I'm not robot

reCAPTCHA

Continue

Wooldridge introductory econometrics 7th edition solutions

This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, i CONTENTS PREFACE iii SUGGESTED COURSE OUTLINES iv Chapter 1 The Nature of Econometrics and Economic Data 1 Chapter 2 The Simple Regression Model 5 Chapter 3 Multiple Regression Analysis: Estimation 15 Chapter 4 Multiple Regression Analysis: Inference 28 Chapter 5 Multiple Regression Analysis: OLS Asymptotics 41 Chapter 6 Multiple Regression Analysis: Further Issues 46 Chapter 7 Multiple Regression Analysis with Qualitative 62 Information: Binary (or Dummy) Variables Chapter 8 Heteroskedasticity 79 Chapter 9 More on Specification and Data Problems 91 Chapter 10 Basic Regression Analysis with Time Series Data 102 Chapter 11 Further Issues in Using OLS with Time Series Data 114 Chapter 12 Serial Correlation and Heteroskedasticity in 127 Time Series Regressions Chapter 13 Pooling Cross Sections Across Time. Simple 139 Panel Data Methods Chapter 14 Advanced Panel Data Methods 154 Chapter 15 Instrumental Variables Estimation and Two Stage 167 Least Squares Chapter 16 Simultaneous Equations Models 183 Chapter 17 Limited Dependent Variable Models and Sample 196 Selection Corrections This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, Chapter 18 Advanced Time Series Topics ii Chapter 19 Carrying Out an Empirical Project Appendix A Basic Mathematical Tools Appendix B Fundamentals of Probability Appendix C Fundamentals of Mathematical Statistics Appendix D Summary of Matrix Algebra Appendix E The Linear Regression Model in Matrix Form This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, iv SUGGESTED COURSE OUTLINES For an introductory, one-semester course, I like to cover most of the material in Chapters 1 through 8 and Chapters 10 through 12, as well as parts of Chapter 9 (but mostly through examples). I do not typically cover all sections or subsections within each chapter. Under the chapter headings listed below, I provide some comments on the material I find most relevant for a first-semester course. An alternative course ignores time series applications altogether, while delving into some of the more advanced methods that are particularly useful for policy analysis. This would consist of Chapters 1 through 8, much of Chapter 9, and the first four sections of Chapter 13. Chapter 9 discusses the practically important topics of proxy variables, measurement error, outlying observations, and stratified sampling. In addition, I have written a more careful description of the method of least absolute deviations, including a discussion of its strengths and weaknesses. Chapter 13 covers, in a straightforward fashion, methods for pooled cross sections (including the so-called "natural experiment" approach) and two-period panel data analysis. The basic cross- sectional treatment of instrumental variables in Chapter 15 is a natural topic for cross-sectional, policy-oriented courses. For an accelerated course, the nonlinear methods used for cross- sectional analysis in Chapter 17 can be covered. I typically do not begin with a review of basic algebra, probability, and statistics. In my experience, this takes too long and the payoff is minimal. (Students tend to think that they are taking another statistics course, and start to drift.) Instead, when I need a tool (such as the summation or expectations operator), I briefly review the necessary definitions and key properties. Statistical inference is not more difficult to describe in terms of multiple regression than in tests of a population mean, and so I briefly review the principles of statistical inference during multiple regression analysis. Appendices A, B, and C are fairly extensive. When I cover asymptotic properties of OLS, I provide a brief discussion of the main definitions and limit theorems. If students need more than the brief review