INTRODUCTORY ECONOMETRICS

ASIA-PACIFIC EDITION

JEFFREY M. WOOLDRIDGE
MOKHTARUL WADUD
JENNY LYE
ROSELYNE JOYEUX

INSTRUCTOR MANUAL

Table of contents

Chapter 1: The nature of econometrics and economic data

TEACHING NOTES

You have substantial latitude about what to emphasise in Chapter 1. We find it useful to talk about the economics of crime example (Example 1.1) and the wage example (Example 1.2) so that students see, at the outset, that econometrics is linked to economic reasoning, even if the economics is not complicated theory.

We like to familiarise students with the important data structures that empirical economists use, focusing primarily on cross-sectional and time series data sets, as these are what we cover in a first-semester course. It is probably a good idea to mention the growing importance of data sets that have both a cross-sectional and a time dimension.

We spend almost an entire lecture talking about the problems inherent in drawing causal inferences in the social sciences. We do this mostly through the agricultural yield, return to education and crime examples. These examples also contrast experimental and non-experimental (observational) data. Students studying business and finance tend to find the term structure of interest rates example more relevant, although the issue there is testing the implication of a simple theory, as opposed to inferring causality. We have found that spending time talking about these examples, in place of a formal review of probability and statistics, is more successful in teaching the students how econometrics can be used. (And, it is more enjoyable for the students and for us.)

We do not use counterfactual notation as in the modern 'treatment effects' literature, but we do discuss causality using counterfactual reasoning. The return to education, perhaps focusing on the return to getting a college degree, is a good example of how counterfactual reasoning is easily incorporated into the discussion of causality.

Solutions to review questions

1 (i) Ideally, we could randomly assign students to classes of different sizes. That is, each student is assigned a different class size without regard to any student characteristics such as ability and family background. We would like substantial variation in class sizes (subject, of course, to ethical considerations and resource constraints).

(ii) A negative correlation means that larger class size is associated with lower performance. We might find a negative correlation because larger class size actually hurts performance. However, with observational data, there are other reasons we might find a negative relationship. For example, children from more affluent families in Australia might be more likely to attend schools with smaller class sizes, and affluent children generally score better on standardised tests. Another possibility is that, within a school, a principal might assign the better students to smaller classes. Or, some parents might insist their children are in the smaller classes, and these same parents tend to be more involved in their children's education.

(iii) Given the potential for confounding factors – some of which are listed in (ii) – finding a negative correlation would not be strong evidence that smaller class sizes actually lead to better performance. Some way of controlling for the confounding factors is needed, and this is the subject of multiple regression analysis.

2 (i) Here is one way to pose the question: If two firms, say *A* and *B*, are identical in all respects except that firm *A* supplies job training one hour per worker more than firm *B*, by how much would firm *A*'s output differ from firm *B*'s?

(ii) Manufacturing firms in Victoria are likely to choose job training depending on the characteristics of workers. Some observed characteristics are years of schooling, years in the workforce and experience in a particular job. Firms might even discriminate based on age, gender or race. Perhaps firms choose to offer training to more or less able workers, where 'ability' might be difficult to quantify but where a manager has some idea about the relative abilities of different employees. Moreover, different kinds of workers might be attracted to firms that offer more job training on average, and this might not be evident to employers.

(iii) The amount of capital and technology available to workers would also affect output. So, two firms with exactly the same kinds of employees would generally have different outputs if they use different amounts of capital or technology. The quality of managers would also have an effect.

(iv) No, unless the amount of training is randomly assigned. The many factors listed in parts (ii) and (iii) can contribute to finding a positive correlation between *output* and *training* even if job training does not improve worker productivity.

3 It does not make sense to pose the question in terms of causality. Economists would assume that students choose a mix of studying and working (and other activities, such as attending class, leisure and sleeping) based on rational behaviour, such as maximising utility subject to the constraint that there are only 168 hours in a week. We can then use statistical methods to measure the association between studying and working, including regression analysis. But we

would not be claiming that one variable 'causes' the other. They are both choice variables of the student.

4 (i) The notion of ceteris paribus indicates circumstances when other factors are equal or remain the same. In economic and econometric analysis, this notion is implicit in many theoretical model-building, estimations and explanations involving relationships of economic variables; and as we identify changes in one variable of interest due to changes in another variable, keeping everything else constant or unchanged. For example, in a typical supply model we may be interested to analyse the effect of price changes on quantity supplied of a product, holding other factors such costs of production, prices of related goods and technology unchanged. As per Example 1.2 of the chapter, we might be interested in the effect of another year of experience on wages, with training and education remaining unchanged. If we allow all the three variables – training, education and experience – to change simultaneously, then the net effect on wages due to changes in experience could not be ascertained and such analysis would have few implications for policy.

(ii)

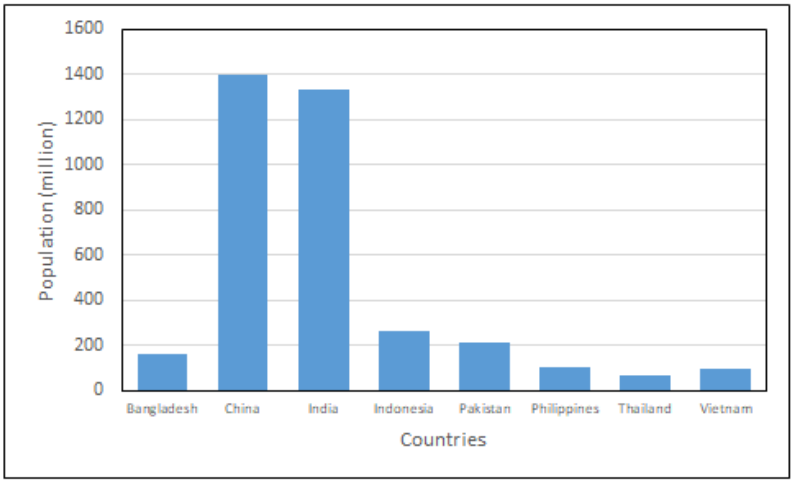


Figure: Total population of selected Asia-Pacific economies, 2018

Multiple-choice questions

1	c	**5**	b
2	d	**6**	d
3	b	**7**	c
4	b	**8**	a

Computer questions

C1 (i) The average of *educ* is about 12.6 years. There are two people reporting zero years of education, and 19 people reporting 18 years of education.

