
全国消費実態調査(2004年)の匿名データから JMP による新擬似マイクロデータの作成

○高橋 行雄¹, 周防 節雄², 宮内 亨³

¹BioStat 研究所(株), ²(公財)統計情報研究開発センター,
³(独)統計センター

2017.11.17 高橋行雄

1

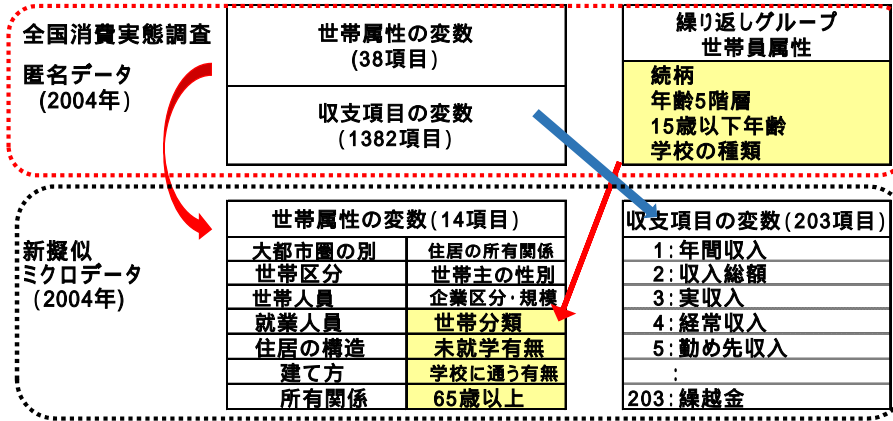
新擬似マイクロデータ開発の経緯

- ◆ 教育用擬似マイクロデータを用いて「SASユーザー総会のデータコンペ」を過去4回開催してきたが、2016年度末で提供打切りとなった。
- ◆ そこで、2004年全国消費実態調査の匿名データを用いて、SASユーザー会世話人有志が新擬似マイクロデータを統計ソフトJMPを用いて新規作成し、2017年のコンテスト用に供した。

2017.11.17 高橋行雄

2

匿名データから新擬似マイクロデータの作成



複数の世帯員属性から世帯分類などの項目を設定

2017.11.17 高橋行雄

3

世帯属性14変数 (匿名データ)

変数名	符号	符号内容	世帯数	変数名	符号	符号内容	世帯数	変数名	符号	符号内容	世帯数
Year		調査年(西暦)		X05		住居の構造		X10		企業区分・規模	
No		世帯番号			1	木造(防火木造含む)	34,266		1	民間・自営1-4人	8,107
X01		大都市圏の別			2	コンクリート他	13,531		2	# 5-29人	6,221
	1	3大都市圏	28,410	X06		住居の建て方			3	# 30-499人	8,851
	0	その他	19,387		1	一戸建	36,132		4	# 500人以上	7,105
X02		世帯区分			2	一戸建以外	11,665		5	官公	5,870
	1	勤労者世帯	27,443	X07		住居の所有関係			6	無職	11,643
	2	勤労者以外の世帯	8,711		1	持ち家	36,132	X11		世帯分類	
	3	無職世帯	11,643		2	持ち家以外	11,665		1	単身世帯	3,936
X03		世帯人員		X08		世帯主の性別			2	夫婦のみ	13,447
	1	1人	27,443		1	男	41,831		3	2世代	20,034
	2	2人	8,711		2	女	5,966		4	2世代(ひとり親)	4,191
	3	3人	11,643	X09		世帯主の年齢			5	3世代	5,139
	4	4人	10,315		5	24歳以下	482		6	その他	1,050
	5	5人以上	6,932		6	25-29歳	1,622	X12		未就学	
X04		就業人員			:	:			1	無	36,531
	0	0人	8,651		16	75歳以上	3,534		2	有	7,330
	1	1人	18,076	X14		65歳以上人員		X13		学校	
	2	2人	15,397		0	0人	29,040		1	無	31,685
	3	3人以上	5,673		1	1人	10,084		2	有	16,112
					2	2人以上	8,673	Weight		集計用乗率	

符号の併合, 再分類, 組合せ再区分を適宜行った.