provided in class, I point them to the appendices. For a master's level course, I include a couple of lectures on the matrix approach to linear regression. This could be integrated into Chapters 3 and 4 or covered after Chapter 4. Again, I do not summarize matrix algebra before proceeding. Instead, the material in Appendix D can be reviewed as it is needed in covering Appendix E. A second semester course, at either the undergraduate or masters level, could begin with some of the material in Chapter 9, particularly with the issues of proxy variables and measurement error. The advanced chapters, starting with Chapter 13, are useful for students who have an interest in policy analysis. The pooled cross section and panel data chapters (Chapters 13 and 14) emphasize how these data sets can be used, in conjunction with econometric methods, for policy evaluation. Chapter 15, which introduces the method of instrumental variables, is also important for policy analysis. Most modern IV applications are used to address the problems of omitted variables (unobserved heterogeneity) or measurement error. I have intentionally separated out the conceptually more difficult topic of simultaneous equations models in Chapter 16. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, v Chapter 17, in particular the material on probit, logit, Tobit, and Poisson regression models, is a good introduction to nonlinear econometric methods. Specialized courses that emphasize applications in labor economics can use the material on sample selection corrections. Duration models are also briefly covered as an example of a censored regression model. Chapter 18 is much different from the other advanced chapters, as it focuses on more advanced or recent developments in time series econometrics. Combined with some of the more advanced topics in Chapter 12, it can serve as the basis for a second semester course in time series topics, including forecasting. Most second semester courses would include an assignment to write an original empirical paper, and Chapter 19 should be helpful in this regard. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, 2SOLUTIONS TO PROBLEMS1.1 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 behavior, such as maximizing 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 that we cover starting in Chapter 2. But we would not be claiming that one variable "causes" the other. They are both choice variables of the student. 1.2 (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. For reasons we will see in Chapter 2, 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 might be more likely to attend schools with smaller class sizes, and affluent children generally score better on standardized 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. 1.3 (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) Firms 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. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, 3(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. SOLUTIONS TO COMPUTER EXERCISESC1.1 (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) To convert 1976 dollars into 2003 dollars, we use the ratio of the CPIs, which is 184 / 56.9 3.23≈. Therefore, the average hourly wage in 2003 dollars is roughly 3.23($5.90) $19.06≈ , which is a reasonable figure. (v) The sample contains 252 women (the number of observations with female = 1) and 274 men. C1.2 (i) There are 1,388 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 1,176 women who did not smoke. Reporting just the average masks the fact that almost 85 percent 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 1,176 zeros. (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. (v) The average and standard deviation of faminc are about 29.027 and 18.739, respectively, but faminc is measured in thousands of dollars. So, in dollars, the average and standard deviation are $29,027 and $18,739. C1.3 (i) The largest is 100, the smallest is 0. (ii) 38 out of 1,823, or about 2.1 percent of the sample. