For Students

Solutions to Odd-Numbered End-of-Chapter Exercises

Chapter 2  
Review of Probability

2.1. (a) Probability distribution function for *Y*

|  |  |  |  |
| --- | --- | --- | --- |
| Outcome (number of heads) | *Y*  0 | *Y*  1 | *Y*  2 |
| Probability | 0.25 | 0.50 | 0.25 |

(b) Cumulative probability distribution function for *Y*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Outcome (number of heads) | *Y*  0 | 0 ≤ *Y*  1 | 1 ≤ Y  2 | *Y* ≥ 2 |
| Probability | 0 | 0.25 | 0.75 | 1.0 |

(c) .

Using Key Concept 2.3: 

and



so that



2.3. For the two new random variables  and  we have:

(a) 

(b) 

(c) 



2.5. Let *X* denote temperature in °F and *Y* denote temperature in °C. Recall that *Y*  0 when *X*  32 and *Y* 100 when *X*  212; this implies  Using Key Concept 2.3, *μX*  70oF implies that  and *σX*  7oF implies 

2.7. Using obvious notation,  thus  and  This implies

(a)  per year.

(b)  , so that  Thus   where the units are squared thousands of dollars per year.

(c)  so that  and  thousand dollars per year.

(d) First you need to look up the current Euro/dollar exchange rate in the Wall Street Journal, the Federal Reserve web page, or other financial data outlet. Suppose that this exchange rate is *e* (say *e*  0.80 Euros per dollar); each 1 dollar is therefore with *e* Euros. The mean is therefore  
*e* × *μC* (in units of thousands of Euros per year), and the standard deviation is *e* × *σC* (in units of thousands of Euros per year). The correlation is unit-free, and is unchanged.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2.9. |  | | **Value of *Y*** | | | | | **Probability Distribution of *X*** |
| **14** | **22** | **30** | **40** | **65** |
|  | **Value of *X*** | 1 | 0.02 | 0.05 | 0.10 | 0.03 | 0.01 | 0.21 |
| 5 | 0.17 | 0.15 | 0.05 | 0.02 | 0.01 | 0.40 |
| 8 | 0.02 | 0.03 | 0.15 | 0.10 | 0.09 | 0.39 |
|  | **Probability distribution of *Y*** | | 0.21 | 0.23 | 0.30 | 0.15 | 0.11 | 1.00 |

(a) The probability distribution is given in the table above.



(b) The conditional probability of *Y*|*X*  8 is given in the table below

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Value of *Y*** | | | | |
| 14 | 22 | 30 | 40 | 65 |
| 0.02/0.39 | 0.03/0.39 | 0.15/0.39 | 0.10/0.39 | 0.09/0.39 |





(c) 





2.11. (a) 0.90

(b) 0.05

(c) 0.05

(d) When  then 

(e)  where thus 

2.13. (a) 

(b) *Y* and *W* are symmetric around 0, thus skewness is equal to 0; because their mean is zero, this means that the third moment is zero.

(c) The kurtosis of the normal is 3, so  ; solving yields  a similar calculation yields the results for *W*.

(d) First, condition on  so that 



Similarly,



From the law of iterated expectations



(e)  thus from part (d). Thus skewness  0. Similarly,  and  Thus,

2.15. (a) 

where *Z* ~ *N*(0, 1). Thus,

(i) *n*  20; 

(ii) *n*  100; 

(iii) *n*  1000; 

(b) 

As *n* get large  gets large, and the probability converges to 1.

(c) This follows from (b) and the definition of convergence in probability given in Key Concept 2.6.

2.17. *μY* = 0.4 and 

(a) (i) *P*( ≥ 0.43)  

(ii) *P*( ≤ 0.37)  

(b) We know Pr(1.96 ≤ *Z* ≤ 1.96)  0.95, thus we want *n* to satisfy  and  Solving these inequalities yields *n* ≥ 9220.

2.19. (a) 

(b) 

(c) When  and  are independent,



so





2.21. (a) 

(b) 

2.23. *X* and *Z* are two independently distributed standard normal random variables, so



(a) Because of the independence between  and   and  Thus 

(b)  and 

(c)  Using the fact that the odd moments of a standard normal random variable are all zero, we have  Using the independence between  and  we have  Thus 

(d) 

2.25. (a) 

(b) 

(c) 

(d) 

2.27 (a) *E*(*W*)  *E*[*E*(*W*|*Z*)]  *E*[*E*(*X *)|*Z*]  *E*[*E*(*X*|*Z*)  *E*(*X*|*Z*)]  0.

(b) *E*(*WZ*)  *E*[*E*(*WZ*|*Z*)]  *E*[*ZE*(*W*)|*Z*]  *E*[ *Z* ×0]  0

(c) Using the hint: *V*  *W*  *h*(*Z*), so that *E*(*V*2)  *E*(*W*2)  *E*[*h*(*Z*)2]  2 × *E*[*W* × *h*(*Z*)]. Using an argument like that in (b), *E*[*W* × *h*(*Z*)]  0. Thus, *E*(*V*2)  *E*(*W*2)  *E*[*h*(*Z*)2], and the result follows by recognizing that *E*[*h*(*Z*)2] ≥ 0 because *h*(*z*)2≤ 0 for any value of *z*.

Chapter 3  
Review of Statistics

3.1. The central limit theorem suggests that when the sample size () is large, the distribution of the sample average () is approximately  with  Given a population   we have

(a)   and



(b)   and



(c)   and



3.3. Denote each voter’s preference by   if the voter prefers the incumbent and  if the voter prefers the challenger.  is a Bernoulli random variable with probability Pr and Pr From the solution to Exercise 3.2,  has mean  and variance 

(a) 

(b) The estimated variance of  is  The standard error is SE

(c) The computed *t*-statistic is



Because of the large sample size  we can use Equation (3.14) in the text to get the  
*p*-value for the test  vs. 



(d) Using Equation (3.17) in the text, the *p*-value for the test  vs.  is



(e) Part (c) is a two-sided test and the *p*-value is the area in the tails of the standard normal distribution outside ± (calculated *t*-statistic). Part (d) is a one-sided test and the *p*-value is the area under the standard normal distribution to the right of the calculated *t*-statistic.

(f) For the test  vs.  we cannot reject the null hypothesis at the 5% significance level. The *p*-value 0.066 is larger than 0.05. Equivalently the calculated *t*-statistic  is less than the critical value 1.64 for a one-sided test with a 5% significance level. The test suggests that the survey did not contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.

3.5. (a) (i) The size is given by where the probability is computed assuming that 



where the final equality using the central limit theorem approximation.

(ii) The power is given by where the probability is computed assuming that *p*  0.53.



where the final equality using the central limit theorem approximation.

(b) (i) so that the null is rejected at the 5% level.

(ii) so that the null is rejected at the 5% level.

(iii) 

(iv) 

(v) 

(c) (i) The probability is 0.95 is any single survey, there are 20 independent surveys, so the probability if 

(ii) 95% of the 20 confidence intervals or 19.

(d) The relevant equation is  Thus *n* must be chosen so that  so that the answer depends on the value of *p*. Note that the largest value that *p*(1 − *p*) can take on is 0.25 (that is, *p*  0.5 makes *p*(1  *p*) as large as possible). Thus if  then the margin of error is less than 0.01 for all values of *p*.

3.7. The null hypothesis is that the survey is a random draw from a population with *p* = 0.11. The *t*-statistic is  where  (An alternative formula for SE() is  which is valid under the null hypothesis that The value of the *t-*statistic is 2.71, which has a *p*-value of that is less than 0.01. Thus the null hypothesis (the survey is unbiased) can be rejected at the 1% level.

3.9. Denote the life of a light bulb from the new process by  The mean of  is  and the standard deviation of  is  hours.  is the sample mean with a sample size  The standard deviation of the sampling distribution of  is  hours. The hypothesis test is  vs.  The manager will accept the alternative hypothesis if  hours.

(a) The size of a test is the probability of erroneously rejecting a null hypothesis when it is valid.

The size of the manager’s test is



where  means the probability that the sample mean is greater than 2100 hours when the new process has a mean of  hours.

(b) The power of a test is the probability of correctly rejecting a null hypothesis when it is invalid. We calculate first the probability of the manager erroneously accepting the null hypothesis when it is invalid:



The power of the manager’s testing is 

(c) For a test with 5%, the rejection region for the null hypothesis contains those values of the   
*t*-statistic exceeding 1.645.



The manager should believe the inventor’s claim if the sample mean life of the new product is greater than 2032.9 hours if she wants the size of the test to be 5%.

