平成 25 年 11 月 22 日
「公的統計のミクロデータの利用に関する研究集会」
@統計数理研究所

統計センター提供の教育用擬似ミクロデータを用いた SAS/JMP によるデータ分析コンテスト

高橋 行雄　BioStat 研究所（株）
周防 節雄　（公財）統計情報研究開発センター
兵庫県立大学 名誉教授

要旨

教育用擬似ミクロデータを用いたデータ分析コンテストを 2013 年 7 月の SAS ユーザー総会にて行ったので，その概要について報告する．今回は若手のアナリストにターゲットを絞るために参加資格を 30 歳未満とし，SAS/JMP の経験年数によって A，B，C クラスに分けた．参加申込は 27 組あったが，論文の最終提出は 15 組となり，その中から 8 組を優秀賞として選抜し，これらの論文は SAS ユーザー総会の論文集に掲載した．総会前日 day 0 にプレゼンテーションによる審査会を開催し，応募クラスごとに最優秀賞を選抜し，SAS ユーザー総会において研究報告がなされた．8 組の優秀賞論文は，解析に使用したプログラムも含めて Web で公開予定である．

キーワード: 教育用擬似ミクロデータ, SAS ユーザー総会, JMP, データ分析コンテスト, 匿名データ

1. はじめに

SAS ユーザー総会は，1981 年以来毎年開催され，2013 年には第 32 回目となり，7 月 18 日（木）～19 日（金）に東京大学伊藤国際学術研究センター（東京大学本郷キャンパス内）にて行なわれた．2013 年の SAS ユーザー総会では，新たな試みとして「Let's データ分析」と題したコンテストを企画した．今回のデータ分析コンテストでは，政府の指定統計調査の一つである全国消費実態調査の平成 16 年データに基づき独立行政法人統計センターが作成した「教育用擬似ミクロデータ」を使用することにした[1].

この擬似ミクロデータは 32,027 件から成る 183 変数，CSV ファイル（約 50M バイト）で，オリジナルの個票データとは異なるが，個票データの特性を失わないように綿密に作成されおり，多次元の集計によっても元の個票データの特性を保っているとの報告もあり[2], 今回のデータ分析コンテストのための「データ」として採用した．応募者に対しては，関連資料を含めて CD-ROM で別途パスワードを付与し提供した．

データ分析コンテストでは規定課題を設定し，これをクリアーできれば，自由課題にチャレンジできることとした．自由課題は，擬似ミクロデータを使って各自の発想で自由な分析処理をし，その結果を論文形式で提出を求めた．

2. 募集要項

開催趣旨：

SAS ユーザー総会では，今年から「Let's データ分析」と題してデータ分析コンテストを行います．2013 年のユーザー総会では，個票データを対象としたデータ分析コンテストを開催します．政府の指定統計調査の一つである全国消費実態調査データ（平成 16 年）に基づき独立行政法人統計センターが作成した「教育用擬似ミクロデータ」を使用します．

「教育用擬似ミクロデータ」は 32,027 件，183 変数，CSV ファイルで約 50M バイトの大きさです．元の個票データとは異なりますが，個票データの特性を失わないように綿密に作成されおり，多次元の集計でも元の個票データの特性を保っているとの報告もあります．そこで，今回のコンテストのための「データ」として採用いたしました．応募された方々には CD-ROM で「データ」を提供いたします．

1）応募資格：

2013 年 3 月 31 日現在で 30 歳未満までの学生および社会人．

2）使用ソフト：

データ分析に際し，SAS または JMP どちらかをデータ分析に使用することを条件とします．分析結果を論文にまとめる段階で Excel などの他のソフトを，結果表の整形やグラフの作成などに使うことは差し支えありません．

3）参加クラス：

A）パワーユーザ（使用経験年数は問わない）

B）SAS または JMP の使用歴 5 年未満

C）SAS または JMP の使用歴 2 年未満

所属する機関や組織の制限はなく，個人でもチームでも応募できます．ただし，全員が同一の参加クラスの条件を満たしていること．使用歴は自主申告です．

4）募集方法：

SAS ユーザー総会のホームページでコンテストの概要を公開します．申し込みは以下のサイトからご登録ください．

Let's データ分析　第一回ミクロデータ分析コンテスト

http://www.sas.com/reg/offer/jp/20130718_sas_academic

（2013 年 2 月 20 日に Web 上で公開）

5）課題：

あらかじめ設定された規定課題，および，自由課題とします．自由課題のみでは，受理しません．自由課題は「教育用擬似ミクロデータ」を使って各自のアイデアで好きな分析処理をして下さい．

