さい場合、そのセルの度数が公表されない場合でも、そのセルの度数が小さいことを知っている者により、そのセルに含まれる調査客体の数値が特定されるリスクが高くなります。また、集計乗率を用いて（加重ありで）作成された加重数量表であっても、ある調査客体が調査に参加していること

を知っている者がおり、かつこの調査客体が含まれるセルの度数が小さい場合、この調査客体の数値が特定されるリスクが高くなります。そのため、数量表の提供を依頼する場合、各セルの加重なしの度数が10以上である必要があります。ただし、度数が0の場合は、当該セルに含まれる調査客体が存在せず、特定は発生しないことから、度数0をチェック内容から除外しています。

## 申出者が提示する資料

### ① 【度数】各セルの度数（加重なし）

提供を依頼する数量表が、チェック内容①を満たすことを示すため、各セルの加重なしの度数を明らかにしてください。

提供を依頼する数量表が、表4のように集計用乗率を用いずに（加重なしで）作成されている場合は、表5のように同じ集計区分で加重なしの度数表を作成するなど、各セルの加重なしの度数を示す補足説明資料を作成してください。

表4 提供を依頼する加重なしの数量表

数量表 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	1075	807	1416	3298
区分B-2	1398	1343	1088	3829
区分B-3	1262	726	1009	2997
計	3735	2876	3513	10124

表5 加重なしの度数表（補足説明資料）

度数表 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	14	28	21	63
区分B-2	11	20	14	45
区分B-3	25	12	12	49
計	50	60	47	157

提供を依頼する数量表が、表6のように集計用乗率を用いて（加重ありで）作成されている場合は、表7のように同じ集計区分で加重なしの度数表を作成するなど、各セルの加重なしの度数を示す補足説明資料を作成してください。加重ありの数量表であっても、加重なしの度数が必要となりますので、ご注意ください。



表 6 提供を依頼する加重ありの数量表

数量表 (加重あり)	区分A-1	区分A-2	区分A-3	計
区分B-1	11850	8391	13883	34124
区分B-2	10736	13606	5472	29814
区分B-3	6995	11119	7971	26085
計	29581	33116	27326	90023

表 7 加重なしの度数表 (補足説明資料)

度数表 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	14	28	21	63
区分B-2	11	20	14	45
区分B-3	25	12	12	49
計	50	60	47	157

チェック内容

- ② 【度数】 行計又は列計の90%超を占めるセルがないこと (加重なし)
- ③ 【度数】 行計又は列計の90%超を占めるセルがないこと (加重あり)

表を構成する属性から、他の属性の情報が高い確率で推測できてしまうことがあります。例えば、表 8 を作成したとします。

表 8 作成した数量表 (加重なし)

年収総和 (万円)	職業1	職業2	職業3	職業4	計
地域1	83505	95013	76248	144704	399470
地域2	16315	3209	4853	4395	28772
地域3	78288	88792	104832	92300	364212
計	178108	187014	185933	241399	792454

同じ集計区分の度数表 (表 9) を見ると、地域 2 において職業 1 の人は 325 人で、地域 2 の約 91.5% を占めています。

表 9 表 8 の各セルの度数（加重なし）

	職業1	職業2	職業3	職業4	計
地域1	150人	170人	240人	320人	880人
地域2	325人	10人	10人	10人	355人
地域3	210人	220人	210人	200人	840人
計	685人	400人	460人	530人	2075人

そのため、各セルの度数を知っている者は、ある人が地域 2 に住んでいることが分かれば、その人が職業 1 であり、その職業の平均年収が  $16315 \div 325 = 50.2$  万円であることを高い確率で推測できてしまいます。このように、提供を依頼する数量表について、各セルの度数が行計又は列計の 90% 超を占めるセルがある場合は注意が必要です。また、集計乗率を用いて（加重ありで）作成された数量表で、度数が 90% 超を占めるセルがない場合でも、集計用乗率を用いずに（加重なしで）作成した場合に 90% 超を占めるセルがある場合もあり得ますし、その逆もあり得ますので、注意が必要です。

ただし、90% 超を占めるセルがある場合でも、その数量表を公開するリスクが高いかどうかは、集計に用いた属性や、集計区分等によります。そのため、事前に提供を依頼する数量表について、集計内容、集計方法や結果表様式等を相談し、チェックの必要性についてお問合せください。

#### 申出者が提示する資料

- ② 【度数】各セルの構成比（行計及び列計に占める割合）（加重なし）
- ③ 【度数】各セルの構成比（行計及び列計に占める割合）（加重あり）

事前に提供を依頼する数量表について相談し、確認の必要があると判断された場合は、求められた内容の補足説明資料を作成してください。集計内容、集計方法や結果表様式により、作成をお願いする資料の内容や様式が変化することがありますので、ご注意ください。例えば、三つの変数から作成した表であっても、そのうちの二つの変数に関する構成比表の作成をお願いすることもあり得ます。また、提供を依頼している表とは別の集計区分で構成比表の作成をお願いすることもあり得ます。

## 2.2 数量表（総和）（事業所・企業調査の場合）

### チェック内容

- ① 【度数】各セルが 1 以上 10 未満の調査客体から算出した値でないこと（加重なし）

数量表に含まれるセル値の算出元である調査客体の数（度数）が小さい場合、そのセルの

度数が公表されない場合でも、そのセルの度数が小さいことを知っている者により、そのセルに含まれる調査客体の数値が特定されるリスクが高くなります。また、集計乗率を用いて（加重ありで）作成された加重数量表であっても、ある調査客体が調査に参加していることを知っている者がおり、かつこの調査客体が含まれるセルの度数が小さい場合、この調査客体の数値が特定されるリスクが高くなります。そのため、数量表の提供を依頼する場合、各セルの加重なしの度数が10以上である必要があります。ただし、度数が0の場合は、当該セルに含まれる調査客体が存在せず、特定は発生しないことから、度数0をチェック内容から除外しています。

### 申出者が提示する資料

#### ① 【度数】各セルの度数（加重なし）

提供を依頼する数量表が、チェック内容①を満たすことを示すため、各セルの加重なしの度数を明らかにしてください。

提供を依頼する数量表が、表10のように集計用乗率を用いずに（加重なしで）作成されている場合は、表11のように同じ集計区分で加重なしの度数表を作成するなど、各セルの加重なしの度数を示す補足説明資料を作成してください。

表10 提供を依頼する加重なしの数量表

数量表 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	1075	807	1416	3298
区分B-2	1398	1343	1088	3829
区分B-3	1262	726	1009	2997
計	3735	2876	3513	10124

表11 表10の各セルの加重なしの度数（補足説明資料）

度数表 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	14	28	21	63
区分B-2	11	20	14	45
区分B-3	25	12	12	49
計	50	60	47	157

提供を依頼する数量表が、表12のように集計用乗率を用いて（加重ありで）作成されている場合は、表13のように同じ集計区分で加重なしの度数表を作成するなど、各セルの加重なしの度数を示す補足説明資料を作成してください。加重ありの数量表であっても、加重なしの度数が必要となりますので、ご注意ください。

表 12 提供を依頼する加重ありの数量表

数量表 (加重あり)	区分A-1	区分A-2	区分A-3	計
区分B-1	1337.9	940.3	1875.1	4153.2
区分B-2	1661.1	1262.8	1386.6	4310.5
区分B-3	1358.0	1096.2	814.5	3268.7
計	4357.0	3299.3	4076.2	11732.5

表 13 表 12 の各セルの加重なしの度数 (補足説明資料)

度数表 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	14	28	21	63
区分B-2	11	20	14	45
区分B-3	25	12	12	49
計	50	60	47	157

チェック内容

② 【数量】各セルにおいて、70%を超えて寄与する調査客体がないこと (変数変換なし)

セルの値に大きく寄与する調査客体が存在した場合、その調査客体の実際の値に対して、セルの値が精度の高い推測値となってしまうことがあります。例えば、表 14 を作成し、各セルの度数が表 15 のとおりだったとします。

表 14 作成した数量表 (加重なし)

研究開発費総和 (百万円)	産業1	産業2	産業3	計
地域1	520	400	360	1280
地域2	210	550	600	1360
地域3	700	350	300	1350
計	1430	1300	1260	3990

表 15 表 14 の各セルの度数 (加重なし)

	産業1	産業2	産業3	計
地域1	20社	15社	15社	50社
地域2	10社	20社	30社	60社
地域3	30社	15社	15社	60社
計	60社	50社	60社	170社

このとき、地域2、産業1のセルについて、セルの値の内訳が表 16 のとおりであったとします。

表 16 地域 2、産業 1 のセルの値の内訳

A社	B社	C社	D社	E社
200	2	1	1	1
F社	G社	H社	I社	J社
1	1	1	1	1

A社の値 200 は、セルの値 210 の約 95%を占めています。そのため、地域 2 にある産業 1 の A 社が研究開発費に多く支出していることを知っている者が、セルの値 210 を A 社の研究開発費の近似値とした場合、実際の A 社の研究開発費 200 を小さな誤差で推測できていることになってしまいます。

一般に、セルの値 $X$ とそれに含まれる調査客体の値 $x_1$ に対し、 $x_1 > 0.7X$ であれば、 $X - x_1 < 0.3X$ となり、 $X$ はセルの値の 30%以内の誤差で $x_1$ を近似します。そのため、数量表の提供を依頼する場合、各セルにおいて、一つの調査客体がセルの値の 70%を超えて寄与していない必要があります。

