



独立行政法人
統計センター

統計データの利活用推進に向けた 統計センターの取組み

—統計的開示抑制のための標準的なチェック内容—

信頼に應えてつくる確かな統計

2022年度統計関連学会連合大会

独立行政法人統計センター
阿部 穂日
統計数理研究所、統計センター
南 和宏

- 統計センターが提供する二次的利用サービス
 - 統計センターの紹介
 - 二次的利用制度の概要
- 標準的なチェック内容の検討
 - 標準的なチェック内容の紹介
 - 中央値・四分位値のチェック内容の検討

本発表資料は、以下に掲載しています

統計センター 学会発表



<https://www.nstac.go.jp/technology/research/society/>

統計センターの紹介



統計利用者・調査対象者



国際機関・外国政府

- 公的統計利用サービスの提供
- オンライン調査利用環境の提供

- 統計作成の技術協力
- 国際会議への参画
- LISとの協定締結
(日本の研究者は無料利用)



学界

- 参画・研究発表
- 大学・研究機関との連携



各府省・地方公共団体

- 統計作成支援
- 業務基盤サービスの提供



調査票情報の二次的利用とは・・・

○一次利用

- 各統計調査の実施前に定めた目的に沿った利用

○二次的利用

- 統計調査により集められた情報を
既存の調査結果のほかに、秘密の保護を図った上で
新たな統計作成や統計的手法を利用した
学術研究等のために活用するもの

● 調査票情報の提供

➤ 電磁的記録媒体による提供

- ・ 調査実施者である行政機関等が電磁的記録媒体により必要な範囲において調査票情報を提供するもの

➤ オンサイト利用

- ・ 統計センターと連携する大学等機関に設置されたオンサイト施設において、調査票情報を利用するもの

● 委託による統計の作成等（オーダーメイド集計）

- ・ 一般からの委託に応じ、統計調査の調査票情報を利用して、統計の作成又は統計的研究を行うもの

● 匿名データの提供

- ・ 一般の利用に供することを目的として調査票情報を特定の個人又は法人その他の団体の識別ができないように加工したデータ（匿名データ）を提供するもの



mi ripo

マイクロデータ利用ポータルサイト



検索

ホーム

マイクロデータ利用

利用実績

お問い合わせ

サイトマップ

統計調査結果 をより 広く活用 いただくための

mi ripo

マイクロデータ 利用のための ポータルサイト

<https://www.e-stat.go.jp/microdata/>

- ミクロデータの利用に関する制度の概要
- 各府省の利用可能なデータ一覧
- 具体的な手続方法
- ミクロデータの利用実績等

を公開

- ・公的統計の主要項目のデータを市区町村別、都道府県・時系列等で提供しているデータセット

<https://www.nstac.go.jp/use/literacy/ssdse/>



SSDSE-Aの解説

SSDSEの解説

国・都道府県 2021	prefecture	municipality	人口				外国人 人口	外国人 比率	出生数	死亡数	自然増減
			A1101	A110101	A110102	A1102					
			年度	2015	2015	2015					
地域コード	都道府県	市区町村	総人口	総人口 (男)	総人口 (女)	日本人					
R01100	北海道	札幌市	1952356	910614	1041742	1937785		5485	287	26613	
R01202	北海道	函館市	265979	120376	145603	264537		700	49	3524	
R01203	北海道	小樽市	121924	54985	66939	121415		348	22	1073	
R01204	北海道	旭川市	339605	156402	183203	335678		879	67	57159	
R01205	北海道	室蘭市	88564	43143	45421	88203		195	10	10395	
R01206	北海道	釧路市	174742	82185	92557	174120		382	26	10717	
R01207	北海道	帯広市	169327	80994	88333	168800		415	31	2774	
R01208	北海道	北見市	121226	58020	63206	120936		220	25	1466	
R47360	沖縄県	伊集り村	1517	821	696	1496		0	1	56	
R47361	沖縄県	久米島町	7755	4086	3669	7721		9	4	23	
R47362	沖縄県	八重瀬町	29064	14247	14817	28915		673	14	1439	
R47375	沖縄県	多良間村	1194	639	555	1180		0	1	23	
R47381	沖縄県	竹富町	3998	2057	1941	3960		2	0	0	
R47382	沖縄県	与那国町	1843	1147	696	1832		0	1	28	



- 結果表から作成した疑似的な個票形式のマイクロデータ

<https://www.nstac.go.jp/use/archives/ippan-microdata/>

調査票情報の新たな提供方法（オンサイト利用）

従来：電磁的記録媒体による提供

- 利用条件** 高度な公益性を有する研究目的
- セキュリティ** セキュリティの確保全般について **研究者**が対応
- 申請内容** 分析内容として、作成する統計表等すべてをあらかじめ具体的・詳細に記述（事前に統計表等の安全性を確認）
- 利用データ** 調査票情報のうち、分析に必要な **最小限の調査事項**に限った提供