6）データの配布：

データに関連するコード表などのメタデータ，CSV 形式のデータ，規定課題を CD-ROM で提供します．なお，CD-ROM の提供前に，配布データの取扱に関する誓約書の提出を求めます．

7）審査方法：

・2013 年 5 月 28 日(火)までに指定のレポート形式（A4 で 4 枚以内）で提出をお願いします．提出されたレポートについて審査を行ない SAS ユーザー総会の論文集への掲載の可否を判定します.

・提出頂いたレポートの結果は，6 月 4 日(火)までに書面にて通知致します.

・各参加クラスで高々 3 組，全体で高々 9 組の優秀レポートを選抜します．選ばれた 9 組のレポートが SAS ユーザー総会の論文集へ掲載されます．なお，SAS ユーザー総会の論文集への掲載の締め切りは 6 月 14 日です．締め切り日まで訂正版の提出が可能です.

・優秀レポートの該当者 9 組には，SAS ユーザー総会前日の 7 月 17 日（Day 0）に SAS 六本木オフィスで開催する公開の審査会でプレゼンテーション（20 分）をして頂きます．その結果，参加クラス（A，B，C）ごとに最優秀賞を 1 件ずつ決定します．それ以外の方には優秀賞を授与します．但し，各参加クラスにおいて最優秀賞が選ばれない場合もあります．その場合は，優秀賞の中からユーザー総会当日にプレゼンテーションする 1 件が選ばれます．なお，プレゼンテーションで用いるパワーポイントなどの事前提出はありません.

・プレゼンテーションの際に，一部の出力結果についてプログラムを実行し再現を求めることがあります.

・優秀レポート受賞者には，SAS ユーザー総会で展示ポスター発表をしていただきます．事前の準備をお願いします.

8）SAS ユーザー総会でのプレゼンテーション：

参加クラス（A，B，C）の最優秀賞は，「Let’s データ分析セッション」で口頭発表をお願いします．なお，その際，審査会（day 0）で用いたプレゼンテーション資料を口頭発表でも使えます.

9）賞金：

最優秀賞： ¥80,000 -

優秀賞： ¥10,000 -

10）SAS プログラム等の公開：

プレゼンテーション資料と SAS のプログラム又は JMP のスクリプトが SAS ユーザー総会の Web 上で公開されることを承諾したことと見なします.

11）参加者に対する支援：

a）SAS ユーザー総会の論文集への掲載が可となった場合は，SAS ユーザー総会参加費が免除されます．ただし，チームで応募された場合は，1 チーム 3 名までとし，ユーザー総会論文集 1 冊がチームに無料配布されます.

b）SAS ユーザー総会の前日（Day 0）の審査会でプレゼンテーションされた口頭発表者は，懇親会にご招待いたします.

c）SAS ユーザー総会の前日(Day 0)の審査会で「発表」する学生・院生には，旅費を援助します.

d）参加者は，「SAS」の一時貸し出しを申請することができます．これは，SAS の使用場所が，職場などに限られている参加者に対する無償サポートです．なお，JMP については，製品版と同様の機能を持つ無料の 30 日間のトライアル版がありますので，各自で SAS 社から入手して下さい．

e）SAS の使用法を援助する目的で SAS の使い方の勉強会「SAS クイックスタート～グラフィカルツールによるデータ加工とレポート作成～」を，4 月 1 日（月）14 時～16 時半に SAS 六本木オフィスにおいて行います．ご希望の方は，以下リンクから申し込みを行ってください．教育機関向けと記載がありますが，当該コンペティション参加者であれば一般の方でも参加可能です．

申し込みウェブサイト以下から：

http://www.sas.com/reg/offer/jp/20130401_sas_academic

12）Let's データ分析　ミクロデータ分析コンテスト説明会について：

日 時：　2013 年 3 月 29 日（金）11:30～13:00

会 場：　SAS Institute Japan 株式会社六本木オフィス

内容：

- 取り扱いデータについての説明（独立行政法人 統計センター）
- ミクロデータ分析コンテスト開催説明（BioStat 研究所株式会社 高橋行雄，兵庫県立大学名誉教授・シンフォニカ客員上席研究員 周防節雄）
- 手続き等の説明（アカデミック・マーケティング）

