全国消費実態調査4回分の匿名データから作成した
擬似ミクロデータを用いたデータ分析コンテスト

BioStat研究所（株）　高橋 行雄
(公財)統計情報研究開発センター　周防 節雄
(独)統計センター 統計情報提供課　宮内 亨

要旨　SAS ユーザー総会では，全国消費実態調査(2004年)のデータおよび匿名データから作成された2種類の擬似ミクロデータ(**教育用擬似ミクロデータ**と**新擬似ミクロデータ**)を用いて2013年から昨年まで「Let's データ分析コンテスト」を開催してきた.本報告では，昨年までの計6回の「Let's データ分析コンテスト」を振り返ると同時に,これまでに応募された42論文の概要に触れる．更に，その後我々が新開発した4年次分の新擬似ミクロデータを用いて，本年9月5～6日に開催された SAS ユーザー総会（＠国際医療福祉大学・東京赤坂キャンパス）で実施した「Let's データ分析コンテスト」の結果についても述べる．

教育用擬似ミクロデータ　全国消費実態調査(2004年)のデータから(独)統計センターが作成し,無料で提供されていたが,現在は提供停止となっている．

新擬似ミクロデータ　教育用擬似ミクロデータの提供停止に伴い,その代わりの擬似ミクロデータとして，我々が2004年全国消費実態調査の匿名データを用いて4種類の統計表を作成し，この統計表だけを用い正規乱数・多次元正規乱数・一様乱数を適宜発生させて復元し,新擬似ミクロデータを作成した[1].

新擬似ミクロデータ4年次分　2004年版新擬似ミクロデータと同様の手法により，全国消費実態調査の1989，1994，1999及び2004年の匿名データを用いて4年次分の擬似ミクロデータを作成した.それらをひとつにまとめ,222変数，271,197レコードから成る SAS データセットを作成後，CSV ファイルにも変換.これら二つの形式の新擬似ミクロデータファイルとメタデータが一つのZIPファイルに収められ一般公開されており,誰でも無償でダウンロードできる．http://mighty.gk.u-hyogo.ac.jp/sas4zensho/

Let's データ分析コンテスト　SAS ユーザー総会のイベントの一つとして2013年から実施しているデータ分析コンテスト．これまでに,教育用擬似ミクロデータを用いて4回(2013～2016年)，新擬似ミクロデータを用いて2回(2017～2018年)のデータ分析コンテストを実施してきた．SAS または JMP の利用者なら誰でも応募（募集は毎年5～6月頃）できる．応募は SAS の利用歴に応じて3クラス(A クラス：利用歴3年以上，B クラス：利用歴3年未満，C クラス：学部学生・修士院生)ある．応募者には規定課題（指定されたクロス表やグラフ等の作成）と自由課題(自由テーマで論文作成)を課す.事前審査で各クラス上位3名が選抜され，SAS ユーザー総会の論文集に収録される[3]．総会前日に開催される公開審査会の口頭発表を経て決定された各クラスの最優秀賞受賞者は SAS ユーザー総会当日に口頭発表できる[2].

擬似ミクロデータの必要性　学術誌などで，報告される論文の実証研究の結果を追試したいと思っても，使用されたデータが非公開であるとか，研究者がデータの公表を望まないことになどより，統計解析の結果の追試ができないことが，統計解析法の進歩の阻害要因の一つと考える．誰もが無償で同じデータが利用できる擬似ミクロデータがあれば，追試が可能となり，オープンな議論ができることに着目した．各種公的統計調査の匿名データから作成される擬似ミクロデータを今後も必要に応じて開発していきたい．

文献

[1] 高橋行雄，周防節雄，宮内亨(2017)，全国消費実態調査(2004 年)の匿名データから JMP による新擬似ミクロデータの作成，http://www.nstac.go.jp/services/pdf/171117_1-2.pdf.
[2] 高橋行雄(2016) 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，https://www.nstac.go.jp/services/pdf/161125_3-3.pdf.
[3] SAS ユーザー総会(1982～2017),論文集，https://www.sas.com/ja_jp/usergroups.html#m=agenda-downloads