

統計センター提供の教育用擬似マイ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)

要旨

教育用擬似マイクロデータを用いたデータ分析コンテストを 2013 年 7 月に SAS ユーザー総会で第 1 回目を開催し、毎年継続し 2016 年 7 月に第 4 回目を開催した。応募者には規定課題および自由課題について論文形式での提出を求め、予備審査で 9 名の優秀論文を選抜し論文集に収録した。総会前日に公開審査会を開催して応募 3 クラスごとの各 1 名の最優秀賞を選抜し SAS ユーザー総会で口頭発表がなされた。第 1~2 回目の規定課題は、利用者の解析用データセットが正しく作成されているかを確認するために統計センターが用意した 4 つの集計表の再現を求めた。第 3 回目は、10 万世帯比での各種の集計表の作成、および、集計乗率を用いたローレンツ曲線の描画およびジニ係数の算出を求めた。第 4 回目は、集計乗率を用いた 10 万世帯比を主体にした各種の集計表、ヒストグラム、分位点の算出、幾何平均を用いた線グラフの作成などを規定課題とした。また、各年度の最優秀賞となった人達には、次年度のユーザー総会で匿名データを用いた実証研究の発表をお願いした。教育用擬似マイクロデータの提供が 2016 年 3 月末に終了したことから、これまでの 4 年分のデータ分析コンテストを総括することとした。

キーワード: 教育用擬似マイクロデータ, 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、2016 年には第 35 回目となり、7 月 21 日~22 日に神戸国際会議場にて行なわれた。Let's データ分析コンテストは、2013~2016 年の 4 年間継続的に SAS ユーザー総会で開催してきた。これまで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「教育用擬似マイクロデータ」を毎年使用してきた。なお、第 1 回目のコンテストについては、2016 年 11 月の本研究集会において概要を報告した[1]。

擬似マイクロデータは 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]、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、擬似マイ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。

提出された論文に基づき事前審査を行い、参加クラス（A, B, C）ごとに3組、全体で9組の優秀レポートを選抜し、SAS ユーザー総会の論文集へ収録し、さらに、SAS ユーザー総会でのポスター発表も合わせてお願いした。SAS ユーザー総会前日（Day 0）に公開の審査会を開催し、20分のプレゼンテーションにより、参加クラスごとに最優秀賞を1件ずつ決定した。さらに、参加クラス的最優秀賞の3組には、ポスター発表と合わせてSAS ユーザー総会の「Let's データ分析セッション」での口頭発表をお願いした。

擬似マイ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であり、最優秀賞の3組に次年度以降のSAS ユーザー総会で匿名データを使用した研究発表をお願いした。ただし、匿名データの利用申請は煩雑でもあり、費用もかかることから、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し、さらに取得費用についてはSAS Institute Japan（株）からの支援を頂いた。

2. 規定課題

規定課題は、擬似マイクロデータを用いて統計解析が行える基本的な素養、集計乗率を用いた10万世帯比の計算法を知った上での各種の集計表の作成を共通課題とした。第3回目では、集計乗率を考慮したローレンツ曲線の描画およびジニ係数の算出を追加とした。第4回目では、5分位数を求めた上での5分位に関する各種集計表を追加課題とした。次に示すように第4回目でのクロス表は、教育的な配慮から結果が正しいかを確認できる最小限の周辺の統計量の例示に留めた。

1) 産業符号・性別 クロス表

産業符号別・性別に集計用乗率なしの単純クロス集計表、および集計用乗率を使って10万世帯あたり（10万世帯比）のクロス集計表を、表1の様式で作成して下さい。（注：産業符号の「不詳」が先頭に来ても構いません。）

産業符号	a) 単純クロス表			b) 10万世帯比でのクロス表		
	男	女	合計	男	女	合計
鉱業						
建設業						
製造業						
電気・ガス・熱供給・水道業						
情報通信業						
運輸業						
卸売・小売業						
金融・保険業						
不動産業						
飲食店・宿泊業						
医療・福祉						
教育・学習支援業						
複合サービス事業						
サービス業(他に分類されないもの)						
公務(他に分類されないもの)						
不詳						
合計	29,381	2,646	32,027	92,000	8,000	100,000

