

2012.11.9

擬似マイクロデータの試行提供

(独) 統計センター 榎田 直木

平成24年11月16日(金) 統計数理研究所
公的統計のマイクロデータの利用に関する研究集会
(午前の部) 「擬似マイクロデータの利用」

構成

1. 公的統計と擬似マイクロデータ
2. 擬似マイクロデータの試行提供
3. 擬似マイクロデータの特徴
4. 擬似マイクロデータの課題

1. 公的統計と擬似マイクロデータ

□ 公的統計

- 行政施策の一つとして作成される統計
- 公的統計は、「公共財」
- 国の予算による統計調査、その規模は大きい。

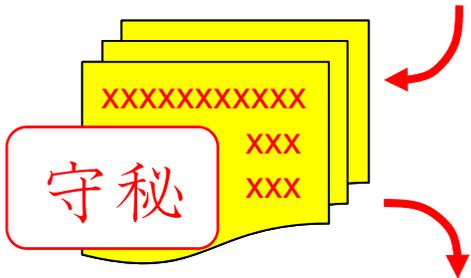
□ 統計法

- 公的統計や統計調査には、法令による権利・義務

1. 公的統計と擬似マイクロデータ

調査票
報告義務
(法第十三条)

調査票情報
適正な管理
(法第四十二条)



出典明記をお願いします。
数値を独自加工の場合は、
その旨も。

各府省ウェブ、
e-Stat、統計報告書など
(利用に制約なし)

統計表

表 1-1 年齢(3区分) 別人口の推移—全国 (大正9年~平成22年)

年 齢	15-14歳			15-14歳			15-14歳		
	総数	15-14歳	15-14歳	総数	15-14歳	15-14歳	総数	15-14歳	15-14歳
大正9年(1920年)	55,983	26,418	31,668	5,943	732	8	108.8	38.5	38.3
16年(1925年)	69,737	31,928	34,792	5,911	899	8	109.9	38.7	38.1
昭和3年(1928年)	84,483	33,579	37,817	5,966	881	8	109.8	38.8	38.1
10年(1935年)	99,294	35,945	46,484	5,225	824	8	109.9	39.0	37.7
15年(1940年)	117,470	38,369	43,232	5,436	894	1	109.9	39.1	37.7
20年(1945年)	151,388	39,477	41,423	5,798	872	4	109.8	38.9	38.1
25年(1950年)	184,118	39,799	38,188	4,181	811	1	109.8	38.4	4.9
30年(1955年)	216,872	38,122	33,147	4,798	1,388	1	109.8	38.4	41.2
35年(1960年)	246,382	36,434	30,409	5,089	1,641	8	109.8	38.2	44.1
40年(1965年)	284,389	35,529	27,444	6,338	1,894	6	109.8	38.7	46.3
45年(1970年)	314,883	35,182	27,119	7,381	2,321	9	109.8	38.9	48.3
50年(1975年)	341,949	27,223	21,997	8,869	2,841	4	109.8	38.9	47.7
55年(1980年)	371,989	27,587	21,620	9,547	3,089	7	109.8	38.9	47.6
60年(1985年)	401,948	26,883	21,488	4,712	41	4	109.8	38.9	48.2
65年(1990年)	428,913	25,489	21,494	5,973	329	10	109.8	38.9	48.7
7年(1995年)	458,376	24,816	21,165	3,281	1,378	12	109.8	38.9	48.8
12年(2000年)	486,926	16,472	16,229	22,949	9,999	22	109.8	38.9	14.4
17年(2005年)	517,768	17,523	16,982	23,472	11,642	48	109.8	38.9	14.2
22年(2010年)	538,937	16,893	16,932	23,349	14,972	97	109.8	38.9	14.1

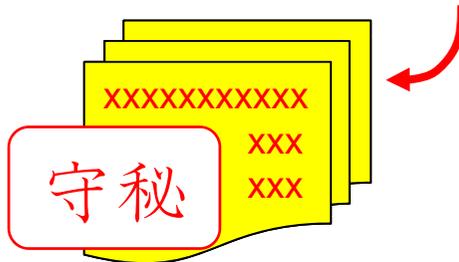
基幹統計の
公表等
(法第八条)




1. 公的統計と擬似マイクロデータ

調査票

調査票情報



調査票情報の提供
(法第三十三条)

行政機関との共同研究など
高度な公益性を有する研究等



調査票情報

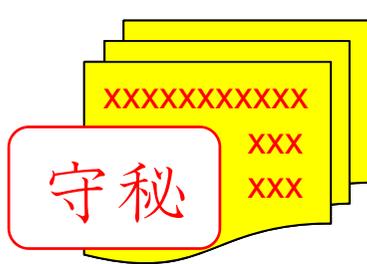
適正な管理
(法第四十二条)
提供を受けた者
の守秘義務
(法第四十三条)

秘密の漏えい 又は 不正な利益を図る
目的で提供・盗用には
罰則 (法第五十七条・第五十九条)

1. 公的統計と擬似ミクロデータ

調査票

調査票情報



匿名化
(法第三十五条)

管理

匿名データ

適正な管理
(法第四十二条)

統計表

表1-1 年齢(3区分)別人口の推移—全国(大正9年—平成22年)

年 次	年齢(千人)			割合(%)		
	総 数	0-14歳	15-64歳	0-14歳	15-64歳	65歳以上
大正9年(1920年)	55,993	20,420	32,600	3.64	58.8	37.5
昭和14年(1939年)	69,737	21,764	39,790	3.12	56.9	39.9
昭和19年(1944年)	64,480	20,070	37,667	3.09	57.6	38.4
昭和24年(1949年)	69,254	20,045	46,462	2.90	67.0	30.1
昭和29年(1954年)	73,073	20,380	48,233	2.78	66.1	29.2
昭和34年(1959年)	71,999	20,477	46,621	2.79	64.9	28.3
昭和39年(1964年)	84,133	20,780	58,150	2.47	69.1	28.4
昭和44年(1969年)	94,110	20,000	68,000	2.13	72.3	25.6
昭和49年(1974年)	104,110	19,000	78,000	1.82	74.9	23.3
昭和54年(1979年)	114,110	18,000	88,000	1.57	77.1	21.3
昭和59年(1984年)	124,110	17,000	98,000	1.37	79.0	19.7
昭和64年(1989年)	134,110	16,000	108,000	1.19	80.5	18.3
平成19年(2007年)	124,110	15,000	109,000	1.21	87.8	10.9
平成22年(2010年)	124,110	14,000	110,000	1.13	88.6	9.3