2017.11.17 高橋行雄

4

14次元クロス表でセル度数を3以上に

X01	X02	...	X14	セル 件数	14次 元番号	セル 件数	セル 行数	セル 件数	セル 行数	セル 件数	セル 行数
1	1		1	19	1	1	2	1	8,406	70	79
1	1		2	1	2	2	1	2	2,261	80	89
1	2		1	29	3	3	1	3	1,011	90	99
1	2		2	1	4	19	1	4	576	110	119
2	1		1	39	5	29	1	5	370	120	129
2	1		2	2	6	39	1	6	292	154	
2	2		1	49	7	49	1	7	220	172	
2	2		2	3	8			8	171	215	
:				:	:			9	114	309	
				14,246		計	8	10	19	509	328
								20	29	149	342
								30	39	69	376
								40	49	28	386
								50	59	23	721
								60	69	12	780
								計			14,246

リスト形式の14次元クロス表
の形式とセル度数の例示

セル度数が1: 8,406セル, セル度数が2: 2,261セル

2017.11.17 高橋行雄

5

マージした匿名データからのコピー

匿名・14次元クロス表					匿名データにマッチ・マージ					匿名データからのコピー									
X 01	X 02	...	X 14	セル 件数	14 次元 番号	セル 件数	14 次元 番号	X 01	X 02	...	X 14	レコ ード 番号	セル 件数	14 次元 番号	X 01	X 02	...	X 14	レコ ード 番号
1	1		1	19	1	1	2	1	1		2	1	1	2	1	1		2	1
1	1		2	1	2	2	1	1	1		2	2	1	4	1	2		2	3
1	2		1	29	3	1	4	1	2		2	3							8,406
1	2		2	1	4	2	6	2	1		2	4	1	2	1	1		2	1
2	1		1	39	5	2	6	2	1		2	5	1	4	1	2		2	3
2	1		2	2	6	2	6	2	1		2	6	1	4	1	2			8,406
2	2		1	49	7	2	6	2	1		2	7							
2	2		2	3	8	3	8	2	2		2	8	2	6	2	1		2	5
:				:	:	3	8	2	2		2	9	2	6	2	1		2	7
				14,246		3	8	2	2		2	10	2	6	2	1			
						3	8	2	2		2	11	2	6	2	1			
						3	8	2	2		2	:							2,261x2=4,522
												47,797							

14変数で匿名データとマッチマージし, 14次元番号, 出現頻度を付与

2017.11.17 高橋行雄

6

203個の支項収目に誤差変動を付与

- ◆ 全抽出 21,334 レコードでの203収支項目
 - 匿名 47,797 レコードと併合した場合でも, 14 次元別の平均値は, 匿名データと同じ値となる
 - 何らかの誤差変動を与える必要がある
- ◆ 常用対数変換後に正規乱数で誤差を付与
 - 平均 0.0, 標準偏差 0.02 の正規乱数
 - $\log_{10}(1,000 \text{ 円}) = 3.00$
 - $3.00 + 0.02 = 3.02$, (0.02 を正規乱数とする)
 - $10^{3.02} = 1,047 \text{ 円}$, $(1047-1000)/1000 = 4.7\%$

2017.11.17 高橋行雄

7

変数を転置して誤差変動の付与

匿名データからの抽出					抽出ファイルの転置・誤差変動の付与							
出現 頻度	14 次元 番号	レコ ード 番号	Y 001	Y ... 203	変数 名	出現 頻度	14 次元 番号	レコ ード 番号	元 データ	常用 対数	正規 乱数 x 0.02	誤差 変動
1	2	1			Y001	1	2	1	1,000	3.00	0.02	3.02
1	4	3			Y001	1	4	3	100	2.00	-0.02	1.98
8,406					Y001	8,406						
1	2	1			Y001	1	2	1	1,000	3.00	-0.01	2.99
1	4	3			Y001	1	4	3	100	2.00	0.01	2.02
8,406					Y001	8,406						
2	6	5			Y001	2	6	5	10,000	4.00	0.02	4.02
2	6	7			Y001	2	6	7	10,000	4.00	-0.02	3.98
2,261x2					Y004	2,261x2						
21,334レコード x 203変数					:							
					Y203	2	6	7	1,000	3.00	0.04	3