3.11. Assume that  is an even number. Then  is constructed by applying a weight of 1/2 to the *n*/2 “odd” observations and a weight of 3/2 to the remaining *n*/2 observations.



3.13 (a) Sample size  sample average   646.2 sample standard deviation  The standard error of  is SE The 95% confidence interval for the mean test score in the population is



(b) The data are: sample size for small classes  sample average  sample standard deviation  sample size for large classes  sample average  sample standard deviation  The standard error of  is  The hypothesis tests for higher average scores in smaller classes is



The *t*-statistic is



The *p*-value for the one-sided test is:



With the small *p*-value, the null hypothesis can be rejected with a high degree of confidence. There is statistically significant evidence that the districts with smaller classes have higher average test scores.

3.15. From the textbook equation (2.46), we know that *E*()  *μY* and from (2.47) we know that   
var()  . In this problem, because *Ya* and *Yb* are Bernoulli random variables,  ,  ,  *pa*(1–*pa*) and  *pb*(1–*pb*). The answers to (a) follow from this. For part (b), note that var(–)  var()  var() – 2cov(,). But, they are independent (and thus have   0 because and are independent (they depend on data chosen from independent samples). Thus var(–)  var()  var(). For part (c), use equation 3.21 from the text (replacing with and using the result in (b) to compute the SE). For (d), apply the formula in (c) to obtain

95% CI is (.859 – .374) ± 1.96 or 0.485 ± 0.017.

3.17. (a) The 95% confidence interval is  where  the 95% confidence interval is (24.98  23.27) ± 0.73 or 1.71 ± 0.73.

(b) The 95% confidence interval is  where  the 95% confidence interval is  or 0.82 ± 0.60.

(c) The 95% confidence interval is   where     The 95% confidence interval is (24.98-23.27) − (20.87−20.05) ± 1.96 × 0.48 or 0.89 ± 0.95.

3.19. (a) No. Thus



(b) Yes. If gets arbitrarily close to *μY* with probability approaching 1 as *n* gets large, then gets arbitrarily close to  with probability approaching 1 as *n* gets large. (As it turns out, this is an example of the “continuous mapping theorem” discussed in Chapter 17.)

3.21. Set *nm*  *nw*  *n*, and use equation (3.19) write the squared SE of  as



Similarly, using equation (3.23)



Chapter 4  
Linear Regression with One Regressor

4.1. (a) The predicted average test score is



(b) The predicted change in the classroom average test score is



(c) Using the formula forin Equation (4.8), we know the sample average of the test scores across the 100 classrooms is



(d) Use the formula for the standard error of the regression (SER) in Equation (4.19) to get the sum of squared residuals:



Use the formula for  in Equation (4.16) to get the total sum of squares:



The sample variance is   Thus, standard deviation is 

4.3. (a) The coefficient 9.6 shows the marginal effect of *Age* on *AWE*; that is, *AWE* is expected to increase by $9.6 for each additional year of age. 696.7 is the intercept of the regression line. It determines the overall level of the line.

(b) *SER* is in the same units as the dependent variable (*Y*, or *AWE* in this example). Thus *SER* is measured in dollars per week.

(c) *R*2 is unit free.

(d) (i) 

(ii) 

(e) No. The oldest worker in the sample is 65 years old. 99 years is far outside the range of the sample data.

(f) No. The distribution of earning is positively skewed and has kurtosis larger than the normal.

(g)  so that  Thus the sample mean of *AWE* is 696.7  9.6 × 41.6  $1,096.06.

4.5. (a) *ui* represents factors other than time that influence the student’s performance on the exam including amount of time studying, aptitude for the material, and so forth. Some students will have studied more than average, other less; some students will have higher than average aptitude for the subject, others lower, and so forth.

(b) Because of random assignment *ui* is independent of *Xi*. Since *ui* represents deviations from average *E*(*ui*)  0. Because *u* and *X* are independent *E*(*ui*|*Xi*)  *E*(*ui*)  0.

(c) (2) is satisfied if this year’s class is typical of other classes, that is, students in this year’s   
class can be viewed as random draws from the population of students that enroll in the class. (3) is satisfied because 0 ≤ *Yi* ≤ 100 and *Xi* can take on only two values (90 and 120).

(d) (i) 

(ii) 

4.7. The expectation of  is obtained by taking expectations of both sides of Equation (4.8):



where the third equality in the above equation has used the facts that *E*(*ui*)  0 and *E*[(−*β*1)]  *E*[(*E*(−*β*1)| )]  because  (see text equation (4.31).)

4.9. (a) With  and  Thus ESS  0 and *R*2  0.

(b) If *R*2  0, then ESS  0, so that  for all *i*. But  so that  for all *i*, whichimplies that or that *Xi* is constant for all *i*. If *Xi* is constant for all *i*, then  and  is undefined (see equation (4.7)).

4.11. (a) The least squares objective function is  Differentiating with respect to *b*1yields  Setting this zero, and solving for the least squares estimator yields 

(b) Following the same steps in (a) yields 

4.13. The answer follows the derivations in Appendix 4.3 in “Large-Sample Normal Distribution of the OLS Estimator.” In particular, the expression for *νi* is now *νi*  (*Xi*  *μX*)*κui*, so that var(*νi*)  *κ*3var[(*Xi*  *μX*)*ui*], and the term *κ*2 carry through the rest of the calculations.

Chapter 5  
Regression with a Single Regressor: Hypothesis  
Tests and Confidence Intervals

5.1 (a) The 95% confidence interval for  is  that is

(b) Calculate the *t*-statistic:



The *p*-value for the test  vs.  is



The *p*-value is less than 0.01, so we can reject the null hypothesis at the 5% significance level, and also at the 1% significance level.

(c) The *t*-statistic is



The *p*-value for the test  vs.  is



The *p*-value is larger than 0.10, so we cannot reject the null hypothesis at the 10%, 5% or 1% significance level. Because  is not rejected at the 5% level, this value is contained in the 95% confidence interval.

(d) The 99% confidence interval for  is  that is, 

5.3. The 99% confidence interval is 1.5 × {3.94 ± 2.58 × 0.31) or 4.71 lbs ≤ WeightGain ≤ 7.11 lbs.

5.5 (a) The estimated gain from being in a small class is 13.9 points. This is equal to approximately 1/5 of the standard deviation in test scores, a moderate increase.

(b) The *t*-statistic is  which has a *p*-value of 0.00. Thus the null hypothesis is rejected at the 5% (and 1%) level.

(c) 13.9 ± 2.58 × 2.5  13.9 ± 6.45.

5.7. (a) The *t*-statistic is  with a *p*-value of 0.03; since the *p*-value is less than 0.05, the null hypothesis is rejected at the 5% level.

(b) 3.2 ± 1.96 × 1.5  3.2 ± 2.94

(c) Yes. If *Y* and *X* are independent, then  but this null hypothesis was rejected at the   
5% level in part (a).

(d) would be rejected at the 5% level in 5% of the samples; 95% of the confidence intervals would contain the value 

5.9. (a)  so that it is linear function of *Y*1, *Y*2, …, *Yn*.

(b) *E*(*Yi*|*X*1, …, *Xn*)  *βXi*, thus



5.11. Using the results from 5.10,  and  From Chapter 3,  and  Plugging in the numbers  and  and 

5.13. (a) Yes, this follows from the assumptions in KC 4.3.

(b) Yes, this follows from the assumptions in KC 4.3 and conditional homoskedasticity

(c) They would be unchanged for the reasons specified in the answers to those questions.

(d) (a) is unchanged; (b) is no longer true as the errors are not conditionally homosckesdastic.

5.15. Because the samples are independent,  and  are independent. Thus    is consistently estimated as  and  is consistently estimated as  so that  is consistently estimated by  and the result follows by noting the SE is the square root of the estimated variance.

Chapter 6  
Linear Regression with   
Multiple Regressors

6.1. By equation (6.15) in the text, we know



Thus, that values of  are 0.175, 0.189, and 0.193 for columns (1)–(3).

6.3. (a) On average, a worker earns $0.29/hour more for each year he ages.

(b) Sally’s earnings prediction is  dollars per hour. Betsy’s earnings prediction is  dollars per hour. The difference is 1.45

6.5. (a) $23,400 (recall that *Price* is measured in $1000s).

(b) In this case Δ*BDR* 1 and Δ*Hsize*  100. The resulting expected change in price is 23.4  0.156 × 100  39.0 thousand dollars or $39,000.

(c) The loss is $48,800.

(d) From the text  so  thus, *R*2  0.727.

