



独立行政法人

統計センター

統計関連学会連合大会

マイクロデータにおける攪乱的手法の有効性に関する比較研究

統計センター統計技術研究課

統計センター統計作成支援課

信頼に応じて作る統計表
総務省統計局総務課

総務省統計研究研修所

吉武 透

菊池 亮

阿久津 文香

伊藤 伸介

目次

1. 本報告の背景と目的
2. スワッピングについて
3. PRAMについて
4. スワッピングとPRAMの比較
5. 侵入者のシナリオの変更
6. おわりに

* 本報告の内容は、個人的見解を示すものであり、統計センター及び
総務省統計局の見解を示すものではありません

* 本報告の内容は、平成29年度に行った研究に基づくものです

1. 本報告の背景と目的

近年、「統計改革」が注目されており、EBPM(=Evidence Based Policy Making, 客観的な事実に基づいて政策立案を行うこと)をわが国でも推進するために、公的統計のデータだけでなく、民間のビックデータや行政記録データの利用可能性が指摘されている。

→2018年3月に刊行された『公的統計の整備に関する基本的な計画』では、調査票情報等の提供と利活用の推進を明記している。



「統計改革」という形で、公的統計データの作成・提供への注目が高まっている。

→public use fileを含む複数のタイプの公的統計の匿名化マイクロデータ(=匿名化技法が適用されたマイクロデータ)の可能性を検討する。

現在、総務省統計局で作成・提供されている 匿名データ(匿名化マイクロデータ)

対象統計調査	調査の年次
国勢調査	平成12年、平成17年
住宅・土地統計調査	平成5年、平成10年、平成15年
全国消費実態調査	平成元年、平成6年、平成11年、平成16年
労働力調査	平成元年1月～平成24年12月
就業構造基本調査	平成4年、平成9年、平成14年、平成19年
社会生活基本調査	平成3年、平成8年、平成13年、平成18年

国勢調査の匿名データの主な特徴

- 地域区分は、都道府県と人口50万以上市区
- データ量は母集団の1%、世帯単位で抽出・提供
- 1種類のみ of 匿名データを作成・提供
- リコーディング、トッピングコーディング、レコード削除といった従来の非攪乱的手法(non-perturbative methods)に加え、攪乱的手法(perturbative methods)として、スワッピングを適用

- ・国勢調査の匿名データの答申（諮問第44号の答申）では、「今後の課題」として、「利用者のニーズを踏まえて、匿名化措置の内容や組合せ、抽出単位の異なる複数の匿名データの作成の可能性について検討する必要がある」とする、「複数の匿名データの作成の可能性に関する検討」を指摘。

- ・匿名データの作成において課題となる、special unique (ex. 全国レベルの結果表のセルで度数1になるケース等)に対しては効果的な匿名化処理の方法が模索されている。

→匿名化マイクロデータにおける攪乱的手法の適用可能性については、さらなる検討の余地があると思われる。

これまで、匿名化マイクロデータの作成に関する基礎研究として、国勢調査の個票データを用いて、スワッピングやPRAMといった攪乱的手法の適用可能性、さらにはリコーディングやトップコーディングにおける区分統合の可能性が検討されてきた(伊藤・星野(2014)、伊藤ほか(2016a, 2016b, 2017))。

→これらの匿名化手法を適用することによって、様々なタイプの匿名化マイクロデータが作成される。

⇒秘匿性の観点から攪乱的手法のさらなる可能性を追究するだけでなく、攪乱的手法が適用された匿名化マイクロデータの有用性を検証することによって、ニーズに応じた匿名化マイクロデータの作成が可能になる。

本報告の目的

- 本報告では、平成22年国勢調査の個票データに基づく匿名化マイクロデータの展開可能性を追究する。
 - 国勢調査のテストデータを作成した上で、攪乱的手法の適用可能性について比較・検討を行う。

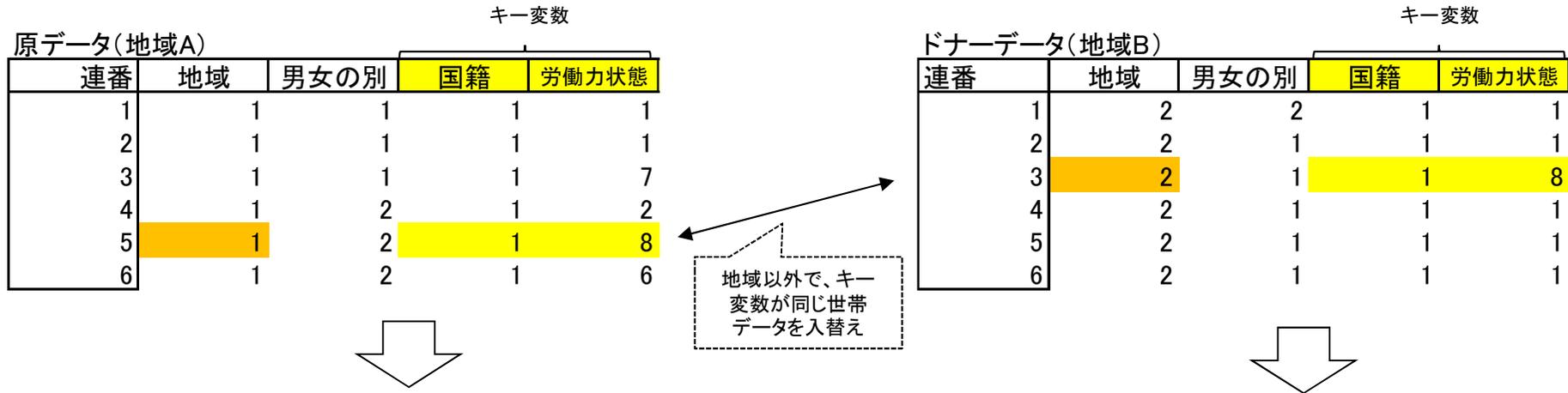
2. スワッピングについて

スワッピング(data swapping)とは・・・マイクロデータに含まれるレコード同士で属性値を入れ替えること (Willenborg and Waal(2001, p.126))。

・スワッピングの可能性に関する議論については少なくとも1970年代に遡ることができる (Dalenius and Reiss(1978))。

スワッピングのイメージ

原データ(秘匿処理を施していない個票データ)



匿名化データ(原データに匿名化技法を適用することによって作成したマイクロデータ)

連番	地域	男女の別	国籍	労働力状態
1	1	1	1	1
2	1	1	1	1
3	1	1	1	7
4	1	2	1	2
5	1(元は2)	1	1	8
6	1	2	1	6

連番	地域	男女の別	国籍	労働力状態
1	2	2	1	1
2	2	1	1	1
3	2(元は1)	2	1	8
4	2	1	1	1
5	2	1	1	1
6	2	1	1	1

上記の例の場合はキー変数(黄色のセル)が同じ世帯データを2地域間で入替えた。スワッピング後の匿名化世帯データにおいて作成された地域別の国籍別、労働力状態別の集計結果は、スワッピング前の原データにおける集計結果と変わらない。

* 本研究において使用するキー変数

- ① 男女の別 (2区分)
- ② 各歳年齢 (114区分)
- ③ 配偶関係 (4区分)
- ④ 国籍 (2区分)
- ⑤ 労働力状態 (8区分)
- ⑥ 従業上の地位 (8区分)
- ⑦ 産業大分類 (21区分)
- ⑧ 職業大分類 (12区分)
- ⑨ 住居の種類 (8区分)

諸外国におけるスワッピングの適用事例

- ・アメリカでは、2000年人口センサスのPUMS (Public Use Microdata Samples)やAmerican Community Surveyにおいて、スワッピングを適用している。

⇒「**スペシャルユニーク(特殊な一意, special uniques)**」の対象となるレコードを探索した上で、別のレコードに置き換えるという処理を行っている。

