

政府統計を用いた社会科学系向けデータ分析教材作成

Development of Teaching Materials on Data Analysis in the field of Social Sciences using Government Statistics

吉根 勝美

Katsumi YOSHINE

南山大学経済学部

Faculty of Economics, Nanzan University

Email: kyoshine@nanzan-u.ac.jp

あらまし：社会科学系の大学新入生を対象にした表計算ソフトウェアによるデータ分析の指導では、架空のデータではなく、後に実際の分析対象となるデータを教材にする方が望ましい。社会科学系の学生には、さまざまな分野を網羅する政府統計がデータ分析の教材に適しているが、統計ごとにデータ構成が異なるため、一律的な教材化が困難である。本稿では、政府統計を用いたデータ分析教材の作成支援手法を、具体例を通して検討する。

キーワード：政府統計，教材開発，データ処理，実際のデータ

1. はじめに

社会科学系の大学新入生を対象にした表計算ソフトウェアによるデータ分析の指導では、のちに実際の分析対象となる可能性があるデータを教材にする方が望ましい。社会科学系の大学生には、問題発見・解決能力向上の一環として、必要なデータを自ら収集し、分析するスキルを獲得させたい。その第1段階で行われる表計算ソフトウェアの操作指導では、架空の数値を使って、きれいな結果が得られるようにすることが多い。しかし、いずれ使いこなすためには、必ずしもきれいな結果が得られない実際のデータにも対応できるようにする必要がある。

大学生や社会人に対する統計教育では、架空のデータではなく、実際のデータを使用することが望ましい。アメリカ統計学会は、「統計教育における評価と指導のガイドライン」のCollege Report 2016⁽¹⁾で、実際のデータを使用することの重要性を示している。同ガイドラインでは6項目の推奨事項が提示されており、その中の1つ Recommendation 3: Integrate real data with a context and a purpose. の前提となるのが“Use real data”である。

社会科学系の大学生、特に経済学部生が実際の分析対象として政府統計を利用することが多い。例えば、景気動向の把握に用いられる主要な経済統計は、GDPをはじめとして、個人消費、設備投資、住宅投資、公共投資、輸出・輸入・国際収支、生産・出荷・在庫、企業収益・業況判断・倒産、雇用、物価、金融の多分野に及ぶ⁽²⁾。これらの経済統計の一部は民間が作成しているが、大多数は政府が作成している。

政府統計は、統計ごとにデータ構成が異なるため、一律的な教材化が困難である。政府統計のポータルサイト e-Stat⁽³⁾には、605調査のデータ、1,531,904件のデータセットが掲載されている(2019年6月18日

現在)。各調査は1個以上の統計表からなり、EXCEL、CSV、PDF形式のファイルをダウンロードするか、データベースの検索結果としてデータを取得する。

本稿では、政府統計から、社会科学系新入生向けのデータ分析教材作成の支援方法を明らかにすることが目的である。具体的には、総務省統計局による小売物価統計調査年報 平成29年のうち「調査品目の月別価格及び年平均価格【県庁所在市及び人口15万以上の市】」⁽⁴⁾を用いて、ヒストグラム作成と散布図作成のそれぞれにふさわしいデータを見つけるためのユーザインタフェースを試作し、この試作を通じて、教材開発の支援手法を検討する。

2. 試作環境

Windows 10 Home で動作する RStudio Desktop Open Source Edition (Version 1.2.1335)で試作し、R version 3.6.0 で動作を確認している。エクセルファイルを読み込むパッケージ readxl と、インタラクティブな Web アプリケーションを容易に構築できるパッケージ shiny を使用している。

R 言語の使用はプログラムの簡潔さをもたらし、プログラミングの経験が少ない教材作成者でも、自らプログラムの修正が可能になることが期待できる。

3. ユーザインタフェースの試作

3.1 前処理

81 都市 560 品目の 2017 年平均小売価格をダウンロードし、欠損値を空欄にした EXCEL 形式のファイルを元データとした。

3.2 ヒストグラム

試作したヒストグラム閲覧用ユーザインタフェースを図1に示す。教材作成者が求めるヒストグラム

の形状を見つけることを支援するため、品目を切り替えると、直ちにヒストグラムが描き換えられるようにした。

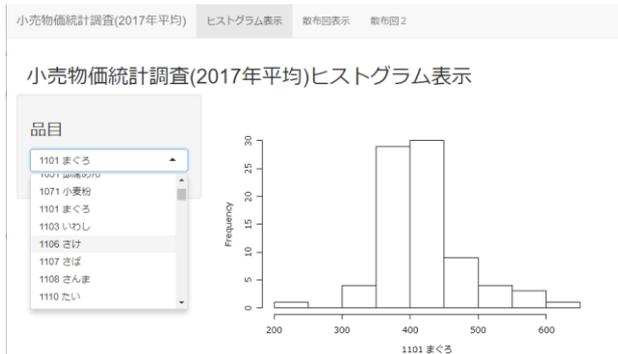


図1 ヒストグラム閲覧用ユーザインターフェース

また、プログラム中、データフレーム `sk` に品目毎に歪度を計算し、`names(sk)[sk > 1]` とすれば、図2のように、右に裾の長い分布の品目だけがインターフェースに現れるようにすることが可能である。