(ii) The average of *wage* is about $5.90, which seems low in the year 2008.

(iii) Using Table B-60 in the 2004 *Economic Report of the President*, the CPI was 56.9 in 1976 and 184.0 in 2003.

(iv) The sample contains 252 women (the number of observations with *female* = 1) and 274 men.

C2 (i) There are 1388 observations in the sample. Tabulating the variable *cigs* shows that 212 women have *cigs* > 0.

(ii) The average of *cigs* is about 2.09, but this includes the 1176 women who did not smoke. Reporting just the average masks the fact that almost 85% of the women did not smoke. It makes more sense to say that the 'typical' woman does not smoke during pregnancy; indeed, the median number of cigarettes smoked is zero.

(iii) The average of *cigs* over the women with *cigs* > 0 is about 13.7. Of course this is much higher than the average over the entire sample because we are excluding 1176 non-smoker women.

(iv) The average of *fatheduc* is about 13.2. There are 196 observations with a missing value for *fatheduc*, and those observations are necessarily excluded in computing the average.

C3 (i) 185/445 ≈ .416 is the fraction of men receiving job training, or about 41.6%.

(ii) For men receiving job training, the average of *re78* is about 6.35, or $6350. For men not receiving job training, the average of *re78* is about 4.55, or $4550. The difference is $1800, which is very large. On average, the men receiving the job training had earnings about 40% higher than those not receiving training.

(iii) About 24.3% of the men who received training were unemployed in 1978; the figure is 35.4% for men not receiving training. This, too, is a big difference.

(iv) The differences in earnings and unemployment rates suggest the training program had strong, positive effects. Our conclusions about economic significance would be stronger if we could also establish statistical significance.

C4 (i) The smallest and largest values of *children* are 0 and 13, respectively. The average is about 2.27.

(ii) Out of 4358 women, only 611 have electricity in the home, or about 14.02%.

(iii) The average of *children* for women without electricity is about 2.33, and for those with electricity it is about 1.90. So, on average, women with electricity have .43 fewer children than those who do not.

(iv) We cannot infer causality here. There are many confounding factors that may be related to the number of children and the presence of electricity in the home; household income and level of education are two possibilities. For example, it could be that women with more education have fewer children and are more likely to have electricity in the home (the latter due to an income effect).

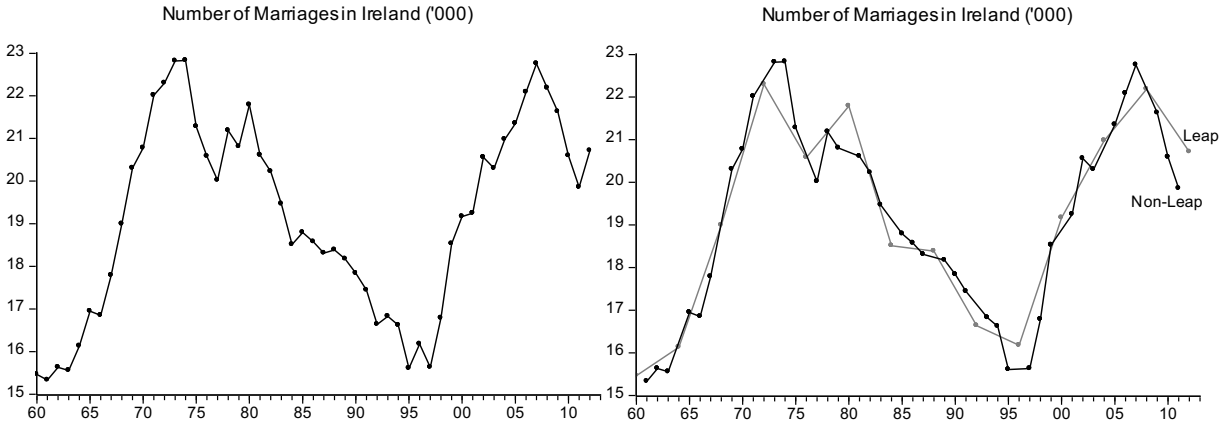
C5 (i)

There appears to be a downward trend in the data.

(ii)