具体的には、非常に粗い地域区分において特定の人口社会的属性群に基づいて一意性を有する世帯のレコードは、露見リスクが非常に高いと考えられることから、別の地域における他の世帯との入れ替えが行われている(Zayatz(2007, p.257))。

⇒また、2000年人口センサスの集計表における秘匿処理として、人口センサスの個別データにスワッピングを適用していることが知られている。

- ・イギリスでは、人口センサスの個別データの作成において、レコードスワッピングが適用されている (Shlomo(2007))。

3. PRAMについて

PRAM (=Post Randomization Method): ノイズ付加等の確率的処理を施して、データのプライバシーを保護する方法

- 攪乱(perturbation)ステップ
 - セル毎に確率的な変更を施す
- 再構築(reconstruction)ステップ
 - ベイズ推定等を用いて分布を推定する

※ 今回の研究では、攪乱のみを対象としており、再構築については今後検討する。

攪乱(一般) (Kooiman(1998))

遷移確率行列に基づいて遷移

任意の属性とその属性値 $\{v_i\}_{0 \leq i < M}$ に対し, 遷移確率行列

$$A := \begin{pmatrix} a_{0,0} & \cdots & a_{0,M-1} \\ \vdots & \ddots & \vdots \\ a_{M-1,0} & \cdots & a_{M-1,M-1} \end{pmatrix} \text{に基づいて遷移するとは,}$$

匿名化処理を Δ としたとき, 元テーブルの属性値を

$a_{i,j} := \Pr[v_j = \Delta(v_i)]$ に従って遷移させること

元データ				攪乱データ		
住所	年齢	職業		住所	年齢	職業
東京都新宿区	51	会社員	確率 $a_{55,22}$ で遷移	東京都新宿区	23	会社員
東京都新宿区	54	会社役員		東京都港区	54	会社役員
東京都品川区	41	自営業		東京都品川区	41	主婦
...

攪乱の具体例: 維持置換攪乱 (Agrawal and Srikant(2000))

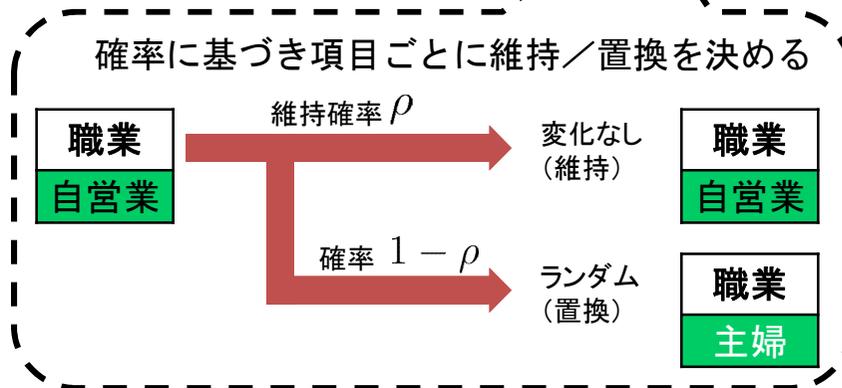
元データ

住所	年齢	職業
東京都新宿区	51	会社員
東京都新宿区	54	会社役員
東京都品川区	41	自営業
...

攪乱

攪乱データ

住所	年齢	職業
東京都新宿区	23	会社員
東京都港区	54	会社役員
東京都品川区	41	主婦
...



置換しないで「主婦」なのか、
置換して「主婦」なのか区別で
きない

$$\text{c.f. } A := \begin{pmatrix} \rho + \frac{1-\rho}{M} & \dots & \frac{1-\rho}{M} \\ \vdots & \ddots & \vdots \\ \frac{1-\rho}{M} & \dots & \rho + \frac{1-\rho}{M} \end{pmatrix}$$

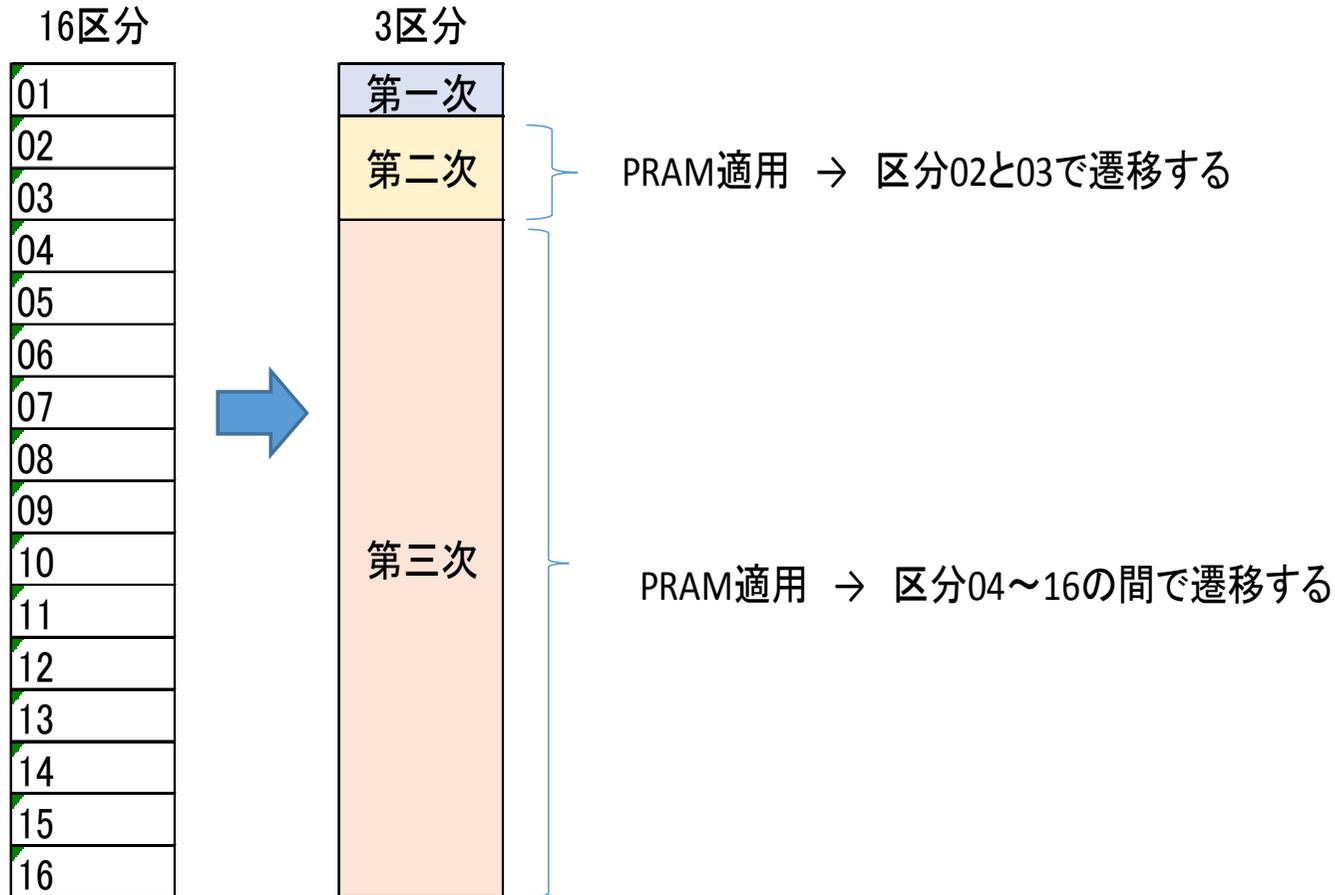
攪乱の効果

- 進入者は攪乱データを見ても、攪乱されているか否かは分からず、また、本来どのようなデータであったのかも分からない
- 種々の安全性を満たせることが知られている
 - Pk匿名性 (Ikarashi et al.(2014))
 - $1/k$ 以上の確率で個人を特定できない
 - Differential privacy(Dwork(2006), Lin et al.(2012), Ikarashi et al.(2014))
 - ある個人が含まれているか否かに出力が依らない