また、集計用乗率を用いて（加重ありで）作成した数量表の場合は、加重ありのセルの値に対する寄与をチェックする必要があります。

その他、変数変換を行った場合、見かけの占有率が変化することがあります。例えば、総和 1000 に対し、それに含まれる最大の値が 500 だとすると、この値はセルの値 50%を占めていることとなりますが、これを対数変換すると総和が $\log 1000 = 6.91$ 、最大値が $\log 500 = 6.21$ となり、見かけ上は総和の約 90%を占めることとなります。

また、例えば表 17 のような値を持つ 10 社があるとします。

表 17 各社の値

A社	B社	C社	D社	E社
1500	300	100	30	20
F社	G社	H社	I社	J社
10	10	10	10	10

このとき、総和 2000 に対し、最大の値 1500 は、総和の値の 75%を占めていることとなりますが、これらの値を対数変換すると、

表 18 対数変換後の各社の値

A社	B社	C社	D社	E社
7.313	5.704	4.605	3.401	2.996
F社	G社	H社	I社	J社
2.303	2.303	2.303	2.303	2.303

となり、対数変換後の値の総和 35.532 に対し、最大の値 7.313 は、見かけ上は総和の値の約 20.6%を占めていることとなります。

このように、変数変換により、正しい確認ができなくなることがありますので、変数変換された数値による数量表であっても、変数変換前の数値を用いてチェックする必要があります。

- ③ 【数量】各セルにおいて、85%を超えて寄与する二つの調査客体の合計値がないこと（変数変換なし）

セルの値に大きく寄与する調査客体が二つ存在した場合、一つの調査客体の実際の値に対して、セルの値からもう一つの調査客体の値を引いたものが精度の高い推測値となってしまうことがあります。例えば、表 19 を作成し、各セルの度数が表 20 のとおりだったとします。

表 19 作成した数量表（加重なし）

研究開発費総和 (百万円)	産業1	産業2	産業3	計
地域1	520	400	360	1280
地域2	210	550	600	1360
地域3	700	350	300	1350
計	1430	1300	1260	3990

表 20 表 19 の各セルの度数（加重なし）

	産業1	産業2	産業3	計
地域1	20社	15社	15社	50社
地域2	10社	20社	30社	60社
地域3	30社	15社	15社	60社
計	60社	50社	60社	170社

このとき、地域2、産業1のセルについて、セルの値の内訳が表 21 のとおりであったとします。

表 21 地域 2、産業 1 のセルの値の内訳

A社	B社	C社	D社	E社
116	86	1	1	1
F社	G社	H社	I社	J社
1	1	1	1	1

A社の値116とB社の値86の合計202は、セルの値210の約96%を占めています。そのため、地域2にある産業1のA社が研究開発費に多く支出していることをB社が知っているときに、セルの値210からB社の値86を引いた124をA社の研究開発費の近似値とした場合、B社は実際のA社の研究開発費116を小さな誤差で推測できていることになってしまいます。

一般に、セルの値 $X$ とそれに含まれる二つの調査客体の値 $x_1, x_2$ に対し、 $x_1 + x_2 > 0.85X$ であれば $(X - x_2) - x_1 < 0.15X$ となり、 $(X - x_2)$ はセルの値の15%以内の誤差で $x_1$ を近似します。そのため、数量表の提供を依頼する場合、各セルにおいて、二つの調査客体の合計値がセルの値の85%を超えて寄与していない必要があります。

また、集計用乗率を用いて(加重ありで)作成した数量表の場合は、集計用乗率を用いて、母集団における一番目及び二番目に大きく寄与する調査客体の値を推定し、加重ありのセルの値に対する寄与をチェックする必要があります。

その他、変数変換を行った場合、見かけの占有率が変化することがあります。例えば、総和1000に対し、それに含まれる最大の値が300、二番目の値が200だとすると、合計値500はセルの値50%を占めていることになりませんが、これを対数変換すると総和が $\log 1000 = 6.91$ 、合計値が $\log 500 = 6.21$ となり、見かけ上は総和の約90%を占めることになります。

また、例えば表22のような値を持つ10社があるとします。

表 22 各社の値

A社	B社	C社	D社	E社
1500	300	100	30	20
F社	G社	H社	I社	J社
10	10	10	10	10

このとき、総和2000に対し、最大の値1500と二番目の値300の合計値1800は、総和の値の90%を占めていることになりませんが、これらの値を対数変換すると、

表 23 対数変換後の各社の値

A社	B社	C社	D社	E社
7.313	5.704	4.605	3.401	2.996
F社	G社	H社	I社	J社
2.303	2.303	2.303	2.303	2.303

となり、対数変換後の値の総和 35.532 に対し、最大の値 7.313 と二番目の値 5.704 の合計値 13.017 は、見かけ上は総和の値の約 36.6%を占めていることとなります。

このように、変数変換により、正しい確認ができなくなることがありますので、変数変換された数値による数量表であっても、変数変換前の数値を用いてチェックする必要があります。

#### 申出者が提示する資料

- ② 【数量】各セルにおいて最も大きく寄与する調査客体の値  $x_1$  及び  $x_1$  がセルの値  $X$  に占める割合 (変数変換なし)
- ③ 【数量】各セルにおいて一番目及び二番目に大きく寄与する調査客体の値  $x_1, x_2$  及び  $x_1, x_2$  の合計値がセルの値  $X$  に占める割合 (変数変換なし)

提供を依頼する数量表が、チェック内容②及び③を満たすことを示すため、各セルにおいて最も大きく寄与する調査客体の値  $x_1$  がセルの値  $X$  に占める割合、及び一番目及び二番目に大きく寄与する調査客体の値  $x_1, x_2$  の合計値がセルの値  $X$  に占める割合を明らかにしてください。

提供を依頼する数量表が、表 24 のように集計用乗率を用いずに (加重なしで) 作成されているとします。

表 24 提供を依頼する数量表 (加重なし)

数量表 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	1075	807	1416	3298
区分B-2	1398	1343	1088	3829
区分B-3	1262	726	1009	2997
計	3735	2876	3513	10124

その場合は、まず表 25、表 26 のように、各セルにおける  $x_1, x_2$  の表を作成し、提示してください。



表 25 表 24 の各セルにおける $x_1$  (補足説明資料)

一番目に 大きい値 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	372	251	477	477
区分B-2	423	444	345	444
区分B-3	400	184	313	400
計	423	444	477	477

表 26 表 24 の各セルにおける $x_2$  (補足説明資料)

二番目に 大きい値 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	219	169	295	372
区分B-2	292	275	225	423
区分B-3	260	150	208	313
計	400	275	345	444

その後、 $x_1$ が $X$ に占める割合の表 (表 27)、及び $x_1 + x_2$ が $X$ に占める割合の表 (表 28) を作成し、提示してください。

表 27  $x_1$ が $X$ に占める割合の表 (補足説明資料)

一番目 占有率 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	34.6%	31.1%	33.7%	14.5%
区分B-2	30.3%	33.1%	31.7%	11.6%
区分B-3	31.7%	25.3%	31.0%	13.3%
計	11.3%	15.4%	13.6%	4.7%

表 28  $x_1 + x_2$ が $X$ に占める割合の表 (補足説明資料)

一番目+二番目 占有率 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	55.0%	52.0%	54.5%	25.7%
区分B-2	51.1%	53.5%	52.4%	22.6%
区分B-3	52.3%	46.0%	51.6%	23.8%
計	22.0%	25.0%	23.4%	9.1%

次に、提供を依頼する数量表が、表 29 のように集計用乗率を用いて (加重ありで) 作成

されているとします。

表 29 提供を依頼する数量表 (加重あり)

数量表 (加重あり)	区分A-1	区分A-2	区分A-3	計
区分B-1	1337.9	940.3	1875.1	4153.2
区分B-2	1661.1	1262.8	1386.6	4310.5
区分B-3	1358.0	1096.2	814.5	3268.7
計	4357.0	3299.3	4076.2	11732.5

その場合は、まず表 30、表 31 のように、各セルにおける加重なしの $x_1, x_2$ の表を作成し、提示してください。

表 30 表 29 の各セルにおける $x_1$  (補足説明資料)

一番目に 大きい値 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	372	251	477	477
区分B-2	423	444	345	444
区分B-3	400	184	313	400
計	423	444	477	477

表 31 表 29 の各セルにおける $x_2$  (補足説明資料)

二番目に 大きい値 (加重なし)	区分A-1	区分A-2	区分A-3	計
区分B-1	219	169	295	372
区分B-2	292	275	225	423
区分B-3	260	150	208	313
計	400	275	345	444

次に、 $x_1$ について、表 32 のように、集計用乗率 $w_1$ の表を作成し、提示してください。

表 32 各 $x_1$ に対する $w_1$  (補足説明資料)

一番目に 大きい値の 集計用乗率	区分A-1	区分A-2	区分A-3	計
区分B-1	1.511	1.674	1.998	1.735
区分B-2	1.888	1.176	1.440	1.987
区分B-3	1.294	1.715	1.172	1.127
計	1.719	1.223	1.446	1.394

次に、以下に従って、各セルにおける加重ありの一番目及び二番目に大きく寄与する調査客体の値 $\hat{x}_1, \hat{x}_2$ を計算します。

- 一番目に大きく寄与する調査客体の値 (加重あり) :  $\hat{x}_1 = x_1$
- 二番目に大きく寄与する調査客体の値 (加重あり) :  $w_1$ について、
  - $(w_1 - 1) \geq 1$ であれば、 $\hat{x}_2 = x_1$
  - $(w_1 - 1) < 1$ であれば、 $\hat{x}_2 = x_1 \times (w_1 - 1) + x_2 \times (1 - (w_1 - 1))$