匿名データの集計結果は統計表と厳密に一致しないが、統計的分布などの情報は十分有効

匿名データの提供
(法第三十六条)



学術研究目的又は高等教育目的等学術研究、高等教育などの目的に限って利用可能(完了時に返却)

不正な利益を図る目的で提供・盗用には罰則(法第六十一条)

1. 公的統計と擬似マイクロデータ

調査票

調査票情報



調査票情報の提供
(法第三十三条)

統計処理



擬似

擬似マイクロ
(利用に法令上の
制約なし)

現在の試行提供では、
申出者及び利用者以外
の者に利用させない、
などの要件あり。

試行
提供



2. 擬似マイクロデータの試行提供

- ミクロデータ利用者の裾野を広げ、公的統計の二次的利用の拡大を図るため、擬似マイクロデータを作成し、試行提供（無償）。
- 教育目的や、公的統計の二次的利用の際のテストデータなどに利用。
- 利用要件
 - 申出者及び利用者以外の者に利用させないこと
 - 集計表から擬似的に作成したデータであるので、分析結果は実証研究の結果と見なすことができないことを理解すること
 - 利用者アンケートの提出

2. 擬似マイクロデータの試行提供

統計センターホームページ

<http://www.nstac.go.jp>

擬似マイクロデータの利用

擬似マイクロデータの利用

統計センターでは、公的統計のマイクロデータの利用を図るため、マイクロデータを用いた実証研究のプログラムテストや大学・高等学校等の授業で利用できるように擬似マイクロデータの提供を行っています。なお、擬似マイクロデータは、集計表から作成したマイクロデータ形式の擬似的なデータセットであることを理解してご利用ください。

また、擬似マイクロデータご利用の際は、アンケートにご協力いただきますようお願いいたします。

■擬似マイクロデータの作成方法

製表技術参考資料「平成24年7月「教育用擬似マイクロデータの開発とその利用 ～平成16年全国消費実態調査を例として～」」をご覧ください。

■利用可能な擬似マイクロデータ

○全国消費実態調査

年次	調査の概要	データレイアウト 及び符号表等	レコード数・ 項目数	擬似マイクロデータ
平成16年 (二人以上の世帯・勤労世帯)	調査ページ (総務省統計局)	大規模データ(XLS) 簡易データ ・CSV形式(XLS) ・EXCEL形式(XLS)	約3万2千 197項目 約8千 25項目	CSV形式(ZIP(パスワード付)) 00001-05000 *1:約4MB 05001-10000 :約4MB 10001-15000 :約4MB 15001-20000 :約4MB 20001-25000 :約4MB 25001-30000 :約4MB 30001-32027 :約2MB CSV形式 *1 (ZIP(パスワード付))約1MB EXCEL形式 *1,2 (ZIP(パスワード付))約1MB

*1 ZIPファイルの中には、擬似マイクロデータのほか、利用の手引、アンケート等も収録されております。

*2 簡易データCSV形式をMicrosoft Excel形式に変換して、変数名を付したものです。

■利用要件

以下の事項に同意していただける方なら、どなたでもご利用いただけます。

1. 擬似マイクロデータを申出者及び利用者以外の者に利用させない。
2. 擬似マイクロデータは集計表から作成したマイクロデータ形式の擬似的なデータセットであるため、

2. 擬似マイクロデータの試行提供

手続きの流れ

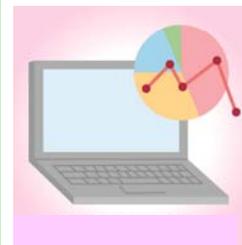
擬似マイクロデータの申出者



利用
申出書

メール
送信

統計センターHPから「擬似マイクロデータ試行提供利用申出書」の様式をダウンロードし、**所定の事項を記入した利用申出書を提出**します。



統計センターHPから擬似マイクロデータの圧縮ファイルをダウンロードし、パスワードを入力して**圧縮ファイルを解凍**します。

圧縮ファイルにある「**利用の手引き**」等を読み、**理解した上で**、擬似マイクロデータを利用してください。



アンケート

メール
送信

利用申出日から**6ヶ月以内**に「**利用者アンケート**」を統計センターに**提出**してください。

統計センター



パスワード
を通知

申出に対して擬似マイクロデータの圧縮ファイルを解凍するための**パスワードを通知**します。



アンケート

利用者アンケートを取りまとめ、今後の検討に利用させていただきます。

2. 擬似マイクロデータの試行提供

□ 提供実績

	申出 件数	利用目的				利用 人数
		大学・高 等学校等 での授業	匿名データ 提供申出前 の演習	データ処理の ためのプログ ラムテスト	その他	
2011 (平成23) 年度	45	45	-	1	-	363
2012 (平成24) 年度※	17	16	1	3	0	565

※10月末現在

2. 擬似マイクロデータの試行提供

- 試行提供中の擬似マイクロデータ
 - 平成16年全国消費実態調査 に基づくもの。

- 平成16年全国消費実態調査
 - 世帯を対象に、家計の収入・支出及び貯蓄・負債、耐久消費財、住宅・宅地などの家計資産を総合的に調査。総務省統計局が5年ごとに実施。
 - 標本の大きさ
 - 二人以上の世帯 約5万5千、単身世帯 約4千
 - 調査票
 - 世帯票、家計簿、耐久財等調査票、年収・貯蓄等調査票

2. 擬似マイクロデータの試行提供

□ 試行提供中の擬似マイクロデータ

平成16年全国消費実態調査 に基づくもの。

	レコード数	収録項目	ファイル形式
大規模データ (CSV形式)	約3万2千 二人以上の 勤労者世帯	197項目	CSV形式 (ZIP (パスワード付)) : 約 4 MB×6 +約2MB×1 (→解凍CSV 約90MB)
簡易データ (CSV形式 Excel形式)	約 8千 世帯人員が4名で 有業人員が 1~2人の世帯のみ	25項目	CSV形式 (ZIP (パスワード付)) : 約1MB (→解凍CSV 約1MB) Excel形式 (ZIP (パスワード付)) : 約1MB (→解凍XLS 約2MB)

ZIPファイルの中には、擬似マイクロデータのほか、利用の手引、アンケート等も収録

2. 擬似マイクロデータの試行提供

□ 試行提供中の擬似マイクロデータ（続き）

	レコード数	収録項目の内訳		
		世帯属性等	支出項目	収入項目
大規模データ (CSV形式)	約3万2千 二人以上の 勤労者世帯	14項目	149項目 用途分類	34項目 年間収入等
簡易データ (CSV形式 Excel形式)	約8千 世帯人員が4名で 有業人員が 1~2人の世帯のみ	世帯主の年齢、 住居の種類な ど	11項目 消費支出及 び十大費目 のみ	なし