本研究におけるPRAMの実験のイメージ

年齢、産業と職業のそれぞれに対して、PRAMを適用している。

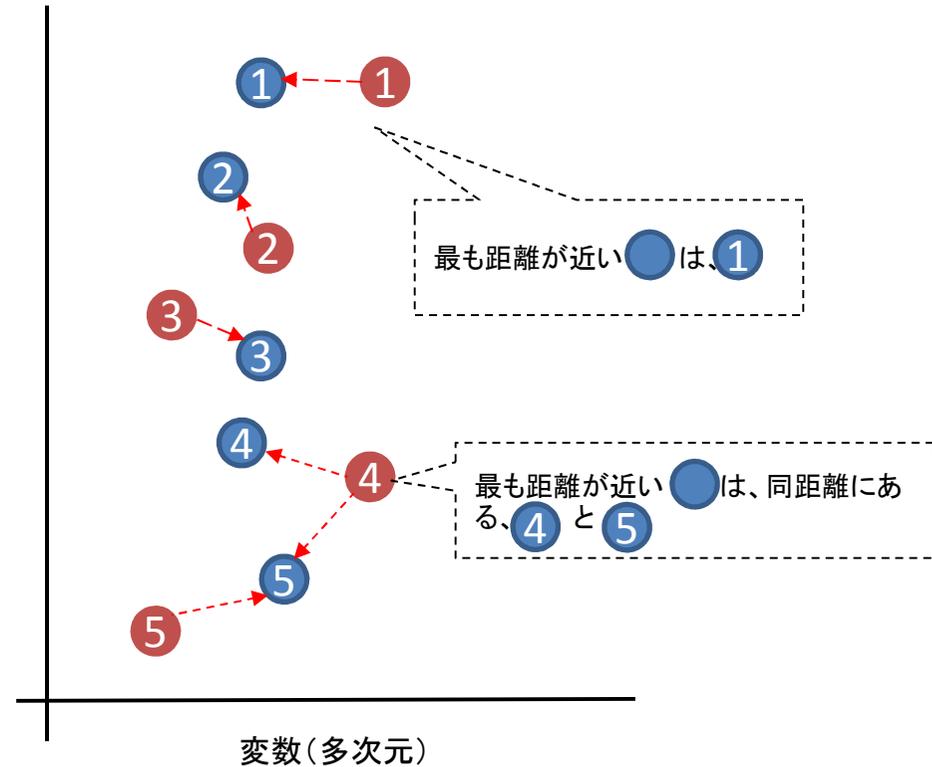
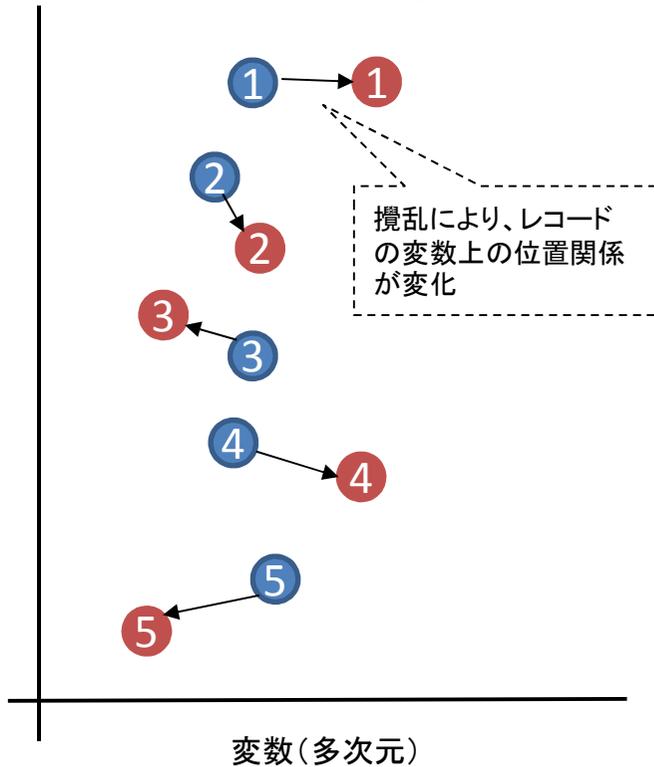
Ex. 産業



