

# 教育用擬似マイクロデータを用いた 授業事例

菅 幹雄

法政大学経済学部・日本統計研究所

東京都町田市相原4342

e-mail: [msuga@hosei.ac.jp](mailto:msuga@hosei.ac.jp)

# 1. 教育用擬似マイクロデータを活用した授業の概要

- 平成23年後期に「現代社会と情報B」という授業で教育用擬似マイクロデータを用いたPC演習を行った。
- 「現代社会と情報B」はマイクロ統計データの利活用と個人情報保護についての授業である。
- 平成23年10月頃、教育用擬似マイクロデータが統計センターから提供されることを知り、マイクロ統計データの利活用の具体例として適切であると考えた。
- 一部授業予定を変更して教育用擬似マイクロデータを用いたPC演習を11月後半から12月にかけて実施した。

- 対象は一般の学部学生であり、統計学を履修済みとは限らず、かつEXCELの操作でさえ覚束ない者もいたので、マイクロデータを体験する程度の内容に留まった。
- 平成24年度にはより体系的な授業を組み立てるため、ゼミナールで教育用擬似マイクロデータを用いたPC演習を行うことにした。
- ゼミナールの学生ならば統計学の授業を履修済みであり、かつEXCEL及びSPSSの操作経験があるからである。

## 2. データの申請・受領・準備

- 平成23年10月に申出書を作成し、統計センターにメールで申出した。
- このときは教員の名前で申出書を作成し、利用にあたっての留意事項にサインした。
- これについて「統計教育という視点からは、教員にとっては面倒ではあるが、演習に参加した学生全員が留意事項を了承したというサインを取った方が良い」と意見を述べたところ、新しい申出書では申出者を除く利用者の所属・氏名の記入欄が設けられた。

- 次にデータが数回に分けてメールで送られてきたが、それを接続するのに手間がかかった。結局、EXCELで読み込んでマクロでつないだ。
- 送られて来たデータは圧縮された状態で29MBであったが、これを解凍すると62MBという膨大なものであった。
- 学生に提供するときには、授業支援システムで履修学生が提供できるようにしたが、ここで問題になったのは20MBまでしか一度には提供できないことであった。

- そこで平成23年度は62MBのファイルの後半の項目（列側）を切って17MBに縮めて提供した。
- なおファイルは授業後すぐに消去し、学生にも保存しないように指示した。
- 平成24年度は必要な部分をあらかじめ抜き出して学生に提供することにした。
- なお、統計センターの新しい提供方法では、ホームページからダウンロードできるようになった。
- またファイルが小さい簡易データも提供されているので、学生へのデータ提供は選択肢が増えかつ楽になった。

### 3. 公表資料との比較

- 教育用擬似マイクロデータは平成16年全国消費実態調査の二人以上の世帯・勤労者世帯に基づいていることから、確認のために両者の1世帯当たり収入・支出を比較した(表1)。
- 乖離率は両者の差を平均で割った値とした。
- 収入の乖離率は小さく、消費支出の乖離率は大きい。
- 特に住居、家具・家事用品の乖離率は大きい。
- 消費支出の中でも教育の乖離率は小さい。

表1 平成16年全国消費実態調査との比較（平均）

	A.教育用擬似 マイクロデータ	B.平成16年全国 消費実態調査	C.乖離率 $(A-B)/(A+B)*2$
世帯人員数	3.52	3.52	-0.1%
年間収入	730	740	-1.4%
実収入	497,656	502,114	-0.9%
消費支出	328,140	339,212	-3.3%
食料	72,883	73,742	-1.2%
住居	17,687	19,393	-9.2%
光熱・水道	19,238	19,398	-0.8%
家具・家事用品	9,204	9,783	-6.1%
被服及び履物	14,138	14,648	-3.5%
保健医療	11,366	11,935	-4.9%
交通・通信	47,961	50,754	-5.7%
教育	22,270	22,330	-0.3%
教養娯楽	31,389	32,472	-3.4%
その他の消費支出	82,003	84,756	-3.3%
非消費支出	75,607	76,600	-1.3%
可処分所得	422,049	425,513	-0.8%



- 表2は教育用擬似マイクロデータの収入・支出の平均、標準偏差、変動係数である。集計用乗率で加重して計算しなければならないため学生には良い練習になる。
- なお教育用擬似マイクロデータの収入・支出の平均及び標準偏差は(独)統計センター「教育用擬似マイクロデータの開発とその利用～平成16年全国消費実態調査を例として～」『製表技術参考資料』16に掲載されており、それと表1及び表2の数値は一致している。