探索的・創造的研究が
困難



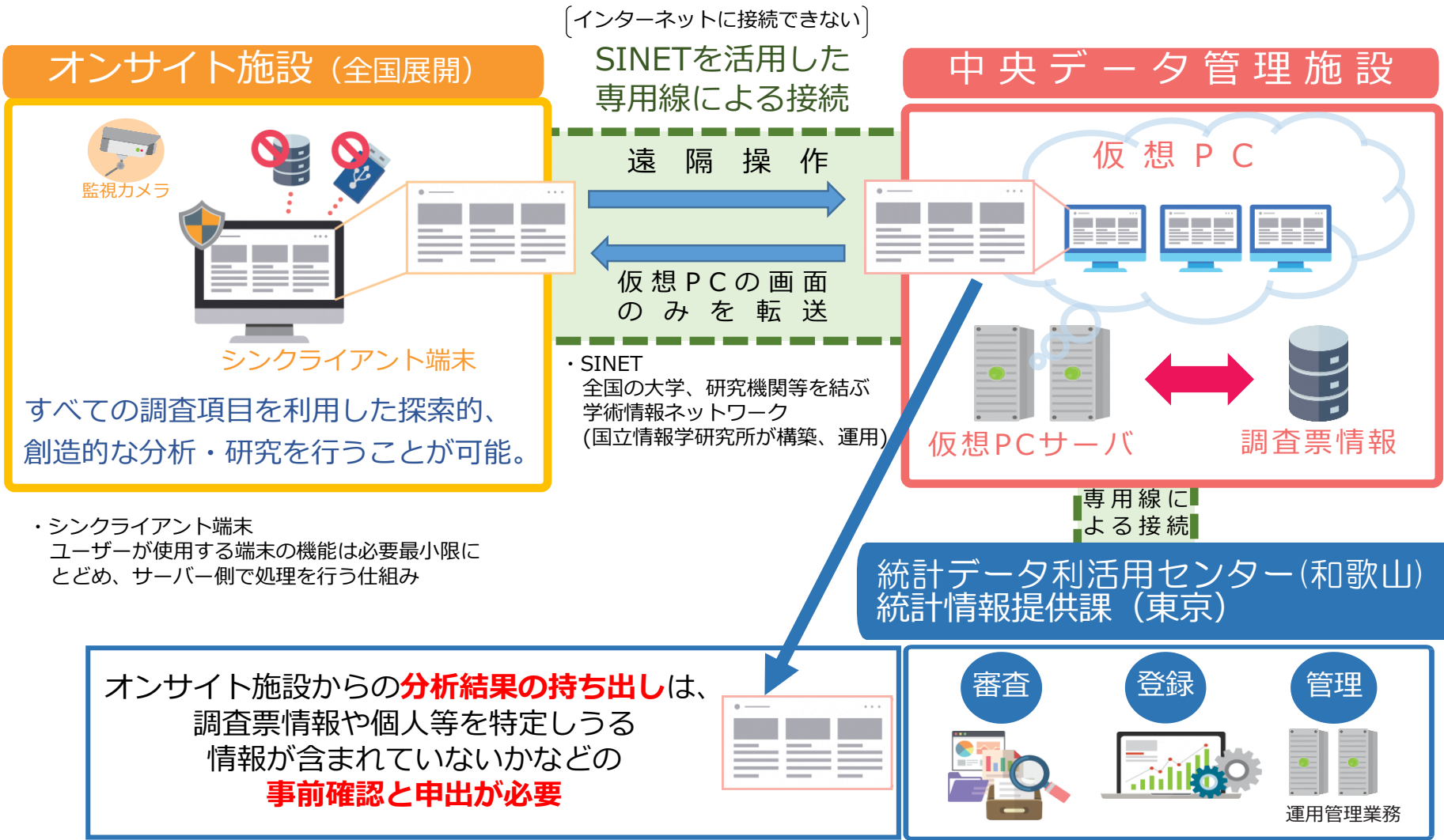
現在：オンサイト利用も可能

- 利用条件** 高度又は相当の公益性を有する学術研究・教育目的
- セキュリティ** 施設のセキュリティ確保は **施設設置者**が対応
- 申請内容** 分析の“概要”を記述（事後的に、統計表等の作成後に安全性を確認（持ち出し審査））
- 利用データ** 調査票情報の **全ての調査事項**を利用