14変数を行方向に並べ替えて誤差変動を一括して付与し, 元に戻す

2017.11.17 高橋行雄

8

匿名データに併合

- ◆ 誤差変動を与えた全抽出データを併合
 - 「匿名 47,797 レコード + コピーした 21,334 レコード = 69,131 レコード」：これが新擬似マイクロデータを作成するための元データとなる
- ◆ 改めて14次元クロス表の作成
 - 出現頻度は, 3, 4, 5, 6, ..., 780 となる
 - 出現頻度: 1 および 2 は出現しない
 - 行方向に展開したクロス表のリスト形式での表の作成

2017.11.17 高橋行雄

9

集計用乗率の調整

匿名 or コピー追加	セル件数	世帯 F or T	14次元番号	匿名 No	集計用乗率	計算法	調整済み集計用乗率	
匿名データ	1	F	4,203	1	913.5	これを1/3	304.5	
	87	F	13,867	2	630.3		630.3	
	1	F	11,582	3	912.3		304.1	
		15	F	7,590	4	420.2		420.2
		2	F	10,427	5	1,312.0	これを1/2	656.0
		7	F	3,290	6	269.7		269.7
					⋮			
セル件数 1 追加	1	F	4,203	1	913.5	これを1/3	304.5	
	1	F	11,582	3	912.3	これを1/3	304.1	
				⋮				
セル件数 1 さらに追加	1	F	4,203	1	913.5	これを1/3	304.5	
	1	F	11,582	3	912.3	これを1/3	304.1	
				⋮				
セル件数 2 追加	2	F	10,427	5	1,312.0	これを1/2	656.0	
	2	F	903	11	269.7	これを1/2	134.9	
				⋮				

14次元番号別に集計乗率の平均値を計算

2017.11.17 高橋行雄

10

14次元クロス表および集計乗率

14次元番号	X01	X02	X03	X04	X05	X06	X07	X08	X09	X10	X11	X12	X13	X14	Weight
	3大都市圏の別	世帯区分	世帯人員	就業人員	住居の構造	住居の建て方	住居の所有関係	世帯主の性別	世帯主の年齢階級	企業区分・規模	世帯分類	未就学	学校	65歳以上人員	集計乗率
1 0 その他	1	1	1	1	1	1	1	1	5	1.1~4人	1	1	1	0	215.9
2 0 その他	1	1	1	1	1	1	1	1	6	3.30~499人	1	1	1	1	845.5
3 0 その他	1	1	1	1	1	1	1	1	6	5.官公	1	1	1	0	1268.3
4 0 その他	1	1	1	1	1	1	1	1	6	5.官公	1	1	1	1	897.6
5 0 その他	1	1	1	1	1	1	1	1	7	1.1~4人	1	1	1	0	845.5
6 0 その他	1	1	1	1	1	1	1	1	7	4.500人以上	1	1	1	0	845.5
7 0 その他	1	1	1	1	1	1	1	1	7	5.官公	1	1	1	0	197
8 0 その他	1	1	1	1	1	1	1	1	8	3.30~499人	1	1	1	0	482.2
9 0 その他	1	1	1	1	1	1	1	1	8	5.官公	1	1	1	0	2317.6
10 0 その他	1	1	1	1	1	1	1	1	9	2.5~29人	1	1	1	1	562.9
11 0 その他	1	1	1	1	1	1	1	1	9	3.30~499人	1	1	1	0	897.6
12 0 その他	1	1	1	1	1	1	1	1	9	4.500人以上	1	1	1	0	796.1
13 0 その他	1	1	1	1	1	1	1	1	9	4.500人以上	1	1	1	1	562.9
14 0 その他	1	1	1	1	1	1	1	1	9	5.官公	1	1	1	0	2576.9
14244	3大都市圏	3.無職	5	3	2.コンクリ	一戸建	1.持ち家	1.男	13	無職	5	1	1	1	779.7
14245	3大都市圏	3.無職	5	3	2.コンクリ	一戸建	1.持ち家	1.男	14	無職	3	1	1	2	779.7
14246	3大都市圏	3.無職	5	3	2.コンクリ	一戸建以	1.持ち家	1.男	14	無職	3	1	1	1	210.1

セル件数は、統計表に含まれているので割愛した

2017.11.17 高橋行雄

11

14次元番号付き統計量

- ◆ 14次元クロス表の全てのセルに付与した連番号
 - 14次元番号: 1, 2, ..., 14,242
 - 統計表作成の元データ(69,131レコード)に世帯属性14変数によってマッチマージして14次元番号を付与
- ◆ 14次元番号別に203個の収支項目の平均, 標準偏差, 0円(対数では欠測値)を除くデータ数を算出