表2 教育用擬似ミクロデータの収入・支出の  
平均、標準偏差、変動係数

	平均	標準偏差	変動係数
世帯人員数	3.52	1.18	0.34
年間収入	729.81	338	0.46
実収入	497,655.92	261,558	0.53
消費支出	328,139.70	192,447	0.59
食料	72,883.42	28,064	0.39
住居	17,687.21	60,587	3.43
光熱・水道	19,237.81	7,690	0.40
家具・家事用品	9,204.04	14,933	1.62
被服及び履物	14,137.63	19,823	1.40
保健医療	11,366.36	19,284	1.70
交通・通信	47,960.92	84,654	1.77
教育	22,269.65	64,157	2.88
教養娯楽	31,389.49	32,723	1.04
その他の消費支出	82,003.18	102,041	1.24
非消費支出	75,606.93	66,378	0.88
可処分所得	422,048.99	233,718	0.55

(独)統計センター「教育用擬似マイクロデータの開発とその利用～平成16年全国消費実態調査を例として～」『製表技術参考資料』16に掲載されている教育用擬似マイクロデータの収入・支出の平均及び標準偏差

	平均			差	標準偏差		
	個票データ	教育用擬似マイクロデータ			個票データ	教育用擬似マイクロデータ	
年間収入	740.18	729.81	-0.01	年間収入	358.18	337.69	-0.06
収入総額	971,789.24	946,779.03	-0.03	収入総額	541,290.74	473,480.73	-0.13
実収入	502,133.73	497,655.92	-0.01	実収入	280,695.92	261,558.27	-0.07
実収入以外の収入	391,823.98	372,130.47	-0.05	実収入以外の収入	353,922.37	263,445.65	-0.26
繰入金	77,831.53	76,992.65	-0.01	繰入金	87,036.21	98,947.04	0.14
支出総額	971,789.24	946,779.03	-0.03	支出総額	541,290.74	473,480.73	-0.13
実支出	415,809.39	403,746.63	-0.03	実支出	224,419.69	219,290.60	-0.02
消費支出	339,199.37	328,139.70	-0.03	消費支出	194,501.15	192,447.21	-0.01
食料	73,738.54	72,883.42	-0.01	食料	30,149.02	28,064.49	-0.07
住居	19,387.99	17,687.21	-0.09	住居	52,962.36	60,587.32	0.14
光熱・水道	19,395.36	19,237.81	-0.01	光熱・水道	8,009.23	7,690.12	-0.04
家具・家事用品	9,783.81	9,204.04	-0.06	家具・家事用品	15,977.65	14,933.13	-0.07
被服及び履物	14,649.44	14,137.63	-0.03	被服及び履物	18,837.04	19,823.09	0.05
保健医療	11,936.01	11,366.36	-0.05	保健医療	19,763.39	19,284.07	-0.02
交通・通信	50,740.68	47,960.92	-0.05	交通・通信	85,021.69	84,654.38	0.00
教育	22,332.15	22,269.65	0.00	教育	51,989.72	64,157.45	0.23
教養娯楽	32,472.95	31,389.49	-0.03	教養娯楽	32,161.60	32,723.04	0.02
その他の消費支出	84,762.44	82,003.18	-0.03	その他の消費支出	95,898.83	102,040.97	0.06
非消費支出	76,610.02	75,606.93	-0.01	非消費支出	56,199.75	66,378.49	0.18
実支出以外の支出	475,947.80	464,318.09	-0.02	実支出以外の支出	394,805.29	334,227.09	-0.15
繰越金	80,032.04	78,714.31	-0.02	繰越金	96,421.45	118,055.82	0.22

### 3. 等価尺度による調整

- 所得不平等度の計算においては、等価尺度による調整が必要になる。
- 等価尺度には以下の3種類が代表的なものである。
  - ① OECD等価尺度 “OECD equivalence scale”
  - ② OECD修正尺度 “OECD-modified scale”
  - ③ 平方根尺度 Square root scale

# ①OECD等価尺度

## “OECD equivalence scale”

- 最初の大人は1とする。
- 2人目以降の14歳以上の世帯員は0.7とする。
- 14歳未満の子供は0.5とする。
- 別名「オクスフォード尺度」“Oxford scale”
- OECD (1982)では「その国独自の等価尺度が出来ていない国に用いることが可能」とされた。
- 「旧OECD尺度」とも呼ばれる。

OECD (1982), *The OECD List of Social Indicators*, Paris.