お問い合わせ窓口：マーケティング本部アカデミック・マーケティング紺屋

（同様の説明会を大阪でも 4 月 5 日に行なった）

13）規定課題：

統計センターは，CSV ファイルで提供した擬似ミクロデータが，利用者の解析用データセットに正しく反映されているかを確認するため，4 つの集計表を示している．そこで，それらの 4 つの表を作成することを規定問題とした．その内の第１－１表，第１－２表，および，第２表を以下に示す．なお，統計センターは，これらの集計表で，世帯人員 4 人かつ有業人員 1 人又は 2 人の世帯に該当する　で示した箇所のセルの数値を確認するように薦めている．

統計センター集計表：第１－１表　集計世帯数（各レコードを、単純にカウントしたもの）

		総数	世帯人員							
			２人	３人	４人	５人	６人	７人	８人	９人	10人
総数		32,027	7,438	8,537	9,944	4,405	1,214	390	81	15	3
有業人員	１人	13,913	4,124	3,908	**4,132**	1,436	256	51	6	0	0
	２人	13,459	3,239	3,391	**4,201**	1,943	494	162	29	0	0
	３人	2,950	0	1,035	1,031	559	232	84	6	3	0
	４人	691	0	0	324	220	104	31	12	0	0
	５人	40	0	0	0	27	6	7	0	0	0
	６人	6	0	0	0	0	3	3	0	0	0
	不詳	968	75	203	256	220	119	52	28	12	3

統計センター集計表：第１－２表 世帯数分布（各レコードを、集計用乗率で重み付けして、カウントしたもの。

		総数	世帯人員							
			２人	３人	４人	５人	６人	７人	８人	９人	10人
総数		495, 465	115, 835	132, 005	155, 850	67, 565	17, 161	5, 611	1, 197	207	33
有業人員	１人	219, 743	64, 691	61, 284	**66, 299**	22, 740	3, 813	831	84	0	0
	２人	205, 723	50, 138	51, 523	**64, 868**	29, 616	6, 801	2, 336	441	0	0
	３人	44, 041	0	15, 912	15, 615	7, 963	3, 302	1, 137	76	36	0
	４人	10, 168	0	0	4, 851	3, 345	1, 354	422	195	0	0
	５人	570	0	0	0	383	80	108	0	0	0
	６人	99	0	0	0	0	49	51	0	0	0
	不詳	15, 120	1, 006	3, 287	4, 216	3, 519	1, 761	727	401	170	33

統計センター集計表：第２表　支出（消費支出 及び 十大費目）　（単位：円）

		集計世帯数	世帯数分布(抽出率調整	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
総数		32, 027	495, 465	328, 140	72, 883	17, 687	19, 238	9, 204	14, 138	11, 366	47, 961	22, 270	31, 389	82, 003
うち世帯人員が４人		9, 944	155, 850	335, 438	76, 362	15, 345	20, 214	8, 885	14, 452	10, 987	47, 894	33, 442	32, 269	75, 588
有業人員	１人	**4, 132**	**66, 299**	**305, 234**	**71, 543**	**17, 556**	**18, 854**	**8, 383**	**13, 579**	**11, 656**	**42, 703**	**31, 202**	**31, 959**	**57, 801**
	２人	**4, 201**	**64, 868**	**347, 740**	**78, 472**	**12, 932**	**20, 621**	**8, 917**	**14, 876**	**10, 538**	**52, 141**	**39, 485**	**33, 681**	**76, 078**
	３人	1, 031	15, 615	380, 521	83, 796	12, 583	23, 045	10, 276	16, 561	10, 331	52, 406	22, 513	27, 852	121, 160
	４人	324	4, 851	399, 962	85, 083	18, 500	22, 490	11, 763	14, 096	10, 812	51, 310	4, 509	32, 769	148, 628
	不詳	256	4, 216	379, 882	82, 134	24, 296	22, 238	7, 843	14, 238	10, 011	43, 521	49, 453	31, 206	94, 942
	(特掲) １人又は２人	**8, 333**	**131, 168**	**326, 256**	**74, 970**	**15, 269**	**19, 728**	**8, 647**	**14, 221**	**11, 103**	**47, 371**	**35, 298**	**32, 810**	**66, 839**

