

経済学部におけるデータ処理実習のための 例題データベースの設計

吉根 勝美*¹

*¹ 南山大学経済学部

Design of a Database of Example Data for Data Processing Exercise at the Faculty of Economics

Katsumi YOSHINE *¹

*¹ Faculty of Economics, Nanzan University

学部新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、学部の専門性にも配慮することが望ましい。例えば、経済学部においては、実際の経済データを用いて、表計算ソフトによる簡単な処理を施し、グラフ化して、ワープロによりレポートを作成するという実習形態が考えられる。本発表では、データ処理実習で使用するため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の設計について議論する。

キーワード：初年次教育，データ分析，実習用例題，実データ，時系列データ

1. はじめに

大学新生、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、実際の経済データを対象にし、表計算ソフトの実習として、四則演算によるデータ加工、関数による統計計算、グラフ作成を繰り返し練習して、経済データの基本的な取り扱いを学び、まとめとして、経済データの分析レポートをワープロで作成させることが望ましい。本研究では、このような実習で用いる例題の作成支援に資することを目的として、経済学部における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の設計について議論する。

2. データ処理実習の内容

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、表計算ソフトを使いながら経済データの基礎的な処理手法を学ぶ。セル番地を含

む数式の作り方を通して、変化率(増減率)・構成比(シェア)・寄与度の計算や、基準年=100とする指数化、物価変動の影響を除く実質化の手法を学ぶ。また表計算ソフトの関数の使い方を通して、平均、中央値、標準偏差、変動係数(=標準偏差÷平均)を学ぶ。

2.1 表計算ソフトの導入

表計算ソフトの導入時に使用する例題は、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)。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。また、このデータが百万円単位であることを伏せて提示し、経済学部の新入生に、経済データの数字に対する感覚を自覚させるねらいもある。なお、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、新入生が最初に学ぶべき操作を、データ入力、縦横集計、書式設定の3つとした。

	A	B	C	D	E	F	G	H
1	○○○○○○○○○○○○○○○○○○○○							
2	単位：××円							
3		紳士服・洋 品	婦人・子供 服・洋品	その他の衣 料	身の回り品	飲食品	その他	合計
4	1月	3,069	12,751	883	6,164	9,003	9,138	41,008
5	2月	1,691	8,323	546	4,124	9,625	8,749	33,058
6	3月	2,475	12,211	901	5,607	9,331	10,587	41,112
7	4月	2,662	9,851	637	5,003	7,192	9,153	34,498
8	5月	2,624	9,524	771	5,066	7,538	9,170	34,693
9	6月	2,328	8,397	626	4,646	9,303	9,901	35,201
10	7月	2,781	10,854	852	6,079	11,939	11,105	43,610
11	8月	1,714	7,274	611	4,430	7,550	9,230	30,809
12	9月	1,946	8,412	663	4,564	7,279	9,334	32,198
13	10月	2,943	9,947	938	4,945	8,175	9,498	36,446
14	11月	3,127	9,647	775	5,041	9,907	9,733	38,230
15	12月	3,766	11,196	716	7,453	17,135	13,462	53,728
16	計	31,126	118,387	8,919	63,122	113,977	119,060	454,591
17	愛知県統計課「平成29年度刊愛知県統計年鑑」より							

図 1 導入時の例題（愛知県百貨店商品別販売額）

2.2 時系列データ分析の実習

経済時系列データの基本的な分析方法を学ぶために使用する例題は、総務省統計局「家計調査」から抽出した、1980年以降の実収入の年次データである(図2)。このデータに対して、対前年変化率(伸び率)や、2015年=100とした指数を求める四則演算の数式を立てるとき、セル番地の相対番地や絶対番地の考え方を合わせて学ぶ。また、経済データでは物価変動の影響を考慮することもあるので、消費者物価指数のデータを用いた実質化の計算方法も学ぶ。