## ②OECD修正尺度 "OECD-modified scale"

- 最初の大人は1とする。
- 2人目以降の14歳以上の世帯員は0.5とする。
- 14歳未満の子供は0.3とする。
- 1980年代から1990年代に初めにかけては旧OECD等価尺度が使用された。
- 1990年代後半にEU統計局(EUROSTAT)が修正等価尺度を用いるようになった。
- Haagenars et al. (1994) が最初に提案。

Haagenars, A., K. de Vos and M.A. Zaidi (1994), *Poverty Statistics in the Late 1980s: Research Based on Micro-data*, Office for Official Publications of the European Communities. Luxembourg.

## ③平方根尺度

### Square root scale

- 近年のOECDの報告書（例えばOECD(2008)）では所得を世帯人員数の平方根で割って比較している。

OECD (2008), *Growing Unequal ? Income Distribution and Poverty in OECD Countries*, Paris.

- 平方根尺度は単純すぎて教育効果がない。
- (教育的な意味で) OECD等価尺度(あるいは OECD修正尺度)を用いたかったが、教育用擬似マイクロデータには世帯人員数の情報はあるが、14歳以上の世帯人員数、14歳未満の世帯人員数の情報がない。
- また世帯員別の情報もない。



- このようなケースは、学生にどうしたら良いか考えさせるよい例である。
- その意味で教育用データはある程度不完全であるほうが良い。
- 教員側で考えた方法は学校給食及び子供用衣類の支出から、子供がいることを推測することである。
- 具体的にはそこで教育用擬似マイクロデータにあわせた等価尺度として次のような方法を考えた。

○「学校給食」、「子供用洋服」、「子供用シャツ・セーター類」、「子供用下着類」のすべてに支出がない場合

- 少なくとも1人は大人であるとし、1とする。
- 2人目以降の世帯員は0.7とする。

○「学校給食」、「子供用洋服」、「子供用シャツ・セーター類」、「子供用下着類」のいずれかの支出があった場合

- 少なくとも1人は大人であるとする。
- 少なくとも1人は子供であると推測、1人を0.5とする。
- 3人目以降の世帯員は0.7とする。

- OECDと本報告の等価尺度を比較したものが表3である。
- ただし実際には子供がいても、9～11月に給食費や子供用衣類の支出がなかったら、子供がいないとされてしまう。
- また子供がいなくても、贈り物用に子供用衣類を購入することも考えられる。

表3 OECDと本報告の等価尺度

	世帯	OECD	OECD	平方根	本報告
	人員数	等価尺度	修正尺度	尺度	等価尺度
大人1人	1	1	1	1	1
大人2人	2	1.7	1.5	1.4	1.7
大人2人子供1人	3	2.2	1.8	1.7	2.2
大人2人子供2人	4	2.7	2.1	2.0	2.9
大人2人子供3人	5	3.2	2.4	2.2	3.6

- 表4は世帯人員1人当たり収入・支出の平均を等価尺度で調整前・後それぞれの場合について示したものである。
- 等価尺度で調整すると1人当たり収入・支出の平均は3割程度大きくなる。
- 収入・支出の中で教育は突出して36%大きくなっており、等価尺度による調整の影響が相対的に大きい。

表4 等価尺度で調整前・後の  
1人当たり収入・支出（平均）

	A.調整前	B.調整後	C.変化率 (B-A)/A
年間収入	226	292	29%
実収入	154,332	199,604	29%
消費支出	101,864	131,755	29%
食料	22,216	28,816	30%
住居	6,042	7,756	28%
光熱・水道	5,804	7,543	30%
家具・家事用品	2,899	3,738	29%
被服及び履物	4,461	5,763	29%
保健医療	3,564	4,610	29%
交通・通信	14,988	19,380	29%
教育	5,483	7,451	36%
教養娯楽	9,891	12,776	29%
その他の消費支出	26,518	33,921	28%
非消費支出	23,674	30,564	29%
可処分所得	130,658	169,040	29%

- 表5は世帯人員1人当たり収入・支出の変動係数を等価尺度で調整前・後それぞれの場合について示したものである。
- 等価尺度で調整すると1人当たり収入・支出の変動係数は2～3割程度小さくなる。

表5 等価尺度で調整前・後の  
1人当たり収入・支出（変動係数）

	A.調整前	B.調整後	C.変化率 (B-A)/A
年間収入	2.30	1.58	-31%
実収入	2.32	1.61	-30%
消費支出	2.32	1.62	-30%
食料	2.32	1.58	-32%
住居	4.03	3.67	-9%
光熱・水道	2.35	1.60	-32%
家具・家事用品	2.75	2.20	-20%
被服及び履物	2.71	2.13	-21%
保健医療	2.90	2.36	-19%
交通・通信	2.86	2.34	-18%
教育	4.29	3.59	-16%
教養娯楽	2.51	1.88	-25%
その他の消費支出	2.50	1.94	-22%
非消費支出	2.45	1.79	-27%
可処分所得	2.34	1.63	-30%