2. 擬似マイクロデータの試行提供

□ CSV形式の内容 (テキストエディタ)

1	4	1,1,1,1,1	2,1,1,2	5	1	9.70000	125503.52963		
29496.07927			2171.59794			17453.18864	3538.19163	4493.31812	
5131.48191			10585.76081			26353.41043	5676.85003	20603.65086	
1	4	1,1,1,1,1	2,1,1,2	5	1	16.23333	255675.88385		
25806.18465			0.00000			18568.59360	5809.80769	6768.05621	
3939.57305			58687.22395			29876.50740	4591.27428	101628.66302	
1	4	1,1,1,1,1	2,1,1,2	5	1	11.83333	175320.42725		
38278.15986			0.00000			17353.30482	6131.40784	11832.53266	
4359.21333			21783.82996			0.00000	12110.00812	63471.97067	
1	4	1,1,1,1,1	2,1,1,3	6	1	12.33333	181085.60342		
74122.05211			0.00000			18132.74714	2513.99975	6548.90369	
1033.04333			18768.08459			21696.58196	8414.43204	29855.78940	
1	4	1,1,1,1,1	2,1,1,3	6	1	11.83333	124470.60227		
33256.83553			329.57746			19593.37742	2192.72008	7918.74866	
2579.98683			13006.90834			0.00000	7617.79925	37975.12171	
1	4	1,1,1,1,1	2,1,1,3	6	1	2.33333	45717.72443		
46992.78709			0.00000			10541.57900	3538.09863	18153.30977	
2217.98300			13822.95746			22233.58608	4286.65482	20930.36576	
1	4	1,1,1,1,1	2,1,1,3	6	1	11.15556	319114.26656		
113177.05416			0.00000			18981.89588	11509.43625	7794.08427	
21715.17824			3936.47340			31640.38994	26268.68240		
47898.76767									
1	4	1,1,1,1,1	2,1,1,3	6	1	11.15556	253685.20384		
67253.62324			341.42010			20883.40328	8921.15560	1611.43356	
8876.36799			50094.30851			20909.89900	32464.40428	42329.18827	
1	4	1,1,1,1,1	2,1,1,3	6	1	11.15556	236447.60216		
61129.75558			55.01982			22640.16161	8468.43328	18744.03002	

2. 擬似マイクロデータの試行提供

□ CSV形式の内容 (Excel)

The screenshot displays a Microsoft Excel spreadsheet titled "GIJI_2004zensho_s_dataset.csv". The spreadsheet contains a table with columns labeled M, N, O, P, Q, R, S, T, and U. The data is organized into rows numbered 1 through 20. A large pink watermark reading "SAMPLE" is overlaid diagonally across the center of the data.

	M	N	O	P	Q	R	S	T	U
1	1	9.7	125503.5	29496.08	2171.598	17453.19	3538.192	4493.318	5000.000
2	1	16.23333	255675.9	25806.18	0	18568.59	5809.808	6768.056	3000.000
3	1	11.83333	175320.4	38278.16	0	17353.3	6131.408	11832.53	4000.000
4	1	12.33333	181085.6	74122.05	0	18132.75	2513.969	6548.904	1000.000
5	1	11.83333	124471	33256.84	329.5775	19593.37	2192.72	7918.749	2000.000
6	1	12.33333	145717.7	46992.79	0	13541.98	3538.099	18153.31	2000.000
7	1	11.15556	319114.3	113177.1	263.3044	18981.9	11509.44	7794.084	2000.000
8	1	11.15556	253685.2	67253.62	341.4201	20883.4	8021.156	1611.434	8000.000
9	1	11.15556	236447.6	61129.76	55.01982	22640.0	8468.433	18744.03	5000.000
10	1	7.1	137315.3	27050.12	9256	12102.0	11367.79	5031.214	4000.000
11	1	10.28889	253393.7	47205.61	0	20075.7	17211.18	3666.61	6000.000
12	1	10.28889	232141.8	52259.7	0	14041.5	14000.91	8397.238	1000.000
13	1	10.28889	214540.4	54921.9	0	16455.4	10619.16	1319.457	6000.000
14	VV	17.56667	21151.4	7491.05	0	12961.09	6112.253	13526.81	4000.000
15	VV	17.56667	276131	78916.1	110.6904	18252.02	12729.08	10836.02	9000.000
16	VV	8.53333	117150.7	73909.58	0	12429.77	9161.74	2833.949	3000.000
17	1	20.0	118895.1	48821.55	408.5446	16905.66	3461.918	2435.489	1000.000
18	1	13.0	130482.8	47798.46	0	12863.9	7354.453	16218.96	1000.000
19	1	7.9	147969.1	50277.93	308.9679	0	5224.538	6803.324	2000.000
20	1	11.73333	150973.7	48290.97	0	16516.59	1177.606	0	6000.000

2. 擬似マイクロデータの試行提供

kyoikugiji_zenshoH16_futari_kinrou_fugo.xls [互換モード] - Microsoft Excel

行番号	項目名	階層	位置	バイト数	配置	型	種別	変数名	符号	符号内容
77	69 企業区分	4	26	1				S1_KigyokuKubun	1	民営
78	70								2	自営
79	71								3	官公
80	72								△	非就業
81	73								v	不詳
82	74	4	27	1				,	,	
83	75 企業規模	4	28	1				S1_KigyokuKibo	1	1~4人
84	76								2	5~29人
85	77								3	30~499人
86	78								4	500~999人
87	79								5	1000人以上
88	80								△	非就業又は官公
89	81								v	不詳
90	82								,	
91	83	4	30	2				S1_Sangyou	1	農業
92	84								△2	林業
93	85								△3	漁業
94	86								△4	鉱業
95	87								△5	建設業
96	88								△6	製造業
97	89								△7	電気・ガス・熱供給・水道業
98	90								△8	情報通信業
99	91								△9	運輸業
100	92								△0	卸売・小売業
101	93								△1	金融・保険業

データレイアウトフォームや
符号票から内容を読み取り

参考 「政府統計個票データレイアウト標準記法」
(平成18年3月31日各府省情報化統括責任者
(CIO) 連絡会議決定)

2. 擬似マイクロデータの試行提供

□ Excel形式の内容 (Excel)

