

計量分析 2：宿題 7

村澤 康友

提出期限：2023 年 1 月 26 日

注意：すべての質問に解答しなければ提出とは認めない。授業の HP の解答例を正確に再現すること（乱数は除く）。グループで取り組んでよいが、個別に提出すること。解答例をコピーしたり、他人の名前で提出した場合は、提出点を 0 点とし、再提出も認めない。すべての結果をワードに貼り付けて印刷し（A4 縦・両面印刷可・手書き不可）、2 枚以上になる場合は必ず左上隅をホッチキスで留めること。

1. (教科書 p. 251, 実証分析問題 10-A) 母親の就業が既婚女性の就業確率に与える平均処置効果 (ATE) を推定したい。データセット「10_2_work.dta」を gretl に読み込み、以下の分析を行いなさい。
 - (a) 両親の学歴・15 歳時の暮らし向き・学業成績・家庭の蔵書数で 15 歳時の母親の就業を説明する線形確率モデルを推定しなさい。また回帰予測値（傾向スコア）を保存しなさい。※推定結果の画面のメニューの「保存」→「理論値」で保存。
 - (b) 傾向スコアの範囲を $(0, 0.65)$, $(0.65, 0.7)$, $(0.7, 0.74)$, $(0.74, 0.78)$, $(0.78, 0.82)$, $(0.82, 1)$ の 6 つの区間に分け、区間ごとに 15 歳時の母親の就業の有無で既婚女性の就業割合を比較しなさい。※メニューの「標本」→「基準に基づいて制限する」で傾向スコアに基づいて標本を制限し、メニューの「表示」→「クロス集計」でクロス表を作成する。
 - (c) 15 歳時の母親の就業と傾向スコアで本人の就業を説明する重回帰モデルを OLS 推定し、母親の就業が既婚女性の就業確率に与える ATE を求めなさい。
2. (教科書 p. 261, 実証分析問題) 少人数学級が算数の学力に与える ATE を推定したい。下記の URL の Angrist and Lavy (1999) の項目からデータセット「final5.dta」を入手して gretl に読み込み、以下の分析を行いなさい。

<https://economics.mit.edu/people/faculty/josh-angrist/angrist-data-archive>

- (a) 分析の前に以下の処理を行う。
 - i. 平均点が 100 点台の観測値は、百の位の誤記と考えられる。メニューの「追加」→「新規変数の定義」で次式を入力し、誤記を修正した従属変数を新たに作成する。
$$y = \text{avgmath} * (\text{avgmath} \leq 100) + (\text{avgmath} - 100) * (\text{avgmath} > 100)$$
 - ii. 算数のテストの受験者が 0 のクラスは分析の対象外。メニューの「標本」→「基準に基づいて制限する」で次式を入力し、受験者が 0 のクラスを除く。
$$\text{mathsize} > 0$$
また Angrist and Lavy (1999) に沿ってクラスサイズが 1 以下と 45 以上、学年生徒数が 5 以下のクラスを除く。
$$\text{classsize} > 1$$

```
classsize < 45
```

```
c_size > 5
```

処理後のデータを用いてクラスサイズ (classsize)・貧困世帯比率 (tipuach)・学年生徒数 (c_size)の一部または全てで算数のクラス平均点を説明する回帰モデルを OLS 推定し、教科書 p.257, 表 11.1 の (1)–(3) の結果を再現しなさい。 ※「頑健標準誤差を使用する」をチェックし、クラスター変数に学校コード (schlcode) を用いると、同一校のクラス間の相関を考慮した標準誤差が得られる。また以下の手順で推定結果の比較表を作成できる。

- i. 推定結果の画面から「ファイル」→「セッションにアイコンとして保存」を選択。
 - ii. 推定結果がアイコン（「モデル 1」「モデル 2」など）に保存される。
 - iii. 保存したアイコンを「モデル比較表」のアイコンにドラッグ。
 - iv. 「モデル比較表」のアイコンをクリック。
- (b) メニューの「追加」→「新規変数の定義」で次式を入力し、操作変数を作成しなさい。

```
z = c_size / (int((c_size - 1) / 40) + 1)
```

その上でクラスサイズ・貧困世帯比率・学年生徒数の一部または全てで算数のクラス平均点を説明する線形モデルを IV 法 (2SLS) で推定し、教科書 p.257, 表 11.1 の (4),(5) の結果を再現しなさい。

- (c) このデータでは学年生徒数 = 40, 80, 120 が分断点となる。分断点周辺の標本に制限するために、まず以下のダミー変数を作成しなさい。

```
d36_45 = (c_size >= 36) * (c_size <= 45)
```

```
d76_85 = (c_size >= 76) * (c_size <= 85)
```

```
d116_125 = (c_size >= 116) * (c_size <= 125)
```

次に 3 つのダミー変数を用いて分断点周辺の標本に制限しなさい。

```
d36_45 + d76_85 + d116_125 = 1
```

その上でクラスサイズ・貧困世帯比率・学年生徒数の一部または全てで算数のクラス平均点を説明する線形モデルを IV 法 (2SLS) で推定し、教科書 p.257, 表 11.1 の (6),(7) の結果を再現しなさい。

参考文献

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114, 533–575.