統計センター集計表（第２表）には，世帯数を 4 人に絞り，その世帯の有業人員別の件数，消費支出，及び十大費目について算術平均が示されている．参考までに，同表の表頭に該当する項目について，擬似ミクロデータからランダムに 30 件分を抽出した結果を表 1 示す．

表 1．擬似ミクロデータリストの一部（32,027 件からランダムに 30 件を抽出）

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	No	V001_世帯区分	V002_世帯人員	V003_有業人員	V014_集計用乗率	Y037_消費支出	Y038_食料	Y079_住居	Y084_光熱・水道	Y089_家具・家事用品	Y099_被服及び履物	Y122_交通・通信	Y129_教育	Y133_教養娯楽	Y169_他の非消費支出
2		1_勤労	2	1	15.317	154641	43597	692.73	23737	8448.7	4083.1	46132	0	7098.2	0
3		1_勤労	2	1	11.367	251248	71111	0	13438	816.4	1513.1	18301	0	12283	0
4		1_勤労	2	1	13.967	206077	78999	0	13000	2281.3	1423.4	31629	0	37039	0
5		1_勤労	2	1	16.456	205672	65003	0	27434	4153.2	3835.5	52485	0	9635	0
6		1_勤労	2	1	25.9	216384	70510	1441.2	18967	4223.9	1843.9	28961	1428	16071	0
7		1_勤労	2	1	13.3	168145	46562	20540	9858.3	13076	8776.2	22956	45.239	19241	0
8		1_勤労	2	2	11.75	381322	74436	16439	19605	6925	1109.4	27406	0	48764	0
9	I	1_勤労	2	2	16.406	236921	71958	17605	18442	1735.2	1925.2	28873	0	35244	0
10	D	1_勤労	2	2	5.6	367619	71179	371.63	12747	3012.5	19706	38274	0	28646	0
11	番	1_勤労	3	1	9.7	188800	23155	30415	15226	2883.5	10435	20451	36967	5700.7	0
12	号	1_勤労	3	1	17.167	255188	51836	45141	19940	3740	8849	30101	0	29199	0
13	割	1_勤労	3	2	8.7167	248634	55865	0	23994	75	20341	30199	0	15580	0
14	愛	1_勤労	3	3	16.5	222442	85345	0	19463	3418	8133	20756	0	37082	0
15		1_勤労	4	1	20.944	377225	82755	0	28365	4094	17137	14429	31204	35382	0
16		1_勤労	4	1	19.95	473659	107741	10530	213	3558.2	14154	46573	123481	25530	0
17		1_勤労	4	1	13.183	373458	90138			2634.5	51026	28229	20939	0	
18		1_勤労	4	2	15.025	217669	73899		1541	16169	22408	32509	5448.3	14368	61.54
19		1_勤労	4	2	14.733	249746	3048	0	3083	11372	6354.6	19984	33159	31587	0
20		1_勤労	4	2	10.8	341637	0	4	13653	2837.5	3309.7	137130	50143	21922	0
21		1_勤労	4	2	13.633	38311	2	57.76	17189	7047.4	11654	35919	19893	25075	0
22		1_勤労	4	2	5.6833	3	1901	0	28719	24781	5605.1	58578	0	35977	0
23		1_勤労	4	2	11.653	3	98010	0	42682	2596.9	22882	24855	24231	46441	0
24		1_勤労	4	2	13.017	212066	77968	3280.3	20981	4618.3	12468	17833	1679.9	39062	0
25		1_勤労	5	2	9.7	266181	82886	0	19192	2803.9	17126	28255	0	12463	0
26		1_勤労	5	2	9.9	201885	55965	0	13116	3281.1	10626	43174	15846	33587	0
27		1_勤労	5	2	15.017	477462	66871	13911	27823	4477.8	1681.5	107771	27727	34943	0
28		1_勤労	5	2	19.533	269028	102457	0	23433	7377.3	13592	15089	40129	35860	0
29		1_勤労	6	1	17.544	595717	190176	22532	40482	24166	8823	66108	124361	37647	1852.3
30		1_勤労	6	2	11.567	354534	95708	1929.4	11349	3525.3	5003.1	74298	1367.9	39238	1065.5
31		1_勤労	6	4	19.367	193681	66440	0	25680	7493.1	1469	27349	11294	13375	0
32	32,027人分のデータからランダムに30人分を抽出した														