GIJI_2004zensho_s_dataset.xls [互換モード] - Microsoft Excel

M	N	O	P	Q	R	S
S1_Shokugyou	Weight	Youto037	Youto038	Youto079	Youto084	Youto088
1	9.70000	125503.52963	29496.07927	2171.59794	17453.18864	3538.18864
2	16.23333	255675.88385	25806.18465	0.00000	18568.59360	5809.18864
3	11.89333	175320.42725	38278.15986	0.00000	17353.30482	6131.18864
4	12.33333	181085.60342	74122.05211	0.00000	18132.74714	2513.18864
5	11.89333	124470.99227	33256.83553	329.57746	19593.37442	2192.18864
6	12.33333	145717.72443	1992.8709	0.00000	1341.97903	3538.18864
7	11.15556	319114.1156	1177.1111	13.3015	18981.89588	11509.18864
8	11.15556	53685.2384	8753.6124	341.42010	20883.40328	8921.18864
9	11.15556	23147.6016	6119.7558	55.01982	22640.16161	8468.18864
10	7.10000	13755.29619	2705.12036	9256.00000	12102.19781	17367.18864
11	10.28889	83333.33333	47205.60506	0.00000	20075.71411	17211.18864
12	10.28889	232141.60008	52259.56985	0.00000	14041.50459	14099.18864
13	10.28889	214540.38409	54920.91358	0.00000	16455.13848	10619.18864
14	9.36667	234151.44962	74993.04794	0.00000	12961.08694	6112.18864
15	9.36667	278431.02315	78916.10390	110.69039	18252.02402	12729.18864
16	13.53333	197180.29633	72903.58386	0.00000	12429.76888	9161.18864
17	13.53333	197180.29633	72903.58386	0.00000	12429.76888	9161.18864
18	13.53333	197180.29633	72903.58386	0.00000	12429.76888	9161.18864
19	13.53333	197180.29633	72903.58386	0.00000	12429.76888	9161.18864
20	13.53333	197180.29633	72903.58386	0.00000	12429.76888	9161.18864
21	11.73333	150373.67450	48290.96808	0.00000	16516.58993	1177.18864
22	11.73333	217478.74110	38505.04816	0.00000	13954.06130	2736.18864
23	13.33333	141490.28719	47930.97663	0.00000	12474.59169	7002.18864

○ 変数名を、第1行目 (ヘッダー) に付与

○ 小数点以下の桁数揃え

○ 符号の内容は、符号表から読み取り

2. 擬似マイクロデータの試行提供

The screenshot shows two overlapping browser windows. The left window displays the main website navigation and a list of links for the survey overview. The right window shows the detailed overview page with a table of contents and a section titled '1 調査の目的' (1. Purpose of the Survey).

平成16年全国消費実態調査 調査の概要

目次

1 調査の目的	5 抽出方法	個人収支簿による調査の概要
2 調査の沿革	6 調査事項	調査の変遷
3 調査の根拠法令	7 調査の時期及び調査の方法	
4 調査の対象	8 調査の流れ	

1 調査の目的

全国消費実態調査は、国民生活の実態について、家計の消費行動（消費、貯蓄、耐久消費財、住宅・宅地などの家計資産を総合的に調査し、全国及び世帯別の消費行動の傾向、構造、分布などを明らかにすることを目的とした調査である。

この種の調査としては毎月実施されている家計調査があるが、その主な目的が全国平均の家計収支の時系列の動きを明らかにすることにあるため、調査規模が約8,000世帯と小さく、詳細な構造分析を行うことができない。そこ

マイクロデータ利用にも調査の概要に関する理解は必須

19

ページが表示されました インターネット | 保護モード: 有効 100%

2. 擬似マイクロデータの試行提供

統計局ホームページ/平成16年全国消費実態調査 調査票 - Wind

http://www.stat.go.jp/data/zensho/2004/form

総務省
統計局・政策統括官(統計基)

サイト内検索

ホーム 統計データ

平成16年全国消費実態調査

- 家計簿A, 家計簿B (PDF:2.365KB)
- 耐久財等調査票 (PDF:1.635KB)
- 年収・貯蓄等調査票 (PDF:1.46KB)
- 世帯票 (PDF:1.882KB)
- 個人収支簿, 家計簿C (PDF:1.0KB)

調査票も一読を

ページが表示されました

統計センター
NSTAC

setai.pdf (保護) - Adobe Acrobat Standard

ファイル(E) 編集(E) 表示(V) 文書(D) 注釈(C) フォーム(O) ツール(I) アドバンス(A) ウィンドウ(W) ヘルプ(H)

秘 平成16年全国消費実態調査 1 郵便庁市区町村番号 世帯番号 一世帯番号 世帯の別 世帯区分 抽出区分 世帯

この調査票は、統計以外の目的には使用しませんから、ありのままを記入してください。

秘 指定統計 第 97 号

平成 16 年 全 国 消 費 実 態 調 査

家 計 簿 A

収入の種類又は支出の品名及び用途	現金収入 〔預貯金引き出し〕 〔借入金を含む〕 (円)	現金支出 〔収入からの控除 預貯金預 け入れ 借入金返済を含む〕 (円)	総務省 統計局 記入欄
1			20
2			
3			

3. 擬似マイクロデータの特徴

□ 基本統計量 (平均、標準偏差)

	平均				標準偏差		
	調査票情報	擬似マイクロ	差率		調査票情報	擬似マイクロ	差率
年間収入	740.18	729.81	-0.01	年間収入	358.18	337.69	-0.06
収入総額	971,789.24	946,779.03	-0.03	収入総額	541,290.74	473,480.73	-0.13
実収入	502,133.73	497,655.92	-0.01	実収入	280,695.92	261,558.27	-0.07
実収入以外の収入	391,823.98	372,130.47	-0.05	実収入以外の収入	353,922.37	263,445.65	-0.26
繰入金	77,831.53	76,992.65	-0.01	繰入金	87,036.21	98,947.04	0.14
支出総額	971,789.24	946,779.03	-0.03	支出総額	541,290.74	473,480.73	-0.13
実支出	415,809.39	403,746.63	-0.03	実支出	224,419.69	219,290.60	-0.02
消費支出	339,199.37	328,139.70	-0.03	消費支出	194,501.15	192,447.21	-0.01
食料	73,738.54	72,883.42	-0.01	食料	30,149.02	28,064.49	-0.07
住居	19,387.99	17,687.21	-0.09	住居	52,962.36	60,587.32	0.14
光熱・水道	19,395.36	19,237.81	-0.01	光熱・水道	8,009.23	7,690.12	-0.04
家具・家事用品	9,783.81	9,204.04	-0.06	家具・家事用品	15,977.65	14,933.13	-0.07
被服及び履物	14,649.44	14,137.63	-0.03	被服及び履物	18,837.04	19,823.09	0.05
保健医療	11,936.01	11,366.36	-0.05	保健医療	19,763.39	19,284.07	-0.02
交通・通信	50,740.68	47,960.92	-0.05	交通・通信	85,021.69	84,654.38	0.00
教育	22,332.15	22,269.65	0.00	教育	51,989.72	64,157.45	0.23
教養娯楽	32,472.95	31,389.49	-0.03	教養娯楽	32,161.60	32,723.04	0.02
その他の消費支出	84,762.44	82,003.18	-0.03	その他の消費支出	95,898.83	102,040.97	0.06
非消費支出	76,610.02	75,606.93	-0.01	非消費支出	56,199.75	66,378.49	0.18
実支出以外の支出	475,947.80	464,318.09	-0.02	実支出以外の支出	394,805.29	334,227.09	-0.15
繰越金	80,032.04	78,714.31	-0.02	繰越金	96,421.45	118,055.82	0.22