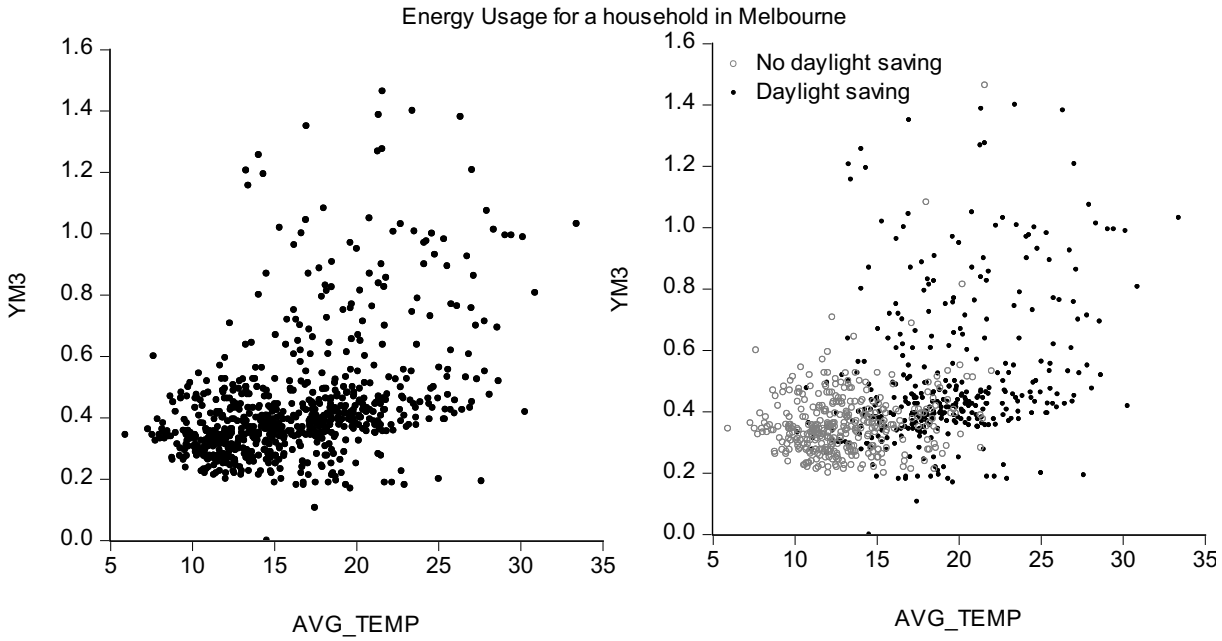
The number of marriages is lower in the leap years although more recently the differential appears to be getting smaller.

(iii)



The number of marriages has fluctuated over time with some periods of downturns and other periods where there has been an increase. There is no obvious difference when we plot the data separately for leap and non-leap years.

C6 (i)



In the first plot we observe that energy usage increases with increases in average temperature. The relationship doesn't appear to be linear.

(ii) In the second plot we divide the data by whether the observation is in the daylight savings period or not. When the observations are not in the daylight savings period these mainly correspond to winter and in the graph we can see these observations are generally clustered in the area of lower average temperatures and lower energy usage. However, the nonlinearity in the data can still be observed.

Chapter 2: Basic mathematical tools

Solutions to review questions

1 The table is extended with required calculations as follows:

Observation	X_i	Y_i	X_i^2	$X_i Y_i$	$(X_i - \bar{X})^2$
1	2	1	4	2	$(2 - 2.6)^2 = .36$
2	0	3	0	0	$(0 - 2.6)^2 = 6.76$
3	–1	–2	1	2	$(-1 - 2.6)^2 = 12.96$
4	5	4	25	20	$(5 - 2.6)^2 = 5.76$
5	7	3	49	21	$(7 - 2.6)^2 = 19.36$
	$\sum_{i=1}^{5} X_i = 13$	$\sum_{i=1}^{5} Y_i = 9$	$\sum_{i=1}^{5} X_i^2 = 79$	$\sum_{i=1}^{5} X_i Y_i = 45$	$\sum_{i=1}^{5} (X_i - \bar{X})^2 = 45.2$

Note that $\bar{X} = \sum_{i=1}^{5} X_i / 5 = 2.6$ and $\bar{Y} = \sum_{i=1}^{5} Y_i / 5 = 1.8$

(i) As the table shows, $\sum_{i=1}^{5} X_i = 13$ and $\sum_{i=1}^{5} Y_i = 9$

(ii) From the table, $\sum_{i=1}^{5} X_i^2 = 79$ and $\left(\sum_{i=1}^{5} X_i \right)^2 = (13)^2 = 169$

(iii) $\sum_{i=1}^{5} X_i Y_i = 45$ and $\left(\sum_{i=1}^{5} X_i \right)\left(\sum_{i=1}^{5} Y_i \right) = 13*9 = 117$

(iv) $\sum_{i=1}^{5} (X_i + Y_i) = 13 + 9 = 22$ and $\left(\sum_{i=1}^{5} X_i + \sum_{i=1}^{5} Y_i \right) = 13 + 9 = 22$. Actually, these are same quantities.

(v) $\sum_{i=1}^{5} (X_i - Y_i) = \sum_{i=1}^{5} X_i - \sum_{i=1}^{5} Y_i = 13 - 9 = 4$

(vi) $\sum_{i=1}^{5} 2X_i = 2\sum_{i=1}^{5} X_i = 2*13 = 26$

(vii) $\sum_{i=1}^{5} 2 = 2*5=10$

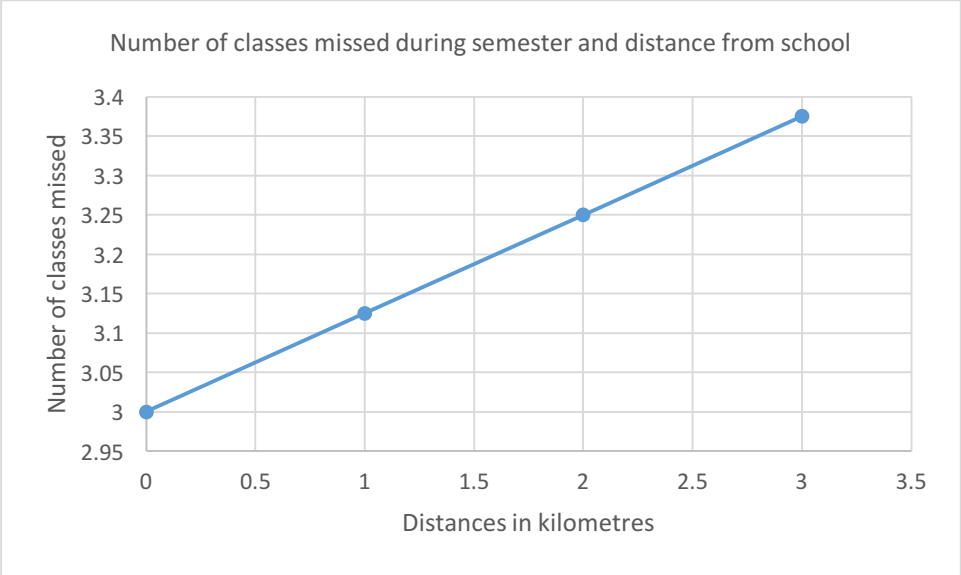
(viii) $\sum_{i=1}^{5}\left(X_i - \bar{X}\right) = 0$

2 (i) The average monthly housing expenditure is \$566.