探索的・創造的研究が
可能

調査票情報のオンライン利用の概要

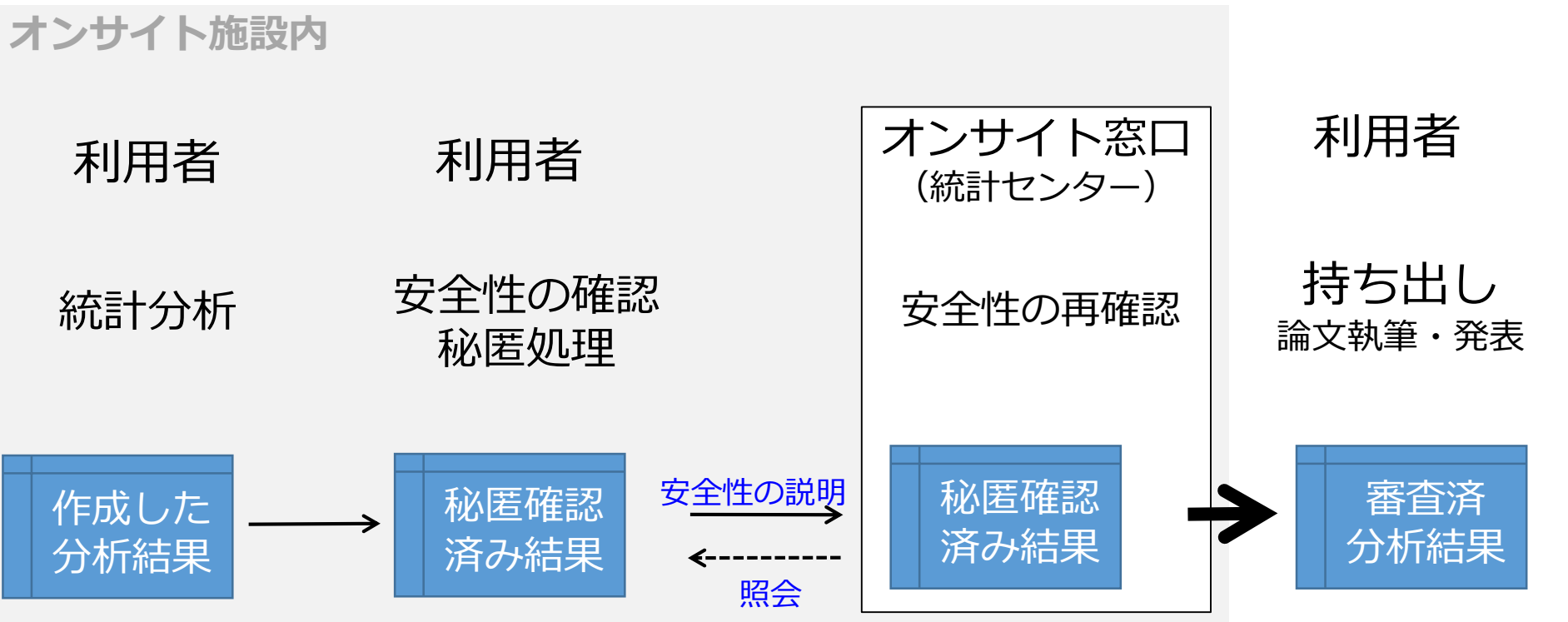


オンサイト施設地図（2022年9月1日現在）



分析結果等の持ち出し

<手順の流れ>



参照すべきマニュアル

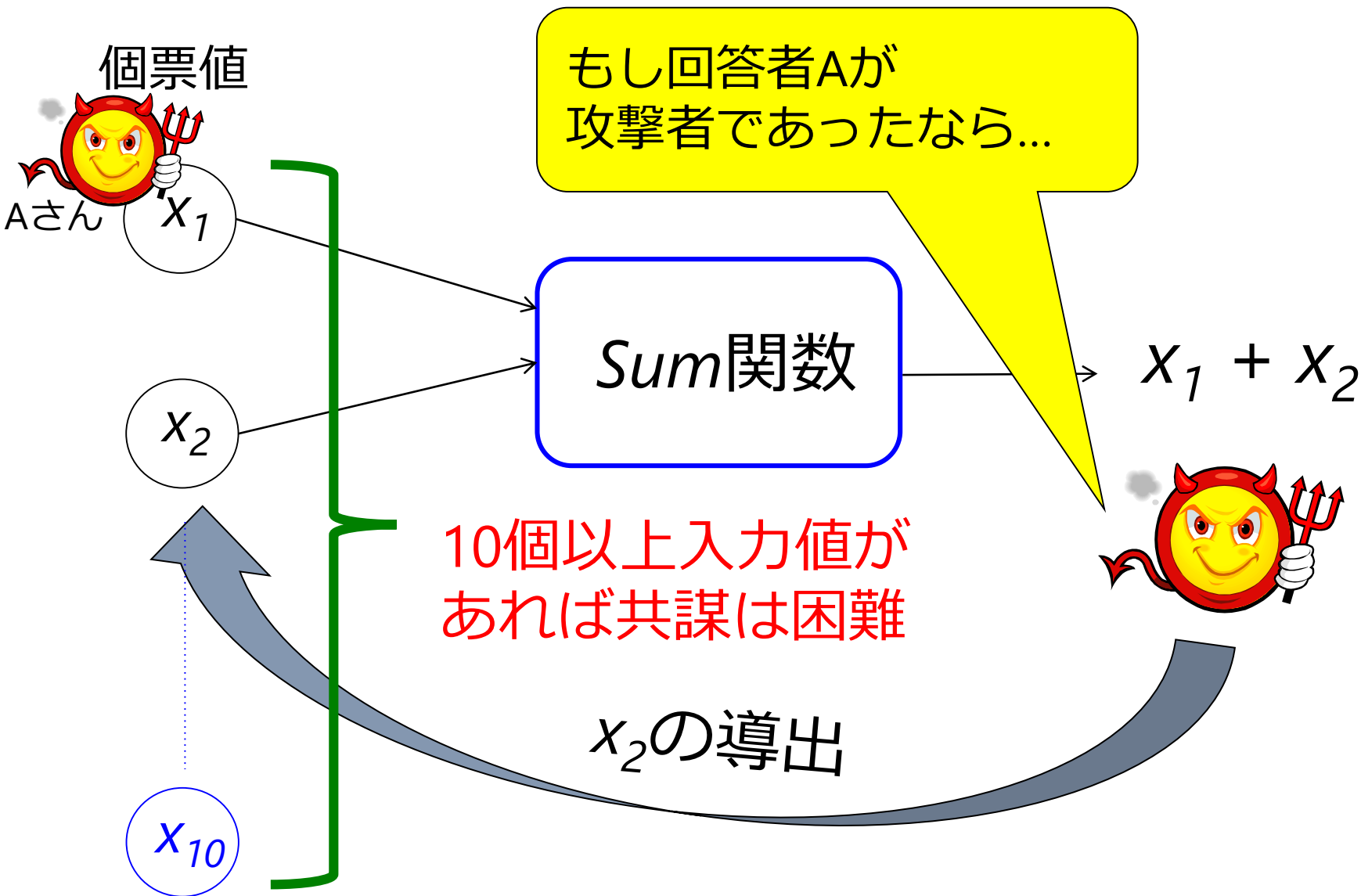
**標準的な
チェック内容**

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

miripoに掲載



客体10の原則 (標準的なチェック内容の例。諸外国も同様)



- 中央値・四分位値については、
記述統計等で広く利用されている
- ある調査客体の値そのまま
又は2つの値から求められる数値であり
客体10の原則を満たさない
- 我が国におけるチェック内容の参考とした
諸外国の書籍・資料においても
中央値・四分位値の確認方法はなかった
- 一方で、一般に全ての調査客体の
順位を正確に知ることは困難
- 中央値・四分位値の安全性の確認方法を検討



Project N°: 262608



ACRONYM: **Data** without **Boundaries**

STANDALONE DOCUMENT
Guidelines for Output Checking

WORK PACKAGE 11
(Improved Methodologies for Managing Risks of Access to Detailed OS Data)