例えば、産業の場合、最も粗い3区分(第一次、第二次、第三次)にデータを分け、グループごとにPRAMの適用を行っている。

4. スワッピングとPRAMの比較

●:原レコード ●:攪乱済みレコード



「原レコード」と「攪乱済みレコード」の各レコード間で、**22変数**(次スライドに示す)による距離を計測し、最も距離の近いレコードとの対応関係を見る。

この関係が1対1である場合、「攪乱済みレコード」は「原レコード」に戻れることから、この関係を「真のリンク」とする(特定リスクが高い)。

一方、データ「4」のように、1対多となる場合は、複数の戻り先があることから、「真のリンク」とはしない。

マッチングに利用した22変数

男女の別

年齢5歳階級

配偶関係

国籍

労働力状態

従業上の地位

産業大分類

職業大分類

住居の種類

住居の建て方

建物の階数

世帯の住んでいる階

延べ床面積

利用交通手段1

利用交通手段2

利用交通手段3

利用交通手段4

利用交通手段5

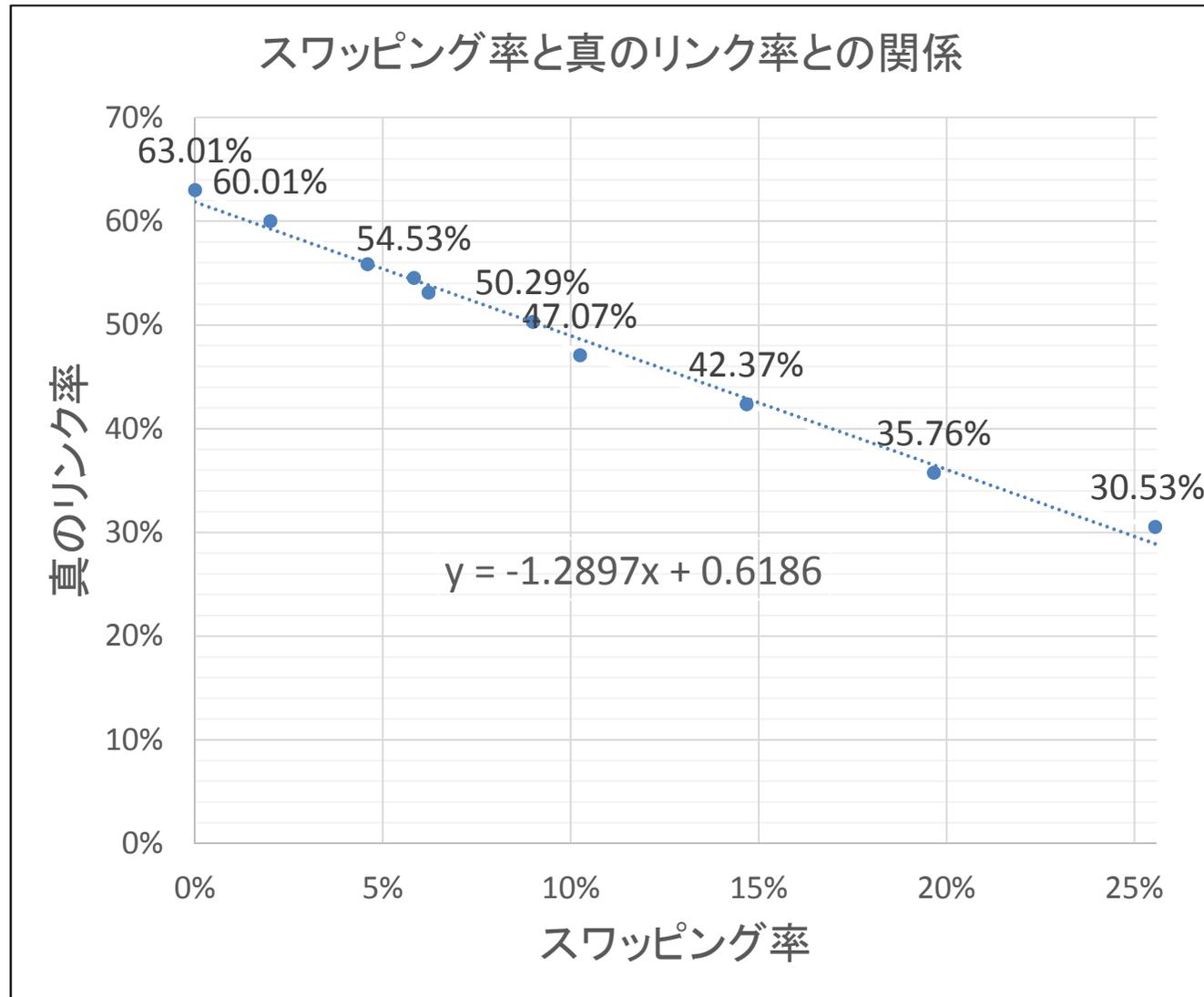
利用交通手段6

利用交通手段7

利用交通手段8

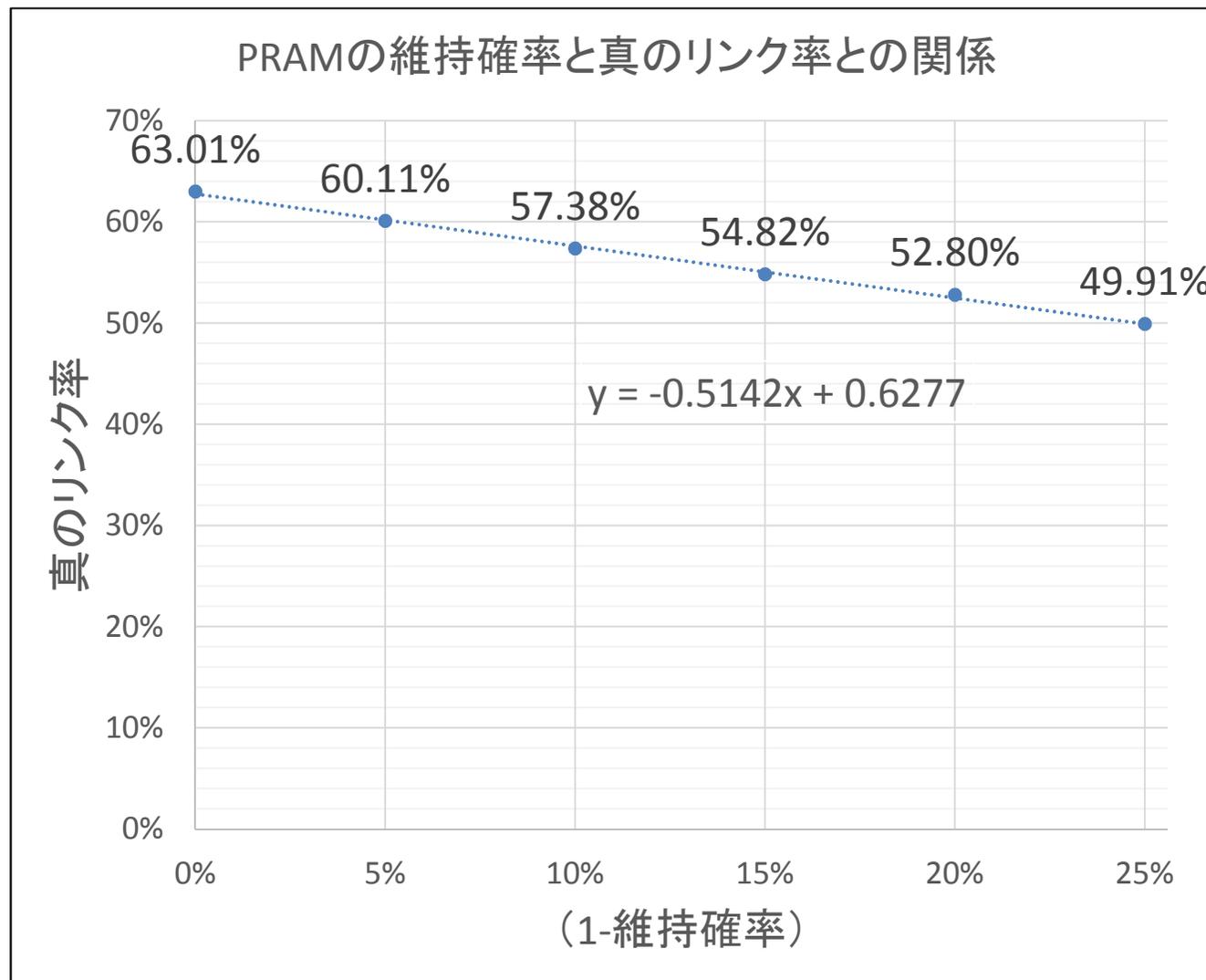
利用交通手段9

スワッピング率を変動させながら、ランダムスワッピングを行い、スワッピング率と真のリンク率との関係を計測



- ・スワッピング率の上昇に伴い、真のリンク率は低下
- ・ほぼ線形

年齢5歳階級、産業大分類、職業大分類に個々にPRAMを適用し、維持確率と真のリンク率との関係を計測



- ・(1-維持確率)の上昇に伴い、真のリンク率は低下

- ・ほぼ線形

情報量損失に基づく有用性の検証

- ・マイクロデータにおける有用性の定量的な評価方法については、クラメールのVといった関連性の指標の算出や原データからの絶対距離の平均値(average absolute distance)の計測等(伊藤・星野(2014))が考えられる。
- ・伊藤・星野・阿久津(2016a)においては、エントロピーに基づいて情報量損失の指標を作成した上で、秘匿の観点から「許容可能」な分類区分の組み合わせについて情報量損失の計測を行った。



本研究においても、原区分と攪乱済みデータ(ex. スワッピング済データ)を行った場合の距離を計測し、情報量損失を計算する。

情報量損失の計測方法

絶対距離の平均値による評価

⇒ 個票データと攪乱済みデータの両方で2変数間のクロス集計表を作成し、双方のクロス集計表の各セルの度数の差の絶対値に関する平均値を求める(Shlomo et al.(2010))。