3. 書類審査

応募は 27 組あった．その内 12 組は論文が期限内に論文の提出がなく失格となった．論文形式のレポートが提出された 15 組について，論文審査を行い優秀賞の選定を行ない，表 2 に示すように，8 組の優秀賞を選抜した．選外となった 7 組の業種は，金融関連 1 組，大学 1 組，IT 関連 5 組であった．なお,これら 8 組の優秀賞の論文は，SAS ユーザー総会論文集に収録した（末尾の付録参照）．

表 2　論文審査の結果

応募クラス	統計ソフト	業種	所属	氏 名
C	SAS	IT 関連	株式会社データフォーシーズ	中島 貴之
C	SAS	IT 関連	株式会社データフォーシーズ	高宗 太輔
C	JMP	製造業	パナソニック株式会社	土井 優子
C	SAS	製薬関連	テルモ株式会社	宇野 慧
B	SAS	大学	東京理科大学	岡村 正太
A	SAS	大学	東京理科大学	魚住 龍史
A	SAS	製薬関連	大鵬薬品	冨里 遼太
A	SAS	製薬関連	AstraZeneca	筒井 杏奈

4. プレゼンテーションによる最優秀賞の選考

4.1. プレゼンテーションの案内文

「Let's データ分析」第一回ミクロデータ分析コンテスト公開審査会を以下の要綱で実施します．

日時：7 月 17 日（水）15 時半～19 時（受付開始 14 時半～）

会場：SAS Institute Japan 株式会社　六本木ヒルズ 11F セミナールーム

連絡先：03-6434-xxxx コンテスト事務局　紺屋宛　xxxxxx@sas.com

・当日スケジュール：プレゼンテーション時間 20 分質疑応答 5 分含む：規定課題及び自由課題のプレゼンテーションを 15 分以内でご用意ください．

・審査結果発表：18 時 50 分，　終了：19 時

・前日までにコンテスト事務局宛にメールで送付するもの：

① 発表用のパワーポイント

② 教育用擬似ミクロデータの CSV ファイルを SAS 又は JMP に取り込むための SAS プログラム又は JMP スクリプト．

③ 統計センター集計表（第 2 表）の作成のための SAS プログラム又は JMP スクリプト．

④ 自由問題として提出した論文に使っている図及び表の作成のための SAS プログラム又は JMP スクリプト．

・当日のプレゼンテーション内容：各自ノート PC を持参してください．

① 統計センター集計表（第 2 表）の SAS プログラム又は JMP スクリプトの表示

② 上記を実行し結果をその場で表示のこと

③ 自由課題に付いて，パワーポイントによるプレゼンテーション

④ 全体で 15 分以内に行ってください．

⑤ 質疑応答は 5 分間.

・最優秀賞に選ばれた 3 組の方には，7 月 18 日（木）2013 年ユーザー総会会場（東京大学・伊藤国際センターでプレゼンテーションをお願いします.

応募クラス	時 間
C	14 時半～15 時
B	15 時～15 時半
A	15 時半～16 時
ポスターセッション	16 時～16 時半

・ユーザー総会当日のポスターセッションに付いて：

公開審査会で発表された方全員に，7 月 18 日 12 時半までにポスターの掲示をお願いいたします．ポスターツアーは，18 日 16 時半から 17 時，ユーザー総会会場（東京大学・伊藤国際学術研究センターの謝恩ホール横）にて行いますので，質疑応答（来場者からの質問に対しての回答）をお願いいたします．ポスターは 19 日の遅くとも 16 時に撤去して下さい.

4.2. 最優秀賞の審査

8 組の書類選考者から募集クラス（A，B，C）別に最優秀賞の選抜を次の 5 名の審査委員で行なった.