これは、集計用乗率から、母集団における一番目及び二番目に大きく寄与する調査客体の推定値を得るためです。このようにして、表 33、表 34 のように、 $\hat{x}_1, \hat{x}_2$ の表を作成し、提示してください。

表 33 表 29 の各セルにおける $\hat{x}_1$  (補足説明資料)

一番目に 大きい値 (加重あり)	区分A-1	区分A-2	区分A-3	計
区分B-1	372	251	477	477
区分B-2	423	444	345	444
区分B-3	400	184	313	400
計	423	444	477	477

表 34 表 29 の各セルにおける $\hat{x}_2$  (補足説明資料)

二番目に 大きい値 (加重あり)	区分A-1	区分A-2	区分A-3	計
区分B-1	297.2	224.3	476.6	449.2
区分B-2	408.3	304.7	277.8	443.7
区分B-3	301.2	174.3	226.1	324.0
計	416.5	312.7	403.9	457.0

最後に、 $\hat{x}_1$ が $X$ に占める割合の表 (表 35)、及び $\hat{x}_1 + \hat{x}_2$ が $X$ に占める割合の表 (表 36) を

作成し、提示してください。

表 35 表 29 の各セルにおいて $\hat{x}_1$ がXに占める割合（補足説明資料）

一番目 占有率 (加重あり)	区分A-1	区分A-2	区分A-3	計
区分B-1	27.8%	26.7%	25.4%	11.5%
区分B-2	25.5%	35.2%	24.9%	10.3%
区分B-3	29.5%	16.8%	38.4%	12.2%
計	9.7%	13.5%	11.7%	4.1%

表 36 表 29 の各セルにおいて $\hat{x}_1 + \hat{x}_2$ がXに占める割合の表（補足説明資料）

一番目+二番目 占有率 (加重あり)	区分A-1	区分A-2	区分A-3	計
区分B-1	50.0%	50.5%	50.9%	22.3%
区分B-2	50.0%	59.3%	44.9%	20.6%
区分B-3	51.6%	32.7%	66.2%	22.2%
計	19.3%	22.9%	21.6%	8.0%

#### チェック内容

- ④ 【度数】 行計又は列計の90%超を占めるセルがないこと（加重なし）
- ⑤ 【度数】 行計又は列計の90%超を占めるセルがないこと（加重あり）

数量表（総和）を構成する属性から、他の属性の情報が高い確率で推測できてしまうことがあります。例えば、表 37 を作成したとします。

表 37 作成した数量表（加重なし）

研究開発費総和 (百万円)	産業1	産業2	産業3	産業4	計
地域1	83505	95013	76248	144704	399470
地域2	16315	3209	4853	4395	28772
地域3	78288	88792	104832	92300	364212
計	178108	187014	185933	241399	792454

同じ集計区分の度数表（表 38）を見ると、地域 2 において産業 1 の事業所は 325 社で、地域 2 の約 91.5%を占めています。

表 38 表 37 における各セルの度数（加重なし）

	産業1	産業2	産業3	産業4	計
地域1	150社	170社	240社	320社	880社
地域2	325社	10社	10社	10社	355社
地域3	210社	220社	210社	200社	840社
計	685社	400社	460社	530社	2075社

そのため、各セルの度数を知っている者は、ある事業所が地域2にあることが分かれば、その事業所が産業1であり、その産業の平均研究開発費が $16315 \div 325 = 50.2$ 百万円であることを高い確率で推測できてしまいます。このように、提供を依頼する数量表について、各セルの度数が行計又は列計の90%超を占めるセルがある場合は注意が必要です。また、集計乗率を用いて（加重ありで）作成された数量表で、度数が90%超を占めるセルがない場合でも、集計用乗率を用いずに（加重なしで）作成した場合に90%超を占めるセルがある場合もあり得ますし、その逆もあり得ますので、注意が必要です。

ただし、90%超を占めるセルがある場合でも、その数量表を公開するリスクが高いかどうかは、集計に用いた属性や、集計区分等によります。そのため、事前に提供を依頼する数量表について、集計内容、集計方法や結果表様式等を相談し、チェックの必要性についてお問合せください。

#### 申出者が提示する資料

- ④ 【度数】各セルの構成比（行計及び列計に占める割合）（加重なし）
- ⑤ 【度数】各セルの構成比（行計及び列計に占める割合）（加重あり）

事前に提供を依頼する数量表について相談し、確認の必要があると判断された場合は、求められた内容の補足説明資料を作成してください。集計内容、集計方法や結果表様式により、作成をお願いする資料の内容や様式が変化することがありますので、ご注意ください。例えば、三つの変数から作成した表であっても、そのうちの二つの変数に関する構成比表の作成をお願いすることもあり得ます。また、提供を依頼している表とは別の集計区分で構成比表の作成をお願いすることもあり得ます。

### 3. 数量表（平均、構成比、集中度）

#### チェック内容

- ① 同じ集計区分で総和の表が、「2.1 又は 2.2 数量表（総和）」のチェック内容に準拠する

こと

平均等、総和と度数から計算できる値による数量表の提供を依頼する場合、計算元となる総和の数量表が「2.1 数量表（総和）（個人・世帯調査の場合）」又は「2.2 数量表（総和）（事業所・企業調査の場合）」のチェック内容を満たす必要があります。

数量表作成に用いた調査票情報が、個人・世帯を対象とした調査のものか、事業所・企業を対象とした調査のものかにより、用いるチェック内容が異なりますので、ご注意ください。

#### 申出者が提示する資料

- ① 「2.1 又は 2.2 数量表（総和）」のチェック内容に準拠すること

提供を依頼する数量表について、同じ集計区分で総和の数量表を提示してください。また、その総和の数量表について、「2.1 数量表（総和）（個人・世帯調査の場合）」又は「2.2 数量表（総和）（事業所・企業調査の場合）」に沿って補足説明資料を提示してください。

数量表作成に用いた調査票情報が、個人・世帯を対象とした調査のものか、事業所・企業を対象とした調査のものかにより、提示する補足説明資料が異なりますので、ご注意ください。

## 4. その他の統計表

### チェック内容

- ① 各セルに含まれる統計量が該当する、「II 統計量」のチェック内容に準拠すること。

度数、総和やそれらから計算される値以外にも、様々な統計量から作成される統計表があり得ます。そのような統計表の提供を依頼する場合、標準的なチェック内容の「II 統計量」以降を参照し、各セルの値が該当する統計量のチェック内容を満たす必要があります。

### 申出者が提示する資料

- ① 該当する「II 統計量」のチェック内容に準拠すること。

提供を依頼する統計表について、各セルの値が該当する統計量のチェック内容を参照し、それに対応する補足説明資料を提示してください。

## II 統計量

### 5. 最頻値

#### チェック内容

- ① 【度数】1以上10未満の調査客体から算出した値でないこと（加重なし）

最頻値の算出元である調査客体の数（度数）が小さい場合、算出元の調査客体が特定されるリスクが高くなります。また、集計乗率を用いて（加重ありで）作成された値であっても、ある調査客体が調査に参加していることを知っている者がいると、この調査客体が含まれる値の度数が小さい場合、この調査客体が特定されるリスクが高くなります。

そのため、最頻値の提供を依頼する場合、その値の加重なしの度数が10以上である必要があります。ただし、度数0の統計量を作成することは通常考えられず、作成しても含まれる調査客体が存在せず特定は発生しないことから、度数0をチェック内容から除外しています。

#### 申出者が提示する資料

- ① 【度数】値の度数（加重なし）

提供を依頼する最頻値が、チェック内容①を満たすことを示すため、最頻値の加重なしの度数を明らかにしてください。加重ありの値であっても、加重なしの度数が必要となりますので、ご注意ください。

#### チェック内容

- ② 【度数】最頻値の値を持つ調査客体数が、算出に用いた調査客体数の90%超を占めないこと（加重なし）
- ③ 【度数】最頻値の値を持つ調査客体数が、算出に用いた調査客体数の90%超を占めないこと（加重あり）

最頻値の値が、算出に用いた調査客体の実際の値と、高い確率で一致してしまうことがあります。例えば、地域1に住む人の年収（10万円区分）の最頻値が100万円だったとします。もし、年収100万円の人が地域1の住人の90%超を占めていた場合、ある人が地域1に住んでいることが分かれば、その人の年収は高い確率で100万円となってしまいます。このように、提供を依頼する最頻値について、最頻値の値を持つ調査客体数が、算出に用いた

調査客体数の 90%超を占める場合は注意が必要です。また、集計乗率を用いて（加重ありで）作成された最頻値で、最頻値の値を持つ調査客体数が 90%超を占めない場合でも、集計用乗率を用いずに（加重なしで）作成した場合に 90%超を占める場合もあり得ますし、その逆もあり得ますので、注意が必要です。

ただし、最頻値の値を持つ調査客体数が 90%超を占める場合でも、その最頻値を公開するリスクが高いかどうかは、算出に用いた属性によります。そのため、事前に提供を依頼する最頻値について、作成方法を相談し、チェックの必要性についてお問合せください。

### 申出者が提示する資料

- ② 【度数】最頻値の値を持つ調査客体数が算出に用いた調査客体数に占める割合（加重なし）
- ③ 【度数】最頻値の値を持つ調査客体数が算出に用いた調査客体数に占める割合（加重あり）

事前に提供を依頼する最頻値について相談し、確認の必要があると判断された場合は、求められた内容の補足説明資料を作成してください。使用するデータにより、作成をお願いする資料の内容や様式が変化することがありますので、ご注意ください。