2017.11.17 高橋行雄

12

14次元番号付き統計量の生成

転置ファイル(例示)				14次元別・変数別の統計量				14次元番号別の統計量						
レコード番号	14次元番号	変数	対数データ	14次元番号	変数	レコード数	データ件数	対数平均	対数SD	14次元番号	ラベル	Y001	Y203	
匿名	1234	99	Y001	1.02	99	Y001	3	2.192	0.005	99	データ数	3	3	
	1234	99	Y002	2.01		Y002	5	1.23	0.12		1	対数平均	2.192	4.719
	:	:	:	:		Y001	5	1.23	0.12		1	対数SD	0.005	0.023
	1234	99	Y203	1.05		Y002	5	3.45	0.23		99	データ数	5	2
	47,797	999	Y203			Y203	5	2	1.78		0.34	99	対数平均	1.23
コピー変動	56789	99	Y001	1.01	14,246	Y001	3	3.112	0.011	14,246	対数SD	0.12	0.34	
	56789	99	Y002	3.05		Y002	3	6.16	0.007		14,246	対数SD	0.011	0.005
	:	:	:	:		:	:	:	:		:	:	:	:
	56789	99	Y203	2.02		Y203	3	5.052	0.005					
	69,131	9999	Y203	2.04										

14次元番号=99 について例示した。転置し、計算し、再転置で復元

2017.11.17 高橋行雄

13

14次元重み付き統計量

14次元番号	ラベル	Y001	Y002	Y003	Y004	Y005	(中略)	Y038	Y039	Y040	Y041	(中略)	Y203
		Z01	Z02	Z03				Z06	Z07	Z08	Z09		Z21
		年間収入	収入総額	実収入	経常収入	勤め先収入	(収入項目)	支出総額	実支出	消費支出	食料	(支出項目)	繰越金
1	データ数	3	3	3	3	3		3	3	3	3		3
1	log10平均	2.192	5.226	5.112	5.119	5.104		5.224	4.764	4.756	3.910		4.719
1	log10_SD	0.005	0.015	0.010	0.034	0.009		0.023	0.011	0.016	0.008		0.023
2	データ数	3	3	3	3	3		3	3	3	3		3
2	log10平均	2.430	5.615	5.338	5.341	5.325		5.627	5.303	5.207	4.557		4.014
2	log10_SD	0.006	0.009	0.010	0.012	0.013		0.010	0.010	0.014	0.011		0.019
14246	データ数	3	3	3	3	3		3	3	3	3		3
14246	log10平均	3.112	6.16	5.921	5.912	5.771		6.162	5.751	5.635	5.076		5.052
14246	log10_SD	0.011	0.007	0.014	0.013	0.002		0.011	0.013	0.027	0.004		0.005

この表のデータ件数は、すべての項目で揃っているが、対数変換により0円は欠測値となるので、データ件数は不揃いとなる。

2017.11.17 高橋行雄

14

主要21項目間の相関行列

- ◆元の匿名データ47,797 レコードから集計用乗率なしで、年間所得の3階級の区分ごとに作成.