解答例

1. (a) 傾向スコアの推定

モデル 1: 最小二乗法 (OLS), 観測: 1–1132

従属変数: mowork15

	係数	Std. Error	<i>t</i> -ratio	p 値
const	0.905716	0.0468994	19.31	0.0000
mocograd	0.0737983	0.0701761	1.052	0.2932
pacograd	−0.105651	0.0357027	−2.959	0.0031
life15	−0.0284479	0.0170811	−1.665	0.0961
academic15	−0.00221650	0.0120106	−0.1845	0.8536
books15	−0.0214158	0.00594938	−3.600	0.0003
Mean dependent var	0.742933	S.D. dependent var		0.437210
Sum squared resid	210.2670	S.E. of regression		0.432132
R^2	0.027413	Adjusted R^2		0.023094
$F(5, 1126)$	6.347301	P-value(F)		8.03e−06
Log-likelihood	−653.4550	Akaike criterion		1318.910
Schwarz criterion	1349.101	Hannan–Quinn		1330.315

(b) 15 歳時の母親の就業の有無による既婚女性の就業割合の比較

i. 傾向スコア：(0,0.65)

	[0]	[1]	計
[0]	49.1%	50.9%	55
[1]	39.5%	60.5%	76
TOTAL	43.5%	56.5%	131

ピアソン (Pearson) のカイ二乗検定 = 1.20072 (1 df, p-value = 0.273178)

ii. 傾向スコア：(0.65,0.7)

	[0]	[1]	計
[0]	61.0%	39.0%	59
[1]	46.5%	53.5%	142
TOTAL	50.7%	49.3%	201

ピアソン (Pearson) のカイ二乗検定 = 3.52464 (1 df, p-value = 0.0604629)

iii. 傾向スコア：(0.7,0.74)

	[0]	[1]	計
[0]	45.5%	54.5%	44
[1]	43.5%	56.5%	85
TOTAL	44.2%	55.8%	129

ピアソン (Pearson) のカイ二乗検定 = 0.0435688 (1 df, p-value = 0.834658)

iv. 傾向スコア：(0.74,0.78)

	[0]	[1]	計
[0]	59.3%	40.7%	54
[1]	43.8%	56.2%	178
TOTAL	47.4%	52.6%	232

ピアソン (Pearson) のカイ二乗検定 = 3.96086 (1 df, p-value = 0.0465699)

v. 傾向スコア：(0.78,0.82)

	[0]	[1]	計
[0]	37.2%	62.8%	43
[1]	43.7%	56.3%	222
TOTAL	42.6%	57.4%	265

ピアソン (Pearson) のカイ二乗検定 = 0.619275 (1 df, p-value = 0.431317)

vi. 傾向スコア：(0,82,1)

	[0]	[1]	計
[0]	47.2%	52.8%	36
[1]	37.0%	63.0%	138
TOTAL	39.1%	60.9%	174

ピアソン (Pearson) のカイ二乗検定 = 1.26384 (1 df, p-value = 0.260925)

(c) 傾向スコアで共変量調整した重回帰モデルによる ATE の推定

モデル 1: 最小二乗法 (OLS), 観測: 1–1132

従属変数: work

	係数	Std. Error	<i>t</i> -ratio	p 値
const	0.331259	0.152207	2.176	0.0297
mowork15	0.0756426	0.0342330	2.210	0.0273
yhat1	0.221641	0.206762	1.072	0.2840
Mean dependent var	0.552120	S.D. dependent var		0.497496
Sum squared resid	278.1980	S.E. of regression		0.496398
R^2	0.006169	Adjusted R^2		0.004408
$F(2, 1129)$	3.504036	P-value(F)		0.030403
Log-likelihood	−811.9095	Akaike criterion		1629.819
Schwarz criterion	1644.914	Hannan–Quinn		1635.522

2. (a) OLS

最小二乘法 (OLS) 推定値
従属変数: y

	(1)	(2)	(3)
const	57.66*** (1.247)	69.81*** (1.174)	70.09*** (1.169)
classsize	0.3217*** (0.04015)	0.07583** (0.03576)	0.01854 (0.04214)
tipuach		-0.3395*** (0.01822)	-0.3317*** (0.01867)
c_size			0.01712** (0.007532)
n	2018	2018	2018
\bar{R}^2	0.0476	0.2473	0.2498
ℓ	-7377	-7139	-7135

丸括弧内は標準誤差

* significant at the 10 percent level

** significant at the 5 percent level

*** significant at the 1 percent level

(b) IV

二段階最小二乗法 (2SLS) 推定値
従属変数: y

	(1)	(2)
const	72.69*** (1.845)	75.96*** (2.355)
classsize	-0.01305 (0.05773)	-0.2311** (0.09860)
tipuach	-0.3546*** (0.01981)	-0.3496*** (0.01997)
c_size		0.04101*** (0.01168)
n	2018	2018
\bar{R}^2	0.2441	0.2334
ℓ	-2.454e+004	-2.445e+004

丸括弧内は標準誤差

* significant at the 10 percent level

** significant at the 5 percent level

*** significant at the 1 percent level

(c) IV (分断点周辺のデータのみ)

二段階最小二乗法 (2SLS) 推定値

従属変数: y

	(1)	(2)
const	78.98*** (5.218)	80.54*** (5.818)
classsize	-0.1855 (0.1553)	-0.4435* (0.2509)
tipuach	-0.4589*** (0.05198)	-0.4347*** (0.05019)
c_size		0.07940** (0.03723)
n	471	471
\bar{R}^2	0.2601	0.2208
ℓ	-5177	-5145

丸括弧内は標準誤差

* significant at the 10 percent level

** significant at the 5 percent level

*** significant at the 1 percent level