2) 経常収入および食料費

経常収入および食料費について、集計用乗率を用いて表2のように5分位点（下側20%点、40%点、60%点および80%点）A～Fを求めて、表2を完成して下さい。更に、経常収入および食糧費についてそれぞれヒストグラムを描いて下さい。ヒストグラムの縦軸はパーセント（又は10万世帯比）、横軸は経常収入および食糧費の階級区分（単位：円）とします。

表2

分位点	経常収入	食料費
20%	310,531	49,701
40%	A	D
60%	B	E
80%	C	F

3) 男女別・経常収入および食料費の5分位分布

男女別に経常収入および食料費について、表2に示す5分位別に、以下の表3に示す10万世帯比のクロス表を作成して下さい。

表3

経常収入					食料費				
5分位	5分位数	男	女	計	5分位	5分位数	男	女	計
1	310,531未満			20,000	1	49,701未満			19,999
2	310,531以上			19,993	2	49,701以上			20,000
3	A			20,009	3	D			20,001
4	B			19,995	4	E			19,998
5	C			20,002	5	F			20,003
	計	92,000	8,000	100,000		計	92,000	8,000	100,000

4) 男女別・経常収入・食料費 についての3重クロス表

男女別・経常収入（5分位）別の食料費（5分位）について、表4に示す3重クロス集計表を作成して下さい。

表4

	経常収入 5分位	食料(5分位)					計
		1	2	3	4	5	
男	1						
	2						
	3						
	4						
	5						
	計						
女	1						
	2						
	3						
	4						
	5						
	計						
男女計		19,999	20,000	20,001	19,998	20,003	100,000

5) 男女別・経常収入別に食料費を常用対数変換した平均値の比較

表5に示すように、男女別・経常収入（5分位）別に、食料費の常用対数を計算し平均値を求めて下さい。その平均値について逆対数変換（ 10^x ）をして幾何平均を計算して下さい。この幾何平均について男女別・経常収入の分位点別に食糧費を比較するために、折れ線グラフを作成して下さい。縦軸を幾何平均（単位：円）、横軸を経常収入（5分位）とします。

表 5

	経常収入	食料		
	5分位	世帯数	常用対数平均	幾何平均
男	1			
	2			
	3			
	4			
	5			
	計	92,000	4.8395	69,103
女	1			
	2			
	3			
	4			
	5			
	計	8,000	4.7325	54,013

規定課題ここまで。

第 3, 4 回目の規定課題は, ミクロデータの統計解析を行うための基礎的な素養を身に着けられるように配慮をした。すなわち, 集計乗率を用いて 10 万世帯比でのクロス表, 分位点の算出, 3 重クロス表, ローレンツ曲線の描画とジニ係数, 性別・経常収入 5 分位別の食料費についての幾何平均などである。

第 4 回目の「5) 男女別・経常収入別に食料費を常用対数変換した平均値の比較」は, 2) から 4) までの前準備を前提に, 繰返しが不揃いな 2 元配置分散分析での交互作用プロットによる結果の吟味を念頭にして出題した。経常収支などの変数についての区分数は, 切れ目のよい値での 10 区分が標準的に用いられているが, 分散分析の要因として用いるためには重すぎる。統計的には 4 分位数がメジャーであるが, 解析の目的に合わせた分位数の算出の経験をしてもらいたいとの意図から, 第 4 回の規定課題では, 5 分位数を求め, 経常収入の 5 分位による食料費の幾何平均の算出を課題とした。

3. 自由課題

優秀賞および最優秀賞に選ばれた年度別の自由課題は, 表 6 に示すように多くの興味がそそられるような内容となっている。優秀賞については, 提出された論文の審査によって決定され, 最優秀賞については, 公開の審査会で優秀賞受賞者全員のプレゼンターを聞いた後に決定した。年度によっては, クラス別の最優秀賞に該当者がいないこともあり, その際には別のクラスから準最優秀賞を選出したこともあった。