6.7. (a) The proposed research in assessing the presence of gender bias in setting wages is too limited. There might be some potentially important determinants of salaries: type of engineer, amount of work experience of the employee, and education level. The gender with the lower wages could reflect the type of engineer among the gender, the amount of work experience of the employee, or the education level of the employee. The research plan could be improved with the collection of additional data as indicated and an appropriate statistical technique for analyzing the data would be a multiple regression in which the dependent variable is wages and the independent variables would include a dummy variable for gender, dummy variables for type of engineer, work experience (time units), and education level (highest grade level completed). The potential importance of the suggested omitted variables makes a “difference in means” test inappropriate for assessing the presence of gender bias in setting wages.

(b) The description suggests that the research goes a long way towards controlling for potential omitted variable bias. Yet, there still may be problems. Omitted from the analysis are characteristics associated with behavior that led to incarceration (excessive drug or alcohol use, gang activity, and so forth), that might be correlated with future earnings. Ideally, data on these variables should be included in the analysis as additional control variables.

6.9. For omitted variable bias to occur, two conditions must be true: *X*1 (the included regressor) is correlated with the omitted variable, and the omitted variable is a determinant of the dependent variable. Since *X*1 and *X*2 are uncorrelated, the estimator of *β*1 does not suffer from omitted variable bias.

6.11. (a) 

(b) 

(c) From (b),  satisfies



or 

and the result follows immediately.

(d) Following analysis as in (c)



and substituting this into the expression for  in (c) yields



Solving for  yields:



(e) The least squares objective function is  and the partial derivative with respect to *b*0is



Setting this to zero and solving for  yields: 

(f) Substituting into the least squares objective function yields , which is identical to the least squares objective function in part (a), except that all variables have been replaced with deviations from sample means. The result then follows as in (c).

Notice that the estimator for *β*1 is identical to the OLS estimator from the regression of *Y* onto *X*1, omitting *X*2. Said differently, when , the estimated coefficient on *X*1 in the OLS regression of *Y* onto both *X*1 and *X*2 is the same as estimated coefficient in the OLS regression of *Y* onto *X*1.

Chapter 7  
Hypothesis Tests and Confidence  
Intervals in Multiple Regression

7.1 and 7.2

|  |  |  |  |
| --- | --- | --- | --- |
| **Regressor** | **(1)** | **(2)** | **(3)** |
| College (*X*1) | 5.46\*\* (0.21) | 5.48\*\* (0.21) | 5.44\*\* (0.21) |
| Female (*X*2) |  2.64\*\* (0.20) |  2.62\*\* (0.20) |  2.62\*\* (0.20) |
| Age (*X*3) |  | 0.29\*\* (0.04) | 0.29\*\* (0.04) |
| Ntheast (*X*4) |  |  | 0.69\* (0.30) |
| Midwest (*X*5) |  |  | 0.60\* (0.28) |
| South (*X*6) |  |  |  0.27 (0.26) |
| Intercept | 12.69\*\* (0.14) | 4.40\*\* (1.05) | 3.75\*\* (1.06) |

(a) The *t*-statistic is 5.46/0.21  26.0, which exceeds 1.96 in absolute value. Thus, the coefficient is statistically significant at the 5% level. The 95% confidence interval is 5.46 ± 1.96 × 0.21.

(b) *t*-statistic is  2.64/0.20  13.2, and 13.2  1.96, so the coefficient is statistically significant at the 5% level. The 95% confidence interval is 2.64 ± 1.96 × 0.20.

7.3. (a) Yes, age is an important determinant of earnings. Using a *t*-test, the *t*-statistic is  with a *p*-value of 4.2 × 1013, implying that the coefficient on age is statistically significant at the 1% level. The 95% confidence interval is 0.29 ± 1.96 × 0.04.

(b) Δ*Age* × [0.29 ± 1.96 × 0.04]  5 × [0.29 ± 1.96 × 0.04]  1.45 ± 1.96 × 0.20  $1.06 to $1.84

7.5. The *t*-statistic for the difference in the college coefficients is  . Because  and  are computed from independent samples, they are independent, which means that  Thus,  = . This implies that  Thus,  There is no significant change since the calculated *t*-statistic is less than 1.96, the 5% critical value.

7.7. (a) The *t*-statistic is  Therefore, the coefficient on BDR is not statistically significantly different from zero.

(b) The coefficient on *BDR* measures the *partial effect* of the number of bedrooms holding house size (*Hsize*) constant. Yet, the typical 5-bedroom house is much larger than the typical  
2-bedroom house. Thus, the results in (a) says little about the conventional wisdom.

(c) The 99% confidence interval for effect of lot size on price is 2000 × [0.002 ± 2.58 × 0.00048] or 1.52 to 6.48 (in thousands of dollars).

(d) Choosing the scale of the variables should be done to make the regression results easy to read and to interpret. If the lot size were measured in thousands of square feet, the estimate coefficient would be 2 instead of 0.002.

(e) The 10% critical value from the  distribution is 2.30. Because 0.08  2.30, the coefficients are not jointly significant at the 10% level.

7.9. (a) Estimate



and test whether *γ*  0.

(b) Estimate



and test whether *γ*  0.

(c) Estimate



and test whether *γ*  0.

7.11. (a) Treatment (assignment to small classes) was not randomly assigned in the population (the continuing and newly-enrolled students) because of the difference in the proportion of treated continuing and newly-enrolled students. Thus, the treatment indicator *X*1is correlated with *X*2. If newly-enrolled students perform systematically differently on standardized tests than continuing students (perhaps because of adjustment to a new school), then this becomes part of the error term *u* in (a). This leads to correlation between *X*1 and *u*, so that *E*(*u*|*X*l) ≠ 0. Because *E*(*u*|*X*l) ≠ 0, the  is biased and inconsistent.

(b) Because treatment was randomly assigned conditional on enrollment status (continuing or newly-enrolled), *E*(*u* | *X*1, *X*2) will not depend on *X*1. This means that the assumption of conditional mean independence is satisfied, and is unbiased and consistent. However, because *X*2 was not randomly assigned (newly-enrolled students may, on average, have attributes other than being newly enrolled that affect test scores), *E*(*u* | *X*1, *X*2) may depend of *X*2, so that  may be biased and inconsistent.

Chapter 8  
Nonlinear Regression Functions

8.1. (a) The percentage increase in sales is  The approximation   
is 100 × [ln (198)  ln (196)]  1.0152%.

(b) When *Sales*2010  205, the percentage increase is  and the approximation is 100 × [ln (205)  ln (196)]  4.4895%. When *Sales*2010  250, the percentage increase is  and the approximation is 100 × [ln (250)  ln (196)]  24.335%. When *Sales*2010  500, the percentage increase is  and the approximation is 100 × [ln (500)  ln (196)]  93.649%.

(c) The approximation works well when the change is small. The quality of the approximation deteriorates as the percentage change increases.

8.3. (a) The regression functions for hypothetical values of the regression coefficients that are consistent with the educator’s statement are: and  When  is plotted against  the regression will show three horizontal segments. The first segment will be for values of the next segment for  the final segment for  The first segment will be higher than the second, and the second segment will be higher than the third.

(b) It happens because of perfect multicollinearity. With all three class size binary variables included in the regression, it is impossible to compute the OLS estimates because the intercept is a perfect linear function of the three class size regressors.

8.5. (a) (1) The demand for older journals is less elastic than for younger journals because the interaction term between the log of journal age and price per citation is positive. (2) There is a linear relationship between log price and log of quantity follows because the estimated coefficients on log price squared and log price cubed are both insignificant. (3) The demand is greater for journals with more characters follows from the positive and statistically significant coefficient estimate on the log of characters.

(b) (i) The effect of ln(*Price per citation*) is given by [0.899  0.141 × ln(*Age*)] × ln(*Price per citation*). Using *Age*  80, the elasticity is [0.899  0.141 × ln(80)]  0.28.

(ii) As described in equation (8.8) and the footnote on page 261, the standard error can be found by dividing 0.28, the absolute value of the estimate, by the square root of the   
*F*-statistic testing *βln*(*Price per citation*)  ln(80) × *βln*(*Age*)×*ln*(*Price per citation*)  0.

(c)  for any constant *a*. Thus, estimated parameter on *Characters* will not change and the constant (intercept) will change.

8.7. (a) (i) ln(*Earnings*) for females are, on average, 0.44 lower for men than for women.

(ii) The error term has a standard deviation of 2.65 (measured in log-points).

(iii) Yes. However the regression does not control for many factors (size of firm, industry, profitability, experience and so forth).