## 4. ジニ係数の計算

- ジニ係数の計算方法は色々とあるが、三角形と大台形の面積の計算方法が分かれば理解できるという意味で数学の不得手な学生でも理解しやすい方法が

$$G_n = 1 - \frac{1}{n} \sum_{i=1}^n (S_i + S_{i-1}) \quad (1)$$

である。ただし  $n$  は世帯数、 $S_i$  は累積所得比率である。

- (1)式はデータをソートしないと計算できないが、ソートしなくても計算できる方法が、

$$G_n = \frac{2 \text{cov}(y, r)}{n\bar{y}} \quad (2)$$

であり、 $y$  は所得、 $r$  は順位、 $\bar{y}$  は平均所得である。

- 教育用擬似マイクロデータの計算では(1)式も(2)式も集計用乗率をウェイトに用いて計算する。計算例を表6に示す。

- 集計用乗率を用いる場合、同じ順位の家帯が集計用乗率の数だけあることになる。
- 表6では家帯番号1の家帯と同じ順位の家帯が100家帯ある。
- 家帯番号1は標本家帯の中で一番所得が低い。
- 家帯番号1の順位を昇順で50位 ( $=100 \div 2$ ) とする。
- 次に家帯番号2の家帯と同じ順位の家帯が300家帯ある。
- 家帯番号2は標本家帯の中で二番目に所得が低い。
- そこで250位 ( $=100 + 300 \div 2$ ) とする。以下同様である。

表6 ジニ係数の計算例

	A	B	C	D	E	F	G
世帯番号	所得	集計用乗率	世帯比率 $S_i$	所得(乗率調整済)	累積所得	累積所得比率	$(S_i + S_{i-1}) \times F$
1	4	100	0.1	400	400	2%	0.002
2	8	300	0.3	2400	2800	16%	0.056
3	16	300	0.3	4800	7600	44%	0.181
4	24	200	0.2	4800	12400	72%	0.233
5	48	100	0.1	4800	17200	100%	0.172
合計	100	1000	1	17200	-	-	0.644
ジニ係数 (1)式		0.356					
	H	J	K	L	M	N	
世帯番号	$A \times C$	所得順位(昇順)	$C \times J$	所得偏差	順位偏差	$C \times L \times M$	
1	0.4	50	5.0	-13	-450.0	594	
2	2.4	250	75.0	-9	-250.0	690	
3	4.8	550	165.0	-1	50.0	-18	
4	4.8	800	160.0	7	300.0	408	
5	4.8	950	95.0	31	450.0	1386	
合計	17.2	-	500.0	-	-	3060	
平均所得			17.2				
所得と順位の共分散			3060				
ジニ係数 (2)式			0.356				

- 表6から(1)式も(2)式も同じ計算結果になることが確認できる。
- (1)式を用いてEXCELでジニ係数を計算した結果が表7である。
- 等価尺度で調整することにより、6%程度ジニ係数が小さくなることが分かる。

表7 等価尺度で調整前・後のジニ係数

	A.調整前	B.調整後	C.変化率 (B-A)/A
実収入	0.287	0.269	-6%
消費支出	0.291	0.274	-6%

(参考) 平方根尺度で調整前・後のジニ係数

	A.調整前	B.調整後	C.変化率 (B-A)/A
実収入	0.287	0.243	-15%
消費支出	0.291	0.248	-15%

## 5. 教育用教材の開発

- 擬似マイクロデータについては、実際のデータを扱える人材の育成という視点が重要。
- その意味では多少、不便さや不親切さがあっても良い。(学生が工夫して計算すれば良い。)
- さらにはわざと異常値が入ったデータを作成して提供してもよいであろう。
- 実は可処分所得についてジニ係数を計算しようとしたところ、可処分所得が負値のケースがあった。(対応を検討中)

- 集計用乗率を考慮しなければならないマイクロデータの場合、可能な分析は限られる。
- 集計用乗率を無視して分析をした場合、ミスリーディングな結論を導く可能性がある。
- 集計用乗率を考慮しないで済むマイクロデータがあると、区間推定や仮説検定の教育に役立つ。(集計用乗率が入ると教えるのはまず困難。)