$$IL = \frac{\sum_c |T^P(c) - T^O(c)|}{n_T}$$

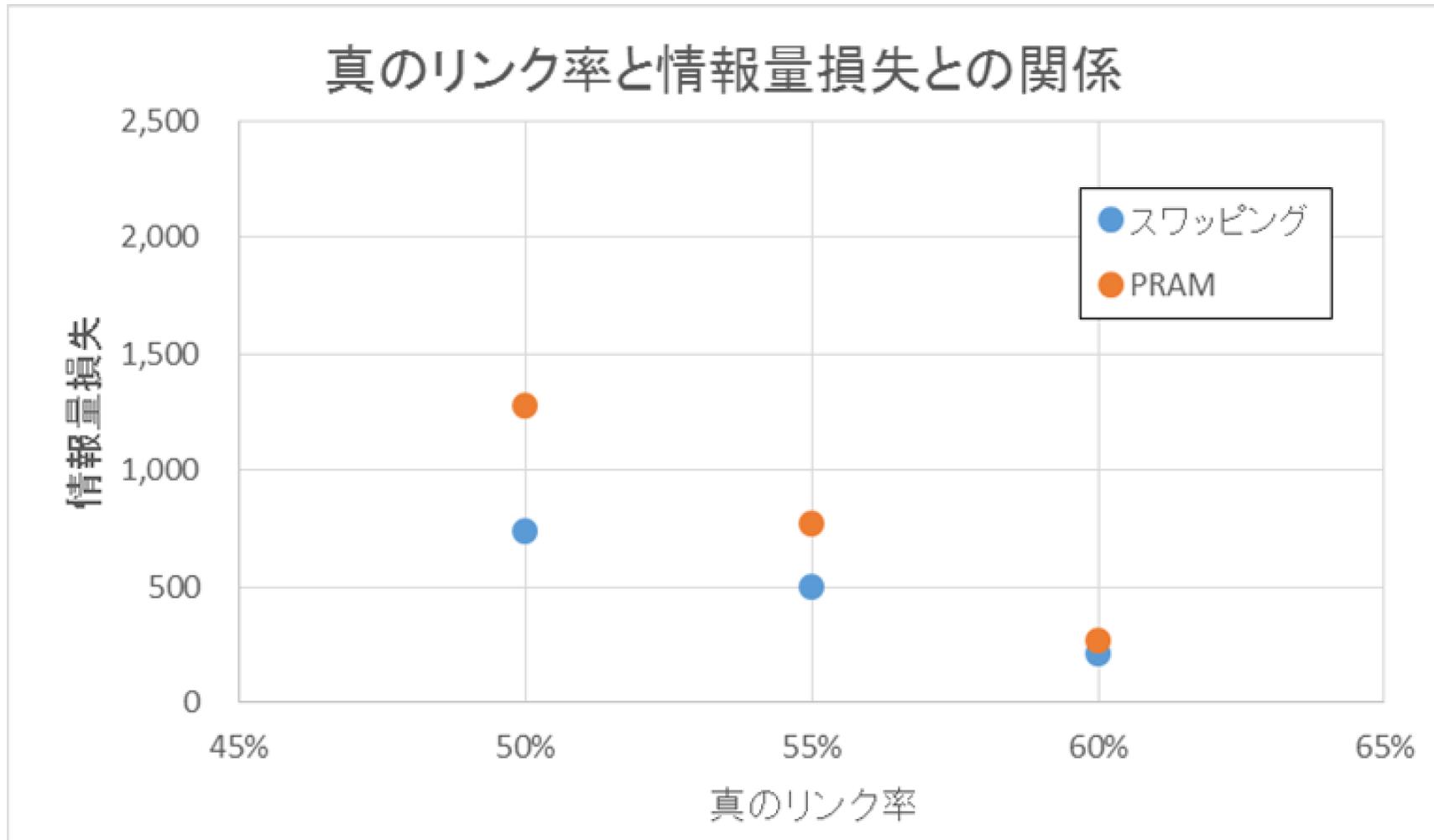
$T^O(c)$: 個票データを用いて作成したクロス表における各セルの度数

$T^P(c)$: 攪乱済みデータをもとに作成したクロス表における各セルの度数

n_T : クロス集計表におけるセルの数

本研究では、これに準じた計算を行った(情報量損失の値そのものには相対的な意味しかないことに注意)。

スワッピング及びPRAMに対し、共に真のリンク率が約50%、55%、60%となる状態での情報量損失を比較。



2変数のクロス表を作成し、各変数を軸とした場合の情報量損失を比較。

真のリンク率	60%	55%	50%	60%	55%	50%
手法	スワッピング			PRAM		
男女の別	70.81	129.03	235.43	30.49	89.82	150.18
年齢5歳階級	10.03	23.65	32.86	44.90	110.60	223.79
配偶関係	39.24	110.52	178.74	16.34	46.85	75.19
国籍	60.06	167.53	239.53	29.09	85.43	143.33
労働力状態(8区分)	16.89	40.02	73.43	7.51	21.77	36.47
従業上の地位(8区分)	27.35	60.54	81.29	9.31	25.16	40.56
産業大分類	11.75	30.10	47.93	162.48	458.02	795.13
職業大分類	21.32	52.99	56.02	206.30	621.49	949.03
住居の種類	41.32	99.33	151.02	9.01	24.27	38.20
住居の建て方	71.96	161.07	217.17	17.18	46.97	74.80
建物の階数	6.47	17.44	26.38	1.20	3.18	5.07
世帯の住んでいる階	4.64	12.82	15.46	1.16	3.14	5.00
延べ床面積(14区分)	36.46	92.78	126.39	6.01	15.05	23.17
利用交通手段1	212.35	457.23	746.60	30.28	88.02	143.49
利用交通手段2	89.20	158.21	271.68	29.72	86.16	144.80
利用交通手段3	60.80	137.25	202.49	29.22	85.37	143.45
利用交通手段4	73.80	134.14	204.56	29.12	85.97	143.88
利用交通手段5	222.31	464.65	750.53	30.26	88.73	143.18
利用交通手段6	57.01	126.80	193.26	28.93	85.19	142.81
利用交通手段7	66.11	135.38	232.29	29.97	85.44	144.37
利用交通手段8	76.17	130.93	207.28	29.63	85.94	144.08
利用交通手段9	61.30	129.14	196.33	29.59	85.68	143.10
延べ数	668.68	1435.79	2243.34	403.85	1164.13	1926.55
延べ数(利用交通手段を除く)	209.16	498.92	740.83	270.49	775.88	1279.96