(iv) No. In isolation, these results do not imply gender discrimination. Gender discrimination means that two workers, identical in every way but gender, are paid different wages. Thus, it is also important to control for characteristics of the workers that may affect their productivity (education, years of experience, etc.) If these characteristics are systematically different between men and women, then they may be responsible for the difference in mean wages. (If this were true, it would raise an interesting and important question of why women tend to have less education or less experience than men, but that is a question about something other than gender discrimination in top corporate jobs.) These are potentially important omitted variables in the regression that will lead to bias in the OLS coefficient estimator for *Female*. Since these characteristics were not controlled for in the statistical analysis, it is premature to reach a conclusion about gender discrimination.

(b) (i) If *MarketValue* increases by 1%, earnings increase by 0.37%

(ii) *Female* is correlated with the two new included variables and at least one of the variables is important for explaining ln(*Earnings*). Thus the regression in part (a) suffered from omitted variable bias.

(c) Forgetting about the effect or *Return*, whose effects seems small and statistically insignificant, the omitted variable bias formula (see equation (6.1)) suggests that *Female* is negatively correlated with ln(*MarketValue*).

8.9. Note that



Define a new independent variable  and estimate



The confidence interval is 

8.11. Linear model: *E*(*Y* | *X*)  *β*0  *β*1*X*, so that  and the elasticity is 

Log-Log Model: *E*(*Y* | *X*)  , where *c*    
*E*(*eu* | *X*), which does not depend on *X* because *u* and *X* are assumed to be independent.

Thus , and the elasticity is *β*1.

Chapter 9  
Assessing Studies Based on  
Multiple Regression

9.1. As explained in the text, potential threats to external validity arise from differences between the population and setting studied and the population and setting of interest. The statistical results based on New York in the 1970s are likely to apply to Boston in the 1970s but not to Los Angeles in the 1970s. In 1970, New York and Boston had large and widely used public transportation systems. Attitudes about smoking were roughly the same in New York and Boston in the 1970s. In contrast, Los Angeles had a considerably smaller public transportation system in 1970. Most residents of Los Angeles relied on their cars to commute to work, school, and so forth. The results from New York in the 1970s are unlikely to apply to New York in 2010. Attitudes towards smoking changed significantly from 1970 to 2010.

9.3. The key is that the selected sample contains only employed women. Consider two women, Beth and Julie. Beth has no children; Julie has one child. Beth and Julie are otherwise identical. Both can earn $25,000 per year in the labor market. Each must compare the $25,000 benefit to the costs of working. For Beth, the cost of working is forgone leisure. For Julie, it is forgone leisure and the costs (pecuniary and other) of child care. If Beth is just on the margin between working in the labor market or not, then Julie, who has a higher opportunity cost, will decide not to work in the labor market. Instead, Julie will work in “home production,” caring for children, and so forth. Thus, on average, women with children who decide to work are women who earn higher wages in the labor market.

9.5. (a) 

and 

(b)  

(c) 

and



(d) (i) 

(ii)  using the fact that *γ*1  0 (supply curves slope up) and *β*1  0  
(demand curves slope down).

9.7. (a) True. Correlation between regressors and error terms means that the OLS estimator is inconsistent.

(b) True.

9.9. Both regressions suffer from omitted variable bias so that they will not provide reliable estimates of the causal effect of income on test scores. However, the nonlinear regression in (8.18) fits the data well, so that it could be used for forecasting.

9.11. Again, there are reasons for concern. Here are a few.

Internal consistency: To the extent that price is affected by demand, there may be simultaneous equation bias.

External consistency: The internet and introduction of “*E*-journals” may induce important changes in the market for academic journals so that the results for 2000 may not be relevant for today’s market.

9.13. (a) . Because all of the *Xi*’s are used (although some are used for the wrong values of *Yj*),  , and . Also, . Using these expressions:



where *n*  300, and the last equality uses an ordering of the observations so that the first   
240 observations ( 0.8 × *n*) correspond to the correctly measured observations (  *Xi*).

As is done elsewhere in the book, we interpret *n*  300 as a large sample, so we use the approximation of *n* tending to infinity. The solution provided here thus shows that these expressions are approximately true for *n* large and hold in the limit that *n* tends to infinity. Each of the averages in the expression for  have the following probability limits:

,

,

, and

,

where the last result follows because  ≠ *Xi* for the scrambled observations and *Xj* is independent of *Xi* for *i* ≠ *j*. Taken together, these results imply.

(b) Because , , so a consistent estimator of *β*1 is the OLS estimator divided by 0.8.

(c) Yes, the estimator based on the first 240 observations is better than the adjusted estimator from part (b). Equation (4.21) in Key Concept 4.4 (page 129) implies that the estimator based on the first 240 observations has a variance that is

.

From part (a), the OLS estimator based on all of the observations has two sources of sampling error. The first is  which is the usual source that comes from the omitted factors (*u*). The second is , which is the source that comes from scrambling the data. These two terms are uncorrelated in large samples, and their respective large-sample variances are:



and

.

Thus



which is larger than the variance of the estimator that only uses the first 240 observations.

Thus



which is larger than the variance of the estimator that only uses the first 240 observations.

Chapter 10  
Regression with Panel Data

10.1. (a) With a $1 increase in the beer tax, the expected number of lives that would be saved is 0.45 per 10,000 people. Since New Jersey has a population of 8.1 million, the expected number of lives saved is 0.45 × 810  364.5. The 95% confidence interval is (0.45  1.96 × 0.22) × 810  [15.228, 713.77].

(b) When New Jersey lowers its drinking age from 21 to 18, the expected fatality rate increases by 0.028 deaths per 10,000. The 95% confidence interval for the change in death rate is 0.028  1.96 × 0.066  [ 0.1014, 0.1574]. With a population of 8.1 million, the number of fatalities will increase by 0.028 × 810  22.68 with a 95% confidence interval [0.1014, 0.1574] × 810  [82.134, 127.49].

(c) When real income per capita in New Jersey increases by 1%, the expected fatality rate increases by 1.81 deaths per 10,000. The 90% confidence interval for the change in death rate is 1.81  1.64 × 0.47  [1.04, 2.58]. With a population of 8.1 million, the number of fatalities will increase by 1.81 × 810  1466.1 with a 90% confidence interval [1.04, 2.58] × 810  [840, 2092].

(d) The low *p*-value (or high *F*-statistic) associated with the *F*-test on the assumption that time effects are zero suggests that the time effects should be included in the regression.

(e) Define a binary variable *west* which equals 1 for the western states and 0 for the other states. Include the interaction term between the binary variable *west* and the unemployment rate, *west* ×(unemployment rate), in the regression equation corresponding to column (4). Suppose the coefficient associated with unemployment rate is *β* and the coefficient associated with *west* ×(unemployment rate) is *γ*. Then *β* captures the effect of the unemployment rate in the eastern states, and *β*  *β* captures the effect of the unemployment rate in the western states. The difference in the effect of the unemployment rate in the western and eastern states is *β*. Using the coefficient estimate  and the standard error  you can calculate the *t*-statistic to test whether *γ*  is statistically significant at a given significance level.

10.3. The five potential threats to the internal validity of a regression study are: omitted variables, misspecification of the functional form, imprecise measurement of the independent variables, sample selection, and simultaneous causality. You should think about these threats one-by-one. Are there important omitted variables that affect traffic fatalities and that may be correlated with the other variables included in the regression? The most obvious candidates are the safety of roads, weather, and so forth. These variables are essentially constant over the sample period, so their effect is captured by the state fixed effects. You may think of something that we missed. Since most of the variables are binary variables, the largest functional form choice involves the *Beer Tax* variable. A linear specification is used in the text, which seems generally consistent with the data in Figure 8.2. To check the reliability of the linear specification, it would be useful to consider a log specification or a quadratic. Measurement error does not appear to a problem, as variables like traffic fatalities and taxes are accurately measured. Similarly, sample selection is a not a problem because data were used from all of the states. Simultaneous causality could be a potential problem. That is, states with high fatality rates might decide to increase taxes to reduce consumption. Expert knowledge is required to determine if this is a problem.

10.5. Let *D*2*i*  1 if *i*  2 and 0 otherwise; *D*3*i*  1 if *i*  3 and 0 otherwise … *Dni*  1 if *i*  *n* and 0 otherwise. Let *B*2*t*  1 if *t*  2 and 0 otherwise; *B*3*t*  1 if *t*  3 and 0 otherwise … *BTt*  1 if *t * *T* and 0 otherwise. Let *β*0 *α*1  *λ*1; *γi*  *αi*  *α*1and *δt*  *λt*  *β*1.

10.7. (a) Average snow fall does not vary over time, and thus will be perfectly collinear with the state fixed effect.