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, 5CHAPTER 2 TEACHING NOTES This is the chapter where I expect students to follow most, if not all, of the algebraic derivations. In class I like to derive at least the unbiasedness of the OLS slope coefficient, and usually I derive the variance. At a minimum, I talk about the factors affecting the variance. To simplify the notation, after I emphasize the assumptions in the population model, and assume random sampling, I just condition on the values of the explanatory variables in the sample. Technically, this is justified by random sampling because, for example, E(ui|x 1 ,x 2 ,...,xn) = E(ui|xi) by independent sampling. I find that students are able to focus on the key assumption SLR.4 and subsequently take my word about how conditioning on the independent variables in the sample is harmless. (If you prefer, the appendix to Chapter 3 does the conditioning argument carefully.) Because statistical inference is no more difficult in multiple regression than in simple regression, I postpone inference until Chapter 4. (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, I use relatively few assumptions to derive the unbiasedness of the OLS slope estimator, followed by the formula for its variance. This is because I 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. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, 6SOLUTIONS TO PROBLEMS2.1 In the equation y = β 0 + β 1 x + u, add and subtract α 0 from the right hand side to get y = (α 0 + β 0) + β 1 x + (u − α 0). Call the new error e = u − α 0 , so that E(e) = 0. The new intercept is α 0 + β 0 , but the slope is still β 1. 2.2 (i) Let yi = GPAi, xi = ACTi, and n = 8. Then x= 25.875, y = 3.2125, 1 n i= ∑(xi – x)(yi – y) = 5.8125, and 1 n i= ∑(xi – x) 2 = 56.875. From equation (2.9), we obtain the slope as β̂ 1 = 5.8125/56.875 ≈ .1022, rounded to four places after the decimal. From (2.17), β̂ 0 = y – β̂ 1 x ≈ 3.2125 – (.1022)25.875 ≈ .5681. So we can write GPAn = .5681 + .1022 ACT n = 8. The intercept does not have a useful interpretation because ACT is not close to zero for the population of interest. If ACT is 5 points higher, GPA n increases by .1022(5) = .511. (ii) The fitted values and residuals — rounded to four decimal places — are given along with the observation number i and GPA in the following table: i GPA GPAn u ̂ 1 2.8 2.7143. 2 3.4 3.0209. 3 3.0 3.2253 –. 4 3.5 3.3275. 5 3.6 3.5319. 6 3.0 3.1231 –. 7 2.7 3.1231 –. 8 3.7 3.6341. You can verify that the residuals, as reported in the table, sum to −.0002, which is pretty close to zero given the inherent rounding error. (iii) When ACT = 20, GPAn= .5681 + .1022(20) ≈ 2.61. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, 8(iii) Families with low incomes do not have much discretion about spending; typically, a low-income family must spend on food, clothing, housing, and other necessities. Higher income people have more discretion, and some might choose more consumption while others more saving. This discretion suggests wider variability in saving among higher income families. 2.6 (i) This derivation is essentially done in equation (2.52), once (1/ SST)x is brought inside the summation (which is valid because SSTx does not depend on i). Then, just define wdii= / SSTx. (ii) Because Cov(,)ββ ̂̂ 11 uu=−E[(β 1)] , we show that the latter is zero. But, from part (i), 11 () 11 E[(̂)] =E nnE( ). ββuwii==iiuuwi uui −=⎡⎤ ⎢⎥ ⎣⎦∑∑ Because the ui are pairwise uncorrelated(they are independent), E(uuii) E(==u n 22 /) σ /n (because E(uuih) 0, = i h≠). Therefore, 22 11 E()(/)(/)0 1. nn n ∑∑ii==wuuii===wiσσn n∑i=wi(iii) The formula for the OLS intercept isβ ̂̂ 0 =−yxβ and, plugging in y=+ +β 01 βxu gives βββ ββ ββ ̂̂ 001 =++−=+−−()x ux u1 0(̂ 11).x (iv) Because