(ii) The average monthly expenditure would be \$5.66, respectively, measured in hundreds of dollars.

(iii) The average monthly housing expenditure increases to \$586.

3 (i) This is just a standard linear equation with intercept equal to 3 and slope equal to .125. The intercept is the number of missed classes for a student who lives on campus.



(ii) The average number of classes missed by students who live 8 kilometres away is: *missed* = 3 + .125(8) = 4.0, or approximately 4 classes.

(iii) The difference between the average number of classes missed by student living 16 kilometres and 32 kilometres away is = [3 + .125(32)] – [3 + .125(16))] = 7 – 5 = 2 classes.

4 If *price* = 15 and *income* = 200, *quantity* = 20 – 1.8(15) + .03(200) = –1, which is nonsense. This shows that linear demand functions generally cannot describe demand over a wide range of prices and income.

5 (i) The percentage point change is 6 – 4 = 2, or a two percentage point increase in the unemployment rate.

(ii) The percentage change in the unemployment rate is 100[(6 – 4)/4] = 50%; i.e. unemployment increased by 50%.

6 The majority shareholder is referring to the percentage point increase in the stock return, while the CEO is referring to the change relative to the initial return of 15%. To be precise, the shareholder should specifically refer to a 3 percentage *point* increase.

7 (i) The person b's salary exceeds that of person B by 100[42 000 – 35 000)/35 000] = 20%.

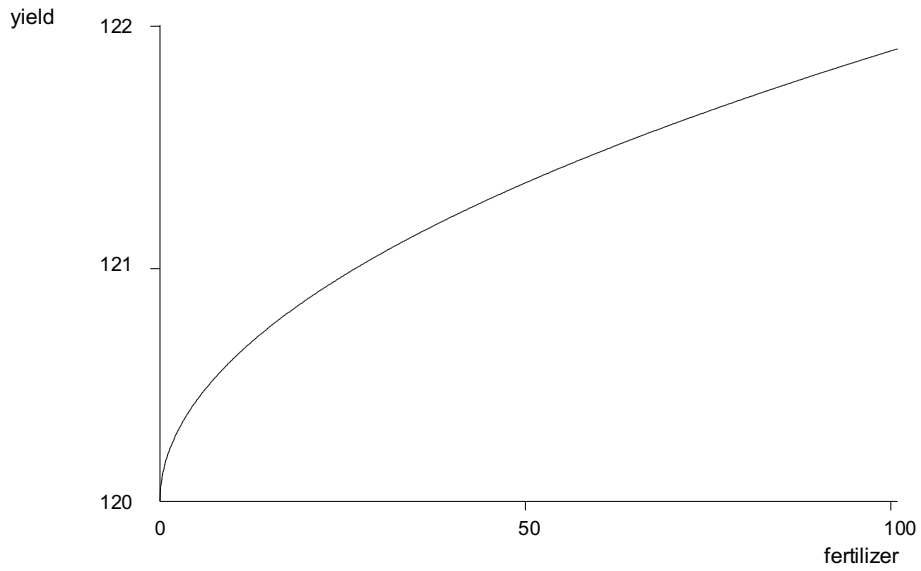
(ii) The approximate proportionate change is log(42 000) – log(35 000) ≈ .182, so the approximate percentage change is 18.2%. [Note: log(·) denotes the natural log.]

8 (i) When *exper* = 0, log(*salary*) = 10.6; therefore, *salary* = exp(10.6) ≈ \$40 134.84. When *exper* = 5, *salary* = exp[10.6 + .027(5)] ≈ \$45 935.80.

(ii) The approximate proportionate increase is .027(5) = .135, so the approximate percentage change is 13.5%.

(iii) 100[(45 935.80 – 40 134.84)/40 134.84) ≈ 14.5%, so the exact percentage increase is about one percentage point higher.

9 (i) The relationship between *yield* and *fertiliser* is graphed.



(ii) Compared with a linear function, the function

$$yield = 120 + .13\sqrt{fertilizer}$$

has a diminishing marginal effect, and the slope approaches zero as *fertiliser* gets large. The initial kilogram of fertiliser has the largest effect, and each additional kilogram has a marginal effect smaller than the previous kilogram.

10 (i) The value 20.5 is the intercept in the equation, so it literally means that if *age* = 0 then the BMI is 20.5. Of course, *age* = 0 measured in years would indicate the BMI of 20.5 as shown by the intercept is the BMI of newborn babies – or more precisely, babies less than a year

old. The intercept by itself is not much of interest since the body fat of babies less than a year old is not usually a concern. Also, the intercept should ideally reflect a dataset on age and BMI of the adult population; so, by itself, 20.5 is not of much interest.