Combination of CP & CSA project funded by the European Community
Under the programme "FP7 - SP4 Capacities"
Priority 1.1.3: European Social Science Data Archives and remote access to Official Statistics

Eurostat (2014, August). Guidelines for the checking of output based on microdata research.

https://ec.europa.eu/eurostat/cros/system/files/dwb_standalone-document_output-checking-guidelines.pdf 14

P.17

- 確認者が考慮すべきは
公表された値から回答者を特定できるかどうか
順位が既知又は推測可能であれば
パーセンタイルは公表できない
- パーセンタイル付近の値の分散が小さい場合
グループ開示の可能性
- パーセンタイル付近の分散が非常に大きい場合
回答者が識別できる可能性
(例えば、2峰性の分布で
中央値の回答者が1人しかない場合)

- 確認者が考慮すべきは
公表された値から回答者を特定できるかどうか
順位が既知又は推測可能であれば
パーセンタイルは公表できない
 - 上位の順位が既知又は推測可能はあり得るが
一般に算出元データが十分大きい場合
中央値・四分位値付近の調査客体の
順位を正確に知ることは困難

- パーセンタイル付近の値の分散が小さい場合
グループ開示の可能性
→ 総和・平均等のチェック内容では、
センシティブな変数等 **個別の事例ごとに確認**

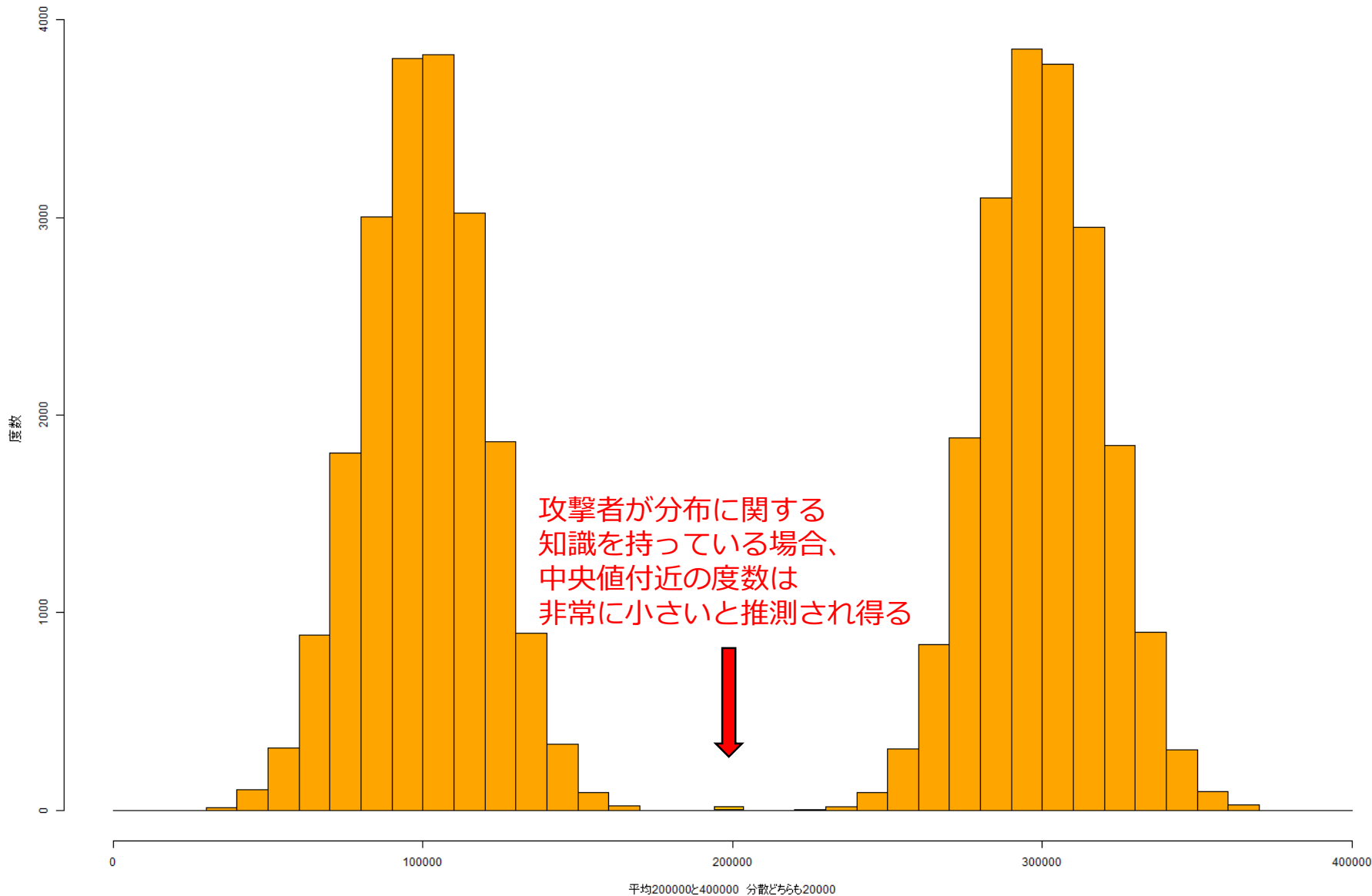
地域別・収入別度数表におけるグループ開示の例

	0～100万円	100～200万円	200～300万円	300万円～	合計
地域 1	20人	20人	30人	25人	95人
地域 2	125人	5人	3人	0人	133人
地域 3	30人	30人	30人	43人	133
合計	175人	55人	63人	68人	361