差率：(擬似マイクロ-調査票情報) / 調査票情報

-0.5より小、0.5より大
-0.5以上-0.1より小、0.1より大0.5以下

単位 年間収入:万円
年間収入以外:円

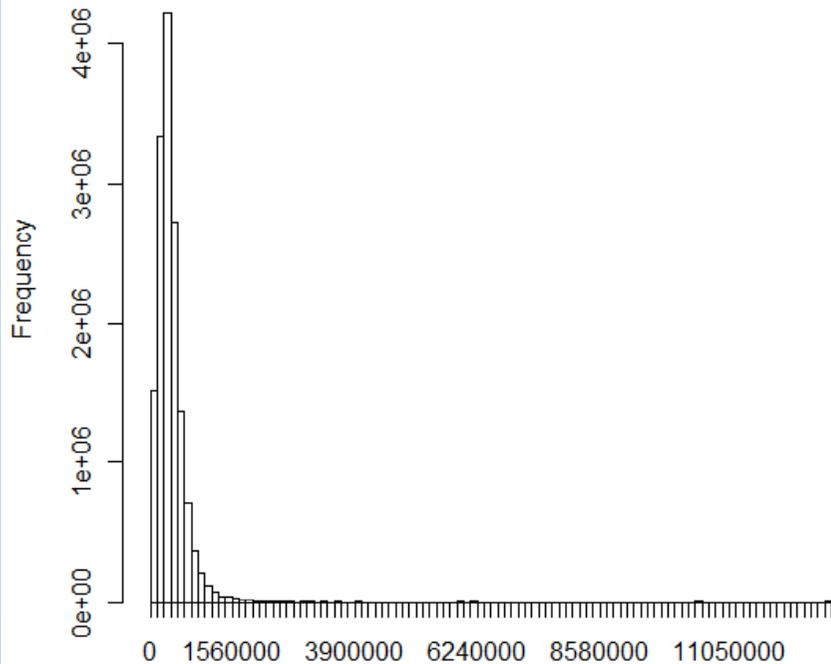
3. 擬似マイクロデータの特徴

- ヒストグラム 標準偏差の差率が最大となる「実収入以外の収入」に関する分布の比較

実収入以外の収入

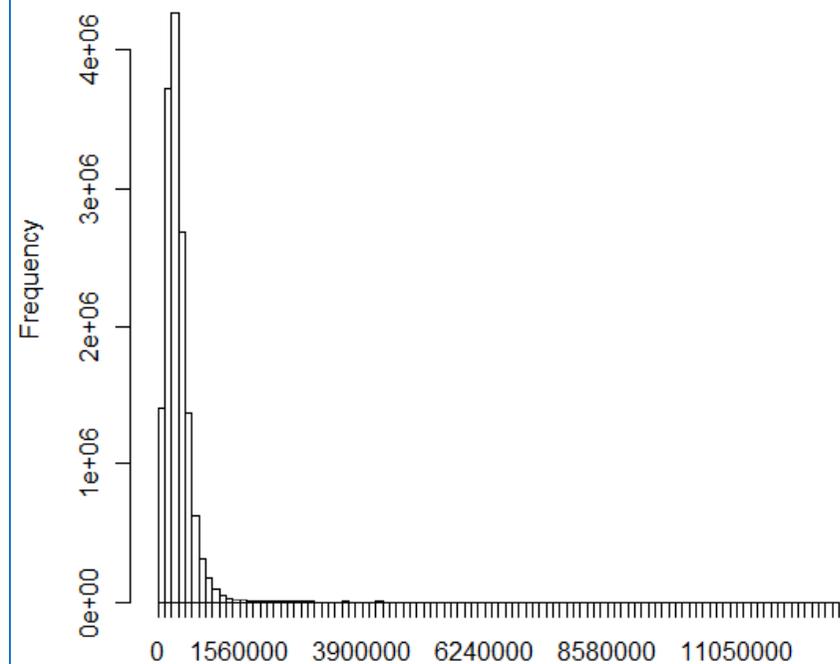
調査票情報

平均 391,823.98
標準偏差 353,922.37



擬似マイクロデータ

平均 372,130.47
標準偏差 263,455.65



3. 擬似マイクロデータの特性

量的属性間の相関係数 調査票情報

赤字 符号が変化
 差：教育用擬似マイクロデーター個票データ
 -0.2より小、0.2より大
 -0.2以上-0.1より小、0.1より大0.2以下

調査票情報

	年間収入	収入総額	実収入	実収入以外の収入	繰入金	支出総額	実支出	消費支出	食料	住居	光熱・水道	家具・家事用品	被服及び履物	保健医療	交通・通信	教育	教養娯楽	その他の消費支出	非消費支出	実支出以外の支出	繰越金	
年間収入	1.00																					
収入総額	0.60	1.00																				
実収入	0.66	0.78	1.00																			
実収入以外の収入	0.35	0.85	0.36	1.00																		
繰入金	0.19	0.26	0.14	0.04	1.00																	
支出総額	0.60	1.00	0.78	0.85	0.26	1.00																
実支出	0.60	0.73	0.56	0.63	0.17	0.73	1.00															
消費支出	0.49	0.66	0.45	0.61	0.16	0.66	0.97	1.00														
食料	0.47	0.42	0.37	0.31	0.17	0.42	0.52	0.50	1.00													
住居	-0.02	0.11	0.00	0.16	0.01	0.11	0.24	0.28	-0.03	1.00												
光熱・水道	0.32	0.24	0.22	0.16	0.11	0.24	0.28	0.27	0.44	-0.07	1.00											
家具・家事用品	0.15	0.25	0.12	0.26	0.09	0.25	0.26	0.27	0.17	0.07	0.10	1.00										
被服及び履物	0.30	0.30	0.24	0.24	0.10	0.30	0.39	0.38	0.29	0.02	0.12	0.16	1.00									
保健医療	0.11	0.16	0.10	0.15	0.07	0.16	0.24	0.25	0.15	0.01	0.07	0.08	0.09	1.00								
交通・通信	0.14	0.33	0.15	0.37	0.04	0.33	0.54	0.57	0.12	0.01	0.05	0.05	0.10	0.06	1.00							
教育	0.18	0.23	0.15	0.23	0.03	0.23	0.37	0.39	0.24	-0.03	0.19	0.02	0.09	0.04	0.07	1.00						
教養娯楽	0.32	0.35	0.27	0.30	0.12	0.35	0.44	0.42	0.32	0.02	0.10	0.15	0.26	0.10	0.10	0.09	1.00					
その他の消費支出	0.39	0.46	0.38	0.37	0.12	0.46	0.66	0.66	0.21	0.01	0.13	0.12	0.19	0.11	0.12	0.04	0.16	1.00				
非消費支出	0.70	0.63	0.70	0.38	0.12	0.63	0.62	0.43	0.35	-0.02	0.19	0.12	0.26	0.08	0.17	0.14	0.29	0.34	1.00			
実支出以外の支出	0.44	0.90	0.72	0.79	0.04	0.90	0.40	0.32	0.25	0.01	0.14	0.18	0.17	0.08	0.14	0.11	0.22	0.23	0.49	1.00		