(b) *Snowit* does vary with time, and so this method can be used along with state fixed effects.

10.9. (a)  which has variance  Because *T* is not growing, the variance is not getting small.  is not consistent.

(b) The average in (a) is computed over *T* observations. In this case *T* is small (*T* 4), so the normal approximation from the CLT is not likely to be very good.

10.11 Using the hint, equation (10.22) can be written as



Chapter 11  
Regression with a Binary   
Dependent Variable

11.1. (a) The *t*-statistic for the coefficient on *Experience* is 0.031/0.009  3.44, which is significant at the 1% level.

(b) 

(c) 

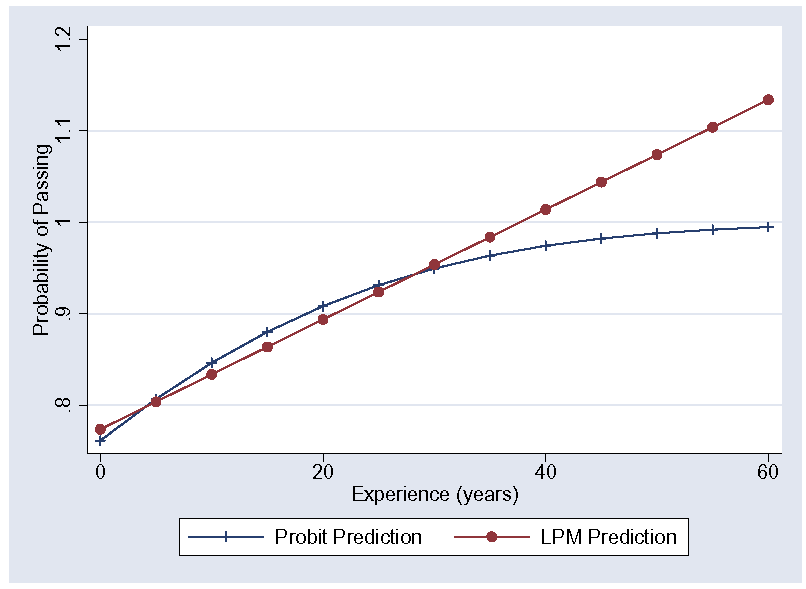
(d)  this is unlikely to be accurate because the sample did not include anyone with more that 40 years of driving experience.

11.3. (a) The *t*-statistic for the coefficient on *Experience* is *t* 0.006/0.002  3, which is significant a the 1% level.

Prob*Matther*  0.774  0.006  10  0.836

Prob*Christopher*  0.774  0.006  0  0.774

(b)



The probabilities are similar except when experience in large ( 40 years). In this case the LPM model produces nonsensical results (probabilities greater than 1.0).

11.5. (a) (0.806  0.041  10  0.174  1  0.015  1  10)  0.814

(b) (0.806  0.041  2  0.174  0  0.015  0  2)  0.813

(c) The *t-*stat on the interaction term is 0.015/0.019  0.79, which is not significant at the 10% level.

11.7. (a) For a black applicant having a P/I ratio of 0.35, the probability that the application will be denied is 

(b) With the P/I ratio reduced to 0.30, the probability of being denied is . The difference in denial probabilities compared to (a) is 4.99 percentage points lower.

(c) For a white applicant having a P/I ratio of 0.35, the probability that the application will be denied is  If the P/I ratio is reduced to 0.30, the probability of being denied is  The difference in denial probabilities is 2.08 percentage points lower.

(d) From the results in parts (a)–(c), we can see that the marginal effect of the P/I ratio on the probability of mortgage denial depends on race. In the logit regression functional form,  
the marginal effect depends on the level of probability which in turn depends on the race  
of the applicant. The coefficient on *black* is statistically significant at the 1% level. The logit and probit results are similar.

11.9. (a) The coefficient on *black* is 0.084, indicating an estimated denial probability that is   
8.4 percentage points higher for the black applicant.

(b) The 95% confidence interval is 0.084 ± 1.96  0.023  [3.89%, 12.91%].

(c) The answer in (a) will be biased if there are omitted variables which are race-related and have impacts on mortgage denial. Such variables would have to be related with race and also be related with the probability of default on the mortgage (which in turn would lead to denial of the mortgage application). Standard measures of default probability (past credit history and employment variables) are included in the regressions shown in Table 9.2, so these omitted variables are unlikely to bias the answer in (a). Other variables such as education, marital status, and occupation may also be related the probability of default, and these variables are omitted from the regression in column. Adding these variables (see columns (4)–(6)) have little effect on the estimated effect of *black* on the probability of mortgage denial.

11.11. (a) This is a censored or truncated regression model (note the dependent variable might be zero).

(b) This is an ordered response model.

(c) This is the discrete choice (or multiple choice) model.

(d) This is a model with count data.

Chapter 12  
Instrumental Variables Regression

12.1. (a) The change in the regressor,  from a $0.50 per pack increase in the retail price is ln(8.00)  ln(7.50)  0.0645. The expected percentage change in cigarette demand is 0.94 ×0.0645 × 100%   .07%. The 95% confidence interval is ( 0.94 ± 1.96 × 0.21) × 0.0645 × 100%  [ 8.72%, 3.41%].

(b) With a 2% reduction in income, the expected percentage change in cigarette demand is  
0.53 × (0.02) × 100%  1.06%.

(c) The regression in column (1) will not provide a reliable answer to the question in (b) when recessions last less than 1 year. The regression in column (1) studies the long-run price and income elasticity. Cigarettes are addictive. The response of demand to an income decrease will be smaller in the short run than in the long run.

(d) The instrumental variable would be too weak (irrelevant) if the *F*-statistic in column (1) was 3.6 instead of 33.6, and we cannot rely on the standard methods for statistical inference. Thus the regression would not provide a reliable answer to the question posed in (a).

12.3. (a) The estimator  is not consistent. Write this as   where  Replacing  with **1,   
as suggested in the question, write this as   The first term on the right hand side of the equation converges to  but the second term converges to something that is non-zero. Thus  is not consistent.

(b) The estimator  is consistent. Using the same notation as in (a), we can write  and this estimator converges in probability to 

12.5. (a) Instrument relevance.  does not enter the population regression for 

(b) *Z* is not a valid instrument. will be perfectly collinear with *W*. (Alternatively, the first stage regression suffers from perfect multicollinearity.)

(c) *W* is perfectly collinear with the constant term.

(d) *Z* is not a valid instrument because it is correlated with the error term.

12.7. (a) Under the null hypothesis of instrument exogeneity, the *J* statistic is distributed as a random variable, with a 1% critical value of 6.63. Thus the statistic is significant, and instrument exogeneity *E*(*ui*|*Z*1*i*, *Z*2*i*)  0 is rejected.

(b) The *J* test suggests that *E*(*ui*|*Z*1*i*, *Z*2*i*) ≠ 0, but doesn’t provide evidence about whether the problem is with *Z*1or *Z*2or both.

12.9. (a) There are other factors that could affect both the choice to serve in the military and annual earnings. One example could be education, although this could be included in the regression as a control variable. Another variable is “ability” which is difficult to measure, and thus difficult to control for in the regression.

(b) The draft was determined by a national lottery so the choice of serving in the military was random. Because it was randomly selected, the lottery number is uncorrelated with individual characteristics that may affect earning and hence the instrument is exogenous. Because it affected the probability of serving in the military, the lottery number is relevant.

Chapter 13  
Experiments and Quasi-Experiments

13.1. For students in kindergarten, the estimated small class treatment effect relative to being in a regular class is an increase of 13.90 points on the test with a standard error 2.45. The 95% confidence interval is 13.90 ± 1.96 × 2.45  [9.098, 18.702].

For students in grade 1, the estimated small class treatment effect relative to being in a regular class is an increase of 29.78 points on the test with a standard error 2.83. The 95% confidence interval is 29.78 ± 1.96 × 2.83  [24.233, 35.327].

For students in grade 2, the estimated small class treatment effect relative to being in a regular class is an increase of 19.39 points on the test with a standard error 2.71. The 95% confidence interval is 19.39 ± 1.96 × 2.71  [14.078, 24.702].

For students in grade 3, the estimated small class treatment effect relative to being in a regular class is an increase of 15.59 points on the test with a standard error 2.40. The 95% confidence interval is 15.59 ± 1.96 × 2.40  [10.886, 20.294].

13.3. (a) The estimated average treatment effect is   1241  1201  40 points.