(ii) We use calculus to obtain the maximum BMI:

$$\frac{dBMI}{dAge} = .2 - .004\,Age \quad \text{and} \quad \frac{d^2BMI}{dAge^2} = -.004 < 0.$$

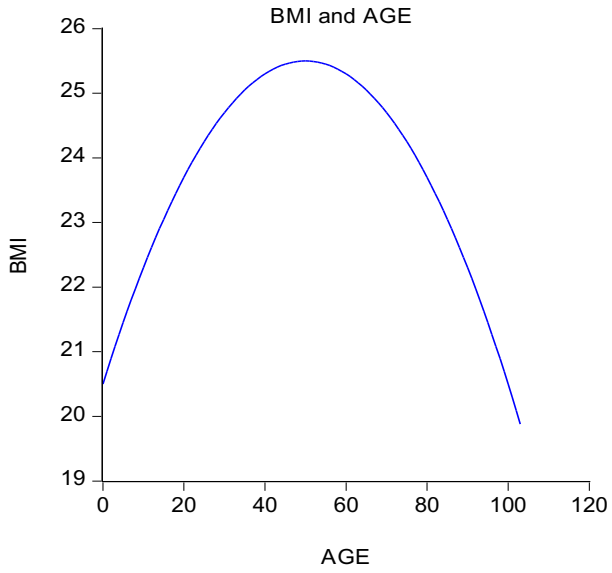
Hence, the BMI function has a maximum. Letting the first derivative equal to 0,

$$\frac{dBMI}{dAge} = .2 - .004\,Age = 0$$

$$Age = \frac{.2}{.004} = 50$$

Therefore, BMI is maximum at the age of 50 years.

(iii) The following graph shows the solution rounded to the nearest integer:



(iv) It is not at all realistic to think that BMI and age will have a deterministic relationship. BMI is also scientifically measured in accordance with the height of a person. Besides, there are many other factors that affect BMI of a person, such as general lifestyle, eating habits, health awareness and income. Multiple regression analysis allows for many observed factors to affect a variable such as BMI, and also recognises that there are unobserved factors that are important and that we can never directly account for.

Multiple-choice questions

1	d	**5**	c	**9**	a
2	b	**6**	b	**10**	a
3	d	**7**	c		
4	b	**8**	c		

Chapter 3: Fundamentals of statistics: a review

Solutions to review questions

1 (i) P(0< Z < 1) = .8413 – .5 = .3413

(ii) P(–1< Z < 1) = .9901 – .9265 = .0636

(iii) P(Z > 2.55) = 1 – .9946 = .0054

(iv) P(Z> –1.92) = 1 – .9726 = .0274

(v) P(Z < –.43) = .3336

2 (i) $P(X \leq 6) = P[(X-5)/2 \leq (6-5)/2] = P(Z \leq .5) \approx .692$, where Z denotes a Normal (0,1) random variable. [we obtain $P(Z \leq .5)$ from Table G.1.]

(ii) $P(X > 4) = P[(X-5)/2 > (4-5)/2] = P(Z > -.5) = P(Z \leq .5) \approx .692$.

(iii) $P(|X-5| > 1) = P(X-5 > 1) + P(X-5 < -1) = P(X > 6) + P(X < 4) \approx (1-.692) + (1-.692) = .616$, where we have used answers from parts (i) and (ii).

(iv) $P(2.5<X<2.8) = P[(X-5)/2] < Z < P(X-5)/2] = P[(2.5-5)/2] < Z < P[(2.8-5)/2] = P(-1.25<Z<-1.1) = .4562 - .1056 = .3506$

(v) $P(4<X<5.74) = P[(X-5)/2] < Z < P(X-5)/2] = P[(4-5)/2] > Z > P[(5.74-5)/2] = P(-.5<Z<.37) = .6443 - .3085 = .3358$

3 Let X denote family income. Then, given the information, we find the required probabilities as shown below:

(i) $$P(X < 30000) = P(Z < \frac{30000-50000}{10000}) = P(Z < -2) = .0228$$

(ii) $$P(X > 70000) = P(Z > \frac{70000-50000}{10000}) = P(Z > 2) = 1 - .9772 = .0228$$

4 Let X represent the marks obtained by the students and X^A and X^{A+} denote the lowest mark that will be awarded an A and A+ grades, respectively. Given that $X \sim N(70,6)$ we first find out the values of standard normal variable Z, such that the probability of Z exceeding this value is 10% or .10 and 5% or .05. That is, we need to find the value of Z that leaves out 10% of the area and 5% of the area under the right tail of the Z distribution.

From the appendix on areas under the standard normal distribution, we find that the relevant value of Z is 1.28 (approximately). Hence we get $\frac{X-70}{6} = 1.28$

$\Rightarrow X^A = (1.28) \cdot 6 + 70 = 77.68$.

Hence, the lowest mark that will be awarded an A grade is 77.68 or 78 (approximately). Similarly, from the standard normal distribution, we find that the relevant value of Z allowing 5% of the area under the normal curve is 1.65 (approximately). Hence we get

$$\frac{X-70}{6}=1.65$$

$$\Rightarrow X^{A+}=1.65\times 6+70=79.9$$

The lowest mark that will be awarded an A+ grade is 79.9 or 80 (approximately).

5 Let Y_{it} be the binary variable equal to one if fund I outperforms the market in year t. By assumption, $P(Y_{it} = 1) = .5$ (a 50-50 chance of outperforming the market for each fund in each year). Now, for any fund, we are also assuming that performance relative to the market is independent across years; but then the probability that fund I outperforms the market in all 10 years – $P(Y_{i,1} = 1, Y_{i,2} = 1, \ldots, Y_{i,10} = 1)$ – is just the product of the probabilities: $P(Y_{i,1} = 1) \cdot P(Y_{i,2} = 1) \ldots P(Y_{i,10} = 1) = (.5)^{10} = 1/1024$ (which is slightly less than .001). In fact, if we define a binary random variable Y_I such that $Y_I = 1$ if and only if fund I outperformed the market in all 10 years, then $P(Y_I = 1) = 1/1024$.

6 In eight attempts, the expected number of free throws is 8(.74) = 5.92, or about six free throws.

7 Tossing three coins gives the following sample space or the possible combinations of events:
HHH, HHT, HTH, HTT, THH, THT,TTH, TTT
Since P(H) = .5 and P(T) = .5, the probability of each event – say of the event HTH – is
P(HTH) = .5*.5*.5 = .125.