## 6. 総和、平均、構成比、集中度

ここでは、総和に対するチェック内容を基本として定めています。平均等、総和と度数から計算できる値の提供を依頼する場合、計算元となる総和がこのチェック基準を満たす必要があります。

公的統計調査には、大きく分けて個人・世帯を対象とした調査と、事業所・企業を対象とした調査の二つがあります。それぞれの調査により、チェック内容及び提示する補足説明資料が異なりますので、ご注意ください。

### 6.1 総和、平均、構成比、集中度（個人・世帯調査の場合）

#### チェック内容

- ① 【度数】1 以上 10 未満の調査客体から算出した値でないこと（加重なし）

総和等の算出元である調査客体の数（度数）が小さい場合、算出元の調査客体が特定されるリスクが高くなります。また、集計乗率を用いて（加重ありで）作成された値であっても、ある調査客体が調査に参加していることを知っている者がいると、この調査客体が含まれ



る値の度数が小さい場合、この調査客体が特定されるリスクが高くなります。

そのため、総和等の提供を依頼する場合、その値の加重なしの度数が10以上である必要があります。ただし、度数0の統計量を作成することは通常考えられず、作成しても含まれる調査客体が存在せず特定は発生しないことから、度数0をチェック内容から除外しています。

#### 申出者が提示する資料

##### ① 【度数】 値の度数 (加重なし)

提供を依頼する値が、チェック内容①を満たすことを示すため、総和等の加重なしの度数を明らかにしてください。加重ありの値であっても、加重なしの度数が必要となりますので、ご注意ください。

## 6.2 総和、平均、構成比、集中度 (事業所・企業調査の場合)

#### チェック内容

##### ① 【度数】 1以上10未満の調査客体から算出した値でないこと (加重なし)

総和等の算出元である調査客体の数(度数)が小さい場合、算出元の調査客体が特定されるリスクが高くなります。また、集計乗率を用いて(加重ありで)作成された値であっても、ある調査客体が調査に参加していることを知っている者がいると、この調査客体が含まれる値の度数が小さい場合、この調査客体が特定されるリスクが高くなります。

そのため、総和等の提供を依頼する場合、その値の加重なしの度数が10以上である必要があります。ただし、度数0の統計量を作成することは通常考えられず、作成しても含まれる調査客体が存在せず特定は発生しないことから、度数0をチェック内容から除外しています。

#### 申出者が提示する資料

##### ① 【度数】 値の度数 (加重なし)

提供を依頼する値が、チェック内容①を満たすことを示すため、総和等の加重なしの度数を明らかにしてください。加重ありの値であっても、加重なしの度数が必要となりますので、ご注意ください。

#### チェック内容

##### ② 【数量】 総和において、70%を超えて寄与する調査客体がないこと (変数変換なし)

セルの値に大きく寄与する調査客体が存在した場合、その調査客体の実際の値に対して、

セルの値が精度の高い推測値となってしまうことがあります。例えば、総和 210 の提供を依頼し、度数が 10 だったとします。

このとき、総和の内訳が表 39 のとおりであったとします。

表 39 各社の値

A社	B社	C社	D社	E社
200	2	1	1	1
F社	G社	H社	I社	J社
1	1	1	1	1

A 社の値 200 は、総和 210 の約 95% を占めています。そのため、算出元のデータの中で A 社が研究開発費に多く支出していることを知っている者が、セルの値 210 を A 社の研究開発費の近似値とした場合、実際の A 社の研究開発費 200 を小さな誤差で推測できていることになってしまいます。

一般に、総和  $X$  とそれに含まれる調査客体の値  $x_1$  に対し、 $x_1 > 0.7X$  であれば、 $X - x_1 < 0.3X$  となり、 $X$  は総和の 30% 以内の誤差で  $x_1$  を近似します。そのため、総和等の提供を依頼する場合、一つの調査客体が値の 70% を超えて寄与していない必要があります。

また、集計用乗率を用いて（加重ありで）作成した総和の場合は、加重ありの総和に対する寄与をチェックする必要があります。

その他、変数変換を行った場合、見かけの占有率が変化することがあります。例えば、総和 1000 に対し、それに含まれる最大の値が 500 だとすると、この値は総和の 50% を占めていることとなりますが、これを対数変換すると総和が  $\log 1000 = 6.91$ 、最大値が  $\log 500 = 6.21$  となり、見かけ上は総和の約 90% を占めることとなります。

また、例えば表 40 のような値を持つ 10 社があるとします。

表 40 各社の値

A社	B社	C社	D社	E社
1500	300	100	30	20
F社	G社	H社	I社	J社
10	10	10	10	10

このとき、総和 2000 に対し、最大の値 1500 は、総和の値の 75% を占めていることとなりますが、これらの値を対数変換すると、

表 41 対数変換後の各社の値

A社	B社	C社	D社	E社
7.313	5.704	4.605	3.401	2.996
F社	G社	H社	I社	J社
2.303	2.303	2.303	2.303	2.303

となり、総和 35.532 に対し、最大の値 7.313 は、見かけ上は総和の値の約 20.6%を占めていることとなります。

このように、変数変換により、正しい確認ができなくなることがありますので、変数変換された数値による総和等であっても、変数変換前の数値を用いてチェックする必要があります。

- ③ **【数量】** 総和において、85%を超えて寄与する二つの調査客体の合計値がないこと（変数変換なし）

セルの値に大きく寄与する調査客体が存在した場合、その調査客体の実際の値に対して、セルの値が精度の高い推測値となってしまうことがあります。例えば、総和 210 の提供を依頼し、度数が 10 だったとします。

このとき、総和の内訳が表 42 のとおりであったとします。

表 42 各社の値

A社	B社	C社	D社	E社
116	86	1	1	1
F社	G社	H社	I社	J社
1	1	1	1	1

A 社の値 116 と B 社の値 86 の合計 202 は、総和 210 の約 96%を占めています。そのため、算出元のデータの中で A 社が研究開発費に多く支出していることを B 社が知っているときに、セルの値 210 から B 社の値 86 を引いた 124 を A 社の研究開発費の近似値とした場合、B 社は実際の A 社の研究開発費 116 を小さな誤差で推測できていることになってしまいます。

一般に、総和 $X$ とそれに含まれる二つの調査客体の値 $x_1, x_2$ に対し、 $x_1 + x_2 > 0.85X$ であれば $(X - x_2) - x_1 < 0.15X$ となり、 $(X - x_2)$ は総和の 15%以内の誤差で $x_1$ を近似します。そのため、総和等の提供を依頼する場合、二つの調査客体の合計値が値の 85%を超えて寄与していない必要があります。

また、集計用乗率を用いて（加重ありで）作成した総和の場合は、集計用乗率を用いて、母集団における一番目及び二番目に大きく寄与する調査客体の値を推定し、加重ありの総

和に対する寄与をチェックする必要があります。

その他、変数変換を行った場合、見かけの占有率が変化することがあります。例えば、総和 1000 に対し、それに含まれる最大の値が 300、二番目の値が 200 だとすると、合計値 500 は総和の 50% を占めていることとなりますが、これを対数変換すると総和が  $\log 1000 = 6.91$ 、合計値が  $\log 500 = 6.21$  となり、見かけ上は総和の約 90% を占めることとなります。

また、例えば表 43 のような値を持つ 10 社があるとします。

表 43 各社の値

A社	B社	C社	D社	E社
1500	300	100	30	20
F社	G社	H社	I社	J社
10	10	10	10	10

このとき、総和 2000 に対し、最大の値 1500 と二番目の値 300 の合計値 1800 は、総和の 90% を占めていることとなりますが、これらの値を対数変換すると、

表 44 対数変換後の各社の値

A社	B社	C社	D社	E社
7.313	5.704	4.605	3.401	2.996
F社	G社	H社	I社	J社
2.303	2.303	2.303	2.303	2.303

となり、総和 35.532 に対し、最大の値 7.313 と二番目の値 5.704 の合計値 13.017 は、見かけ上は総和の値の約 36.6% を占めていることとなります。

このように、変数変換により、正しい確認ができなくなることがありますので、変数変換された数値による総和等であっても、変数変換前の数値を用いてチェックする必要があります。

#### 申出者が提示する資料

- ② 【数量】総和  $X$ 、最も大きく寄与する調査客体の値  $x_1$  及び  $x_1$  が  $X$  に占める割合
- ③ 【数量】総和  $X$ 、一番目及び二番目に大きく寄与する調査客体の値  $x_1, x_2$  及び  $x_1, x_2$  の合計値が  $X$  に占める割合

提供を依頼する値が、チェック内容②及び③を満たすことを示すため、総和  $X$  に対して最も大きく寄与する調査客体の値  $x_1$  が  $X$  に占める割合、及び一番目及び二番目に大きく寄与する調査客体の値  $x_1, x_2$  の合計値が  $X$  に占める割合を明らかにしてください。

提供を依頼する総和（加重なし）が 1075 だったとします。その場合は、まず総和におけ

る $x_1, x_2$ を提示してください。例えば、

$$x_1 = 372, x_2 = 219$$

だったとします。その後、 $x_1$ が $X$ に占める割合、及び $x_1 + x_2$ が $X$ に占める割合を提示してください。ここでは、一番目の占有率は、

$$x_1 \div X = 372 \div 1075 = 34.6\%$$

一番目と二番目の合計値の占有率は、

$$(x_1 + x_2) \div X = 591 \div 1075 = 55.0\%$$

となります。

次に、提供を依頼する総和が、集計用乗率を用いて（加重ありで）作成されているとします。例えば、加重ありの総和 $\hat{X}$ が1337.9だったとします。その場合は、まず加重なしの総和における加重なしの $x_1, x_2$ を提示してください。例えば、

$$x_1 = 372, x_2 = 219$$

だったとします。次に、 $x_1$ について集計用乗率 $w_1$ を提示してください。例えば、

$$w_1 = 1.511$$

だったとします。次に、以下に従って、各セルにおける加重ありの一番目及び二番目に大きく寄与する調査客体の値 $\hat{x}_1, \hat{x}_2$ を計算し、提示してください。