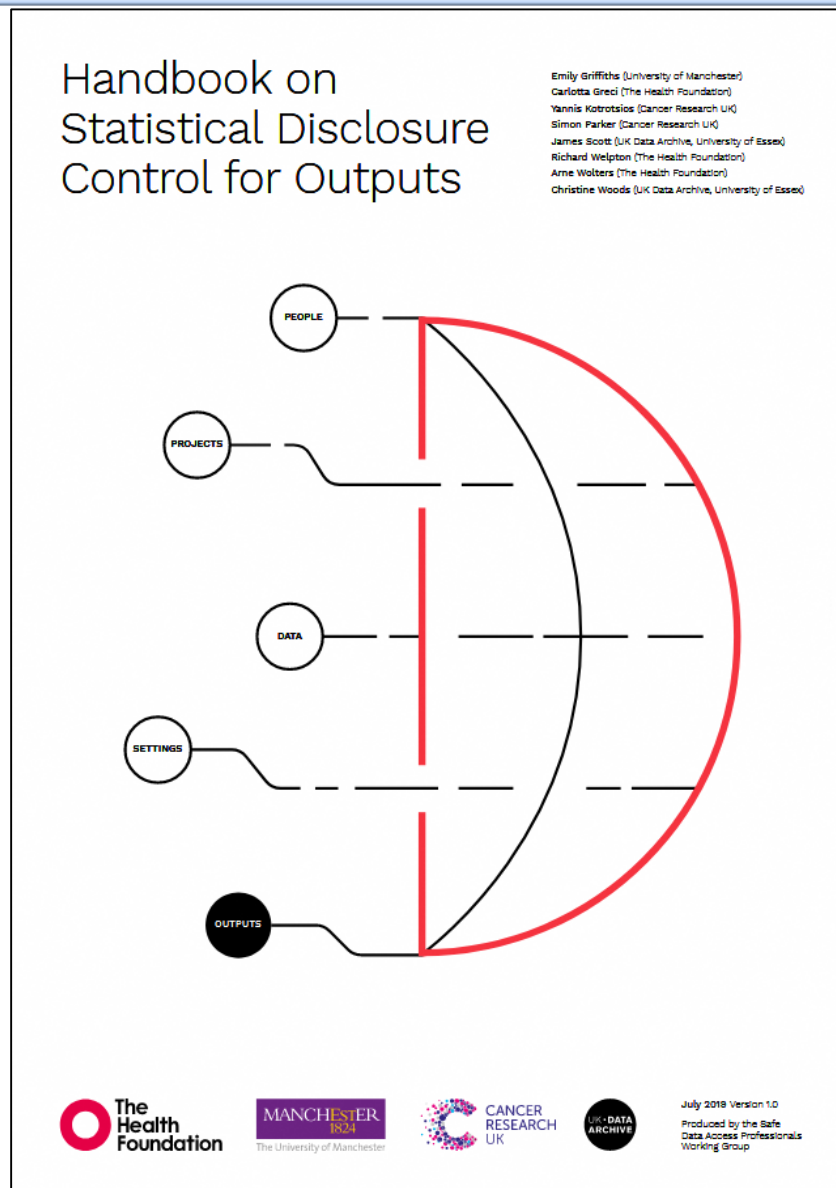
「地域2」の人なら収入階級が「最低」と高い確率で推測できる。

Eurostatの事例（2峰性分布の例）（4/5）

パーセンタイル付近の分散が非常に大きい場合については検討が必要



UK Data Serviceの事例 (1/2)



UK Data Service (2019, July). Handbook on Statistical Disclosure Control for Outputs.

https://ukdataservice.ac.uk/app/uploads/thf_datareport_aw_web.pdf

「丸め秘匿」

P.33-34

- 中央値・四分位値ごとに、
丸めた値と同じ丸め値を持つ回答者の度数が10以上となるまで
その値と個別の回答者の値を丸める桁数を増やしていく

真値

	第1四分位	中央値	第3四分位
算出した値	3804.9	5503.7	7983.6

丸めた値
(公表値)