東京大学 教授	大橋 靖雄	BioStat 研究所株式会社	高橋 行雄
北海道大学 准教授	伊藤 陽一	独立行政法人統計センター	宮内 亨
兵庫県立大学 名誉教授	周防節雄		

審査結果を表 3 に示す．C クラスの最優秀賞は宇野慧氏，B クラスの最優秀賞は該当者がなかった．A クラスは，筒井杏奈氏と魚住龍史氏が最優秀賞候補に選ばれたが，厳正な審査の結果，表 3 に示すように，筒井氏を最優秀賞，魚住氏を準最優秀賞とした．この 3 名の方々には SAS ユーザー総会での研究発表をお願いした．（本日の研究集会には，宇野慧氏，魚住龍史氏の研究報告が組み込まれている.）

表 3　審査結果

氏 名	応募クラス	演 題	審査結果
中島 貴之	C	シニア世代の消費特徴分析	
高宗 太輔	C	食費の分析による食に対する意識と食生活の杷握	
土井 優子	C	家計簿から見た「世帯人員ごとの幸福度」に関する研究	
宇野 慧	C	世帯主の就業状況が貯蓄性保険需要に与える影響についての考案 ～擬似ミクロデータを用いたTobit/Hurdleモデル推定～	C 最優秀賞
岡村 正太	B	健康因子に対する支出傾向に関する解析	
魚住 龍史	A	擬似ミクロデータによる国内旅行費支出と世帯情報の関連の検討	A 準最優秀賞
富里 遼太	A	教育用擬似ミクロデータを用いた収入・消費傾向の考察	
筒井 杏奈	A	変曲点を用いた最低生活費の推定 －疑似ミクロデータ（平成16年全国消費実態調査）による－	A 最優秀賞

写真 1　プレゼンテーション審査会

写真 2　発表者および審査委員

4.3. SASユーザー総会での研究発表

SAS ユーザー総会当日は，最終選考に残った 8 組全員のパネルの展示も行なわれた．口頭による研究発表は，宇野慧氏（C クラス最優秀賞），魚住龍史氏（A クラス準最優秀賞），筒井杏奈氏（A クラス最優秀賞）の順に，東京大学・伊藤国際学術研究センターの大ホールで行なわれ，聴衆は約 250 人と大盛況であった．審査会の翌日ではあったが，持ち時間 30 分内で，前日の審査会で指摘された事項について，追加・訂正がなされており，主催者として喜ばしいことであった．

審査会当日の審査結果発表の際，選考委員を代表して周防（本稿の共著者）から，「米国のデータ・サイエンティストは，SAS を使うことが当たり前．SAS を使っていないと仲間に入れてもらえないという感覚すらあるのですが，日本ではそこまで浸透していない歯がゆさがあります．今回のコンテストを機に，皆さんには SAS の若き伝道師になってほしいです．若い人たちにどんどん出てきて頂くと，より意義のあるコンテストになります」との講評が行なわれた．

5. 考察

日本の SAS ユーザー会が 1981 年に発足した当時は，多様な分野からの研究発表がなされていたが，1990 年代になると，医薬系の発表が徐々に多くなり，それ以外の分野の発表が伸び悩み，2000 年代後半から医薬系の発表が半数に迫るようになった．これは，日本のみならず世界中で SAS が医薬系の統計解析業務の事実上の標準ソフトとなったからであるが，その反面，日本では医薬関連以外の産業界や学界で SAS が標準的には使われてなくなってきたことを反映している．

そこで，この偏りを是正するための企画として，経営科学系研究部会連合協議会が主催している「データ解析コンペティション」（http://jasmac-j.jimdo.com/）を参考に， 2013 年 7 月の SAS ユーザー総会で「データ分析コンテスト」を開催する決定をした．用いるデータとしては，政府の指定統計の一つである全国消費実態調査データに基づき独立行政法人統計センターが作成した「教育用擬似ミクロデータ」を使用することにした．このデータは，今回のコンテスト参加者のみならず，誰でも統計センターに申請さえすれば，無料で使うことができ，CSV ファイルで約 50M バイトと適度な大きさであるので採用を決めた．

参加者のモチベーションを上げるために SAS Institute Japan（株）に賞金を提供して頂き，総額 30 万円が予算化された．配分方法について議論が白熱したが，クラスごとの最優秀賞を手厚くすることとした．今回は，若手の SAS ユーザーの拡大という趣旨から，年齢制限を 30 歳未満とし，経験年数で 3 クラスに分け，それぞれ最優秀賞を出すことにした．なお，次回からはこの年齢制限を外すことも検討している．