繰越金	0.16	0.24	0.13	0.06	0.86	0.24	0.13	0.12	0.13	0.02	0.10	0.07	0.07	0.05	0.02	0.02	0.08	0.10	0.10	0.01	1.00	

3. 擬似マイクロデータの特徴

赤字 符号が変化
差：教育用擬似マイクロデーター個票データ
-0.2より小、0.2より大
-0.2以上-0.1より小、0.1より大0.2以下

量的属性間の相関係数 擬似マイクロ

	年間収入	収入総額	実収入	実収入以外の収入	繰入金	支出総額	実支出	消費支出	食料	住居	光熱・水道	家具・家事用品	被服及び履物	保健医療	交通・通信	教育	教養娯楽	その他の消費支出	非消費支出	実支出以外の支出	繰越金	
年間収入	1.00																					
収入総額	0.58	1.00																				
実収入	0.63	0.85	1.00																			
実収入以外の収入	0.38	0.83	0.48	1.00																		
繰入金	0.12	0.32	0.15	0.05	1.00																	
支出総額	0.58	1.00	0.85	0.83	0.32	1.00																
実支出	0.52	0.71	0.59	0.64	0.14	0.71	1.00															
消費支出	0.42	0.63	0.49	0.60	0.14	0.63	0.96	1.00														
食料	0.46	0.40	0.36	0.32	0.13	0.40	0.45	0.43	1.00													
住居	-0.05	0.08	0.04	0.09	0.03	0.08	0.24	0.28	-0.06	1.00												
光熱・水道	0.32	0.25	0.23	0.18	0.09	0.25	0.26	0.25	0.44	-0.07	1.00											
家具・家事用品	0.12	0.15	0.11	0.14	0.04	0.15	0.19	0.19	0.15	0.00	0.10	1.00										
被服及び履物	0.21	0.23	0.19	0.20	0.06	0.23	0.29	0.28	0.20	0.01	0.08	0.12	1.00									
保健医療	0.07	0.13	0.09	0.13	0.04	0.13	0.19	0.20	0.11	0.00	0.06	0.05	0.05	1.00								
交通・通信	0.12	0.30	0.17	0.35	0.04	0.30	0.50	0.54	0.10	-0.01	0.05	0.03	0.06	0.04	1.00							
教育	0.14	0.24	0.18	0.24	0.02	0.24	0.38	0.41	0.18	-0.02	0.16	0.01	0.04	0.02	0.04	1.00						
教養娯楽	0.26	0.30	0.24	0.28	0.06	0.30	0.35	0.34	0.26	-0.01	0.06	0.12	0.18	0.07	0.07	0.05	1.00					
その他の消費支出	0.33	0.44	0.38	0.37	0.11	0.44	0.63	0.65	0.17	-0.02	0.11	0.07	0.11	0.06	0.09	0.04	0.10	1.00				
非消費支出	0.50	0.50	0.52	0.35	0.07	0.50	0.53	0.26	0.24	-0.04	0.14	0.07	0.14	0.05	0.09	0.07	0.18	0.21	1.00			
実支出以外の支出	0.45	0.85	0.77	0.74	0.07	0.85	0.32	0.25	0.25	-0.05	0.15	0.08	0.13	0.05	0.09	0.09	0.18	0.18	0.35	1.00		
繰越金	0.10	0.28	0.14	0.05	0.82	0.28	0.07	0.07	0.09	0.00	0.08	0.03	0.03	0.02	0.01	0.00	0.02	0.06	0.04	0.00	1.00	