	第1四分位	中央値	第3四分位
秘匿した値	3800	5504	7980

丸めた値と
同じ丸め値を持つ
調査客体の度数

度数	62	10	35
----	----	----	----

「丸め秘匿」を自動処理するソースコードの例（Rによる）



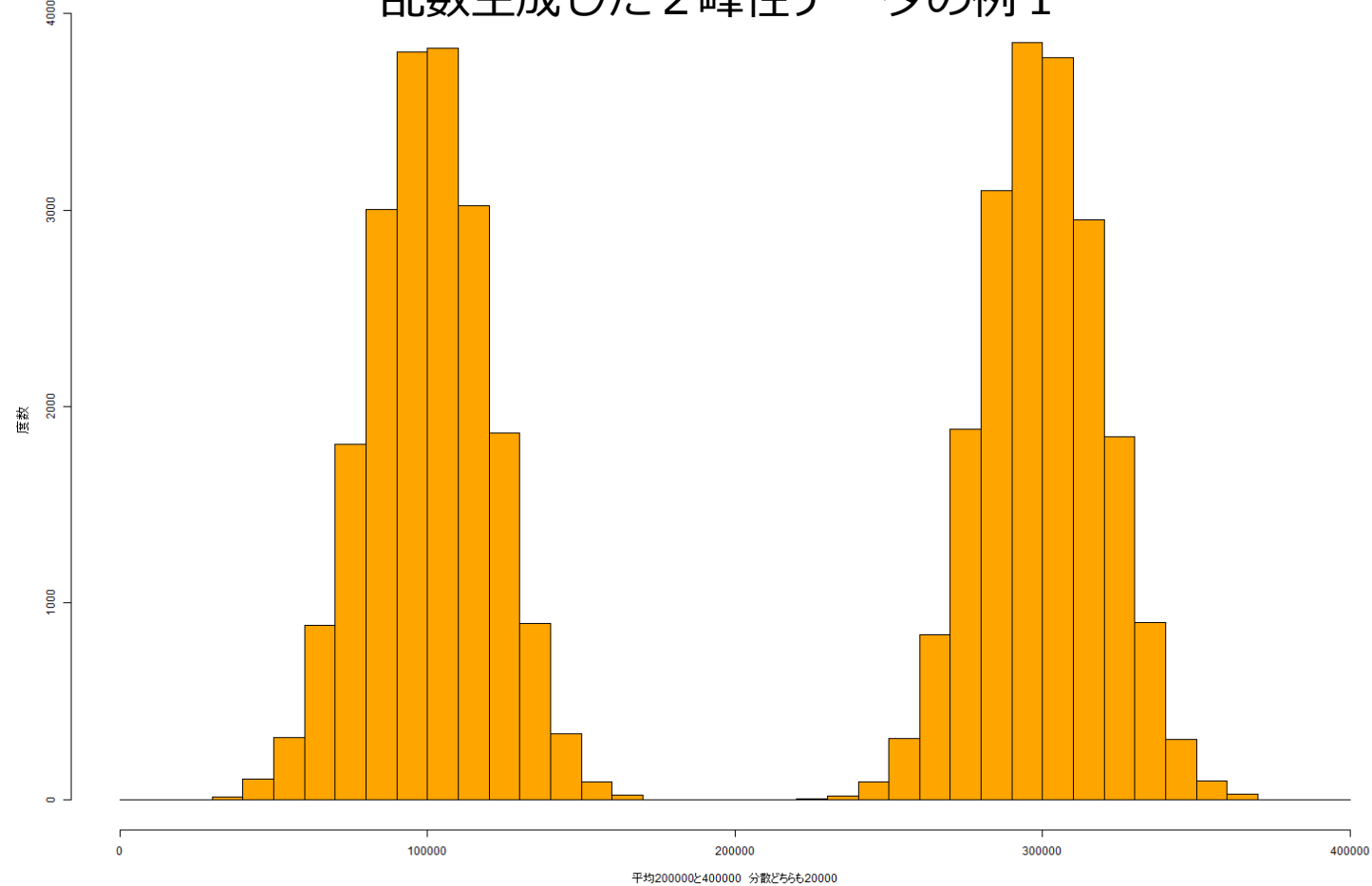
```
# 丸めると秘匿結果と同じ値になる度数の初期値
n <- 0
# 丸める桁数の初期値
k <- 0
# 丸めることができる桁数の最大値
k.max <- floor(log10(quartile))
# quartileが0だった場合はk.maxを0にする
if (is.infinite(k.max)) {
  k.max <- 0
}

# 同じ値になる度数が10以上になるまでループ
while (n < 10) {
  # k桁で丸めた四分位数の計算
  Q <- round(quartile, digits = -k)
  # k桁で丸めるとQと同じ値になる度数の計算
  n <- length(which(round(data, digits = -k) == Q))
  # 丸める桁数を増やす
  k <- k + 1
  # 上限まで丸めてもnが10未満ならQをNAにして終了
  if (n < 10 && k > k.max) {
    Q <- NA
    break
  }
}
```