(b) There would be nonrandom assignment if men (or women) had different probabilities of being assigned to the treatment and control groups. Let *pMen* denote the probability that a male is assigned to the treatment group. Random assignment means *pMen*  0.5. Testing this null hypothesis results in a *t*-statistic of  so that the null of random assignment cannot be rejected at the 10% level. A similar result is found for women.

13.5. (a) This is an example of attrition, which poses a threat to internal validity. After the male athletes leave the experiment, the remaining subjects are representative of a population that excludes male athletes. If the average causal effect for this population is the same as the average causal effect for the population that includes the male athletes, then the attrition does not affect the internal validity of the experiment. On the other hand, if the average causal effect for male athletes differs from the rest of population, internal validity has been compromised.

(b) This is an example of partial compliance which is a threat to internal validity. The local area network is a failure to follow treatment protocol, and this leads to bias in the OLS estimator of the average causal effect.

(c) This poses no threat to internal validity. As stated, the study is focused on the effect of dorm room Internet connections. The treatment is making the connections available in the room; the treatment is not the use of the Internet. Thus, the art majors received the treatment (although they chose not to use the Internet).

(d) As in part (b) this is an example of partial compliance. Failure to follow treatment protocol leads to bias in the OLS estimator.

13.7. From the population regression



we have



By defining *Yi*  *Yi*2  *Yi*1, *Xi*  *Xi*2  *Xi*1 (a binary treatment variable) and *ui*  *vi*2  *vi*1, and using *D*1  0 and *D*2  1, we can rewrite this equation as



which is Equation (13.5) in the case of a single *W* regressor.

13.9. The covariance between  and *Xi* is



Because *Xi* is randomly assigned, *Xi* is distributed independently of **1*i*. The independence means



Thus  can be further simplified:



So



13.11. Following the notation used in Chapter 13, let *π*1*i* denote the coefficient on state sales tax in the “first stage” IV regression, and let *β*1*i* denote cigarette demand elasticity. (In both cases, suppose that income has been controlled for in the analysis.) From (13.11)



where the first equality uses the uses properties of covariances (equation (2.34)), and the second equality uses the definition of the average treatment effect. Evidently, the local average treatment effect will deviate from the average treatment effect when  ≠ 0. As discussed in Section 13.6, this covariance is zero when *β*1*i* or *π*1*i* are constant. This seems likely. But, for the sake of argument, suppose that they are not constant; that is, suppose the demand elasticity differs from state to state (*β*1*i* is not constant) as does the effect of sales taxes on cigarette prices (*π*1*i* is not constant). Are *β*1*i* and *π*1*i* related? Microeconomics suggests that they might be. Recall from your microeconomics class that the lower is the demand elasticity, the larger fraction of a sales tax is passed along to consumers in terms of higher prices. This suggests that *β*1*i* and *π*1*i* are positively related, so that  Because *E*(*π*1*i*)  0, this suggests that the local average treatment effect is greater than the average treatment effect when *β*1*i* varies from state to state.

Chapter 14  
Introduction to Time Series Regression  
and Forecasting

14.1. (a) Since the probability distribution of *Yt* is the same as the probability distribution of *Yt–*1 (this is the definition of stationarity), the means (and all other moments) are the same.

(b) *E*(*Yt*)  **0 **1*E*(*Yt–*1)  *E*(*ut*), but *E*(*ut*)  0 and *E*(*Yt*)  *E*(*Yt–*1). Thus *E*(*Yt*)  **0 **1*E*(*Yt*), and solving for *E*(*Yt*) yields the result.

14.3. (a) To test for a stochastic trend (unit root) in ln(*IP*), the ADF statistic is the *t*-statistic testing the hypothesis that the coefficient on ln(*IPt –* 1) is zero versus the alternative hypothesis that the coefficient on ln(*IPt –* 1) is less than zero. The calculated *t*-statistic is  From Table 14.4, the 10% critical value with a time trend is 3.12. Because 2.5714  3.12, the test does not reject the null hypothesis that ln(*IP*) has a unit autoregressive root at the 10% significance level. That is, the test does not reject the null hypothesis that ln(*IP*) contains a stochastic trend, against the alternative that it is stationary.

(b) The ADF test supports the specification used in Exercise 14.2. The use of first differences in Exercise 14.2 eliminates random walk trend in ln(*IP*).

14.5. (a) 

(b) Using the result in part (a), the conditional mean squared error



with the conditional variance This equation is minimized when the second term equals zero, or when (An alternative is to use the hint, and notice that the result follows immediately from exercise 2.27.)

(c) Applying Equation (2.27), we know the error *ut* is uncorrelated with *ut –* 1 if *E*(*ut*|*ut –* 1)  0. From Equation (14.14) for the AR(*p*) process, we have



a function of *Yt –* 1 and its lagged values. The assumption  means that conditional on *Yt –* 1 and its lagged values, or any functions of *Yt* – 1 and its lagged values, *ut* has mean zero. That is,



Thus *ut* and *ut –* 1 are uncorrelated. A similar argument shows that *ut* and *ut – j* are uncorrelated for all *j* ≥ 1. Thus *ut* is serially uncorrelated.

14.7. (a) From Exercise (14.1) *E*(*Yt*)  2.5  0.7*E*(*Yt* –1)  *E*(*ut*), but *E*(*Yt*)  *E*(*Yt* – 1) (stationarity) and *E*(*ut*)  0, so that *E*(*Yt*)  2.5/(1  0.7). Also, because *Yt*  2.5  0.7*Yt* – 1  *ut*, var(*Yt*)   
0.72var(*Yt* – 1)  var(*ut*) 2 × 0.7 × cov(*Yt* – 1, *ut*). But cov(*Yt* – 1, *ut*)  0 and var(*Yt*)  var(*Yt* – 1) (stationarity), so that var(*Yt*)  9/(1  0.72)  17.647.

(b) The 1st autocovariance is



The 2nd autocovariance is



(c) The 1st autocorrelation is



The 2nd autocorrelation is



(d) The conditional expectation of  given *YT* is

 2.5  0.7YT  2.5  0.7 × 102.3  74.11.

14.9. (a) *E*(*Yt*)  **0 *E*(*et*)  *b*1*E*(*et–*1)  ⋅⋅⋅  *bqE*(*et*–*q*)  **0 [because *E*(*et*)  0 for all values of *t*].

(b) 

where the final equality follows from var(*et*)  for all *t* and cov(*et*, *ei*)  0 for *i≠* *t*.

(c) *Yt*  **0 *et*  *b*1*et–*1 *b*2*et* – 2 ⋅⋅⋅  *bqet* – *q* and *Yt–j*  **0 *et – j*  *b*1*et –* 1 – *j*  *b*2*et –* 2 – *j*  ⋅⋅⋅

 *bqet – q – j* and cov(*Yt*, *Yt* – *j*)   where *b*0  1. Notice that

cov(*et–k*, *et–j–m*)  0 for all terms in the sum.

(d)   and  for *j*  1.

14.11. Write the model as *Yt*  *Yt* – 1 **0 **1(*Yt –* 1 *Yt –* 2)  *ut*. Rearranging yields

Yt  0  (1  1)Yt – 1  1Yt – 2  ut.

Chapter 15  
Estimation of Dynamic Causal Effects

15.1. (a) See the table below. *βi* is the dynamic multiplier. With the 25% oil price jump, the predicted effect on output growth for the *i*th quarter is 25*βi* percentage points.

|  |  |  |  |
| --- | --- | --- | --- |
| Period ahead          (*i*) | Dynamic multiplier (*βi*) | Predicted effect on output growth (25*βi*) | 95% confidence interval 25 × [*βi* ±1.96SE (*βi*)] |
| 0 | 0.055 | 1.375 | [4.021, 1.271] |
| 1 | 0.026 | 0.65 | [3.443, 2.143] |
| 2 | 0.031 | 0.775 | [3.127, 1.577] |
| 3 | 0.109 | 2.725 | [4.783, 0.667] |
| 4 | 0.128 | 3.2 | [5.797, 0.603] |
| 5 | 0.008 | 0.2 | [1.025, 1.425] |
| 6 | 0.025 | 0.625 | [1.727, 2.977] |
| 7 | 0.019 | 0.475 | [2.386, 1.436] |
| 8 | 0.067 | 1.675 | [0.015, 0.149] |

(b) The 95% confidence interval for the predicted effect on output growth for the *i*th quarter from the 25% oil price jump is 25 × [*βi* ± 1.96SE (*βi*)] percentage points. The confidence interval is reported in the table in (a).

(c) The predicted cumulative change in GDP growth over eight quarters is

25 × (0.055  0.026  0.031  0.109  0.128  0.008  0.025  0.019)  8.375%.