擬似マイクロデータ

3. 擬似マイクロデータの特徴

□ 散布図

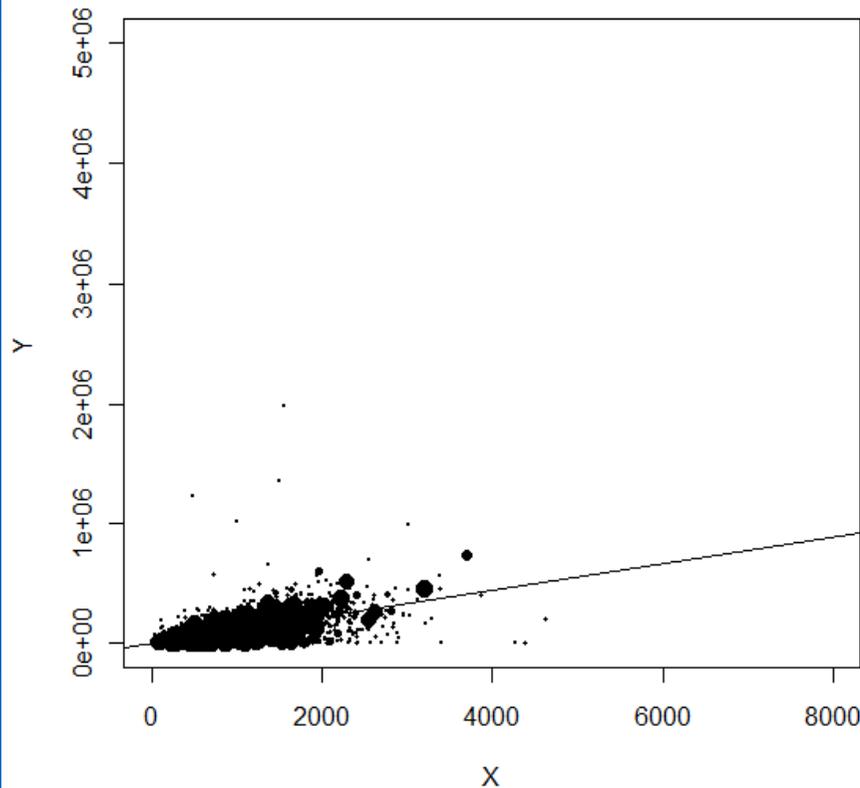
調査票情報と擬似マイクロの間で、相関係数の差が最も大きい「年間収入」と「非消費支出」の散布図。

年間収入(x) × 非消費支出(y)

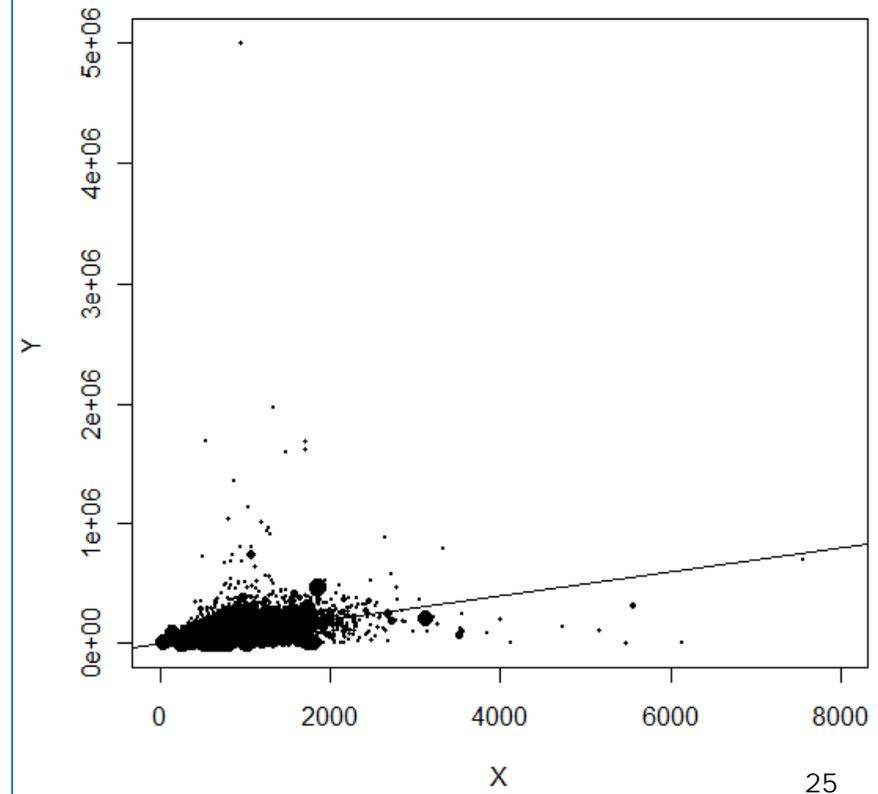
調査票情報

擬似マイクロデータ

相関係数: 0.70



相関係数: 0.50



3. 擬似マイクロデータの特性

□ クロス集計表

公表値（平成16年 全国消費実態
調査報告第13表）を参考に作成

差：（教育用擬似マイクロデータ－個票データ）／個票データ
 -0.5より小、0.5より大
 -0.5以上-0.1より小、0.1より大0.5以下

世帯人員2人、年間収入階級別1世帯当たり1か月間の支出

	平均	年間収入階級（万円）									
		200未満	200 ～300	300 ～400	400 ～500	500 ～600	600 ～800	800 ～1000	1000 ～1250	1250 ～1500	1500以上
食料	60,984	38,439	46,341	51,449	55,430	59,147	64,313	68,487	71,419	81,540	92,303
住居	23,905	25,703	22,623	23,331	26,316	24,298	22,386	24,143	27,417	22,171	15,331
光熱・水道	15,096	12,304	13,709	14,515	14,544	14,654	15,349	15,822	16,155	16,640	20,022
家具・家事用品	9,286	4,398	5,549	7,483	7,368	8,698	10,255	12,056	10,972	13,492	14,796
被服及び履物	13,654	5,904	6,825	7,955	8,821	11,513	14,737	17,283	22,484	26,940	37,843
保健医療	11,134	5,501	8,175	9,444	10,006	11,713	11,314	13,163	12,805	14,661	16,623
交通・通信	45,703	20,996	25,966	32,570	38,339	46,288	46,286	55,709	63,407	78,109	81,492
教育	1,213	3,641	2,530	1,260	836	1,153	785	697	1,940	887	1,142
教養娯楽	30,863	11,051	16,564	18,288	23,541	26,479	34,068	37,126	51,438	56,794	67,421
その他の消費支出	89,539	27,955	38,910	51,071	60,584	76,484	97,967	120,000	153,095	163,964	208,602

	平均	年間収入階級（万円）									
		200未満	200 ～300	300 ～400	400 ～500	500 ～600	600 ～800	800 ～1000	1000 ～1250	1250 ～1500	1500以上
食料	60,342	40,970	44,783	50,789	55,577	59,450	62,807	68,987	72,671	77,010	87,092
住居	22,241	23,467	25,408	22,322	24,755	24,034	19,762	25,761	15,321	15,644	9,864
光熱・水道	15,062	12,565	13,113	14,135	14,410	15,090	15,190	16,060	16,440	17,481	19,122
家具・家事用品	8,550	4,413	6,072	7,048	7,472	8,272	9,069	9,775	10,675	13,353	13,103
被服及び履物	13,215	7,535	6,748	7,746	9,586	11,361	14,166	18,306	20,846	24,587	29,413
保健医療	10,386	7,824	7,702	8,618	9,970	9,575	10,846	12,150	12,488	13,335	13,873
交通・通信	43,772	23,271	24,048	34,239	35,225	44,714	46,207	53,821	61,657	59,127	79,587
教育	1,337	4,640	4,141	1,196	1,489	830	744	398	634	0 ²⁶	10,052
教養娯楽	30,194	11,654	15,620	19,395	23,153	28,587	32,610	40,740	43,390	50,752	60,672
その他の消費支出	85,264	26,578	38,004	52,043	59,657	77,723	90,088	116,778	142,395	160,150	216,446