コンテストの優秀賞者全員の論文は, SAS ユーザー総会論文集に収録した。SAS ユーザー会では, 1982 年の年会以後の全ての論文集を無料公開としているので, 全ての優秀賞者全員の論文は, ダウンロード可能となっている。ただし, 論文単位ではなく年度ごとの論文集の一括ダウンロードとなっている。

表 6 年度別の優秀賞および最優秀賞者リスト

年度	代表者氏名	応募クラス	演題	審査結果
第1回	中島 貴之	C	シニア世代の消費特徴分析	
2013年	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把握	
	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 ～擬似マイ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	擬似マイ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	教育用擬似マイ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 ～擬似マイクロデータ(平成16年全国消費実態調査)による～	A 最優秀賞
第2回目	芥川 麻衣子	C	マイクロデータを用いたセルフメディケーションの実態把握に関する検討	
2014年	吉田 大祐	C	ゲーム・インターネット好きの世帯は、どのような食生活をしているのか	
	笹谷 知輝	C	年代別の魚食傾向に関する考察～教育用擬似マイクロデータを用いて～	C 最優秀賞
	谷口 裕明	C	住宅ローン返済中における家計の逼迫と消費行動に関する分析	
	宇野 慧	B	同時方程式モデルを用いた健康/不健康支出の分析	
	木下 陽介	B	ジニ係数による所得格差の解析	B 最優秀賞
第3回	辻智 康	C	「理想の生活を手に入れるための職業選択と貯蓄法」 ～趣味にお金を使いたい人を例に～	
2015年	中川 雄貴	C	家計係数の妥当性と特性の検討～Geek and Angel～	C 最優秀賞
	竹田 雅裕	B	世帯年代別からみた賃貸・持ち家の動向	
	谷 和泰	B	酒とタバコ: 収入への関連性	
	中松 健	B	コメ消費量の減少と食事づくりの簡素化に関して	B 最優秀賞
	谷口 裕明	A	家計の支出行動から探る無担保ローンの利用者像	
	井上 貴博	A	40歳未満の夫婦共働き世帯における一子あり世帯と 子どもなし世帯の比較: 横断研究	A 最優秀賞
第4回	賈 慧玲	C	シングルマザー世帯の家計の実態 ～片働き世帯と共働き世帯との比較から～	C 最優秀賞
2016年	村木 翔太	C	収入と医療費の関係から見える「医療格差」に関して	
	田中 健一郎	B	世代別及び経常収入別に見た官民の平均貯蓄率の 差異に関する考察	B 最優秀賞
	田中 基史	B	一戸建て・共同住居住まい世帯における傾向の比較	
	内田 晃秀	B	マイナス金利下におけるローン事業の発展性と 疑似マイクロデータに見るローン利用者の行動特性	
	四方 克成	B	貧困・中流・富裕世帯における教育費の実態調査	
	野田 斉	B	業種で見る中小企業の収入 ～大企業との比較～	
	松本 光右	A	消費税率の引上げ及び軽減税率制度の導入は低所得者に どのような影響を与えるか	A 該当者なし

第1回目 2013年のコンテストの論文が収録されている論文集は、

<http://www.sascom.jp/download/pdf/SUGJ2013.pdf>

によりダウンロードでき、この507-548ページにコンテストの規定課題および自由課題の論文が掲載されている。第2回 2014年以降も同様にダウンロードができる。

第2回目：<http://www.sascom.jp/download/pdf/SUGJ2014.pdf>, 847-904.

第3回目：<http://www.sascom.jp/download/pdf/SUGJ2015.pdf>, 591-676.

第4回目：SAS ユーザー総会論文集（Web 準備中），365-479.