- 一番目に大きく寄与する調査客体の値（加重あり）： $\hat{x}_1 = x_1$
- 二番目に大きく寄与する調査客体の値（加重あり）： $w_1$ について、
  - $(w_1 - 1) \geq 1$ であれば、 $\hat{x}_2 = x_1$
  - $(w_1 - 1) < 1$ であれば、 $\hat{x}_2 = x_1 \times (w_1 - 1) + x_2 \times (1 - (w_1 - 1))$

これは、集計用乗率から、母集団における一番目及び二番目に大きく寄与する調査客体の推定値を得るためです。ここでは、

$$\begin{aligned}\hat{x}_1 &= 372, \\ \hat{x}_2 &= x_1 \times (w_1 - 1) + x_2 \times (1 - (w_1 - 1)) \\ &= 372 \times (1.511 - 1) + 219 \times (1 - (1.511 - 1)) \\ &= 372 \times 0.511 + 219 \times 0.489 \\ &= 297.2\end{aligned}$$

となります。最後に、 $\hat{x}_1$ が $\hat{X}$ に占める割合、及び $\hat{x}_1 + \hat{x}_2$ が $\hat{X}$ に占める割合を計算し、提示してください。ここでは、一番目の占有率は、

$$\hat{x}_1 \div \hat{X} = 372 \div 1337.9 = 27.8\%$$

一番目と二番目の合計値の占有率は、

$$(\hat{x}_1 + \hat{x}_2) \div \hat{X} = 669.2 \div 1337.9 = 50.0\%$$

となります。

## 7. 線形回帰係数、非線形回帰係数

### チェック内容

#### ① 【自由度】残差の自由度が10以上であること

回帰モデルの残差（残差平方和）の自由度が小さい場合、各調査客体の推定残差が特定されるリスクが高くなります。推定残差は線形回帰式などの統計モデルと実際の値の差であることから、これが特定された場合、統計モデルと組み合わせることで実際の調査客体の値が特定できてしまいます。

そのため、線形回帰係数又は非線形回帰係数の提供を依頼する場合、回帰モデルの残差の自由度が10以上である必要があります。

一般的に、自由度は以下の式で計算できます。

$$\text{自由度} = \text{使用したデータのサイズ} - \text{説明変数の数} - 1$$

### 申出者が提示する資料

#### ① 【自由度】残差の自由度

提供を依頼する線形回帰係数又は非線形回帰係数が、チェック内容①を満たすことを示すため、残差の自由度を提示してください。

例えば、Excelで線形回帰分析を行った場合、表45のように、結果概要の一部として残差の自由度が出力されますので、この結果概要を提示してください。

表 45 Excel による線形回帰分析結果

分散分析表					
	自由度	変動	分散	観測された分散比	有意 F
回帰	3	87.72292842	29.24097614	295.539138	8.5881E-62
残差	146	14.44540491	0.09894113		
合計	149	102.1683333			

その他の統計ソフトの場合でも、同様に残差の自由度を示す結果概要を提示するか、結果概要内でデータサイズや目的変数の数が確認できる場所を示し、自由度を計算して提示してください。

#### チェック内容

- ② **【データ】** 一つの調査客体のみに関するデータから作成していないこと（例：一つの調査客体のみに関する時系列データから作成していない）

一つの調査客体のみに関するデータであっても、長期間の時系列データなど、見かけのデータサイズを増やして回帰分析を行った場合、見かけの自由度がチェック内容①を満たしていたとしても、その調査客体の情報が漏洩するリスクが高くなります。そのため、一つの調査客体のみに関するデータを用いて回帰分析を行わないよう、ご注意ください。

#### 申出者が提示する資料

- ② **【データ】** 目的変数及び説明変数の説明資料

提供を依頼する回帰係数が、チェック内容②を満たすことを示すため、目的変数及び説明変数の説明を提示してください。

また、使用したデータの説明もチェックのために必要ですので、提供を依頼する様式の所定の欄に詳細を記載してください。

## 8. 分布の高次モーメント、相関係数、要約統計量及び検定統計量

#### チェック内容

- ① **【自由度】** 自由度が 10 以上であること

分布の高次モーメント、相関係数、要約統計量又は検定統計量が公開されると、その統計量は算出元のデータの制約条件となります。統計量の自由度が小さい場合、制約条件から算出元のデータが特定されるリスクが高くなります。そのため、分布の高次モーメント、相関係数、要約統計量及び検定統計量の提供を依頼する場合、自由度が 10 以上である必要があ

ります。

## 申出者が提示する資料

### ① 【自由度】 自由度

提供を依頼する値が、チェック内容①を満たすことを示すため、分布の高次モーメント、相関係数、要約統計量又は検定統計量の自由度を提示してください。

## 9. 最大値、最小値

### ① 持ち出し不可

最大値や最小値は、通常ただ一つの調査客体の値です。ある調査客体について、その値が不明であっても、最大値又は最小値であることが推測される場合、その値が特定されるリスクが高くなります。例えば、ある地域内で一つだけ規模が大きい工場があった場合、その売上高は、その地域の工場の中で最大であろうことは容易に推測できるため、最大値の公表が値の特定につながります。そのため、原則としてこれらの提供を受けることはできません。

## 10. 推定残差

### ① 持ち出し不可

推定残差は線形回帰式などの統計モデルと実際の値の差であることから、統計モデルと組み合わせることで実際の調査客体の値が特定できてしまいます。そのため、推定残差の提供を受けることはできません。

## 11. グラフ

### ① 持ち出し不可

グラフについては、通常は提供が承諾された分析結果等から作成することを想定しており、原則としてグラフそのものの提供を受けることはできません。すなわち、グラフの元データの提供後に、オンサイト施設外にてグラフを作成することになります。そのため、グラフの元データを適切なチェック内容を用いてチェックし、提供を依頼してください。



### III その他

#### 12. 分析プログラム、操作ログ等

##### チェック内容

##### ① 統計表又は統計量が含まれていないこと

分析プログラムや操作ログ等、調査票情報から作成されていない成果物については、他の事項でチェックすべき統計表や統計量が含まれていなければ提供を受けることができます。そのため、以下の例のような記述はしないでください。

- 平均値以上のデータを抽出する処理において、平均値の値を直接条件式内に記述すること
- 条件に合致するデータを抽出する処理のコメントとして、抽出されたデータの度数や構成比を記述すること
- 外れ値を除く処理のコメントとして、除いた個別データの詳細を記述すること

## 統計表における秘匿措置

標準的なチェック内容の「表2 統計表における秘匿措置（以下「チェック内容表2」という。）」では、チェック内容表1を満たさない分析結果について、提供を受けるために申出者が講ずるべき秘匿措置を例示しています。

### 1. 集計区分の変更

#### 秘匿措置

各セルに集計される区分を変更して再度集計を行い、表1の内容を満たすようにすること。集計区分の変更方法には、既存の区分の分割、他の区分と統合、新たな区分の設定がある。

例として、「他の区分と結合」とは、下記のような方法になります。

表46では、職種2の年齢45～49歳のセルの値が5となっており、チェック内容表1の「I 統計表」「1. 度数表、度数の構成比表」「①【度数】各セルが1以上10未満の調査客体から算出した値でないこと（加重なし）」を満たしていません。

表46 チェック内容を満たさない度数表

年齢	...	35～39	40～44	45～49	50～54	55～60	...
職種1	...	20	30	15	20	10	...
職種2	...	15	20	5	20	20	...
職種3	...	25	20	10	10	25	...
合計	...	60	70	30	50	55	...

このとき、例えば、45～49歳の区分を50～54歳の区分と結合し、45～54歳として集計すると、チェック内容表1を満たす表47になります。

表 47 表 46 の集計区分を結合することでチェック内容を満たした度数表

年齢	…	35～39	40～44	45～54	55～60	…
職種 1	…	20	30	35	10	…
職種 2	…	15	20	25	20	…
職種 3	…	25	20	20	25	…
合計	…	60	70	80	55	…

このように、集計区分の結合などによって、チェック内容表 1 を満たすように提供を受ける分析結果を変更する方法が集計区分の変更になります。新たに作成された統計表がチェック内容表 1 を満たすことを確認し、提供を依頼してください。

## 2. 集計対象の変更

### 秘匿措置

集計対象の範囲を拡大又は縮小して再度集計を行い、表 1 の内容を満たすようにすること。

(例：集計対象が、あるグループ X に属する調査客体のみから作成した統計表の場合、

- ①グループ Y に属する調査客体を集計対象に加えて新たな統計表を作成する (拡大)。
- ②グループ X に属する調査客体のうち、他の調査客体から大きく外れた値を持つ調査客体などを除外して新たな統計表を作成する (縮小)。

例として、①の「拡大」とは、下記のような方法になります。

A、B、C 市の 3 市を対象に集計した表 48 では、職種 2 の年齢 45～49 歳のセルの値が 5 となっており、チェック内容表 1 の「I 統計表」「1. 度数表、度数の構成比表」「①【度数】各セルが 1 以上 10 未満の調査客体から算出した値でないこと (加重なし)」を満たしていません。

