

経済データ処理実習のための例題データベース

Database of Examples of Data File for Exercises on Economic Data Processing

吉根 勝美^{*1}

Katsumi YOSHINE^{*1}

^{*1} 南山大学経済学部

^{*1} Faculty of Economics, Nanzan University

Email: kyoshine@nanzan-u.ac.jp

あらまし：学部新入生に対する情報リテラシー教育において、オフィスソフトの基本的操作を習得させる際には、学部の専門性に配慮することが望ましい。例えば、経済学部においては、表計算ソフトを用いて実際の経済データに簡単な処理を施し、グラフ化して、短いレポートを作成するということである。本研究では、例題作成支援に資することを目的にして、データ処理で使用する経済データのデータベース構築の要件を議論する。

キーワード：初年次教育，データ分析，実習用例題，実データ，経済時系列データ

1. はじめに

特に文系学部の新入生に対する情報リテラシー教育において、オフィスソフトの基本的操作を習得させる際には、学部の専門性に配慮することが望ましい。例えば、経済学部においては、表計算ソフトの実習として、実際の経済データを分析させることが望ましい。本研究では、この実習で用いる例題の作成支援に資することを目的にして、データ処理の実習で使用する経済データのデータベース構築の要件を議論する。

2. 実際の実習内容

筆者が所属する経済学部の新入生は、必修科目「データ処理入門」において表計算ソフトを用いた経済データの基礎的な処理手法を学ぶ。セル番地を含む数式の作り方は、変化率(増減率)・構成比(シェア)・寄与度の算出や、基準年を=100とする指数化、物価変動の影響を除く実質化を通して、関数の使い方は、平均、中央値、標準偏差、変動係数(=標準偏差÷平均)の算出を通して学ぶ。

授業で使用する最初の例題は、平成29年度刊愛知県統計年鑑の11-3表「百貨店・スーパーの事業所数、従業者数及び販売額等」から、百貨店における商品別の販売額の平成28年の月次データを抽出したものである(図1)。新入生の表計算ソフトへの習熟度には大きな開きがあるのが現状であり、初学者が学ぶべき基礎操作として縦横集計を採用した。また、経済学部の新入生が、のちに受講する授業において目にする機会が多いのは時系列データであることを考慮した。さらに、このデータが百万円単位であることを伏せて、経済学部の新入生に経済データに対する感覚を自覚させるねらいもある。

セル番地を含む数式を作る実習のために、総務省統計局「家計調査」および「消費者物価指数」から抽出した図2の例題を用意している。経済時系列データの基本的な処理として、変化率・構成比・寄与度、指数化、実質化の説明の後、図2のデータを使って実際に計算してみる。

表計算ソフトの関数の使い方は、基本統計量の学習を兼ねて、図1の時系列方向に対して、平均、中央値、標準偏差、変動係数を算出する実習を行う。時系列データに対して平均や標準偏差を求めることは、経済学分野での基本統計量の応用例であるリターンとリスクの考え方を知る上でも有用である。

	A	B	C	D	E	F	G	H
		紳士服・洋品	婦人・子供服・洋品	その他の衣料品	身の回り品	飲食料品	その他	販売額計
1								
2	1月	3069	12751	883	6164	9003	9138	
3	2月	1691	8323	546	4124	9625	8749	
4	3月	2475	12211	901	5607	9331	10587	
11	10月	2943	9947	938	4945	8175	9498	
12	11月	3127	9647	775	5041	9907	9733	
13	12月	3766	11196	716	7453	17135	13462	
14	年間合計							

図1 最初の例題(百貨店商品別販売額)

	A	B	C	D	E
		1世帯当たり年平均1か月の消費支出(円)	消費支出のうち食料(円)	消費支出のうち被服及び履物(円)	消費者物価指数(総合)
1					
2	1980年	238126	66245	17914	74.5
3	1981年	251275	69032	18417	78.1
4	1982年	266063	71046	18915	80.3
37	2015年	315428	74405	13598	100
38	2016年	310389	74939	13153	99.9
39	2017年	312926	74634	13206	100.4

図2 セル番地を含む数式のための例題

3. 例題データベースの必要性

新入生が表計算ソフトの基本操作を習得するために実習を繰り返すとき、例題となる経済データをその都度入れ替える必要がある。例えば、図1の代わりとなるデータとして、総務省統計局「家計調査」より作成した菓子類への支出額（和生菓子、洋生菓子、スナック菓子、チョコレート、アイスクリーム、その他）を用いたり、財務省「貿易統計」より作成した主要商品別輸出額（化学製品、原料別製品、一般機械、電気機器、輸送用機器、その他）を用いたりした実習を行っている。図2の代わりに、厚生労働省「賃金構造基本統計調査」の「第1表 企業規模別新規学卒者の初任給の推移」から抽出した大卒初任給額を、総務省統計局「小売物価統計調査」の東京都区部小売価格から抽出したある品目の価格を用いて、物価変動を除く実質化を実施するレポート課題を出題している。