	A	B	C	D	E	F	G	H
1		実収入(円)		伸び率(%)		2015年=100		消費者物
2		名目	実質	名目	実質	名目	実質	価指数
3	1980年	349,686	469,377			66.5	89.2	74.5
4	1981年	367,111	470,052	5.0	0.1	69.8	89.4	78.1
5	1982年	393,014	489,432	7.1	4.1	74.7	93.1	80.3
6	1983年	405,517	495,742	3.2	1.3	77.1	94.3	81.8
7	1984年	424,025	507,207	4.6	2.3	80.6	96.4	83.6
36	2013年	523,736	542,170	0.9	0.5	99.6	103.1	96.6
37	2014年	520,256	524,452	-0.7	-3.3	98.9	99.7	99.2
38	2015年	525,955	525,955	1.1	0.3	100.0	100.0	100.0
39	2016年	528,103	528,632	0.4	0.5	100.4	100.5	99.9
40	2017年	533,936	531,809	1.1	0.6	101.5	101.1	100.4

図 2 時系列データ分析の基礎を学ぶための例題（家計調査と消費者物価指数）

2.3 記述統計量の実習

平均や標準偏差などの記述統計量を学ぶために使用する例題は、総務省統計局「家計消費状況調査」から抽出した地方別インターネットを利用した1世帯当たり1か月間の支出(2017年)である(図3)。このデータの1~12月の12個の数値に対して、平均、中央値、標準偏差、変動係数を、表計算ソフトの関数を用いて

求める。時系列データに対して平均や標準偏差を求めることは、経済学分野での記述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
1	関東	食料	家電	衣類・履物	旅行関係費	チケット	その他	支出総額
2	1月	3,211	1,043	2,067	3,931	452	6,862	17,566
3	2月	2,855	1,043	1,725	3,175	601	6,396	15,795
4	3月	2,900	1,258	1,547	5,085	676	6,682	18,148
11	10月	2,543	912	2,152	3,431	761	6,529	16,328
12	11月	2,664	986	2,293	3,731	761	7,348	17,783
13	12月	3,359	1,440	2,508	4,108	798	8,813	21,026
14	年間合計	32,946	13,145	22,428	49,537	8,949	81,477	208,482
15	平均	2,745.5	1,095.4	1,869.0	4,128.1	745.8	6,789.8	17,373.5
16	中央値	2,646.5	1,061.5	1,746.5	3,763.0	778.0	6,605.5	17,190.0
17	標準偏差	278.6	190.3	306.3	1,050.6	128.6	685.0	1,548.3
18	変動係数(%)	10.1	17.4	16.4	25.4	17.2	10.1	8.9

図 3 記述統計量を学ぶための例題（インターネットを利用した1世帯当たり1か月間の支出）

2.4 総合的な実習

授業のまとめとして、長期間にわたる月次データを与え、学生自身で分析方法を考えて、実際に分析した結果をレポートにまとめ、試験の代わりとしている。総合的な実習で出題するデータは、学期ごとに出題データを変更しており、2018年度第1クォーターでは、経済産業省「特定サービス産業動態統計調査」から抽出した、2010年1月以降のゴルフ場の売上高、利用者数を出題した(図4)。合計欄を空欄のまま出題しているのは、横方向に並んでいるどの項目が合計の対象になるかを、学生自身に気付いてもらうためである。

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		売上高(百万円)								利用者数(人)			
2		利用料金収入				店売上高				会員			
3		計	平日	土・日・祝日	計	平日	土・日・祝日	計	平日	土・日・祝日	計	平日	土・日・祝日
4	2010年	1月	1355	1732	951	769	68752	118931	156455	117031			
5		2月	1099	1352	753	609	56539	87848	141447	97593			
6		3月	1749	1915	1080	964	77257	97440	212819	141326			
7		4月	2437	2647	1568	1246	91613	137461	303678	203139			
8		5月	2608	4062	2159	1608	103410	206982	328633	347947			
9		6月	3103	2775	1988	1515	130859	138812	403099	227886			
97		10月	2894	2751	1620	1547	107913	117148	406518	236044			
98		11月	2820	2778	1426	1583	104127	122870	371552	218980			
99		12月	2168	2172	1067	1141	87919	107280	270260	162647			
100	2018年	1月	1232	1413	697	686	65331	88046	156864	112112			
101		2月	1069	1201	645	598	55542	76420	150816	96992			
102		3月	1906	2129	1023	1093	80181	108676	254891	166714			