(d) The 1% critical value for the *F*-test is 2.407. Since the HAC *F*-statistic 3.49 is larger than the critical value, we reject the null hypothesis that all the coefficients are zero at the 1% level.

15.3. The dynamic causal effects are for experiment A. The regression in exercise 15.1 does not control for interest rates, so that interest rates are assumed to evolve in their “normal pattern” given changes in oil prices.

15.5. Substituting



into Equation (15.4), we have



Comparing the above equation to Equation (15.7), we see *δ*0  *β*0, *δ*1  *β*1, *δ*2  *β*1  *β*2,  
*δ*3  *β*1  *β*2 *β*3,…, and *δr*  1  *β*1  *β*2 ⋅⋅⋅ *βr*  *βr*  1.

15.7. Write 

(a) Because  for all *i* and *t*, *E*(*ui*|*Xt*)  0 for all *i* and *t*, so that *Xt* is strictly exogenous.

(b) Because for *j* ≥ 0, *Xt* is exogenous. However *E*(*ut+*1|)   so that *Xt* is not strictly exogenous.

15.9. (a) This follows from the material around equation (3.2).

(b) Quasi-differencing the equation yields *Yt* – *φ*1*Yt* – 1 (1  *φ*1)*β*0  and the GLS estimator of (1  *φ*1)*β*0is the mean of *Yt* – *φ*1*Yt* – 1  . Dividing by (1*φ*1) yields the GLS estimator of *β*0.

(c) This is a rearrangement of the result in (b).

(d) Write  so that  and the variance is seen to be proportional to 

Chapter 16  
Additional Topics in Time  
Series Regression

16.1. *Yt* follows a stationary AR(1) model,  The mean of *Yt* is   
and 

(a) The *h*-period ahead forecast of  is



(b) Substituting the result from part (a) into *Xt* gives



16.3. *ut* follows the ARCH process with mean *E* (*ut*)  0 and variance 

(a) For the specified ARCH process, *ut* has the conditional mean  and the conditional variance.



The unconditional mean of *ut* is *E* (*ut*)  0, and the unconditional variance of *ut* is



The last equation has used the fact that  which follows because *E* (*ut*)  0. Because of the stationarity, var(*ut*–1)  var(*ut*). Thus, var(*ut*)  1.0  0.5var(*ut*) which implies 

(b) When  The standard deviation of *ut* is *σt*  1.01. Thus



When *ut*–1  2.0,  The standard deviation of *ut* is *σt*  1.732. Thus



16.5. Because 



So



Note that  because *Y*0  0. Thus:



16.7.  Following the hint, the numerator is the same expression as (16.21) (shifted forward in time 1 period), so that  The denominator is  by the law of large numbers. The result follows directly.

16.9. (a) From the law of iterated expectations



where the last line uses stationarity of *u*. Solving for  gives the required result.

(b) As in (a)



so that .

(c) This follows from (b) and the restriction that > 0.

(d) As in (a)



(e) This follows from (d) and the restriction that 

Chapter 17  
The Theory of Linear Regression  
with One Regressor

17.1. (a) Suppose there are *n* observations. Let *b*1 be an arbitrary estimator of *β*1. Given the estimator *b*1, the sum of squared errors for the given regression model is



 the restricted least squares estimator of *β*1, minimizes the sum of squared errors. That is,  satisfies the first order condition for the minimization which requires the differential of the sum of squared errors with respect to *b*1 equals zero:



Solving for *b*1 from the first order condition leads to the restricted least squares estimator



(b) We show first that  is unbiased. We can represent the restricted least squares estimator  in terms of the regressors and errors:



Thus



where the second equality follows by using the law of iterated expectations, and the third equality follows from



because the observations are i.i.d. and *E*(*ui*|*Xi*)  0. (Note, *E*(*ui*|*X*1,…, *Xn*)  *E*(*ui*|*Xi*) because the observations are i.i.d.

Under assumptions 13 of Key Concept 17.1, is asymptotically normally distributed. The large sample normal approximation to the limiting distribution of  follows from considering



Consider first the numerator which is the sample average of *vi*  *Xiui*. By assumption 1 of Key Concept 17.1, *vi* has mean zero:  By assumption 2, *vi* is i.i.d. By assumption 3, var(*vi*) is finite. Let  Using the central limit theorem, the sample average



or



For the denominator,  is i.i.d. with finite second variance (because *X* has a finite fourth moment), so that by the law of large numbers



Combining the results on the numerator and the denominator and applying Slutsky’s theorem lead to



(c)  is a linear estimator:



The weight *ai* (*i*  1,…, *n*) depends on *X*1,…, *Xn* but not on *Y*1,…, *Yn*.

Thus



 is conditionally unbiased because



The final equality used the fact that



because the observations are i.i.d. and E (*ui*|*Xi*)  0.

(d) The conditional variance of  given *X*1,…, *Xn*, is



(e) The conditional variance of the OLS estimator  is



Since



the OLS estimator has a larger conditional variance:   
The restricted least squares estimator  is more efficient.

(f) Under assumption 5 of Key Concept 17.1, conditional on *X*1,…, *Xn*,  is normally distributed since it is a weighted average of normally distributed variables *ui*:



Using the conditional mean and conditional variance of  derived in parts (c) and (d) respectively, the sampling distribution of , conditional on *X*1,…, *Xn*, is



(g) The estimator



The conditional variance is



The difference in the conditional variance of  is



In order to prove  we need to show



or equivalently



This inequality comes directly by applying the Cauchy-Schwartz inequality



which implies



That is 

Note: because  is linear and conditionally unbiased, the result  follows directly from the Gauss-Markov theorem.

17.3. (a) Using Equation (17.19), we have



by defining *vi*  (*Xi*  *μX*)*ui*.

(b) The random variables *u*1,…, *un* are i.i.d. with mean *μu*  0 and variance  By the central limit theorem,



The law of large numbers implies  By the consistency of sample variance, converges in probability to population variance, var(*Xi*), which is finite and non-zero. The result then follows from Slutsky’s theorem.

(c) The random variable *vi*  (*Xi*  *μX*) *ui* has finite variance:



The inequality follows by applying the Cauchy-Schwartz inequality, and the second inequality follows because of the finite fourth moments for (*Xi*, *ui*). The finite variance along with the fact that *vi* has mean zero (by assumption 1 of Key Concept 15.1) and *vi* is i.i.d. (by assumption 2) implies that the sample average  satisfies the requirements of the central limit theorem. Thus,



satisfies the central limit theorem.

(d) Applying the central limit theorem, we have



Because the sample variance is a consistent estimator of the population variance, we have



Using Slutsky’s theorem,



or equivalently



Thus



since the second term for  converges in probability to zero as shown in part (b).

17.5. Because *E*(*W*4)  [*E*(*W*2)]2  var(*W*2), [*E*(*W*2)]2 ≤ *E* (*W*4)  ∞. Thus *E*(*W*2) < ∞.

17.7. (a) The joint probability distribution function of *ui*, *uj*, *Xi*, *Xj* is *f* (*ui*, *uj*, *Xi*, *Xj*). The conditional probability distribution function of *ui* and *Xi* given *uj* and *Xj* is *f* (*ui*, *Xi*|*uj*, *Xj*). Since *ui*, *Xi*, *i*  1,…, *n* are i.i.d., *f* (*ui*, *Xi*|*uj*, *Xj*)  *f* (*ui*, *Xi*). By definition of the conditional probability distribution function, we have



(b) The conditional probability distribution function of *ui* and *uj* given *Xi* and *Xj* equals



The first and third equalities used the definition of the conditional probability distribution function. The second equality used the conclusion the from part (a) and the independence between *Xi* and *Xj*. Substituting



into the definition of the conditional expectation, we have



(c) Let *Q*  (*X*1, *X*2,…, *Xi* – 1, *Xi* + 1,…, *Xn*), so that *f* (*ui*|*X*1,…, *Xn*)  *f* (*ui*|*Xi*, *Q*). Write



where the first equality uses the definition of the conditional density, the second uses the fact that (*ui*, *Xi*) and *Q* are independent, and the final equality uses the definition of the conditional density. The result then follows directly.

(d) An argument like that used in (c) implies



and the result then follows from part (b).

17.9. We need to prove



Using the identity 



The definition of  implies



Substituting this into the expression for  yields a series of terms each of which can be written as *anbn* where  and  where *r* and *s* are integers. For example,  and so forth. The result then follows from Slutksy’s theorem if  where *d* is a finite constant. Let  and note that *wi* is i.i.d. The law of large numbers can then be used for the desired result if  There are two cases that need to be addressed. In the first, both *r* and *s* are non-zero. In this case write



and this term is finite if *r* and *s* are less than 2. Inspection of the terms shows that this is true. In the second case, either *r*  0 or *s*  0. In this case the result follows directly if the non-zero exponent (*r* or *s*) is less than 4. Inspection of the terms shows that this is true.