データが入れ替わっても戸惑いなく表計算ソフトの基本操作を学ぶ実習ができるよう、データの枠組みは変更しない。図1は、縦方向は合計可能な1年分の月次データであり、横方向は複数の部分を合計する意味のあるデータであるという枠組みを有している。図2は、ある金額を示す長期の時系列データと物価を示す同時期の時系列データからなるという枠組みである。

実習を繰り返すときの練習用データは、まったくの架空データや、実際のデータを模して作られた模擬データを使うことも考えられる。こうしたデータを使えば、学生ごとに異なるデータを用いたレポートを出題することも可能になるという利点がある。

しかし、経済学部の新入生には、実際の経済データを用いた実習を行う方が、架空データや模擬データを用いるよりも有効であると考え、本研究では、実習用の例題作成に資することを目的として、経済データのデータベース化を試みることにした。新入生が表計算ソフトの基本操作を学ぶ目的は、その後の大学での各分野における学習や研究においてツールとして使用するためである。架空データより分野に合わせたデータの方を用いる方が実習は効果的になる。経済学部生なら、模擬データを用いて経済データの分析の仕方を学ぶことも可能だが、実社会と結びついている経済データは、実データのもつ数値自体についても学ぶべきである。

4. 例題データベースに必要な要件

本研究で議論する例題データベースでは、図1、2の枠組みに当てはまる経済データの単なる寄せ集めではなく、指定した枠組みに当てはまるようにデータを再構築してくれる機能を必要とする。図1、2の枠組みに当てはまらない経済データの分析手法を学ぶ可能性もあり、実際「データ処理入門」では、図1を縦方向に拡張し、7~8年に渡る月次データを分析対象とするレポート課題を出題している。

経済時系列データには、毎月末時点での自動車保有台数のように、時系列方向の合計に意味がないものがあるので、データベースに格納するときには、そのことを示す属性を付す。リターンとリスクを学ぶのに欠かせない株式・為替データにも、時系列方向の合計ができない旨の属性を持たせる。

実際の経済データは、さまざまなフォーマットで公開されており、例題データベースの内部ではデータの統一的な表現が望ましい。Tidy data⁽¹⁾⁽²⁾⁽³⁾は内部表現の1候補であり、例えば、「主要商品別輸出額」は図3のように表現できる。

また、用語や単位を統一的に扱うため、用語集の整備が望ましい。また、「百万円」単位で公表されている金額データは、単位「円」と乗数（ $\times 10^6$ ）を分けて考えるとよい。用語集の構築には、できるだけ既存のものを使う。例えば、総務省統計局は、統計のLOD(Linked Open Data)化を進めており⁽⁴⁾、この中で単位や単位乗数の定義をしている⁽⁵⁾。

図3のデータから、図1の枠組みに合う例題を作るには、特定の地域向けの1年分のデータを抽出し、主要な商品5つを選択して、残りはその他としてまとめる。また、金額と人数のように異なる単位を持つデータを組み合わせる例題も考えられる。

5. まとめ

経済時系列データを分析する実習のための例題データベースの構築について議論した。経済データに特有の属性を考慮するほか、Tidy dataの導入、用語や単位の語彙の統一を提案した。今後は、授業で使用したデータからデータベースの実装を試みる。

参考文献

- (1) Hadley Wickham: "Tidy data", Journal of Statistical Software, Vol.59, No.10 (2014) doi:10.18637/jss.v059.i10
- (2) Hadley Wickham and Garrett Grolemund: "Tidy Data with tidy", R for Data Science, pp.147-169, O'Reilly (2016)
- (3) 西原史暁: "整然データとは何か", 情報の科学と技術, Vol.67, No.9, pp.448-453 (2017)
- (4) 西村正貴: "Linked Open Data (LOD) による統計データの提供: 政府統計データ (e-Stat) の新しい形", 情報管理, Vol.59, No.12, pp.812-821 (2017)
- (5) 総務省統計局, "属性", 統計LOD, <https://data.e-stat.go.jp/lodw/rdfschem/observation/>

1988年	1月	米国	総額	輸出金額	738613800	千円
1988年	2月	米国	1. 食料品	輸出金額	2591713	千円
2018年	3月		9. その他	輸出金額	131358703	千円
2018年	4月	米国	総額	輸出金額	1286361863	千円
2018年	4月	米国	1. 食料品	輸出金額	7833692	千円
2018年	4月	米国	2. 原料品	輸出金額	4842822	千円
2018年	4月	米国	3. 鉱物性燃料	輸出金額	6537869	千円
2018年	4月	米国	4. 化学製品	輸出金額	74821743	千円
2018年	4月	米国	5. 原料別製品	輸出金額	79644933	千円
2018年	4月	米国	6. 一般機械	輸出金額	292916426	千円
2018年	4月	米国	7. 電気機器	輸出金額	175929526	千円
2018年	4月	米国	8. 輸送用機器	輸出金額	507381512	千円
2018年	4月	米国	9. その他	輸出金額	136453340	千円

図3 データの内部表現