この擬似ミクロデータを用いる場合に，統計センターは 4 つの集計表を作成することにより，元の CSV ファイルから正しくデータセットが作成されたことを確認するように薦めている．そこで，この集計表を作成できることが最低限の参加資格であることとした．同じ結果を出すにしても多様なアプローチがあり，これらを互いに知ることは参加者にとって有意義であると考えた．また，審査の際のプレゼンテーションでは，実際に各自のプログラムをその場で実行し，結果を再現することを求めたが，審査員からも感嘆の声が出るような見事な SAS プログラムもあった．

自由課題には，何ら制限を定めずに，論文による事前審査を行ない，8 組を選んだ．統計センターの Web サイトには，擬似ミクロデータに関する Q&A に以下の記述がある．「擬似ミクロデータは、我が国の行政機関が実施した統計調査の集計表から作成したミクロデータ形式の擬似的なデータです。擬似ミクロデータを用いた分析結果は、実証研究の結果とみなすことはできませんので、あくまでも教育目的やテストデータとしてご利用ください。」　しかしながら，最優秀賞の審査の際に，よくここまでの分析を擬似ミクロデータで行なったものだ，との賞賛の声が上がるほどの研究発表もあった．これが，もし，統計センターが提供している「匿名データ」（公的統計のミクロデータから特定の個人又は法人が特定できないようにデータの加工を行なったもの）であれば，もっとすばらしい実証研究となっていたはずだが，との感想も審査員からあった．

学術誌などで報告される実証研究の場合，結果を追試したいと思っても，使用されたデータは非公開であるとか，研究者がデータの公表を望まないことなどの理由により，統計解析の結果の吟味ができないことが，統計解析技法の進歩の阻害要因の一つと思われる．誰でも申請すれば無料で同じデータが入手できる擬似ミクロデータならば，追試が自由に行え，オープンな議論ができること

は，大いに意義があると考えている．

これまでの SAS ユーザー総会の研究発表では，他のユーザーが結果を再現できるように SAS プログラムを論文に添付することが慣例的に行なわれてきた．そこで，今回のデータ分析コンテストでは，規定問題および自由課題で用いた SAS のプログラム又は JMP スクリプトを Web 上で公開することを事前に承諾して頂いた．規定問題は，統計センターがデータ確認のために集計することを薦めているクロス表を出力する SAS プログラム又は JMP スクリプトを開示・実行して再現を求めたが，さまざまなアプローチがあり大変興味深かった．

擬似ミクロデータを使って解析を行なうことは，匿名データを用いた本格的なデータ分析を行なうためのトレーニングとして最適と考える．そこで，次回の SAS ユーザー総会では，匿名データを用いた研究発表も推進するために，著者らが正式に匿名データの提供を統計センターに申請し，次回の SAS ユーザー総会で研究発表を行なうことにした．さらに，本年度のデータ分析コンテストの最優秀賞と優秀賞の 8 組の SAS ユーザーの方々に対して，次回の SAS ユーザー総会で匿名データを使用した研究発表のお願いをする予定である．さらに，社会科学系の研究者に対しても同様のお願いをすることも考えている．この匿名データの利用をより一層推進するために，参加者の統計センターへの利用申請に際し，SAS Institute Japan（株）の紺屋恭子氏に申請代理人をお願いすることを予定している．これにより参加者の利用申請に伴う事務手続きや経費の負担を軽減できるので，匿名データを使用した多くの研究発表が行われることを期待している．

今回の SAS ユーザー総会でのデータ分析コンテストでは，平成 16 年全国消費実態調査データに基づき統計センターが作成した教育用擬似ミクロデータを使用したが，全国消費実態調査だけではなく，匿名データとしてすでに提供されている他の統計調査（労働力調査，住宅・土地統計調査，就業構造基本調査，社会生活基本調査）についても，教育用擬似ミクロデータを順次提供して頂くことを統計センターに強くお願いしたい．