Given that X represents the number of tails, we can construct the probability distribution of X that takes the values of 0 (no tail), 1 (one tail), 2(two tails) and 3 (three tails).

X	0	1	2	3
Prob.	.125	.375	.375	.125

$E(X) = 0*.125 + 1*.375 + 2*.375 + 3*.125 = 1.5$

$E(X^2) = 0^2*.125 + 1^2*.375 + 2^2*.375 + 3^2*.125 = 3$

Profit = $(X^2 + X) - 5$

$E(\text{Profit}) = E(X^2 + X) - 5 = E(X^2) + E(X) - 5 = 3 + 1.5 - 5 = -.5$, hence a loss of 50 cents.

8 If Y is salary in dollars then $Y = 1000 \cdot X$, and so the expected value of Y is 1000 times the expected value of X, and the standard deviation of Y is 1000 times the standard deviation of X. Therefore, the expected value and standard deviation of salary, measured in dollars, are $57 000 and $14 600, respectively.

9 (i) P(male wins) = 40/60 = .667 apx

(ii) P(married/ male) = $\frac{P(married\ \&\ male)}{P(male)}=\frac{10/60}{40/60}=\frac{10}{60}*\frac{60}{40}=\frac{10}{40}=.25$

10 E(*GRADE*|*ATAR=65*) = 10.5 + .85 (65) = 65.75. Similarly, E(*GRADE*|*ATAR=95*) = 10.5 + .85 (95) = 91.25. The difference in expected grade obtained in the subject is substantial, but the difference in ATAR scores is also rather large.

11 (i) This is just a special case of what we covered in the text, with $n = 4$: E($\bar{Y}$) = μ and Var($\bar{Y}$) = $\sigma^2/4$.

(ii) E(W) = E(Y_1)/8 + E(Y_2)/8 + E(Y_3)/4 + E(Y_4)/2 = μ[(1/8) + (1/8) + (1/4) + (1/2)] = μ(1 + 1 + 2 + 4)/8 = μ, which shows that W is unbiased. Because the Y_i are independent,

Var(W) = Var(Y_1)/64 + Var(Y_2)/64 + Var(Y_3)/16 + Var(Y_4)/4
= σ^2[(1/64) + (1/64) + (4/64) + (16/64)] = σ^2(22/64) = σ^2(11/32).

(iii) Because 11/32 > 8/32 = 1/4, Var(W) > Var($\bar{Y}$) for any $\sigma^2 > 0$, so $\bar{Y}$ is preferred to W because each is unbiased.

12 (i) E(W_a) = a_1E(Y_1) + a_2E(Y_2) + … + a_nE(Y_n) = ($a_1 + a_2 + \ldots + a_n$)μ. Therefore, we must have $a_1 + a_2 + \ldots + a_n = 1$ for unbiasedness.

(ii) Var(W_a) = a_1^2 Var(Y_1) + a_2^2 Var(Y_2) + … + a_n^2 Var(Y_n) = ($a_1^2 + a_2^2 + \ldots + a_n^2$)$\sigma^2$.

(iii) From the hint, when $a_1 + a_2 + \ldots + a_n = 1$ – the condition needed for unbiasedness of W_a – we have $1/n \leq a_1^2 + a_2^2 + \ldots + a_n^2$.

But then Var($\bar{Y}$) = $\sigma^2/n \leq \sigma^2(a_1^2 + a_2^2 + \ldots + a_n^2)$ = Var(W_a).

13 (i) E(W_1) = [(n – 1)/n]E($\bar{Y}$) = [(n – 1)/n]μ, and so Bias(W_1) = [(n – 1)/n]$\mu - \mu = -\mu/n$. Similarly, E(W_2) = E($\bar{Y}$)/2 = μ/2, and so Bias(W_2) = μ/2 – $\mu = -\mu/2$. The bias in W_1 tends to zero as $n \to \infty$, while the bias in W_2 is –μ/2 for all n. This is an important difference.

(ii) plim(W_1) = plim[(n – 1)/n] · plim($\bar{Y}$) = 1 · $\mu = \mu$. plim(W_2) = plim($\bar{Y}$)/2 = μ/2. Because plim(W_1) = μ and plim(W_2) = μ/2, W_1 is consistent whereas W_2 is inconsistent.

(iii) Var(W_1) = [(n – 1)/n]^{2}Var($\bar{Y}$) = [(n – 1)^2/n^3]σ^2 and Var(W_2) = Var($\bar{Y}$)/4 = $\sigma^2/(4n)$.

(iv) Because $\bar{Y}$ is unbiased, its mean squared error is simply its variance. On the other hand, MSE(W_1) = Var(W_1) + [Bias(W_1)]^2 = [(n – 1)^2/n^3]$\sigma^2 + \mu^2/n^2$. When $\mu = 0$, MSE(W_1) = Var(W_1) = [(n – 1)^2/n^3]$\sigma^2 < \sigma^2/n$ = Var($\bar{Y}$) because (n – 1)/n < 1. Therefore, MSE(W_1) is smaller than Var($\bar{Y}$) for μ close to zero. For large n, the difference between the two estimators is trivial.

14 (i) While the expected value of the numerator of G is E($\bar{Y}$) = θ, and the expected value of the denominator is E(1 – $\bar{Y}$) = 1 – θ, the expected value of the ratio is not the ratio of the expected value.

(ii) By Property PLIM.2(iii), the plim of the ratio is the ratio of the plims (provided the plim of the denominator is not zero):
plim(G) = plim[$\bar{Y}$/(1 – $\bar{Y}$)] = plim($\bar{Y}$)/[1 – plim($\bar{Y}$)] = $\theta/(1 - \theta) = \gamma$.