表 48 チェック内容を満たさない度数表

A, B, C 市	…	35～39	40～44	<b>44～49</b>	50～54	55～60	…
職種 1	…	20	30	15	20	10	…
<b>職種 2</b>	…	15	20	<b>5</b>	20	20	…
職種 3	…	25	20	10	10	25	…
合計	…	60	70	30	50	55	…

このとき、チェック内容表 1 を満たすように、例えば、A、B、C 市に隣接する D 市を集計対象に加え、A、B、C、D 市の 4 市を対象として集計すると、チェック内容表 1 を満たす表 49 になります。

表 49 表 48 の集計対象を拡大することでチェック内容を満たした度数表

A, B, C, D 市	…	35～39	40～44	<b>44～49</b>	50～54	55～60	…
職種 1	…	30	45	20	25	15	…
<b>職種 2</b>	…	20	30	<b>10</b>	30	30	…
職種 3	…	35	30	15	15	40	…
合計	…	85	105	45	70	85	…

このように、集計対象の拡大などによって、チェック内容表 1 を満たすように提供を受ける分析結果を変更する方法が集計対象の変更になります。新たに作成された統計表がチェック内容表 1 を満たすことを確認し、提供を依頼してください。

### 3. セルの値を秘匿

#### 秘匿措置

##### ① 一次秘匿

表 1 の内容を満たさないセルの値を「X」などのマークに置き換え、具体的な値を掲載しないようにすること。

表 50 では、職種 2 の年齢 65 歳～のセルの値が 8 となっており、チェック内容表 1 の「I 統計表」「1. 度数表、度数の構成比表」「①【度数】各セルが 1 以上 10 未満の調査客体か

ら算出した値でないこと（加重なし）」を満たしていません。

表 50 チェック内容を満たさない度数表（秘匿前）

度数表	～39 歳	40～64 歳	65 歳～	合計
職種 1	20	24	28	72
職種 2	38	38	8	84
職種 3	40	39	42	121
合計	98	101	78	277

このセルを、「×」で置き換え、以下のように一次秘匿を行った表 51 を作成します。

表 51 表 50 に一次秘匿を行った結果

度数表	～39 歳	40～64 歳	65 歳～	合計
職種 1	20	24	28	72
職種 2	38	38	×	84
職種 3	40	39	42	121
合計	98	101	78	277

この表では、職種 2 の行の合計値 84 から「～39 歳」、「40～64 歳」の値 38,38 を引くことで「×」の値が 8 だと計算できてしまいます。同様に「65 歳～」の列の合計からも「×」が計算できてしまいます。

そこで、次の「② 二次秘匿」を行います。

## ② 二次秘匿

一次秘匿を行ったセルの値が他のセルの値や行計・列計等から算出できる場合、これを算出できないように一次秘匿の対象となるセル以外のセルの値を「X」などのマークに置き換え、具体的な値を掲載しないようにすること。

上記の表 51 のように、「① 一次秘匿」を行っただけではセルの値が算出できてしまうため、追加で他のセルを「×」で置き換え二次秘匿を行います。

表 52 表 51 に二次秘匿を行った結果

度数表	～39 歳	40～64 歳	65 歳～	合計
職種 1	×	24	×	72
職種 2	×	38	×	84
職種 3	40	39	42	121
合計	98	101	78	277

二次秘匿は、一次秘匿したセルの値が算出できないように行う必要があります。例えば、 $x$  の位置を二次秘匿した表 53 の場合、太線で囲った「職種 1」行と「職種 2」行の合計 156 から、秘匿されていないセルの値 28 を引くことで、秘匿された 5 セルの合計が 128 であることがわかります。一方、点線で囲った「～39 歳」列と「40～64 歳」列の合計 199 から、秘匿されていないセルの値 40 と 39 を引くことで、秘匿された 4 セルの合計が 120 であることがわかり、差分から一次秘匿したセル  $x$  の値は 8 と算出できます。

表 53 誤った二次秘匿を行った結果

度数表	～39 歳	40～64 歳	65 歳～	合計
職種 1	x	x	28	72
職種 2	x	x	X	84
職種 3	40	39	x	121
合計	98	101	78	277

このように、一見秘匿できていても、実際は一次秘匿したセルの値が算出できてしまうこともあり得ます。そのため、二次秘匿後には一次秘匿したセルの値が算出できないことを確認してください。

二次秘匿により、一次秘匿したセルの値は正確には分からなくなります。しかし、場合によっては一次秘匿したセルの値がかなり狭い範囲で推測できてしまうことがあります。そこで、以下のとおり秘匿インターバルを計算し、一次秘匿したセルの値を相当の範囲でしか推測できないことを確認します。

③ 秘匿インターバル（度数表の場合）

一次秘匿した各セルが取り得る値の最大と最小の差（秘匿インターバル）が度数 10 以上であること。

④ 秘匿インターバル（数量表の場合）

一次秘匿した各セルが取り得る値の最大と最小の差（秘匿インターバル）が、当該セル値の 30% 以上であること。

上述の一次秘匿、二次秘匿を行った度数表について秘匿した各セルの値を表 54 のように、 $x_{11}$ 、 $x_{13}$ 、 $x_{21}$ 、 $x_{23}$  とします。

表 54 秘匿された度数表

度数表	～39 歳	40～64 歳	65 歳～	合計
職種 1	$x_{11}$	24	$x_{13}$	72
職種 2	$x_{21}$	38	$x_{23}$	84
職種 3	40	39	42	121
合計	98	101	78	277

このとき、行計と列計の値から以下の関係式が成り立ちます。

$$\begin{aligned} x_{11} + x_{13} &= 48 \\ x_{21} + x_{23} &= 46 \\ x_{11} + x_{21} &= 58 \\ x_{13} + x_{23} &= 36 \end{aligned}$$

この関係式から、一次秘匿した $x_{23}$ の値の範囲が次のようになることが分かります。また、それに伴い二次秘匿した $x_{11}$ 、 $x_{13}$ 、 $x_{21}$ の値も計算できます。表 55 のとおり、 $x_{23}$ の最小値は 0 となります。

表 55  $x_{23}$ の最小値

度数表	～39 歳	40～64 歳	65 歳～	合計
職種 1	12	24	36	72
職種 2	46	38	0	84
職種 3	40	39	42	121
合計	98	101	78	277

また、表 56 のとおり  $x_{23}$  の最大値は 36 となります。

表 56  $x_{23}$ の最大値

度数表	～39 歳	40～64 歳	65 歳～	合計
職種 1	48	24	0	72
職種 2	10	38	36	84
職種 3	40	39	42	121
合計	98	101	78	277

以上から、秘匿インターバル（一次秘匿したセルが取り得る値の最大と最小の差）は、

$$36 - 0 = 36 \geq 10$$

となり、秘匿インターバルが 10 以上である条件を満たしています。これで、この二次秘匿した表が標準的なチェック内容を満たすことの確認ができました。

なお、秘匿インターバルについての条件を満たしていない場合、二次秘匿するセルを変更する、二次秘匿するセルを追加するなどして、秘匿インターバルについての条件を満たすようにする必要があります。

### 申出者が提示する資料

#### ① 秘匿前の統計表

申出者が行った秘匿の内容について確認するため、秘匿前の統計表を提示してください。上記の例では、「表 50 チェック内容を満たさない度数表（秘匿前）」になります。



② 一次秘匿した各セルの位置を明示する情報

二次秘匿した表からでは、チェック内容表1を満たさなかったセルである一次秘匿したセルの場所が分からないため、一次秘匿したセルの位置を明示する情報を提示してください。上記の例では、提供を受けようとする「職種2」・「65歳～」のセルが一次秘匿したセルであることを、秘匿前の統計表上などで示してください。

③ (度数表の場合)

一次秘匿した各セルが取り得る最大値、最小値及び最大値と最小値の差

④ (数量表の場合)

一次秘匿した各セルが取り得る最大値、最小値及び最大値と最小値の差を当該セル値で除した割合

秘匿インターバルが十分な大きさであることを確認するために、一次秘匿した各セルの最大値、最小値、最小値と最小値の差を提示してください。上記の例では、一次秘匿した職種2、65歳～のセルが取り得る最大値が「36」、最小値が「0」、最大値と最小値の差が「36」であることを提示してください。

## チェックの具体例

### 【事例 1】

ある地域の 500 企業をグループ A～C の 3 グループに分け、3 グループそれぞれと全体について三つの変数の平均値等を計算し、以下の分析結果を作成した。

様々な統計量が含まれている表のため、この分析結果の提供を依頼するにあたっては、各項目についてチェック内容に照らし合わせてチェックしていく必要がある。

なお、この分析結果は全数調査の調査票情報を用いて作成したものである。

グループ A (企業数 465)	平均値	標準偏差	最頻値	最小値	最大値
変数 1	100	25	110	40	160
変数 2	150	10	140	115	180
変数 3	20	18	15	0	80

グループ B (企業数 15)	平均値	標準偏差	最頻値	最小値	最大値
変数 1	90	30	95	45	145
変数 2	150	15	130	120	180
変数 3	25	20	16	0	50

グループ C (企業数 20)	平均値	標準偏差	最頻値	最小値	最大値
変数 1	120	25	120	50	160
変数 2	160	20	150	135	210
変数 3	18	18	18	0	130

合計 (企業数 500)	平均値	標準偏差	最頻値	最小値	最大値
変数 1	101	25	110	40	160
変数 2	150	11	140	115	210
変数 3	20	18	15	0	130