17.11. Note: in early printing of the third edition there was a typographical error in the expression for *μY*|*X*. The correct expression is.

(a) Using the hint and equation (17.38)



Simplifying yields the desired expression.

(b) The result follows by noting that *fY*|*X*=*x*(*y*) is a normal density (see equation (17.36)) with *μ*  *μT*|*X* and *σ*2   .

(c) Let *b*  *σXY*/ and *a*  *μY* −*bμX*.

17.13 (a) The answer is provided by equation (13.10) and the discussion following the equation. The result was also shown in Exercise 13.10, and the approach used in the exercise is discussed in part (b).

(b) Write the regression model as *Yi*  *β*0  *β*1*Xi*  *vi*, where *β*0  *E*(*β*0*i*), *β*1  *E*(*β*1*i*), and *vi*  *ui*   
(*β*0*i*  *β*0)  (*β*1*i*  *β*1)*Xi*. Notice that

E(vi|Xi)  E(ui|Xi)  E(β0i  β0|Xi)  XiE(β1i − β1|Xi)  0

because *β*0*i* and *β*1*i* are independent of *Xi*. Because *E*(*vi* | *Xi*) = 0, the OLS regression of *Yi* on *Xi* will provide consistent estimates of *β*0  *E*(*β*0*i*) and *β*1  *E*(*β*1*i*). Recall that the weighted least squares estimator is the OLS estimator of *Yi*/*σi* onto 1/*σi* and *Xi*/*σi* , where . Write this regression as

.

This regression has two regressors, 1/*σi* and *Xi*/*σi*. Because these regressors depend only on   
*Xi*, *E*(*vi*|*Xi*)  0 implies that *E*(*vi*/*σi* | (1/*σi*), *Xi*/*σi*)  0. Thus, weighted least squares provides a consistent estimator of *β*0  *E*(*β*0*i*) and *β*1  *E*(*β*1*i*).

Chapter 18  
The Theory of Multiple Regression

18.1. (a) The regression in the matrix form is

**Y**  **X**  **U**

with





(b) The null hypothesis is H0: ***R*** ** ***r*** versus H1: **R** ≠ **r,** with



The heteroskedasticity-robust *F*-statistic testing the null hypothesis is



With *q*  1. Under the null hypothesis,

.

We reject the null hypothesis if the calculated *F*-statistic is larger than the critical value of the  distribution at a given significance level.

18.3. (a) 

where the second equality uses the fact that *Q* is a scalar and the third equality uses the fact that *μQ*  **c**′***μ*w**.

(b) Because the covariance matrix  is positive definite, we have  for every non-zero vector from the definition. Thus, var(*Q*) > 0. Both the vector **c** and the matrix  are finite, so var(*Q*)   is also finite. Thus, 0 < var(*Q*) < ∞.

18.5. **PX**  **X** (**X′X**)1**X′**, **MX**  **I***n*  **PX**.

(a) **PX** is idempotent because

***PXPX*** ** ***X****(****X****′****X****)1* ***X****′****X****(****X****′****X****)1* ***X****′* *********X****(****X****′****X****)1****X****′ * ***PX****.*

**MX** is idempotent because



**PXMX ** 0*nxn* because



(b) Because  we have



which is Equation (18.27). The residual vector is



We know that **MXX** is orthogonal to the columns of **X**:

**MXX**  (**In**  **PX**) **X**  **X**  **PXX**  **X** **X** (**X′X**)1 **X′X** **X**  **X**  **0**

so the residual vector can be further written as



which is Equation (18.28).

18.7. (a) We write the regression model, *Yi*  *β*1*Xi*  *β*2*Wi*  *ui*, in the matrix form as

**Y**  **Xβ**  **Wγ**  **U**

with





The OLS estimator is



By the law of large numbers    (because *X* and *W* are independent with means of zero);  (because *X* and *u* are independent with means of zero);  Thus



(b) From the answer to (a)  if *E*(*Wu*) is nonzero.

(c) Consider the population linear regression *ui* onto *Wi*:

ui  λWi  ai

where λ  E(Wu)/E(W2). In this population regression, by construction, E(aW)  0. Using this equation for ui rewrite the equation to be estimated as



where  A calculation like that used in part (a) can be used to show that



where *S*1 is distributed  Thus by Slutsky’s theorem



Now consider the regression that omits *W*, which can be written as:



where *di*  *Wiθ*  *ai*. Calculations like those used above imply that



Since  the asymptotic variance of  is never smaller than the asymptotic variance of 

18.9. (a) 

The last equality has used the orthogonality **MWW  0**. Thus



(b) Using **MW**  **I***n*  **PW** and **PW**  **W**(**W**′**W**)1**W**′ we can get



First consider  The (*j*, *l*) element of this matrix is  By Assumption (ii), **X***i* is i.i.d., so *XjiXli* is i.i.d. By Assumption (iii) each element of **X***i* has four moments, so by the Cauchy-Schwarz inequality *XjiXli* has two moments:



Because *XjiXli* is i.i.d. with two moments,  obeys the law of large numbers, so



This is true for all the elements of *n*1 **X**′**X**, so



Applying the same reasoning and using Assumption (ii) that (**X***i*, **W***i*, *Yi*) are i.i.d. and Assumption (iii) that (**X***i*, **W***i*, *ui*) have four moments, we have



and



From Assumption (iii) we know  are all finite non-zero, Slutsky’s theorem implies



which is finite and invertible.

(c) The conditional expectation



The second equality used Assumption (ii) that are i.i.d., and the third equality applied the conditional mean independence assumption (i).

(d) In the limit



because 

(e)  converges in probability to a finite invertible matrix, and  converges in probability to a zero vector. Applying Slutsky’s theorem,



This implies



18.11. (a) Using the hint ***C*** [***Q***1 ***Q***2], where ***Q***′***Q*** ***I***. The result follows with ***A******Q***1.

(b) ***W******A***′***V***  N(***A***′***0***, ***A***′***I****n****A***) and the result follows immediately.

(c) ***V***′***CV******V***′***AA***′***V***(***A***′***V***)′(***A***′***V***)  ***W****’****W*** and the result follows from (b).

18.13. (a) This follows from the definition of the Lagrangian.

(b) The first order conditions are

**(\*) X′(YX)  R′λ  0**

and

**(\*\*) R  r  0**

Solving (\*) yields

(\*\*\*)    (**X**′**X**)–1**R**′**λ**.

Multiplying by ***R*** and using (\*\*) yields ***r*** ***R******R***(***X***′***X***)–1***R***′***λ***, so that

**λ  [R(X′X)–1R′]1(R  r).**

Substituting this into (\*\*\*) yields the result.

(c) Using the result in (b), ***Y*** **** ***X*** **** **(Y**  ***X*)**  ***X*(*X***′***X*)**–**1*R***′**[** ***R***(***X***′***X*)**–**1*R***′]–**1**(***R*** so that

**(Y  X)′(Y  X)  (Y X)′(Y  X)  (R′[R(X′X)–1R′]–1(R  r)**

** 2(Y  X)′ X(X′X)–1R′[R(X′X)–1R′]–1(R  r).**

But **(*Y***  ***X*)**′***X******* **0**, so the last term vanishes, and the result follows.

(d) The result in (c) shows that (***R***  ***r***)′[***R***(***X***′***X***)–1***R***′]–1(***R***  ***r***)  *SSRRestricted*  *SSRUnrestricted*. Also  *SSRUnrestricted*/(*n* **** *kUnrestricted* – 1), and the result follows immediately.

18.15. (a) This follows from exercise (18.6).

(b) , so that



(c)  where  are i.i.d. with mean  and finite variance (because *Xit* has finite fourth moments). The result then follows from the law of large numbers.

(d) This follows the Central limit theorem.

(e) This follows from Slutsky’s theorem.

(f)  are i.i.d., and the result follows from the law of large numbers.

(g) Let . Then



and 

Because , the result follows from (a)  and (b)  Both (a) and (b) follow from the law of large numbers; both   
(a) and (b) are averages of i.i.d. random variables. Completing the proof requires verifying that  has two finite moments and  has two finite moments. These in turn follow from 8-moment assumptions for (*Xit*, *uit*) and the Cauchy-Schwartz inequality. Alternatively, a “strong” law of large numbers can be used to show the result with finite fourth moments.

18.17 The results follow from the hints and matrix multiplication and addition.