今回のコンテストでは，最初に擬似ミクロデータを使って規定問題をクリアーことにより，データの特性を十分に理解した上で，自由課題の設定に取りかかるという手順を踏んだ．擬似ミクロデータの利用申請に際しては，事前に研究計画を提出する必要がないので、これが可能であった．これも，多くの SAS ユーザーの関心を引きつけ，多数の応募に繋がったと要因の一つと思われる．

これに対し，匿名データの場合には，利用申請の際に，かなり綿密な研究計画を提出する必要があり，その分野に精通している研究者でなければ，その作業は煩雑で容易ではない．従って，匿名データを用いたデータ分析コンテストは，データ保管など，利用上の種々の制約もあることを考慮すれば，実施することがかなり困難であることは十分承知している[3]．そこで，匿名データを使用した研究報告については，コンペティション形式ではなく，SAS ユーザー総会で研究発表をすることを応募上の条件とし，「匿名データ研究優秀賞」を出すことも検討している．

SAS ユーザー総会において教育用擬似ミクロデータを用いたデータ分析コンテスト，それに続く匿名データを用いた研究発表を継続的に支援することによって，日本における統計分析能力の底上げになることを期待している．また，多くの教育・研究機関で教育用擬似ミクロデータおよび匿名データを用いて実践的な統計教育[4]や本格的な実証研究が幅広く行なわれるように願っている．

ビッグデータの時代が到来しているが、複数の大容量ファイルのマージを伴う公的ミクロデータを分析処理するツールとしては、世界標準の SAS 以外にはあり得ないと考えている。

謝 辞

今回の試みに賛同して頂いた SAS ユーザー会の世話人各位をはじめ，事務局の業務を担当して頂いた SAS Institute Japan（株）の紺屋恭子氏，必要な予算上の措置を賜った SAS Institute Japan（株），および，教育用擬似ミクロデータの利用に便宜を図って頂いた(独)統計センターの関係者各位に深く感謝申し上げます.

参 考 文 献

[1] 槙田直木（2012），「擬似ミクロデータの試行提供」，http://www.nstac.go.jp/services/pdf/121116_1-1.pdf

[2] 秋山 裕美，山口 幸三，伊藤 伸介ほか（2012），「教育用擬似ミ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～」，『統計センター 製表技術参考資料』，http://www.nstac.go.jp/services/pdf/sankousiryou2407.pdf

[3] 小林良行（2012），「公的統計ミクロデータ提供の現状と展望， 一橋大学での取り組みをもとに」，『日本統計学会誌』，第 41 巻，第 2 号，401-420.

[4] 松田芳郎，伴金美，美添泰人（2000），「第 6 章ミクロデータと統計分析」，『講座ミクロ統計分析 ミクロ統計の集計解析と技法』，日本評論社，305-379.

付録：2013 年 SAS ユーザー総会論文集に収録された Let's データ分析コンテスト最優秀賞・優秀賞受賞論文一覧

1） 中島貴之（2013），「シニア世代の消費特徴分析」，『SAS ユーザー総会 2013 論文集』，549-552.

2） 高宗太輔（2013），「食費の分析による食に対する意識と食生活の杷握」，『SAS ユーザー総会 2013 論文集』，523-528.

3） 土井優子（2013），「家計簿から見た「世帯人員ごとの幸福度」に関する研究」，『SAS ユーザー総会 2013 論文集』，537-544.

4） 宇野慧（2013），「世帯主の就業状況が貯蓄性保険需要に与える影響についての考案 ～擬似ミクロデータを用いた Tobit/Hurdle モデル推定～」，『SAS ユーザー総会 2013 論文集』，515-518.

5） 岡村正太，若林将史，東川正晃（2013），「健康因子に対する支出傾向に関する解析」，『SAS ユーザー総会 2013 論文集』，519-522.

6） 魚住龍史（2013），「擬似ミクロデータによる国内旅行費支出と世帯情報の関連の検討」，『SAS ユーザー総会 2013 論文集』，507-510.

7） 富里遼太，土生敏明，米倉孝俊（2013），「教育用擬似ミクロデータを用いた収入・消費傾向の考察」，『SAS ユーザー総会 2013 論文集』，545-548.

8） 筒井杏奈（2013），「変曲点を用いた最低生活費の推定 －疑似ミクロデータ（平成 16 年全国消費実態調査）による－」，『SAS ユーザー総会 2013 論文集』，529-536.