PRAMの対象となる3変数を軸とした場合では、情報量損失はPRAMの方が大きい。
それ以外の変数を軸とした場合では、情報量損失はPRAMの方が小さい。

単変数での度数分布を作成し、原データと攪乱済データとの相対度数の差の絶対値を見る。

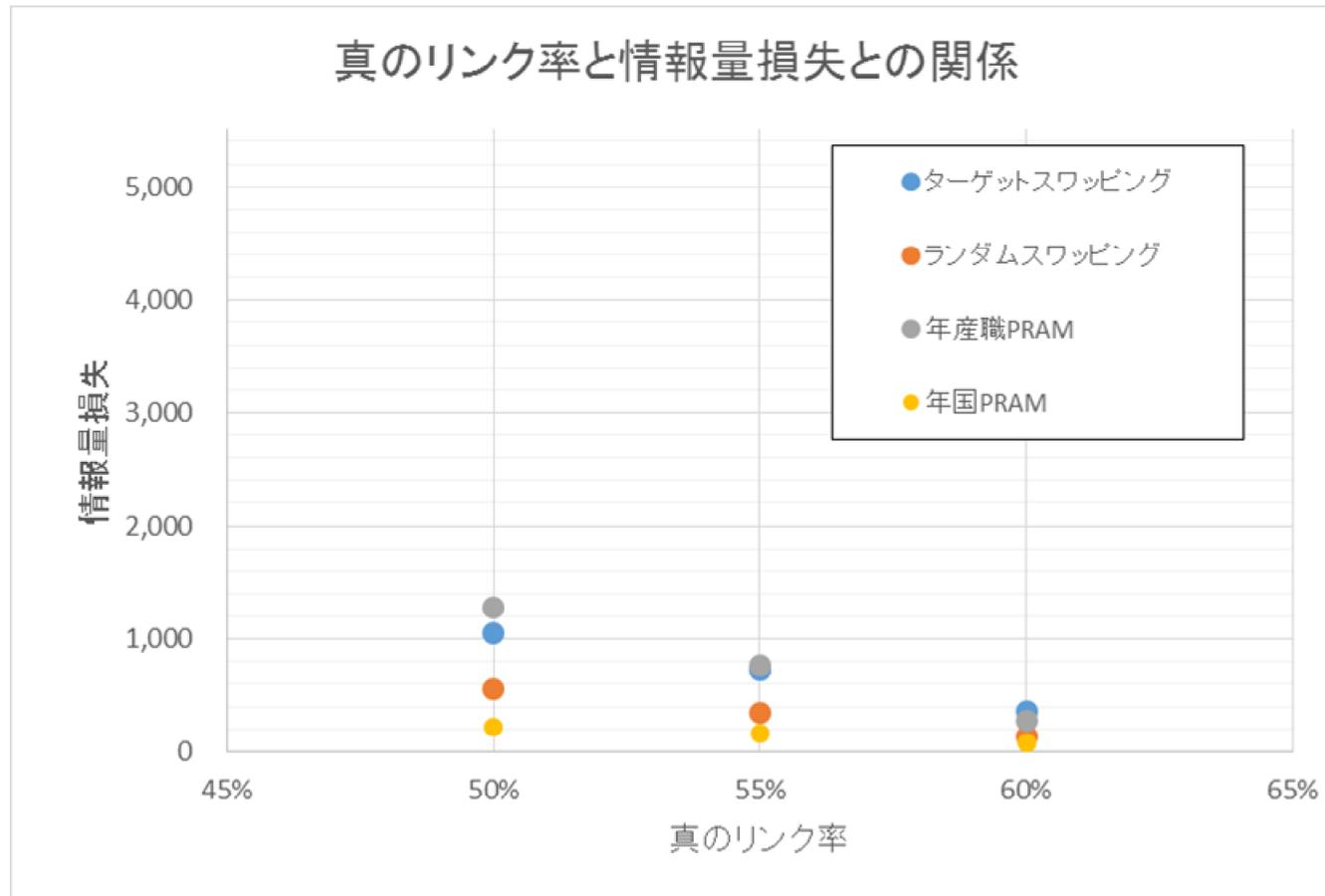
	スワッピング			PRAM		
	真のリンク率			真のリンク率		
	約60%	約55%	約50%	約60%	約55%	約50%
年齢5歳階級	相対度数の差(絶対値)			相対度数の差(絶対値)		
不詳	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
15歳未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
15～19歳	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%
20～24歳	0.01%	0.01%	0.00%	0.06%	0.26%	0.51%
25～29歳	0.00%	0.01%	0.01%	0.06%	0.26%	0.51%
30～34歳	0.00%	0.00%	0.02%	0.11%	0.13%	0.31%
35～39歳	0.00%	0.00%	0.01%	0.11%	0.13%	0.31%
40～44歳	0.00%	0.00%	0.01%	0.05%	0.07%	0.21%
45～49歳	0.00%	0.00%	0.01%	0.05%	0.07%	0.21%
50～54歳	0.00%	0.01%	0.01%	0.00%	0.04%	0.16%
55～59歳	0.00%	0.01%	0.01%	0.00%	0.04%	0.16%
60～64歳	0.00%	0.01%	0.00%	0.09%	0.35%	0.74%
65～69歳	0.00%	0.01%	0.00%	0.09%	0.35%	0.74%
70～74歳	0.00%	0.00%	0.00%	0.05%	0.09%	0.23%
75～79歳	0.00%	0.00%	0.00%	0.05%	0.09%	0.23%
80～84歳	0.00%	0.00%	0.00%	0.03%	0.01%	0.04%
85歳以上	0.00%	0.00%	0.00%	0.03%	0.01%	0.04%

	スワッピング			PRAM		
	真のリンク率			真のリンク率		
	約60%	約55%	約50%	約60%	約55%	約50%
産業大分類	相対度数の差(絶対値)			相対度数の差(絶対値)		
不詳	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%
A 農業	0.00%	0.02%	0.01%	0.06%	0.18%	0.30%
U 林業	0.00%	0.01%	0.00%	0.04%	0.08%	0.16%
B 漁業	0.00%	0.00%	0.01%	0.03%	0.10%	0.14%
C 鉱業, 砕石業, 砂利採取業	0.00%	0.00%	0.07%	0.76%	2.30%	3.84%
D 建設業	0.01%	0.01%	0.28%	0.51%	1.28%	2.06%
E 製造業	0.03%	0.07%	0.02%	1.27%	3.58%	5.90%
F 電気・ガス・熱供給・水道業	0.00%	0.00%	0.05%	0.15%	0.40%	0.71%
G 情報通信業	0.02%	0.03%	0.05%	0.15%	0.37%	0.67%
H 運輸業, 郵便業	0.00%	0.02%	0.03%	0.08%	0.18%	0.63%
I 卸売業, 小売業	0.00%	0.03%	0.03%	0.13%	0.46%	0.95%
J 金融業, 保険業	0.02%	0.02%	0.04%	0.14%	0.35%	0.79%
K 不動産業, 物品賃貸業	0.00%	0.02%	0.14%	0.13%	0.36%	0.80%
L 学術研究, 専門・技術サービス業	0.01%	0.07%	0.01%	0.65%	1.84%	3.15%
M 宿泊業, 飲食サービス業	0.02%	0.02%	0.04%	0.03%	0.26%	0.32%
N 生活関連サービス業, 娯楽業	0.01%	0.01%	0.01%	0.06%	0.11%	0.36%
O 教育, 学習支援業	0.02%	0.05%	0.04%	0.06%	0.19%	0.16%
P 医療, 福祉	0.00%	0.03%	0.00%	0.06%	0.11%	0.03%
Q 複合サービス事業	0.02%	0.02%	0.01%	0.09%	0.46%	0.97%
R サービス業(他に分類されないもの)	0.00%	0.04%	0.00%	0.03%	0.05%	0.19%
S 公務(他に分類されるものを除く)	0.00%	0.03%	0.00%	0.02%	0.09%	0.11%
T 分類不能の産業	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

	スワッピング			PRAM		
	真のリンク率			真のリンク率		
	約60%	約55%	約50%	約60%	約55%	約50%
職業大分類	相対度数の差(絶対値)			相対度数の差(絶対値)		
不詳	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%
0 管理的職業従事者	0.01%	0.03%	0.01%	0.58%	1.74%	2.40%
1 専門的・技術的職業従事者	0.05%	0.08%	0.12%	0.57%	1.62%	2.54%
2 事務従事者	0.01%	0.03%	0.04%	0.02%	0.13%	0.14%
3 販売従事者	0.03%	0.06%	0.05%	0.21%	0.26%	0.45%
4 サービス職業従事者	0.00%	0.05%	0.01%	0.01%	0.33%	0.47%
5 保安職業従事者	0.00%	0.03%	0.01%	0.22%	0.69%	1.09%
6 農林漁業従事者	0.00%	0.02%	0.01%	0.22%	0.74%	1.50%
7 生産工程従事者	0.07%	0.14%	0.13%	1.01%	3.23%	5.07%
8 輸送・機械運転従事者	0.03%	0.04%	0.02%	0.19%	0.52%	0.83%
9 建設・採掘従事者	0.00%	0.00%	0.01%	0.09%	0.22%	0.35%
V 運搬・清掃・包装等従事者	0.00%	0.05%	0.02%	0.09%	0.47%	0.38%
T 分類不能の職業	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

全体として、攪乱することによって度数分布の変化が大きいのはPRAMである。特に、ランダムな入れ替えを行っていることから、もともとの度数が平均に比べて相対的に少ないものや、多いものについて、より大きな変化が見られる。

「年齢・産業・職業」のPRAM のほか、「年齢・国籍」のPRAMとも比較し、変数の組み合わせによる、PRAMの特性を比較。



「年齢・産業・職業」3変数では、ターゲットスワッピングやランダムスワッピングよりも情報量損失が大きかったが、「年齢・国籍」2変数では、そこまで大きな情報量損失がない。