15 (i) H_0: $\mu = 0$.

(ii) H_1: $\mu < 0$.

(iii) The standard error of $\overline{y}$ is $s/\sqrt{n}$ = 13.8/30 = .46. Therefore, the *t*-statistic for testing H_0: $\mu = 0$ is $t = \overline{y}/\text{se}(\overline{y}) = -.97/.46 \approx -2.11$. We obtain the *p*-value as $P(Z \leq -2.11)$, where $Z \sim$ Normal(0,1). These probabilities are in the appendix of statistical tables. *p*-value = .0174. Because the *p*-value is below .05, we reject H_0 against the one-sided alternative at the 5% level. We do not reject at the 1% level because *p*-value = .0174 > .01.

(iv) The estimated reduction, about .97 litres, does not seem large for an entire year's consumption. If the alcohol is beer, .97 litres is less than three 375-mL cans of beer. Even if this is hard liquor, the reduction seems small. (On the other hand, when aggregated across the entire population, alcohol distributors might not think the effect is so small.)

(v) The implicit assumption is that other factors that affect alcohol consumption – such as income, or changes in price due to transportation costs – are constant over the two years.

16 (i) The average increase in wage is $\overline{D}$ = .24, or 24 cents. The sample standard deviation is about .451, and so, with *n* = 15, the standard error of $\overline{D}$ is $.451/\sqrt{15} \approx .1164$.

From Table A.2, the 97.5th percentile in the t_{14} distribution is 2.145.
So the 95% CI is $.24 \pm 2.145(.1164)$, or about –.010 to .490.

(ii) If $\mu = E(D_i)$ then H_0: $\mu = 0$. The alternative is that management's claim is true: H_1: $\mu > 0$.

(iii) We have the mean and standard error from part (i): $t = .24/.1164 \approx 2.062$. The 5% critical value for a one-tailed test with *df* = 14 is 1.761, while the 1% critical value is 2.624. Therefore, H_0 is rejected in favour of H_1 at the 5% level but not the 1% level.

17 (i) For each player, θ is estimated using $\overline{Y}$ in the table below.

Player	**Goals**	**TSG**	$\overline{Y}$
Nick Riewoldt	44	73	0.603
Luke Breust	49	63	0.778
Jarryd Roughhead	55	99	0.556
Lance Franklin	52	99	0.525
Jack Riewoldt	48	93	0.516
Travis Cloke	38	82	0.463

(ii) $\text{Var}(\overline{Y}) = \theta(1-\theta)/n$ [because the variance of each Y_i is $\theta(1-\theta)$ and so $\text{sd}(\overline{Y}) = \sqrt{\theta(1-\theta)/n}$.

(iii) The asymptotic *t*-statistic is $(\overline{Y} - .5)/\text{se}(\overline{Y})$; when we plug in the estimate for each player we obtain the se($\overline{y}$) for each player. These are calculated and presented in the third column of the following table. The critical value (based on the standard normal distribution) with 5% level of significance for a one-tailed test with the alternate hypothesis H_1: $\theta > .5$ is 1.645. So the null hypothesis that the probability of kicking any particular goal = .5 or $\theta = .5$ is rejected for the first two players, as shown in the table.

Player	$\overline{Y}$	**se(** $\overline{y}$ **)** = $\sqrt{\overline{y}(1-\overline{y})/n}$	***t*-statistics**	**Test outcome**
Nick Riewoldt	0.603	0.057	(.603 – .5)/.057 ≈ 1.807	Reject H_0
Luke Breust	0.778	0.052	(.778 – .5)/.052 ≈ 5.346	Reject H_0
Jarryd Roughhead	0.556	0.050	(.556 – .5)/.050 ≈ 1.12	Do not reject H_0
Lance Franklin	0.525	0.050	(.525 – .5)/.050 ≈ 0.503	Do not reject H_0
Jack Riewoldt	0.516	0.052	(.516 – .5)/.052 ≈ 0.308	Do not reject H_0
Travis Cloke	0.463	0.055	(.463 – .5)/.055 ≈ –0.6734	Do not reject H_0

18 We need to conduct a hypothesis test of the mean price of new houses in Sydney.

The hypotheses are:
$H_0: \mu = \$370\,000$
$H_1: \mu > \$370\,000$

The test statistic is:
$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{375500 - 370000}{\sqrt{\frac{16000^2}{256}}} = 5.5$$

At 5% level of significance, the critical value of *t* with upper one tailed test with n–1 = 256–1 = 255 *df* is $t_{.05}$ = 1.645, which is the same as the value of a standard normal value (given the large sample).

As $t = 5.5 > Z_{crit} = c = 1.645$, hence we reject the null hypothesis and we conclude that, based on this sample, the average new house price in Sydney is significantly higher than the national average price of new homes.

Multiple-choice questions

1	c	**5**	c	**9**	a
2	c	**6**	b	**10**	b
3	b	**7**	d	**11**	c
4	a	**8**	d	**12**	a

Chapter 4: The simple regression model

TEACHING NOTES

This is the chapter where we expect students to follow most, if not all, of the algebraic derivations. In class, we like to derive at least the unbiasedness of the OLS slope coefficient, and usually, we derive the variance. At a minimum, we talk about the factors affecting the variance. To simplify the notation, after we emphasise the assumptions in the population model, and assume random sampling, we just condition on the values of the explanatory variables in the sample. Technically, this is justified by random sampling because, for example, $E(u_i|x_1, x_2, ..., x_n) = E(u_i|x_i)$ by independent sampling. We find that students are able to focus on the key assumption SLR.4 and subsequently take our word about how conditioning on the independent variables in the sample is harmless. Because statistical inference is no more difficult in multiple regression than in simple regression, we postpone inference until Chapter 6. (This reduces redundancy and allows you to focus on the interpretive differences between simple and multiple regression.)