## 平均値のチェック

平均値については、チェック内容表 1 の「3.数量表（平均、構成比、集中度）」を確認する。「同じ集計区分の総和の表が、「2.1 又は 2.2 数量表（総和）」のチェック内容を満たすこと。」とあるので、事業所・企業調査の場合は「2.2」の内容を確認する。

この分析結果は全数調査の調査票情報を用いて作成されたものであるから、集計用の乗率は存在しないので、チェック内容にある「(加重あり)」「(加重なし)」の条件は考慮しなくてよい。また、変数変換により作成されたものでもないので、「(変数変換なし)」の条件も考慮しなくてよい。

集計対象客体数は、各表の企業数が最低でも 15 あることから、「① 【度数】各セルが 1 以上 10 未満の調査客体から算出した値でないこと」を満たす。

さらに、事業所・企業調査の場合、「② 各セルにおいて、70%を超えて寄与する調査客体がないこと」「③ 各セルにおいて、85%を超えて寄与する二つの調査客体の合計値がないこと」を満たすことを確認するため、例えば以下のような補足説明資料を作成する必要がある（※）。この例の場合、1 位の調査客体の値の占有率が 70%を、1 位の客体と 2 位の客体の合計値の占有率が 85%を超えていないので、チェック事項を満たす。

総和	グループ A	グループ B	グループ C	合計
変数 1	46500	1350	2400	50250
変数 2	69750	2250	3200	75200
変数 3	9300	375	360	10035

1 位 (占有率(%))	グループ A	グループ B	グループ C	合計
変数 1	160 (0.34)	145 (10.74)	160 (6.67)	160 (0.32)
変数 2	180 (0.26)	180 (8.00)	210 (6.56)	210 (0.28)
変数 3	80 (0.86)	50 (13.33)	130 (36.11)	130 (1.30)

1位+2位 (占有率(%))	グループA	グループB	グループC	合計
変数1	318 (0.68)	285 (21.11)	290 (12.08)	320 (0.64)
変数2	355 (0.51)	340 (15.11)	395 (12.34)	395 (0.53)
変数3	145 (1.56)	90 (24.00)	190 (52.78)	210 (2.09)

(※) 各グループの2位の値は分析結果からは読み取れないので、元データである調査票情報からこの表を作成する必要がある。

「行計又は列計の90%超を占めるセルがないこと」については、「集計対象客体数」を参照のこと。

#### 標準偏差

標準偏差については、チェック内容表1「4.その他の統計表」を確認する。標準偏差は「8.分布の高次モーメント、相関係数、要約統計量及び検定統計量」に該当する。集計対象客体数が最低でも15あるので、「自由度が10以上あること」を満たす。

#### 最頻値

最頻値については、チェック内容表1「4.その他の統計表」を確認する。「5.最頻値」のチェック内容を満たすか、この分析結果だけでは判断できないので、「値の度数及びその構成比」(最頻値の値を取る調査客体の割合など)の補足説明資料を作成する。同資料により、チェック内容を満たしていないことが確認された場合は、最頻値の項目自体を削除するか、持ち出しを希望するならば提供依頼書の該当欄にて安全性の説明が必要となる。

#### 最小値、最大値

最小値、最大値については、チェック内容表1「4.その他の統計表」を確認する。「9.最大値、最小値」にあるとおり、原則「提供不可」である。最小値、最大値の項目自体を削除するか、持ち出しを希望するならば提供依頼書の該当欄にて安全性の説明が必要となる。

#### 集計対象客体数

集計対象客体数は度数であり、3グループと合計の企業数は、実際には以下の度数表を持ち出すことと同じであるため、「1.度数表、度数の構成比表」を確認する。

必須のチェック事項ではないが、集計内容等を勘案し、安全性の確認を求められた場合に

は、「行計又は列計の 90%超を占めるセルがないこと」のチェックが必要となる。

	グループ A	グループ B	グループ C	合計
客体数	465	15	20	500
構成比 (%)	93	3	4	100

この例では、グループ A の客体数は、「行計又は列計の 90%超を占めるセルがないこと」のチェック内容を満たしていない。そのため、集計客体数と合計の表を落とす（※）などの何らかの秘匿処置をするか、持ち出しを希望するならば提供依頼書の該当欄にて安全性の説明が必要となる。

（※）この例のような場合、集計客体数のみを落としても、グループ A の客体数を  $a$ 、グループ B の客体数を  $b$ 、グループ C の客体数を  $c$  とすると、表の平均値の値から、

$$100a + 90b + 120c = 101(a + b + c)$$

$$150a + 150b + 160c = 150(a + b + c)$$

$$20a + 25b + 18c = 20(a + b + c)$$

という三つの関係式が成り立ち、各グループの度数の構成比が分かってしまう（グループ A が 90%超を占めることが分かってしまう）可能性がある。

（この例の場合端数処理しているため、正確な値は分からない。）

## 【事例 2】

ある研究者は、性別・年収別の度数表を公開したいと考え、以下の内容を論文に掲載した。「性別・年収別の表は以下のとおりである。男性の行と女性の行にそれぞれ 1 以上 10 未満のセルがあったため、100～199 万円のセルと 200～299 万円のセルを合算した（秘匿表 1）。ただし、男女計のみの表は 1 以上 10 未満のセルがないため、合算せず記載する（秘匿表 2）。」

秘匿表 1

	100～299 万円	300 万円以上	計
男性	44	63	107
女性	12	58	70
男女計	56	121	177

秘匿表 2

	100～199 万円	200～299 万円	300 万円以上	計
男女計	20	36	121	177

このとき、秘匿前の表が下記の方法で復元できてしまう。

秘匿表 1 及び秘匿表 2 から、以下の表の  $a$ 、 $b$ 、 $c$ 、 $d$  の値を特定できれば、秘匿前の表が復元できたことになる。

	100～199 万円	200～299 万円	計
男性	$a$	$b$	44
女性	$c$	$d$	12
男女計	20	36	56

前提条件として、論文の記載から、少なくとも  $a$  か  $b$  のどちらかが 1 以上 10 未満であり、少なくとも  $c$  か  $d$  のどちらかが 1 以上 10 未満であることがわかる。

$c$  が 1 以上 10 未満すなわち  $1 \leq c \leq 9$  だと仮定すると、 $a + c = 20$  から  $11 \leq a \leq 19$  となる。

$a$  が 10 以上のため、前提条件から、 $b$  が 1 以上 10 未満すなわち  $1 \leq b \leq 9$  ということになるが、 $12 \leq a + b \leq 28$  となり、 $a + b = 44$  と矛盾する。

すなわち、 $c$  は 10 未満ではない。

$c$ が10以上なので、前提条件から $d$ が1以上10未満すなわち $1 \leq d \leq 9$ となり、 $c + d = 12$ から、 $10 \leq c \leq 11$ である。

すると、 $a + c = 20$ から $9 \leq a \leq 10$ となり、 $a + b = 44$ から $34 \leq b \leq 35$ となる。 $b$ が10以上であるため、前提条件から $a$ は10未満となるが、 $9 \leq a \leq 10$ から、 $a = 9$ である。

$a = 9$ から計算して、秘匿前の表は以下のとおりである。

	100～199万円	200～299万円	300万円以上	計
男性	9	35	63	107
女性	11	1	58	70
男女計	20	36	121	177

この事例は、

- ① 一つの表に異なる秘匿措置を講じて作成した複数の表
- ② 一次秘匿対象であるセルの位置に関する情報や秘匿理由に関する情報

を公開することによる危険性を示している。

- ① この問題では、本来一つの表である性別・年収階級別集計結果を、100～199万円と200～299万円の2列を合算して秘匿した表1と、年収階級はそのまま男性と女性の2行を合算して秘匿した表2を公開しているが、行の合算と列の合算という、異なる方法で秘匿した2表を組み合わせることで、秘匿が破られている。
- ② この問題では、男女行それぞれに秘匿すべきセルがあったことや、その秘匿理由が度数10未満であったことを公開することで、秘匿が破られている。

このような公表をしないよう、十分な注意が必要である。

また、公開したい情報が残るようにするためには、事前に秘匿方法について打合せを行い、一つの秘匿結果だけの公表となるようにすることが重要である。

## 参考 オンサイト利用における分析結果等の提供に関する標準的なチェック内容

以下の表1及び表2は、オンサイト利用における分析結果等の提供に関する標準的なチェック内容を定めたものである。分析結果等が調査票情報の全部又は一部を含む場合は、提供を依頼することはできない。申出者は、分析結果等の提供を依頼するに当たり、原則として、当該結果等が表1の内容を満たすこと、表1. Iの内容を満たさない統計表の場合には、表2に例示する方法その他の方法により秘匿措置を講ずることが求められる。ここで、分析結果等が「内容を満たす」とは、当該結果等が該当する表1の「チェック内容」に記載する全ての事項を満たす場合をいう。

### 【用語の解説・留意事項】

(表1)

#### \*1 分析結果等の種類

提供を依頼する分析結果等のことをいい、I統計表、II統計量、IIIその他に分けられる。

#### \*2 チェック内容

提供を依頼する分析結果等をチェックする事項を定めたもの。

#### \*3 加重なし

提供を依頼する分析結果等が、集計用の乗率等を乗じて作成したものであっても、乗率等を乗じない値でチェックする事項

#### \*4 加重あり

提供を依頼する分析結果等が、集計用の乗率等を乗じないで作成したものであっても、乗率等を乗じた値でチェックする事項

#### \*5 変数変換なし

提供を依頼する分析結果等が、変数変換（対数変換、累乗根変換など）し作成したものであっても、変数変換前の値でチェックする事項

(表1、表2共通)