他の変数の組み合わせも検討したが、矛盾が生じず、情報量損失の少ない2変数を選定した。

5. 侵入者のシナリオの変更

特定個人の情報を持った者が、数地域(ここでは2地域の場合と3地域の場合)の匿名化データと変数を照合した場合の、「真のリンク率」を考える。

原データ(地域A)				
連番	変数1	変数2	変数23
1	1	1	4
2	1	1	1
3	1	1	7
.....
20000	1	2	6

匿名化データ(地域A)				
連番	変数1	変数2	変数23
1	1	4	1
2	1	3	1
3	3	1	7
.....
20000	2	4	1

匿名化データ(地域B)				
連番	変数1	変数2	変数23
1	2	3	3
2	1	1	4
3	1	1	7
.....
20000	2	1	2

匿名化データ(地域C)				
連番	変数1	変数2	変数23
1	3	2	1
2	3	1	7
3	1	2	5
.....
20000	3	1	2

マッチングに利用した23変数

地域情報

男女の別

年齢5歳階級

配偶関係

国籍

労働力状態

従業上の地位

産業大分類

職業大分類

住居の種類

住居の建て方

建物の階数

世帯の住んでいる階

延べ床面積

利用交通手段1

利用交通手段2

利用交通手段3

利用交通手段4

利用交通手段5

利用交通手段6

利用交通手段7

利用交通手段8

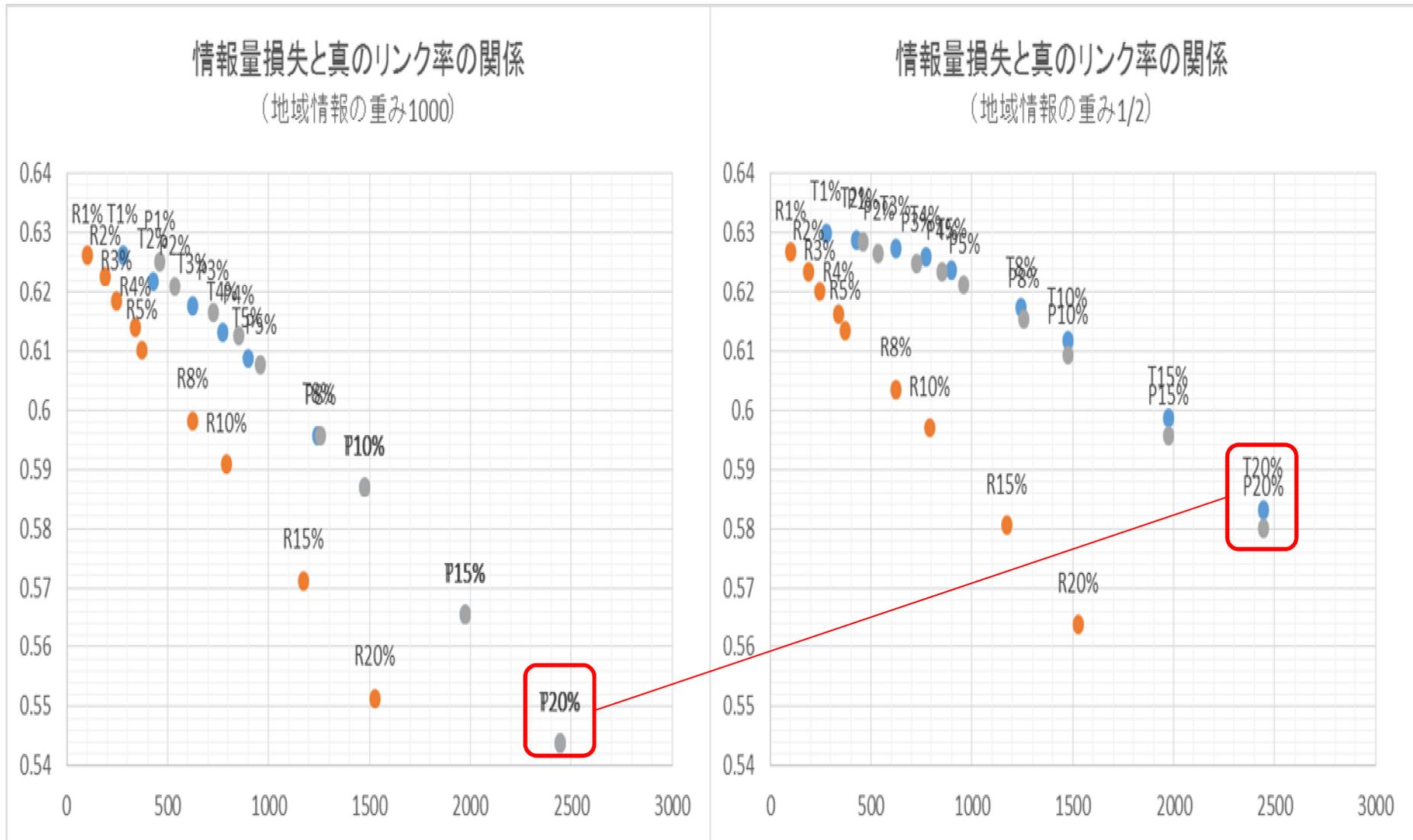
利用交通手段9

先ほどまでの22変数に、「地域情報」を追加した。

スワッピングのもともとの考え方は「地域情報が入れ替わったなら、当該レコードは安全である」という前提に立っている。

ここでは、地域情報も単なる1変数でしかないと考える侵入者のシナリオを想定し、個人が特定されるリスクを考える。

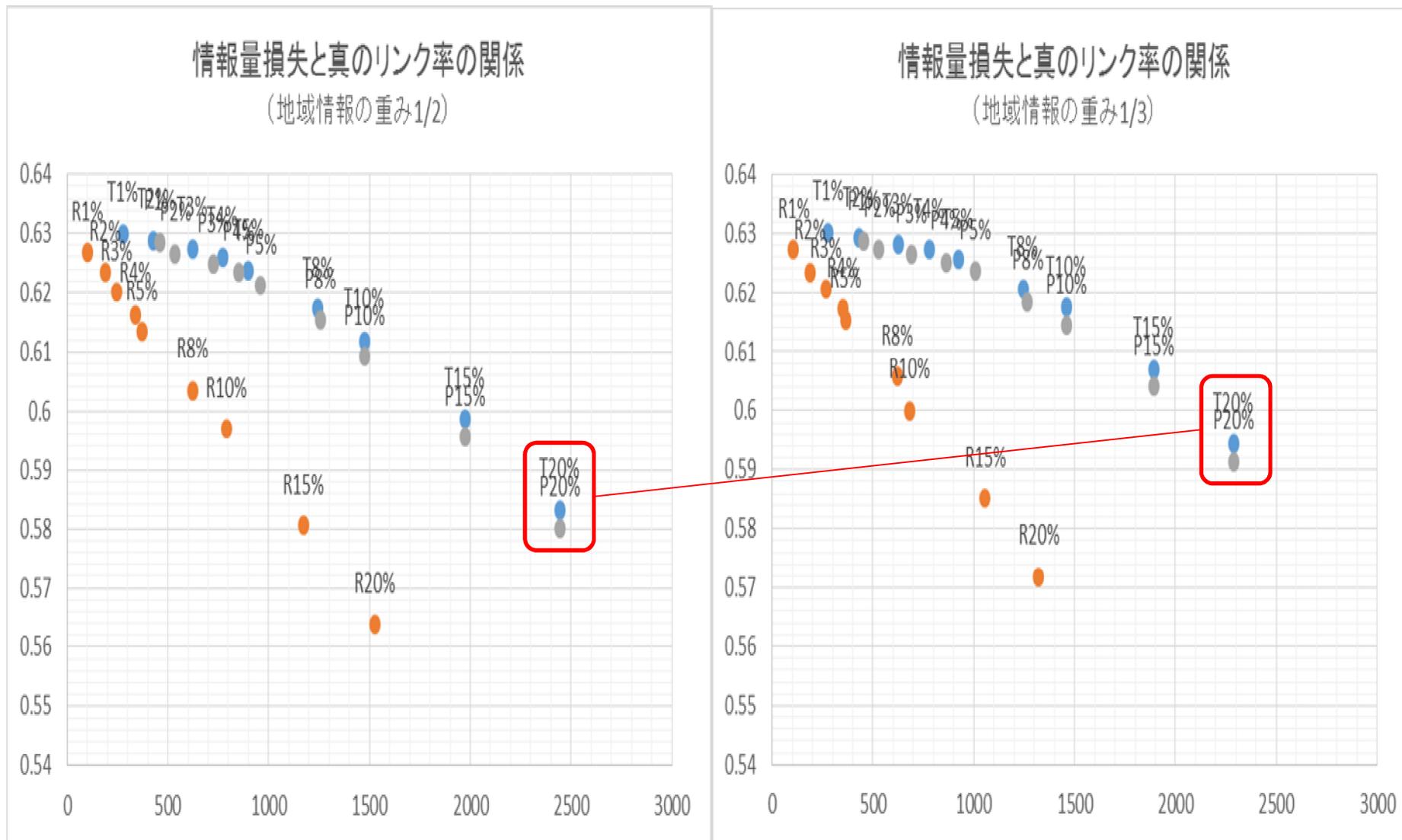
地域情報の重みMAXの場合と1/2(2地域)の場合の比較



横軸: 情報量損失 縦軸: 真のリンク率。

T: ターゲットスワッピング R: ランダムスワッピング P: ターゲットスワッピング + 年齢国籍PRAM
各アルファベットの添字は、スワッピング率を示す。PRAMの対象は全体の4.0%のスペシャルユニークレコード。

地域情報の重み1/2(2地域)の場合と1/3(3地域)の場合の比較



横軸: 情報量損失 縦軸: 真のリンク率。

T: ターゲットスワッピング R: ランダムスワッピング P: ターゲットスワッピング + 年齢国籍PRAM
各アルファベットの添字は、スワッピング率を示す。PRAMの対象は全体の4.0%のスペシャルユニークレコード。

6. おわりに

- ・本報告では、国勢調査の匿名化マイクロデータの作成方法に関する新たな取り組みとして、攪乱的手法であるスワッピングとPRAMに焦点を当て、攪乱的手法が適用された匿名化マイクロデータの可能性を追究した。
- ・PRAMは、対象となる変数に関して情報量損失が相対的に大きくなることが特徴である。矛盾を生じさせないという性質と、情報量損失を大幅に増やさないという性質を併せ持つ組み合わせとして、「年齢5歳階級、国籍」の2変数を組み合わせた。
- ・また、今回は、これまでのように「匿名データ」側から本人に戻れるか否かだけでなく、セキュリティ分野において使われる侵入者のシナリオを取り入れ、「個人情報を持っている人が、スワッピングされている可能性も考慮した場合」の安全性についても検討した。
- ・ターゲットスワッピングのみの場合よりも、PRAMを組み合わせることで、一定程度「真のリンク」を回避することができた。この効果は、ターゲットスワッピング2～3%分程度の効果があった。
- ・引き続き、攪乱的手法について検討を重ねる余地がある。

主要参考文献

- Agrawal, R., Srikant, R.(2000) Privacy-Preserving Data Mining, SIGMOD 2000..
- Agrawal, R., Srikant, R., and Thomas, D. (2005) Privacy preserving OLAP. SIGMOD, ACM..
- Dalenius, T and Reiss, S. P. (1978) Data-swapping: a technique for disclosure control (extended abstract). In Proc. Section on Survey Research Methods, pp. 191-194. Washington, D.C.: American Statistical Association.
- DePersio, M, Lemons, M., Ramanayake, K. A., Tsay, J., Zayatz, L.(2012) n-Cycle Swapping for the American Community Survey, In Privacy in Statistical Databases UNESCO Chair in Data Privacy International Conference, PSD 2012 Palermo, Italy, September, 2012 Proceedings(eds Domingo-Ferrer, J. and Tinnirello, I.), Springer, pp.143-164.
- de Wolf, P.P., Gouweleeuw, J.M., Kooiman, P., and Willenborg, L.C.R.J.(1998) Reflections on PRAM, Statistical Data Protection.
- de Wolf, P.P. and van Gelder, I.(2004) An empirical evaluation of PRAM, Discussion paper 04012, Statistics Netherlands.