疑似マイクロデータをもちいたコンテストで最優秀賞を獲得した人達に、次年度の SAS ユーザー

総会で匿名データを用いた研究発表をお願いしてきた。その結果として研究発表となった人達のリストを 表 7 に示す。

表 7 SAS ユーザー総会で匿名データを用いた演題

年度	代表者氏名	演題
2014	魚住 龍史	Let's匿名データ分析:パック旅行費支出と世帯情報の関連の検討
	宇野 慧	全国消費実態調査の匿名データを用いた2人以上世帯の保険需要の分析
2015	#高橋 行雄	統計センターの匿名データ13万件を用いた統計解析の実践
	#周防 節雄	全国消費実態調査の匿名データとその符号表から自動的にSASのデータセット, 変数ラベルと変数フォーマットを作成するSASプログラム
2016	*池田 瑞穂	全国消費実態調査の匿名データから見た食料品目別消費傾向による食生活の変化について
	井上 貴博	匿名データを用いた夫婦共働き世帯における子あり世帯と夫婦のみ世帯の比較:横断研究
	中松 建	匿名データを用いたコメ離れと食事づくりの簡素化についての検証
	# 主催者, * コンテスト共同研究者	

魚住氏の 2014 年の「Let's 匿名データ分析：パック旅行費支出と世帯情報の関連の検討」は、前年 2013 年の「擬似マイ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」を実証研究とした結果である。宇野氏の 2014 年の「全国消費実態調査の匿名データを用いた 2 人以上世帯の保険需要の分析」は、前年 2014 年の「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察」を実証研究とした結果である。

池田氏の 2016 年の「全国消費実態調査の匿名データから見た食料品目別消費傾向による食生活の変化について」は、前々年 2014 年の笹谷氏らとの共同研究「年代別の魚食傾向に関する考察－教育用疑似マイクロデータを用いて－」をさらに発展させた実証研究である。井上氏の 2016 年の「匿名データを用いた夫婦共働き世帯における子あり世帯と夫婦のみ世帯の比較:横断研究」は、前年 2015 年の「40 歳未満の夫婦共働き世帯における一子あり世帯と子どもなし世帯の比較:横断研究」を実証研究とした結果である。中松氏の 2016 年の「匿名データを用いたコメ離れと食事づくりの簡素化についての検証」は、前年の「コメ消費量の減少と食事づくりの簡素化に関して」を実証研究とした結果である。

擬似マイクロデータをもちいたコンテストで最優秀賞を獲得した人達に、SAS ユーザー総会だけでなく、表 8 に示すように統計センターが主催している公的統計のマイクロデータ研究集会での研究発表をお願いしてきた。

魚住氏の 2013 年の「SAS による擬似マイクロデータを用いたパック旅行費支出と世帯情報の関連の検討」発表は、同年の SAS ユーザー総会での擬似マイクロデータを用いた結果をさらに充実させた発表である。同様に宇野氏の「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察」も同年の SAS ユーザー総会の結果である。

池田氏の 2014 年の「世代別魚介類消費の推移に関する考察」は、前年の笹谷氏との共同研究「年代別の魚食傾向に関する考察－教育用疑似マイクロデータを用いて－」を実証研究とした結果である。魚住氏の 2014 年の「SAS による匿名データ分析：パック旅行費支出と世帯情報の関連の検討」は、前年 2013 年の擬似マイクロデータでの結果に基づいた実証研究である。