#### \*6 四角数字（□、▣など）の事項

当該チェック事項が必須ではなく、チェック・情報提示を求められる場合のある事項（該当するか否かは事前に要相談）。

#### \*7 申出者が提示する情報

申出者が提供を依頼する際に、当該結果等のチェックの確認のために添付する情報。この外、分析結果等の概要、使用したデータ（申出者自身が新たに作成した指標・変数等のデータを含む）、原変数及び申出者自身が作成した変数の説明等、求められた情報を提示すること。



表1 標準的なチェック内容

	分析結果等の種類 <sup>*1</sup>	チェック内容 <sup>*2</sup>	申出者が提示する情報 <sup>*7</sup>
統計表	1.度数表、度数の構成比表	①【度数】各セルが1以上10未満の調査客体から算出した値でないこと（加重なし <sup>*3</sup> ）	①【度数】各セルの度数（加重なし <sup>*3</sup> ）
		② <sup>*6</sup> 【度数】行計又は列計の90%超を占めるセルがないこと（加重なし <sup>*3</sup> ）	② <sup>*6</sup> 【度数】各セルの構成比（行計及び列計に占める割合）（加重なし <sup>*3</sup> ）
		③ <sup>*6</sup> 【度数】行計又は列計の90%超を占めるセルがないこと（加重あり <sup>*4</sup> ）	③ <sup>*6</sup> 【度数】各セルの構成比（行計及び列計に占める割合）（加重あり <sup>*4</sup> ）
	2.1 数量表（総和）（個人・世帯調査の場合）	①【度数】各セルが1以上10未満の調査客体から算出した値でないこと（加重なし <sup>*3</sup> ）	①【度数】各セルの度数（加重なし <sup>*3</sup> ）
		② <sup>*6</sup> 【度数】行計又は列計の90%超を占めるセルがないこと（加重なし <sup>*3</sup> ）	② <sup>*6</sup> 【度数】各セルの構成比（行計及び列計に占める割合）（加重なし <sup>*3</sup> ）
		③ <sup>*6</sup> 【度数】行計又は列計の90%超を占めるセルがないこと（加重あり <sup>*4</sup> ）	③ <sup>*6</sup> 【度数】各セルの構成比（行計及び列計に占める割合）（加重あり <sup>*4</sup> ）
	2.2 数量表（総和）（事業所・企業調査の場合）	①【度数】各セルが1以上10未満の調査客体から算出した値でないこと（加重なし <sup>*3</sup> ）	①【度数】各セルの度数（加重なし <sup>*3</sup> ）
	②【数量】各セルにおいて、70%を超えて寄与する調査客体がないこと（変数変換なし <sup>*5</sup> ）	②【数量】各セルにおいて最も大きく寄与する調査客体の値 $x_1$ 及び $x_1$ がセルの値 $X$ に占める割合（変数変換なし <sup>*5</sup> ）	
	③【数量】各セルにおいて、85%を超えて寄与する二つの調査客体の合計値がないこと（変数変換なし <sup>*5</sup> ）	③【数量】各セルにおいて一番目及び二番目に大きく寄与する調査客体の値 $x_1, x_2$ 及び $x_1, x_2$ の合計値がセルの値 $X$ に占める割合（変数変換なし <sup>*5</sup> ）	
	④ <sup>*6</sup> 【度数】行計又は列計の90%超を占めるセルがないこと（加重なし <sup>*3</sup> ）	④ <sup>*6</sup> 【度数】各セルの構成比（行計及び列計に占める割合）（加重なし <sup>*3</sup> ）	
	⑤ <sup>*6</sup> 【度数】行計又は列計の90%超を占めるセルがないこと（加重あり <sup>*4</sup> ）	⑤ <sup>*6</sup> 【度数】各セルの構成比（行計及び列計に占める割合）（加重あり <sup>*4</sup> ）	

	3.数量表（平均、構成比、集中度）	①同じ集計区分の総和の表が、「2.1 又は 2.2 数量表（総和）」のチェック内容を満たすこと。	①「2.1 又は 2.2 数量表（総和）」のチェック内容に準拠すること
	4.その他の統計表	①各セルに含まれる統計量が該当する、「Ⅱ統計量」のチェック内容に準拠すること。	①該当する「Ⅱ統計量」のチェック内容に準拠すること。
Ⅱ 統計量	5.最頻値	①【度数】1以上10未満の調査客体から算出した値でないこと（加重なし <sup>*3</sup> ） ② <sup>*6</sup> 【度数】最頻値の値を持つ調査客体数が、算出に用いた調査客体数の90%超を占めないこと（加重なし <sup>*3</sup> ） ③ <sup>*6</sup> 【度数】最頻値の値を持つ調査客体数が、算出に用いた調査客体数の90%超を占めないこと（加重あり <sup>*4</sup> ）	①【度数】値の度数（加重なし <sup>*3</sup> ） ② <sup>*6</sup> 【度数】最頻値の値を持つ調査客体数が算出に用いた調査客体数に占める割合（加重なし <sup>*3</sup> ） ③ <sup>*6</sup> 【度数】最頻値の値を持つ調査客体数が算出に用いた調査客体数に占める割合（加重あり <sup>*4</sup> ）
	6.1 総和、平均、構成比、集中度（個人・世帯調査の場合）	①【度数】1以上10未満の調査客体から算出した値でないこと（加重なし <sup>*3</sup> ）	①【度数】値の度数（加重なし <sup>*3</sup> ）
	6.2 総和、平均、構成比、集中度（事業所・企業調査の場合）	①【度数】1以上10未満の調査客体から算出した値でないこと（加重なし <sup>*3</sup> ） ②【数量】総和において、70%を超えて寄与する調査客体がないこと（変数変換なし <sup>*5</sup> ） ③【数量】総和において、85%を超えて寄与する二つの調査客体の合計値がないこと（変数変換なし <sup>*5</sup> ）	①【度数】値の度数（加重なし <sup>*3</sup> ） ②【数量】総和 $X$ 、最も大きく寄与する調査客体の値 $x_1$ 及び $x_1$ が $X$ に占める割合（変数変換なし <sup>*5</sup> ） ③【数量】総和 $X$ 、一番目及び二番目に大きく寄与する調査客体の値 $x_1, x_2$ 及び $x_1, x_2$ の合計値が $X$ に占める割合（変数変換なし <sup>*5</sup> ）
	7.線形回帰係数、非線形回帰係数	①【自由度】残差の自由度が10以上であること ②【データ】一つの調査客体のみに関するデータから作成していないこと（例：一つの調査客体のみに関する時系列データから作成していない）	①【自由度】残差の自由度 ②【データ】目的変数及び説明変数の説明資料

	8.分布の高次モーメント <sup>1</sup> 、相関係数、要約統計量及び検定統計量 <sup>2</sup>	①【自由度】自由度が10以上であること	①【自由度】自由度
	9.最大値、最小値	①持ち出し不可（通常ただ一つの調査客体を指していることから）	
	10.推定残差	①持ち出し不可（観測値を推定できるリスクが大きいことから）	
	11.グラフ	①持ち出し不可（通常承諾された他の分析結果から作成できることから）	
Ⅲ その 他	12.分析プログラム、操作ログ等	①統計表又は統計量が含まれていないこと	

<sup>1</sup> 分散、歪度、尖度等

<sup>2</sup> 決定係数、変動係数、分散、情報量規準、t検定、F検定、 $\chi^2$ 検定、Wald検定、Hausman検定等

表2 統計表における秘匿措置

秘匿方法	秘匿措置	
1.集計区分の変更	各セルに集計される区分を変更して再度集計を行い、表1の内容を満たすようにすること。 集計区分の変更方法には、既存の区分の分割、他の区分と統合、新たな区分の設定がある。	
2.集計対象の変更	集計対象の範囲を拡大又は縮小して再度集計を行い、表1の内容を満たすようにすること。 (例：集計対象が、あるグループXに属する調査客体のみから作成した統計表の場合、 ①グループYに属する調査客体を集計対象に加えて新たな統計表を作成する(拡大)。 ②グループXに属する調査客体のうち、他の調査客体から大きく外れた値を持つ調査客体などを除外して新たな統計表を作成する(縮小。))	
3.セルの値を秘匿	<p style="text-align: center;">秘匿措置</p> <p>以下の一次秘匿、二次秘匿、秘匿インターバルの各内容を満たすようにすること。</p> <p>①一次秘匿 表1の内容を満たさないセルの値を「X」などのマークに置き換え、具体的な値を掲載しないようにすること。</p> <p>②二次秘匿 一次秘匿を行ったセルの値が他のセルの値や行計・列計等から算出できる場合、これを算出できないように一次秘匿の対象となるセル以外のセルの値を「X」などのマークに置き換え、具体的な値を掲載しないようにすること。</p> <p>③<sup>*6</sup>秘匿インターバル(度数表の場合) 一次秘匿した各セルが取り得る値の最大と最小の差(秘匿インターバル)が度数10以上であること。</p> <p>④<sup>*6</sup>秘匿インターバル(数量表の場合) 一次秘匿した各セルが取り得る値の最大と最小の差(秘匿インターバル)が、当該セル値の30%以上であること。</p>	<p style="text-align: center;">申出者が提示する情報<sup>*7</sup></p> <p>①秘匿前の統計表</p> <p>②一次秘匿した各セルの位置を明示する情報</p> <p>③(度数表の場合) 一次秘匿した各セルが取り得る最大値、最小値及び最大値と最小値の差</p> <p>④(数量表の場合) 一次秘匿した各セルが取り得る最大値、最小値及び最大値と最小値の差を当該セル値で除した割合</p>