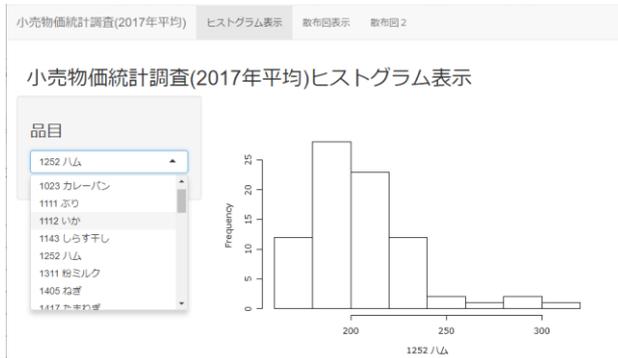


図2 図1で「歪度 > 1」に限定した場合

3.3 散布図

試作した散布図閲覧用ユーザインターフェースを図3に示す。教材作成者が求める分布を見つけることを支援するため、2種類の品目を切り替えると、直ちに散布図が描き換えられるようにした。



図3 散布図閲覧用ユーザインターフェース

また、例えば「やや正の相関」がみられる分布を見つけやすくするため、相関係数行列から特定の範囲にある品目の組合せを作成しておいて、組合せを切り替えると直ちに散布図が描き換えられるような

インターフェースを図4のように作成した。本試作では、コードが1100番台で欠損値のない21品目について相関係数行列 `cor.df` を作成し、相関係数の値が0.3以上の品目の組合せのリスト `pair` を下のよう

```
if(cor.df[i,j] > 0.3) {
  pair<-c(pair,paste(rownames(cor.df)[i],"x",colnames(c
or.df)[j]))
}
```

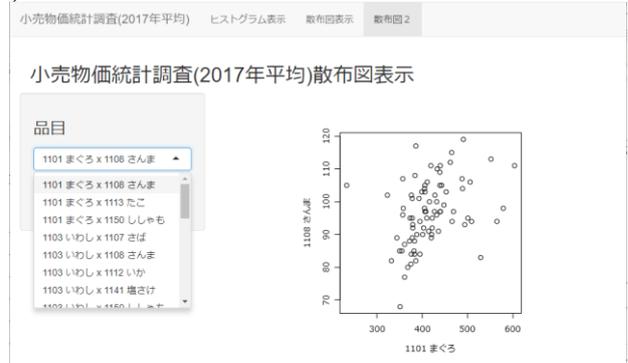


図4 指定範囲内の相関係数の散布図表示

4. おわりに

本稿では、都市別小売物価のデータを具体例として、社会科学系新入生向けのデータ分析教材作成を支援することを目的に、ヒストグラム作成と散布図作成のそれぞれにふさわしいデータを見つけるためのユーザインターフェースを試作した。

この試作にはR言語を用い、インタラクティブなWebアプリケーションを容易に構築できるパッケージ `shiny` を使用することにより、簡潔なプログラムで実現することができた。

プログラムの簡潔さは、プログラミングの経験が少ない教材作成者でも、`names(sk)[sk > 1]` や `cor.df[i,j] > 0.3` の部分を書き換えれば自分が望む分布が得られることに気付くという利点をもたらしている。

今回の試作では、教材作成者がさまざまなデータに対するグラフを次々と閲覧しながら、教材にふさわしいものを見つけるということを想定している。今後の検討課題として、例えば、特徴が類似しているデータ、まったく特徴が異なるデータ、対照的な特徴をもつ2群に分けられるデータのように、教材作成者の要求に沿うデータを探索する手法の開発が挙げられる。

参考文献

- (1) GAISE College Report ASA Revision Committee: "Guidelines for Assessment and Instruction in Statistics Education College Report 2016", <http://www.amstat.org/education/gaise> (2016)
- (2) 内閣府: "マクロ経済統計リンク集", <https://www5.cao.go.jp/keizai3/getsurei/macro/>
- (3) 総務省統計局, 独立行政法人統計センター: "政府統計の総合窓口(e-Stat)" <https://www.e-stat.go.jp/>
- (4) 総務省統計局: "小売物価統計調査(動向編)について(2018年4月現在)" <http://www.stat.go.jp/data/kouri/doukou/1.html> (2018)