図 4 総合的な実習で出題するデータ（ゴルフ場の売上高、利用者数）

3. 例題データベースの考え方

前章で示した実習を十分に実施するには、例題データベースの存在が望ましい。実習後、理解を深めるために実習内容を繰り返すには、代替となるデータを用意する必要がある。また、授業科目「データ処理入門」は、新入生をクラス分けして異なる学期で実施するため、成績評価の対象として出題する課題で使用するデータは、直近の数期間は重複しないようにしている。

教材作成者は、実習や課題で使用するデータをその都度作成する必要がある、この際、例題データベースの存在は有用である。

3.1 代替となるデータ

実習内容を繰り返したり、課題として出題したりするときには、代替となるデータを用意する必要がある。時系列データ分析の実習で使用した例題（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。記述統計量の実習で使用した例題（図 3）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）を用いたりして、レポート課題を出題している。

3.2 データの「枠組み」

データが入れ替わっても、学生が戸惑いなく実習に取り組めるように、同じ授業の実習で使用するデータの枠組みは変更しない。図 2 のデータは、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を有している。図 3 のデータは、縦方向に見ると集計可能な 1 年分の月次データであり、横方向に並んでいる 6 項目は、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。なお、表計算ソフト導入時に使用する図 1 のデータは、図 3 と同じ枠組みに当てはまる。

総合的な実習で使用する図 4 のデータには、図 2、3 のように明確な枠組みがあるわけではないが、少なくとも縦方向に見ると長期間の月次データであり、かつ集計する意味があるという枠組みはある。

3.3 実データの使用

American Statistical Association が示した統計教育ガイドラインでは、“Use real data.” を強く推奨し

ている¹⁾。同ガイドラインでは、純粋に計算するだけで、意味のある解釈ができない単なる数値の使用を推奨していない。文脈を与えて多少現実感を持たせたデータも推奨していない。また、同じ real data でも、テストの点数のようなデータではなく、実際の研究から得られたデータで、魅力的なリサーチクエスチョンへと導くデータが望ましいとしている。

これまでに、授業科目「データ処理入門」では、データ分析の実習以外で使用するものも含めて、次のような実際の経済データを教材として使用している。

愛知県百貨店商品別販売額

全国百貨店・スーパー販売額

1 世帯当たり年平均 1 か月間の収入と支出

消費者物価指数、東京都区部小売価格

所得金額階級別世帯数、金融資産保有額

為替レート、日経平均株価

車種別新車販売台数

データ処理の実習を行う授業は、経済学部の新入生には、実際の経済データにはじめて接触する機会でもある。高度で専門的な経済データの分析や解釈は、のちに学ぶことであるが、少なくとも、それがどのようなデータであるかを説明できたり、データの大まかな動向を説明できたりすることは、むしろ新入生にこそ必要な能力である。

また、経済学部生なら、経済データの数値自体についても学ぶ必要がある。経済データによく見られるパターンを学ぶために模擬的なデータを作成しても、それは実社会を反映しない架空のデータになってしまう。

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。本研究では、経済学部生を対象とするデータ分析の実習で用いる例題の作成に資することを目的として、実際の経済データのデータベース化を試みることにした。

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 2、3、4 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。また、これら

の枠組みに当てはまらない経済データも扱えるような拡張可能性も考慮する必要がある。

4.1 経済時系列データに特有の属性

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。

4.2 さまざまなフォーマットへの対応

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いる。例えば、「インターネットを利用した1世帯当たり1か月間の支出」は、図5のようなフォーマットで公開されている。

例題データベースの内部ではデータの統一した表現が望ましい。Tidy data⁽²⁾⁽³⁾⁽⁴⁾は内部表現の1候補であり。図5をTidy data化したものが図6である。