Q : 秘匿結果

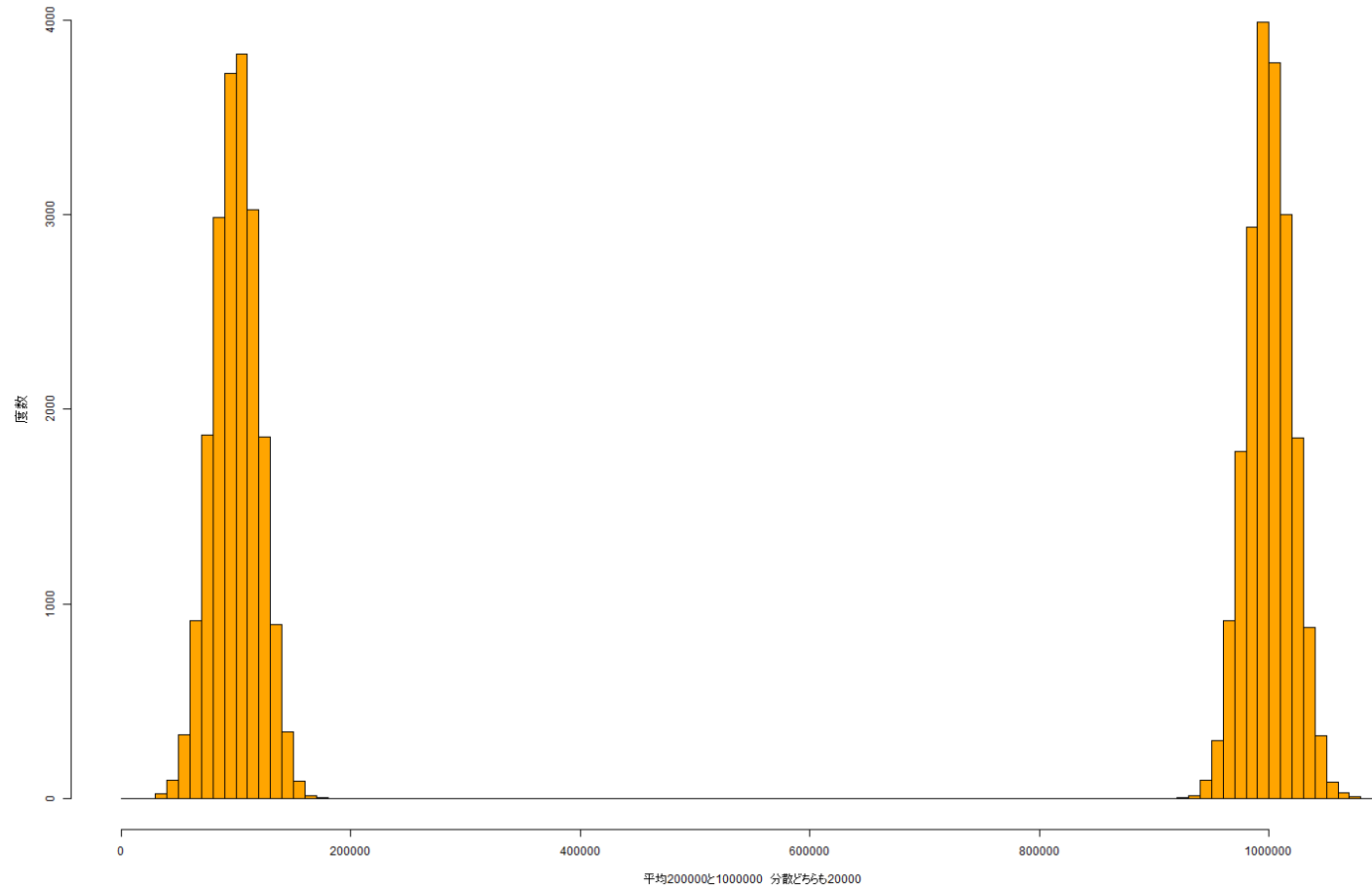
n : 丸めた中央値・四分位値と同じ丸め値を持つデータの度数

乱数生成した 2 峰性データの例 1



	第 1 四分位	中央値	第 3 四分位
真値	100178.4	196825.0	299731.9
丸めた値	100200	200000	299700
同じ丸め値の度数	32	228	36

乱数生成した 2 峰性データの例 2



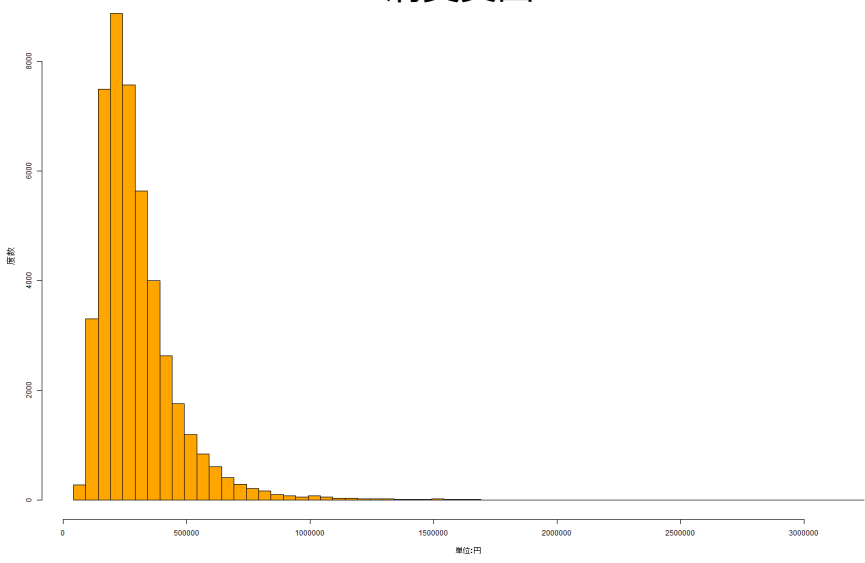
	第 1 四分位	中央値	第 3 四分位
真値	100136.5	550353.1	999868.4
丸めた値	100100	NA	999900
同じ丸め値の度数	39	0	32

一般用ミクロデータを用いた検証 (1/3)

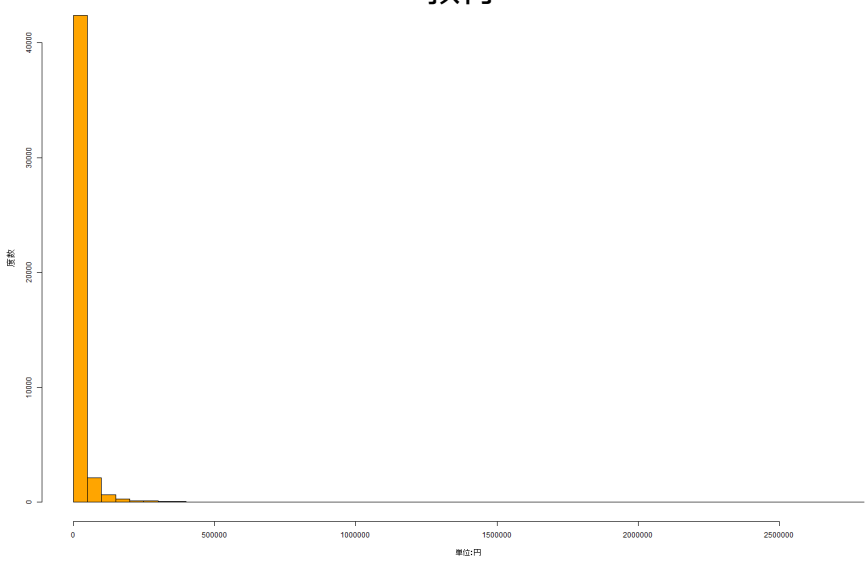
一般用ミクロデータ
 (2009年全国消費実態調査) 十大費目
 全データの度数45,811

	平均	歪度
年間収入	6401.9千円	2.12
消費支出	298373.7円	3.01
食料	68740.2円	1.26
住居	16127.9円	12.25
光熱・水道	19421.0円	1.36
家具・家事用品	9374.0円	7.63
被服及び履物	12054.8円	6.41
保健医療	13281.0円	7.17
交通・通信	44692.4円	7.99
教育	15014.9円	14.59
教養娯楽	31099.3円	4.59
その他の消費支出	68568.3円	5.94

消費支出



教育



一般用ミクロデータを用いた検証 (2/3)