表 8 公的統計のマイクロデータ研究集会での演題

年度	代表者氏名	演題
2013	#高橋 行雄	統計センター提供の教育用擬似マイクロデータを用いた SAS/JMPによるデータ分析コンテストの概要
	魚住 龍史	SASによる擬似マイクロデータを用いたパック旅行費支出と世帯情報の関連の検討
	宇野 慧	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察
2014	*池田 瑞穂	世代別魚介類消費の推移に関する考察
	#周防 節雄	教育用擬似マイクロデータの符号表からSASデータセット用の変数ラベルと変数フォーマットを自動作成するSASプログラム
	魚住 龍史	SASによる匿名データ分析: パック旅行費支出と世帯情報の関連の検討
2015	#高橋 行雄	全消匿名データに対する多次元分散分析の活用
	#周防 節雄	全国消費実態調査の匿名データからSASのリレーショナルデータベースを完全自動構築するシステムの開発
	*大野 嵩護	食生活から見た食料消費傾向に関する考察－匿名データを用いて－
2016	*矢田 徹	全国消費実態調査の匿名データを用いた40歳未満の共働き核家族世帯における子あり世帯と夫婦のみ世帯の比較: 横断研究
	#高橋 行雄	教育用擬似マイクロデータを用いた4回のデータ分析コンテストの総括
	# 主催者, * コンテスト共同研究者	

大野氏の 2015 年の「食生活から見た食料消費傾向に関する考察－匿名データを用いて－」は、前年の共同研究者としての「年代別の魚食傾向に関する考察－教育用擬似マイクロデータを用いて－」をさらに発展させた実証研究である。

矢田氏の 2016 年の「全国消費実態調査の匿名データを用いた 40 歳未満の共働き核家族世帯における子あり世帯と夫婦のみ世帯の比較: 横断研究」は、前年 2015 年の共同研究「40 歳未満の夫婦共働き世帯における一子あり世帯と子どもなし世帯の比較: 横断研究」についての実証研究である。

4. 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。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」を開催することとなった。用いるデータは、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「教育用擬似マイクロデータ」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。第1回目は、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。第2回目以後も 3 クラス制を継続したが、資格要件については微調整を毎回行った。

この擬似マイ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、第1回目と2回目は、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、原則として各クラス 3 組を優秀賞として選んだが、優秀賞の水準に達していなければ、3 組にはこだわらないこととした。統計センターの Web サイトには、擬似マイクロデータに関する Q&A に以下の記述がある。「擬似マイクロデータは、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マイクロデータ形式の擬似的なデータです。擬似マイ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。」との注意がなされている。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マイ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」（公的統計のマイ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少なからずあった。

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マイクロデータならば、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た。

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、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を得た。

擬似マイ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、次回以降の SAS ユーザー総会で、匿名データを用いた研究発表も推進する必要性を強く感じた。このため、筆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とした。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した。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てきた。

今回のコンテストでは、最初に擬似マイ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。擬似マイ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る、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」を出してきた。

SAS ユーザー総会において教育用擬似マイ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た。また、多くの教育・研究機関で教育用擬似マイ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きた。

残念ながら、教育用疑似マイクロデータは、2016年3月末で統計センターからの提供が打ち切られた。これに代わるマイクロデータとして、平成21年度の全国消費実態調査についての一般マイクロデータが提供されているが、その内容は教育用疑似マイクロデータの197変数に対して20変数と1/10と大幅に減少し、世帯人員など世帯に関するデータが(2人, 3人以上)のように極端な2区分化が行われている。一般マイクロデータを用いたデータ分析コンテストの実施可能性を検討してはみたが、消費に関する変数として十大費目に限定されており、公募方式のコンテストに用いるマイクロデータとしては適さないと判断した。

これまでの教育用疑似マイクロデータを用いたデータ分析コンテストに多数の応募者があり、匿名データを用いた実証研究を行う切っ掛けとなったことは、喜ばしいことでもあり何らかの方法で新たに“疑似マイクロデータ”の提供が再開されることを切望する。

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、教育用擬似マイクロデータおよび匿名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] 高橋行雄, 周防節雄(2013), 統計センター提供の教育用擬似マイクロデータを用いた SAS/JMP によるデータ分析コンテスト, http://www.nstac.go.jp/services/pdf/131122_4-1.pdf.
- [2] 槇田直木(2012), 「擬似マイクロデータの試行提供」, http://www.nstac.go.jp/services/pdf/121116_1-1.pdf
- [3] 小林良行(2012), 「公的統計マイ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章マイクロデータと統計分析」, 『講座マイクロ統計分析 ミクロ統計の集計解析と技法』, 日本評論社, 305-379.