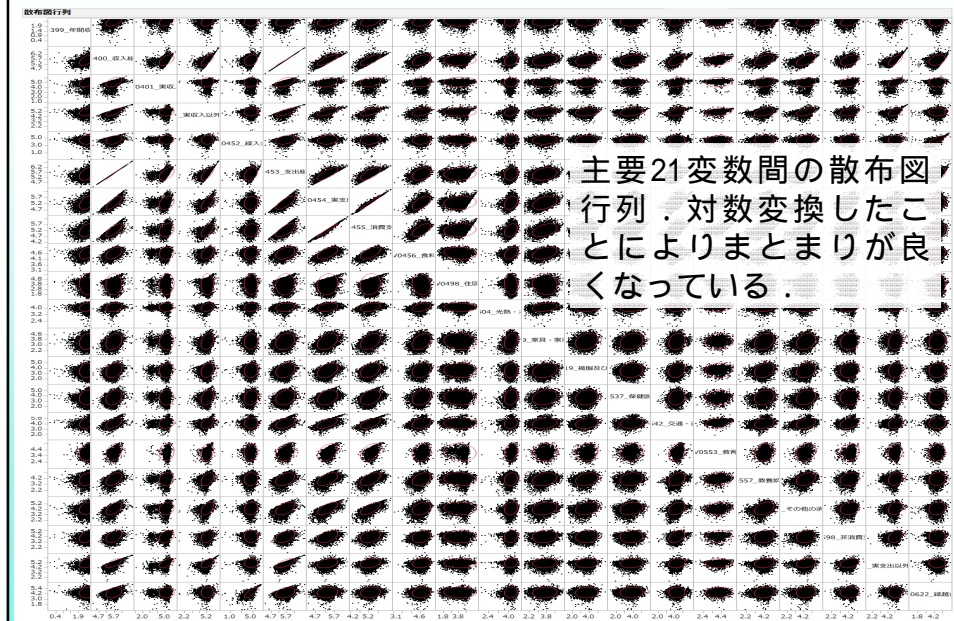
Y001	Z01	年間収入	Y040	Z08	消費支出	Y126	Z15	交通・通信
Y002	Z02	収入総額	Y041	Z09	食料	Y137	Z16	教育
Y003	Z03	実収入	Y083	Z10	住居	Y141	Z17	教養娯楽
Y024	Z04	実収入以外の収入	Y088	Z11	光熱・水道	Y151	Z18	その他の消費支出
Y037	Z05	繰入金	Y093	Z12	家具・家事用品	Y179	Z19	非消費支出
Y038	Z06	支出総額	Y103	Z13	被服及び履物	Y190	Z20	実支出以外の支出
Y039	Z07	実支出	Y121	Z14	保健医療	Y203	Z21	繰越金

- ◆年間収入と消費支出の相関関係は、低所得世帯で 0.40, 中所得世帯で 0.19, 高所得世帯で 0.29 と階層間に違いがみられた.

2017.11.17 高橋行雄

15

散布図行列 (年間収入3階級: 低)



69,131世帯から統計表の作成

- ◆ 以下の統計表(Excel)をウェブ上に公開
 - 世帯に関する情報14項目と集計乗率
 - 14次元クロス表の14246個のセルに対して, 203個の収支項目の対数変換値の平均値と標準偏差
 - 年間収入3階級別の主要21項目間の相関行列
- ◆ この公開情報のみから新擬似マイクロデータを作成し, ウェブ上に公開

正規乱数を用いた復元

- ◆ 14次元表の各14,246レコードについて出現頻度分をコピー
- ◆ 順次縦方向に連結, 正規乱数を用いて擬似マイクロデータ

14次元 番号	世帯 番号	複製 番号	項目名	log10 平均	log10 SD	正規 乱数	loga10 擬似	金額 擬似
1	1	1	年間収入	2.192	0.005	-0.840	2.190	154.92
1	2	2				-1.380	2.189	154.45
1	3	3				0.400	2.193	156.06
2	4	1		2.430	0.006	0.360	2.431	269.96
2	5	2				-1.580	2.426	266.38
2	6	3				-0.480	2.429	268.41
⋮								
14246	61311	1		3.112	0.011	0.31	3.114	1299.57
14246	61312	2				-0.37	3.110	1288.84
14246	61313	3				-0.61	3.109	1284.99

公表した統計表を14次元番号毎に転置し, 複製番号ファイルにマージして正規乱数を用いたlog10擬似データを計算.

互いに相関を持つ主要21項目

- ◆ 公表した主要21項目の年間収入3階級別に 21×21 の相関係数行列
- ◆ 互いに相関を持つ多変量正規乱数の作成
- ◆ 相関係数行列をコレスキー分解した相関行列 (21×21) と正規乱数行列 ($21 \times 61,131$) の積が、多変量正規乱数となる
- ◆ 行列を転置し、ファイルとして保存

互いに相関を持つ正規乱数

- ◆ (年間収入3階級:低) 69,131行 21列

	列1	列2	列3	列4	列5	列19	列20	列21
1	0.23	0.26	1.29	-0.02	-0.46	1.36	-0.46	-1.02
2	-0.29	0.47	1.07	0.92	-1.17	-0.23	0.16	-1.04
3	-0.37	0.67	0.33	1.38	1.84	-0.99	-0.04	1.92
4	-1.22	-0.28	-0.16	-0.99	-0.46	-1.86	-1.67	0.75
5	0.60	0.90	0.27	0.52	0.16	1.32	0.54	-0.62