調査票情報（公表値）

擬似マイクロデータ

3. 擬似マイクロデータの特性

□ 回帰

用途項目の1か月当たり支出金額(円) = f (消費支出(円))

	消費支出			定数			決定係数	調整済決定係数	F値
	係数	標準誤差	P値	係数	標準誤差	P値			
食料	0.07681	0.00075	0.000	47,683.0	294.2	0.000	0.2456	0.2456	10,425
住居	0.07649	0.00146	0.000	-6,558.3	571.0	0.000	0.0789	0.0789	2,744
光熱・水道	0.01123	0.00022	0.000	15,585.8	86.6	0.000	0.0744	0.0744	2,574
家具・家事用品	0.02205	0.00044	0.000	2,305.1	172.9	0.000	0.0720	0.0720	2,486
被服及び履物	0.03669	0.00050	0.000	2,202.6	195.8	0.000	0.1436	0.1435	5,368
保健医療	0.02551	0.00055	0.000	3,284.6	214.9	0.000	0.0630	0.0630	2,153
交通・通信	0.25071	0.00200	0.000	-34,299.9	782.4	0.000	0.3289	0.3289	15,699
教育	0.10323	0.00138	0.000	-12,682.5	538.7	0.000	0.1491	0.1491	5,613
教養娯楽	0.06986	0.00084	0.000	8,775.9	327.5	0.000	0.1785	0.1785	6,959
その他の消費支出	0.32741	0.00206	0.000	-26,296.4	805.5	0.000	0.4410	0.4410	25,262
	消費支出			定数			決定係数	調整済決定係数	F値
	係数	標準誤差	P値	係数	標準誤差	P値			
食料	0.06232	0.00074	0.000	52,433.5	280.3	0.000	0.1826	0.1826	7,156
住居	0.08949	0.00169	0.000	-11,679.6	641.7	0.000	0.0808	0.0808	2,815
光熱・水道	0.01002	0.00022	0.000	15,951.0	82.2	0.000	0.0628	0.0628	2,147
家具・家事用品	0.01509	0.00043	0.000	4,251.4	161.8	0.000	0.0378	0.0378	1,259
被服及び履物	0.02908	0.00055	0.000	4,594.6	210.1	0.000	0.0797	0.0797	2,774
保健医療	0.01958	0.00055	0.000	4,940.7	208.9	0.000	0.0382	0.0382	1,271
交通・通信	0.23652	0.00207	0.000	-29,652.4	788.4	0.000	0.2891	0.2891	13,023
教育	0.13728	0.00170	0.000	-22,777.2	645.8	0.000	0.1696	0.1695	6,539
教養娯楽	0.05808	0.00089	0.000	12,331.1	339.7	0.000	0.1167	0.1166	4,230
その他の消費支出	0.34253	0.00226	0.000	-30,392.9	860.4	0.000	0.4173	0.4173	22,934

調査票情報

擬似マイクロデータ

3. 擬似マイクロデータの特徴

□ 参考

- 「教育用擬似マイクロデータの開発とその利用
～平成16年全国消費実態調査を例として～」
(製表技術参考資料 平成24年7月 (独) 統計センター)

<http://www.nstac.go.jp/services/reference.html>

□ 擬似マイクロデータ作成における検討課題

- 多変量対数正規乱数を発生させた場合に、実際の個票データでは出現していない範囲を超えた数値が作成されてしまうこと。
- 高次元の集計表の中で度数が少ないセルのデータについて、個票データの分布との乖離が大きくなってしまうこと。
- セル内の度数に対応するレコードの属性値群のすべてが、基本的な量的属性値の内訳に関して同一の構成比とならないように、レコードによって構成比を変える必要性あり。
- 新たな擬似マイクロデータの作成。
(今回は、質的属性が多い統計調査に関するもの)

4. 擬似マイクロデータの課題

□ 公的統計における制度上の位置付け

「統計データの二次的利用促進に関する研究会 平成23年度報告書」(抄)
平成24年7月 総務省政策統括官室(統計基準担当)

さらに、諸外国における政府統計データの活用の状況について幅広くとらえてみると、ニュージーランド統計局では、公表されている集計結果表から作成されたマイクロデータが利用制限なしに提供されていることがわかった。日本では、統計センターにおいて、同様の観点から作成する擬似マイクロデータの研究が行われている。

(中略)

このため、このような新たな技術的手法による擬似マイクロデータについて、その研究を進めるとともに、総務省においてその制度上の位置づけ等について検討を行うことが求められる。

4. 擬似マイクロデータの課題

□ 公的統計における制度上の位置付け

「平成23年度統計法施行状況に関する審議結果報告書」(抄)
平成24年9月25日 内閣府 統計委員会

(イ) 施策の進捗状況等に対する評価

○また、法第33条第2号による調査票情報の利用については、調査対象の識別可能性や情報漏えいのリスクへの対応として厳格な運用が求められている中で、オンサイト利用や、匿名データ・教育用擬似マイクロデータとの関係整理も必要である。

(ウ) 今後の施行の方向性等についての基本的考え方

○また、二次的利用を取り巻く諸課題については、総務省の研究会における検討情報を注視していくこととする。なお、オンサイト利用や教育用擬似マイクロデータの検討に当たっては、コストやその負担、国民の理解や研究者に対する国民の信頼感にも密接に関係することに留意が必要である。

4. 擬似マイクロデータの課題

- 擬似マイクロデータの利活用のために
 - 擬似マイクロデータを用いて行った授業の資料
 - 擬似マイクロデータを利用するために作成したプログラム（CSVデータインポート、クロス集計、回帰計算等） など

上のようなコンテンツへ、統計センターウェブ（擬似マイクロのコーナー）からリンクを張らせていただきたいと思います。その際は、ご理解、ご協力をお願いします。

- ※ 擬似マイクロデータの全部又は一部などをウェブで提供する行為は、控えて下さい。
統計センターが試行提供している擬似マイクロデータは、統計センターから直接入手してください（申出者・利用者による第三者提供は禁止）。

ご清聴ありがとうございました。

意見、見解に及ぶものは、本人のものであり、
所属する組織を代表するものではありません。