4.3 用語と単位

データベースで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(×10⁶)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⁽⁵⁾、この中で単位や単位乗数の定義をしている⁽⁶⁾。

4.4 データの抽出

図6のデータから、図3の枠組みに合う例題を作るには、特定の地域向けの1年分の月次データを抽出し、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、図4のように、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た例題も考えられる。

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設計について議論した。実習に使うのは、実際の経済時系列データとし、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、授業で使用してきたデータからデータベースの実装を試みる。

G	H	I	J	K	L	M
3						
4						第1-1表 全
5			Table 1-1	Monthly Expenditure on Goods a		
6						
7	平成30年(2018年)4月					
8	April 2018					
9		全国				
10			北海道	東北	関東	北陸
11	項 目					
12		All Japan	Hokkaido	Tohoku	Kanto	Hokuriku
14	世帯数分布(抽出率調整)	10,000	481	880	3,671	3
15	集計世帯数	19,971	1,025	1,528	6,549	8
16	世帯人員(人)	2.99	2.78	3.18	2.95	3.
17	有業人員(人)	1.49	1.53	1.59	1.47	1.
18	世帯主の年齢(歳)	69.0	69.4	61.1	69.3	61
19						
20	インターネットを利用した支出総額(22品目計)	11,888	10,478	7,409	14,349	10.7
21	贈答品	304	282	207	406	1
22	自家用					
24	自家用計	11,384	10,182	7,202	13,842	10.6
25	5.2~5.4計(食料)	1,405	1,134	834	2,008	8
26	5.2 食料品	888	708	340	1,438	6
27	5.3 飲料	303	248	212	434	1
28	5.4 出前	113	178	82	136	
29	5.5 家電	855	446	189	868	9
30	5.6 家具	292	145	163	395	3
31	6.7~9.9計(衣類・履物)	1,271	1,003	1,081	1,564	1.2
32	5.7 紳士用衣類	254	184	165	270	3
33	5.8 婦人用衣類	889	484	670	880	6
34	5.9 履物・その他の衣類	328	335	245	414	1
35	6.0~6.6計(娯楽・娯楽)	775	480	400	544	4

図5 公開データのフォーマットの例(インターネットを利用した1世帯当たり1か月間の支出)

	A	B	C	D	E	F
1	年	月	地域	調査項目	単位	調査値
2	2017年	1月	北海道	インターネットを利用した支出総額(22品目計)	円	9225
3	2017年	1月	北海道	5.1 贈答品	円	528
4	2017年	1月	北海道	自家用計	円	8697
5	2017年	1月	北海道	5.2~5.4計(食料)	円	1385
6	2017年	1月	北海道	5.2 食料品	円	913
7	2017年	1月	北海道	5.3 飲料	円	311
8	2017年	1月	北海道	5.4 出前	円	160
9	2017年	1月	北海道	5.5 家電	円	488
10	2017年	1月	北海道	5.6 家具	円	285
26	2017年	1月	北海道	6.9~7.0計(旅行関係費)	円	2389
27	2017年	1月	北海道	6.9 宿泊料、運賃、バック旅行費(インターネット)	円	1581
28	2017年	1月	北海道	7.0 宿泊料、運賃、バック旅行費(上記以外の決済)	円	808
29	2017年	1月	北海道	7.1 チケット	円	55
30	2017年	1月	北海道	7.2 上記に当てはまらない商品・サービス	円	1198

図6 Tidy Data化の例(インターネットを利用した1世帯当たり1か月間の支出)

参考文献

- (1) American Statistical Association, "Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016", p.60 (2016)
- (2) Hadley Wickham: "Tidy data", Journal of Statistical Software, Vol.59, No.10 (2014)
doi:10.18637/jss.v059.i10
- (3) Hadley Wickham and Garrett Grolemund: "Tidy Data with tidy", R for Data Science, pp.147-169, O'Reilly (2016)
- (4) 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)
- (5) 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)
- (6) 総務省統計局, "属性", 統計 LOD, <https://data.e-stat.go.jp/lodw/rdfs/schema/observation/>