発生させた乱数は、ファイルに保存し、必要に応じて計算用のファイルにマージして使い、結果の再現性を確保した。

多次元正規乱数による誤差変動

14次元 番号	世帯 番号	セル 件数	複製 番号	変数名 項目名	デー タ数	log10 平均	log10 SD	年収階層別乱数			年間収 入区分	log10 疑似	金額 疑似
								低	中	高			
1	1	3	1	Z01 年間収入	3	2.192	0.005	0.23	0.23	0.23	1 低	2.194	156
1	1	3	1	Z02 収入総額	3	5.226	0.015	0.26	0.21	0.24	1 低	5.230	169,742
1	1	3	1	Z03 実収入	3	5.112	0.010	1.29	1.27	1.18	1 低	5.125	133,328
1	1	3	1	Z04 実収入以外の収入	3	4.550	0.004	-0.02	-0.10	-0.09	1 低	4.550	35,463
1	1	3	1	Z05 繰入金	3	2.726	0.010	-0.46	-0.49	-0.47	1 低	2.721	526
1	1	3	1	Z06 支出総額	3	5.224	0.023	0.04	-0.07	0.20	1 低	5.225	167,865
1	1	3	1	Z07 実支出	3	4.764	0.011	1.82	1.80	1.68	1 低	4.784	60,838
1	1	3	1	Z08 消費支出	3	4.756	0.016	1.82	1.56	1.71	1 低	4.785	60,985
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	1	3	1	Z21 繰越金	3	4.719	0.023	-1.02	-1.10	-1.08	1 低	4.696	49,605
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
14246	69131	3	3	Z19 非消費支出	3	5.210	0.006	-0.36	-0.19	-0.85	3 高	5.205	160,361
14246	69131	3	3	Z20 実支出以外の支出	3	5.859	0.003	-1.16	-0.75	-0.47	3 高	5.857	720,031
14246	69131	3	3	Z21 繰越金	3	5.052	0.005	0.41	0.49	0.43	3 高	5.054	113,157

21変数を行方向に、データ数・log10平均・SDを列方向に、年収3階級毎の乱数を行方向に展開しマージし、log10疑似を計算。

2017.11.17 高橋行雄

21

0円の割合を保持

- ◆ $N_i=5$ で、データ数 $n_{ij}=3$ 、世帯中2世帯は欠測値
- ◆ 一様乱数 u_{ij} を発生、 $3/5=0.60$ 未満 疑似データあり
0.60以上 欠測値

14次元 番号	世帯 番号	セル件 数 N_i	複製番 号 j	変数名 項目名	デー タ 数 n_{ij}	対数 平均	対数 SD	一様 乱数	n_{ij}/N_i	判 定	扱 い
1	1	3	1	Y001 年間収入	3	2.192	0.005	0.70	1.00	○	採用
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
888	999	5	1	Y137 教育費	3	4.00	0.10	0.70	0.60	×	欠測値
888	999	5	2	Y137 教育費	3	4.00	0.10	0.20	0.60	○	採用
888	999	5	3	Y137 教育費	3	4.00	0.10	0.80	0.60	×	欠測値
888	999	5	4	Y137 教育費	3	4.00	0.10	0.39	0.60	○	採用
888	999	5	5	Y137 教育費	3	4.00	0.10	0.50	0.60	○	採用
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
14,246	69,131	3	3	Y203 繰越金	3	5.052	0.005	0.30	1.00	○	採用

14次元番号=888 はセル件数=5だが、教育費は3世帯だけ、一様乱数を用いて、2世帯を欠測値とする。

2017.11.17 高橋行雄

22

足し上げ構造

- ◆元の匿名データには、下位項目の合計が上位項目と一致する足し上げ構造がある。
- ◆公表した統計表の平均値とした段階で、足し上げ構造に乱れが生ずる。
- ◆公表した統計表から正規乱数で元のデータ数分の復元に際し、更に乱れが加算される。
- ◆下位項目の金額の合計を上位項目の金額とすれば、足し上げ構造を復元することはできるが、誤差変動が上位項目で拡大する。

2017.11.17 高橋行雄

23

逆足し上げ

- ◆最上位項目の収入総額と支出総額から下位項目方向へ逆足し上げを行う。
- ◆支出総額は、実支出・実支出以外の支出・繰越金の合計であるが、合計が支出総額になるように、下位項目の金額を調整する。
- ◆実支出は、消費支出・非消費支出の合計であるが、合計が実支出になるように下位項目の金額を調整する。