You might notice how, compared with most other texts, we use relatively few assumptions to derive the unbiasedness of the OLS slope estimator, followed by the formula for its variance. This is because we do not introduce redundant or unnecessary assumptions. For example, once SLR.4 is assumed, nothing further about the relationship between u and x is needed to obtain the unbiasedness of OLS under random sampling.

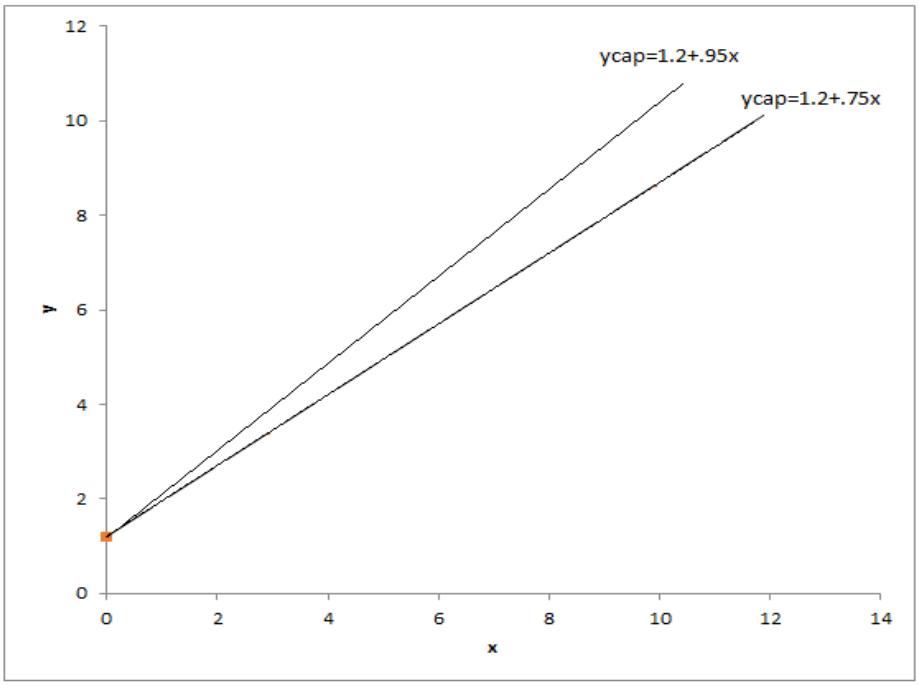
Incidentally, one of the uncomfortable facts about finite-sample analysis is that there is a difference between an estimator that is unbiased conditional on the outcome of the covariates and one that is unconditionally unbiased. If the distribution of the x_t is such that they can all equal the same value with positive probability – as is the case with discreteness in the distribution – then the unconditional expectation does not really exist. Or, if it is made to exist, then the estimator is not unbiased. We do not try to explain these subtleties in an introductory course, but we have had instructors ask about the difference.

Solutions to review questions

1 (i) Income, age, and family background (such as number of siblings) are just a few possibilities. It seems that each of these could be correlated with years of education. (Income and education are probably positively correlated; age and education may be negatively correlated because women in more recent cohorts have, on average, more education; and number of siblings and education are probably negatively correlated.)

(ii) A simple regression as shown will not be sufficient if the factors we listed in part (i) are correlated with *educ*. Because we would like to hold these factors fixed, they are part of the error term. However as per one of the basic assumptions of simple regression, we know that u is not to be correlated with the explanatory variable. Hence, if u is correlated with *educ* then $E(u|educ) \neq 0$, and so SLR.4 fails.

2 The estimated regression models are shown in the following figure.



It is clear that the estimates of the intercept, β_1 is same for both the samples. However, the slopes are different. The estimates of the slope β_2 from the first and second samples are 0.75 and 0.95, respectively. The above figure shows that while the fitted line representing the second sample is steeper due to its higher slope estimate, both the lines have the same intercept (1.2).

3 (i) Let $y_i = MARKS_i$, $x_i = HOURS_i$, and $n = 8$. Then $\bar{x} = 25.875$, $\bar{y} = 66.125$, $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 120.125$, and $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 56.875$. From equation (4.17), we obtain the slope as $\hat{\beta}_1 = 120.125/56.875 \approx 2.112$, rounded to three places after the decimal. From (4.18),
$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \approx 66.125 - (2.112)25.875 \approx 11.477$. So we can write

$$MARKS = 11.447 + 2.112\, HOURS$$

$$n = 8.$$

The intercept does not have a useful interpretation because the value of the variable *HOURS* is not close to zero for the population of interest. If *HOURS* is 5 units higher, *MARKS* increases by 2.112(5) = 10.56.

(ii) The fitted values and residuals — rounded to three decimal places — are given along with the observation number *i* and *MARKS* in the following table:

i	*MARKS*	$\widehat{MARKS}$	$\hat{u}$
1	58	55.829	2.171
2	69	62.165	6.835
3	62	66.389	–4.389
4	73	68.501	4.499
5	74	72.725	1.275
6	62	64.277	–2.277
7	55	64.277	–9.277
8	76	74.837	1.163

You can verify that the residuals, as reported in the table, sum to 0.

(iii) When *HOURS* = 25, *MARKS* = 11.447 + 2.112(25) ≈ 64.25.

(iv) The sum of squared residuals, $\sum_{i=1}^{n}\hat{u}_i^2$, is about 185.1604 (rounded to four decimal places), and the total sum of squares, $\sum_{i=1}^{n}(y_i - \bar{y})^2$, is about 438.875. So the *R*-squared from the regression is

$$R^2 = 1 - SSR/SST \approx 1 - (185.1604/438.875) \approx .578.$$

Therefore, about 57.8% of the variation in *MARKS* is explained by *HOURS* in this small sample of students.