- Domingo-Ferrer, J. and Torra, V. (2001) Disclosure control methods and information loss for microdata. In *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies* (eds Doyle P. *et al.*), pp. 91-110. Amsterdam: Elsevier Science.
- Elliot, M., Manning, A. M., and Ford, R. W. (2002) A computational algorithm for handling the special uniques problem. *Int. J. Uncertainty Fuzziness Knowledge Based Syst.*, 10(5), 493-509.
- Gomatam, S. and Karr, A. F. (2003) Distortion measures for categorical data swapping. Technical Report, No. 131, National Institute of Statistical Sciences.
- Gouweleeuw, J.M., Kooiman, P., Willenborg, L.C.R.J., and de Wolf, P.P.(1997) Post Randomisation for Statistical Disclosure Control: Theory and Implementation, Research paper no. 9731, Statistics Netherlands.

伊藤伸介・星野なおみ(2014)「国勢調査マイクロデータを用いたスワッピングの有効性の検証」『統計学』107号, 1～16頁

伊藤伸介・星野なおみ・阿久津文香(2016a)「国勢調査における匿名化マイクロデータの有用性と秘匿性の定量的な評価」『製表技術参考資料』No.32, 1～33頁

伊藤伸介・星野なおみ・阿久津文香(2016b)「国勢調査マイクロデータに対する匿名化措置の可能性に関する研究」『製表技術参考資料』No.34, 1～59頁

伊藤伸介・星野なおみ・阿久津文香・菊池亮(2017)「国勢調査の匿名化マイクロデータの作成方法に関する新たな取り組み」『製表技術参考資料』No.37, 1～27頁

Kooiman, P., L. Willenborg and J. Gouweleeuw (1998) “PRAM: A Method for Disclosure Limitation of Microdata”, Research Paper, No. 9705, Statistics Netherlands, Voorburg.

Moore, R. A. (1996) Controlled data-swapping techniques for masked public use microdata sets. Statistical Research Division Report Series, RR 96-04, U.S. Bureau of the Census, Statistical Research Division, Washington, D. C.

Müller, W., Blien, U., and Wirth, H. (1995) Identification risks of micro data: evidence from experimental studies. *Socio. Meth. Res.*, 24(2), 131-157.

Nin, J., Herranz, J., and Torra, V. (2008) Rethinking rank swapping to decrease disclosure risk. *Data Knowl. Engineering*, 64(1), 346-364.

Shlomo, N., Tudor, C., and Groom, P. (2010) Data swapping for protecting census tables. In *Privacy in Statistical Databases UNESCO Chair in Data Privacy International Conference, PSD 2010 Corfu, Greece, September, 2010 Proceedings* (J. Domingo-Ferrer and E. Magkos, eds), New York: Springer, pp. 41-51.

Takemura, A. (2002) Local recoding and record swapping by maximum weight matching for disclosure control of microdata set. *J. Off. Stat.*, 18(2), 275-289.

Zayatz, L. (2007) “Disclosure Avoidance Practices and Research at the U.S. Census Bureau: An Update”, *Journal of Official Statistics*, Vol.23, No.2, pp.253-265.

Willenborg, L. and De Waal, T. (2001) *Elements of Statistical Disclosure Control*, New York: Springer. .

統計データの二次的利用の利用相談ブース

(独)統計センターでは、統計データの二次的利用（匿名データ、オーダーメイド集計、調査票情報のオンサイト利用等）について職員による利用相談（無料）を行っています。

ご関心のある方は、お気軽にお立ち寄りください。

- 日時：9月10日（月）～13日（木）9時30分～17時30分（13日は16時まで）
- 場所：5号館1階 5136教室 入口左手奥のブース