2017.11.17 高橋行雄

24

「逆」足し上げの例示

階層	新擬似Y名	項目名	レベル6		調整前			合計		調整後	
			区分	足上	データ	0の計	1の計	データ	足し上げ		
6	Y042	穀類	602	0	400	400	360	x	400	400	o
7	Y043	米	602	1	200	400	360		222		
7	Y044	パン	602	1	100	400	360		111		
7	Y045	めん類	602	1	50	400	360		56		
7	Y046	他の穀類	602	1	10	400	360		11		
6	Y047	魚介類	603	0	250	250	300	x	250	250	o
7	Y048	生鮮魚介	603	1	150	250	300		125		
7	Y049	塩干魚介	603	1	100	250	300		83		
7	Y050	魚肉練製品	603	1	0	250	300		0		
7	Y051	他の魚介加工品	603	1	50	250	300		42		

- ◆ 穀類の下位項目の合計は360円，合計が400円になるように，下位の項目の金額に1.111倍する
- ◆ 魚介類の下位の金額に，0.833倍する

注) 収支項目間に階層構造がある場合に，「足し上げ構造」があると言う

2017.11.17 高橋行雄

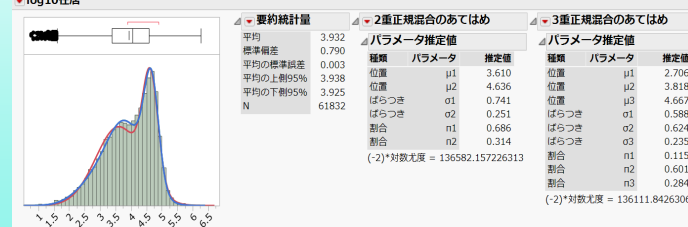
25

3重正規混合分布の再現

匿名
データ



新擬似
マイクロ
データ



やや平滑となるが，分布の特性は保持されている

2017.11.17 高橋行雄

26

203項目のデータ数と平均の比較

変数	平均(データ)			平均(log10)			
	匿名	新擬似	差異%	匿名	新擬似	log10 差異%	逆対数 差異%
Y001 年間収入	643	652	1.3	2.730	2.734	0.1	0.8
Y002 収入総額	679,579	653,741	-3.8	5.852	5.848	-0.1	-1.0
Y003 実収入	324,750	315,927	-2.7	5.497	5.495	0.0	-0.4
Y004 経常収入	314,359	307,482	-2.2	5.512	5.503	-0.2	-1.9
Y005 勤め先収入	261,033	259,507	-0.6	5.541	5.518	-0.4	-5.2
Y006 世帯主の勤め先	210,906	210,715	-0.1	5.509	5.485	-0.4	-5.2
Y007 世帯主の配偶者	33,509	30,416	-9.2	5.008	4.984	-0.5	-5.3
Y008 他の世帯員の勤	16,618	18,375	10.6	4.991	4.946	-0.9	-9.7
Y009 事業・内職収入	1,686	1,788	6.0	4.427	4.410	-0.4	-3.9
Y010 農林漁業収入	354	325	-8.0	4.228	4.132	-2.3	-19.9
Y201 財産購入	3,186	2,271	-28.7	5.868	5.269	-10.2	-74.8
Y202 その他の実支出	1,152	1,433	24.4	3.376	3.360	-0.5	-3.5
Y203 繰越金	62,555	60,122	-3.9	4.656	4.655	0.0	-0.1
総平均	28,794	27,923	-3.0	3.811	3.765	-1.2	-10.0

「新擬似」が「匿名」に比べて、若干小さ目となっている。

2017.11.17 高橋行雄

原因については、検討中。

27

今後の展望

- ◆ 2017年の第5回目コンテストの実施に間に合うように、新擬似マイクロデータを提供できた。
- ◆ 公表した統計表の作成、統計表から擬似マイクロデータの作成について、今回この手順を整理しまとめることができた。
- ◆ 2004年の全消匿名データに加え、1989年、1994年、1999年の全消匿名データについても新擬似マイクロデータを開発中であり、2018年以降の「Let'sデータ分析コンテスト」に供したい。

2017.11.17 高橋行雄

28