	第1四分位	丸めた 第1四分位	誤差率	中央値	丸めた 中央値	誤差率	第3四分位	丸めた 第3四分位	誤差率
年間収入	3804.9	3800	0.13%	5503.7	5504	0.01%	7983.6	7980	0.05%
消費支出	194216.5	194200	0.01%	260255.8	260300	0.02%	354990.9	355000	0.00%
食料	48104.4	48100	0.01%	63497.8	63500	0.00%	83502.0	83500	0.00%
住居	659.3	660	0.11%	2876.9	2880	0.11%	17461.6	17500	0.22%
光熱・水道	13589.7	13590	0.00%	17905.0	17900	0.03%	23544.7	23540	0.02%
家具・家事用品	2877.6	2880	0.08%	5613.9	5610	0.07%	11057.8	11060	0.02%
被服及び履物	3893.7	3890	0.09%	7686.1	7690	0.05%	14515.6	14500	0.11%
保健医療	3849.0	3850	0.03%	7687.3	7690	0.03%	15434.6	15430	0.03%
交通・通信	10372.4	10400	0.27%	23464.7	23460	0.02%	48760.9	48800	0.08%
教育	0.0	0	0.00%	1761.2	1760	0.07%	13535.3	13500	0.26%
教養娯楽	12182.3	12180	0.02%	21707.7	21710	0.01%	38310.9	38300	0.03%
その他の消費支出	24793.2	24800	0.03%	44634.2	44600	0.08%	80909.8	80900	0.01%

誤差率 := |真値 - 丸め値| / 真値

一般用ミクロデータを用いた検証 (3/3)

各サンプルサイズでランダムに抽出した第1四分位の丸めによる誤差率

	100/45811	200/45811	300/45811	400/45811	500/45811	600/45811	700/45811	800/45811	900/45811	1000/45811
年間収入	7.47%	3.96%	5.72%	7.04%	5.19%	7.83%	0.50%	0.16%	7.67%	0.26%
消費支出	18.67%	19.66%	1.03%	18.03%	1.79%	1.64%	1.63%	1.48%	0.59%	1.75%
食料	6.58%	4.15%	7.22%	4.06%	7.32%	6.12%	5.10%	0.50%	0.06%	0.68%
住居	NA	6.42%	10.74%	9.27%	11.62%	12.99%	11.92%	17.66%	7.54%	16.05%
光熱・水道	9.34%	1.37%	0.40%	1.68%	1.37%	1.29%	1.02%	1.07%	1.65%	0.70%
家具・家事用品	3.25%	1.97%	7.81%	6.98%	1.31%	4.48%	1.58%	7.63%	6.90%	1.64%
被服及び履物	NA	5.23%	0.99%	1.51%	3.84%	6.47%	5.72%	5.12%	6.13%	3.92%
保健医療	NA	0.54%	0.56%	1.59%	6.57%	3.18%	3.99%	3.07%	5.45%	2.08%
交通・通信	17.45%	9.20%	14.02%	9.22%	19.47%	8.00%	0.68%	15.09%	13.46%	1.56%
教育	NA	33.01%	0.52%	19.85%	9.57%	6.17%	28.54%	30.90%	21.75%	1.01%
教養娯楽	12.72%	15.50%	7.43%	2.30%	0.68%	1.59%	1.72%	2.34%	2.26%	1.62%
その他の消費支出	5.11%	3.05%	6.84%	5.15%	7.43%	0.83%	9.20%	0.64%	9.46%	0.98%

まとめ

- UK Data Serviceの秘匿方法は、**シンプル**でありながら
客体10の原則を満たす方法と考えられる
- 一方で、算出元データの度数のしきい値等**追加のチェック内容**が必要である可能性
- ただし、おなじ丸め値を持つ客体に対する**グループ開示の可能性**については
依然として考慮する必要がある
- 今後は、攻撃者の順位推定能力を想定し
それに対して安全を確保する方法を検討する

本発表資料は、以下に掲載しています

統計センター 学会発表



統計マイクロデータで、 新たな発見を！

国勢調査などの公的統計のマイクロデータを、
公益性のある学術研究等において、利用することができます。
随時相談を受け付けておりますので、下記＜お問い合わせ先＞まで、
お気軽にご相談ください。

オンサイト 利用

学術研究目的など、各府
省の統計調査の調査票
情報を、大学、研究機関
などと連携して設置され
たオンサイト施設*で、ご
利用いただけます

*入退室管理やデータ管理など、高い
情報セキュリティを確保した施設
(令和4年7月時点:18か所)

オーダー メイド集計

研究や教育目的など、
ご利用者の委託により、
公表されていない集計表
を、調査票情報を用いて
作成し、提供いたします

匿名 データ

研究や教育を目的とした
多変量解析などの実証分
析が可能な特定の個人や
団体等が識別できないよ
うに、匿名化措置を施した
マイクロデータを提供いたし
ます

利用者の声

マイクロデータを利用した
研究について、今後、
さらに活用できると思わ
れることから、ぜひ若手
研究者にも、このサービ
スについて、知ってもら
うと良いと思います。

利用可能な調査や申請手続については、
「マイクロデータ利用ポータルサイト」をご覧ください。



miripo

で

検索、



独立行政法人
統計センター

＜お問い合わせ先＞

統計技術・提供部 統計情報提供課 二次の利用担当
〒 162-8668 東京都新宿区若松町19-1 総務省第2庁舎
☎ 03-5273-1205 ✉ nijiriyounstac.go.jp
🌐 <https://www.nstac.go.jp/>

ご清聴ありがとうございました



mi ri po

マイクロデータ利用ポータルサイト



検索

ホーム

マイクロデータ利用

利用実績

お問い合わせ

サイトマップ

統計調査結果 をより 広く活用 いただくための

mi ri po

マイクロデータ 利用のための ポータルサイト



- ・マイクロデータに関する制度の概要
- ・具体的な手続方法
- ・各府省の利用可能なデータ一覧
- ・マイクロデータの実績一覧
- ・よくある質問

などを公開しています！

<https://www.e-stat.go.jp/microdata/>