β ̂ 1 and u are uncorrelated, 22 2 22 22 01 Var(̂̂) Var( ) Var( ) / (/ SST) / / SST ββ= + =+uxn xnσσx =+σσxx, which is what we wanted to show. (v) Using the hint and substitution gives Var(βσ̂ 0)=+ 22 [SST /(x nx)] / SSTx () ( ) 21 22 2 21 2 11 /SST /SST. nn σσnxiiiixxx nxx −− == =−⎡⎤+=⎢ ⎥⎣⎦∑∑2.7 (i) Yes. If living closer to an incinerator depresses housing prices, then being farther away increases housing prices. (ii) If the city chose to locate the incinerator in an area away from more expensive neighborhoods, then log(dist) is positively correlated with housing quality. This would violate SLR.4, and OLS estimation is biased. (iii) Size of the house, number of bathrooms, size of the lot, age of the home, and quality of the neighborhood (including school quality), are just a handful of factors. As mentioned in part (ii), these could certainly be correlated with dist [and log(dist)]. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, 92.8 (i) We follow the hint, noting that cy 1 = cy 1 (the sample average of cy 1 i is c 1 times the sample average of yi) and cx 2 = cx 2. When we regress c 1 yi on c 2 xi (including an intercept) we use equation (2.19) to obtain the slope: 2211 12 11 1 222 22 2 11 111 1 222 1 ()() ()()() ()()( )̂.()nn ii ii ii nn ii ii n ii i n i i cx cx cy cy cc x x y y cx cx c x x xxyy cc cc xx β β == == = == −− −−==−−−−−·· =−∑∑∑∑∑∑From (2.17), we obtain the intercept as β 0 = (c 1 y) – β 1 (c 2 x) = (c 1 y) – [(c 1 /c 2)β̂ 1](c 2 x) = c 1 (y – β̂ 1 x) = c 1 β̂ 0) because the intercept from regressing yi on xi is (y – β̂ 1 x). (ii) We use the same approach from part (i) along with the fact that ()cy 1 + = c 1 + y and ()cx 2 + = c 2 + x. Therefore, ()cy cy 11 +−+i () = (c 1 + yi) – (c 1 + y) = yi – y and (c 2 + xi) – ()cx 2 + = xi – x. So c 1 and c 2 entirely drop out of the slope formula for the regression of (c 1 + yi) on (c 2 + xi), and β 1 = β̂ 1. The intercept is β 0 = ()cy 1 + – β 1 ()cx 2 + = (c 1 + y) – β̂ 1 (c 2 + x) = (yx−β̂ 1) + c 1 – c 2 β̂ 1 = β̂ 0 + c 1 – c 2 β̂ 1 , which is what we wanted to show. (iii) We can simply apply part (ii) because log(cy 11 ii) log( ) log( )= c+ y. In other words, replace c 1 with log(c 1), yi with log(yi), and set c 2 = 0. (iv) Again, we can apply part (ii) with c 1 = 0 and replacing c 2 with log(c 2) and xi with log(xi). If β̂̂ 01 and β are the original intercept and slope, then β 11 =β̂ and β 00 =−ββ̂̂log( )c 21. 2.9 (i) The intercept implies that when inc = 0, cons is predicted to be negative $124.84. This, of course, cannot be true, and reflects that fact that this consumption function might be a poor predictor of consumption at very low-income levels. On the other hand, on an annual basis, $124.84 is not so far from zero. (ii) Just plug 30,000 into the equation: consn = –124.84 + .853(30,000) = 25,465.16 dollars. (iii) The MPC and the APC are shown in the following graph. Even though the intercept is negative, the smallest APC in the sample is positive. The graph starts at an annual income level of $1,000 (in 1970 dollars). This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, 11Conditional on the xi, we have E(β 1) = β 0 1 n i i x⎛⎞ ⎜⎟ ⎝⎠∑ /2 1 n i i x⎛⎞ ⎜⎟ ⎝⎠∑ + β 1because E(ui) = 0 for all i. Therefore, the bias in β 1 is given by the first term in this equation. This bias is obviously zero when β 0 = 0. It is also zero when 1 n i i x∑ = 0, which is the same asx = 0. In the latter case, regression through the origin is identical to regression with an intercept. (ii) From the last expression for β 1 in part (i) we have, conditional on the xi, Var(β 1) = 2 2 1 n i i x − = ⎛⎞ ⎜⎟ ⎝⎠∑ Var1 n ii i xu⎛⎞ ⎜⎟ ⎝⎠∑ =2 2 1 n i i x − = ⎛⎞ ⎜⎟ ⎝⎠∑2 1 Var( ) n ii i x u⎛⎞ ⎜⎟ ⎝⎠∑=2 2 1 n i i x − = ⎛⎞ ⎜⎟ ⎝⎠∑22 1 n i i σ x = ⎛⎞ ⎜⎟ ⎝⎠∑ =σ 2 / 2 1 n i i x⎛⎞ ⎜⎟ ⎝⎠∑.(iii) From (2.57), Var(β̂ 1) = σ 2 / 2 1 ()n i i x x⎛⎞ ⎜⎟ −⎝⎠∑. From the hint,2 1 n i i x∑ ≥2 1 ()n i i x x∑ − , and soVar(β 1) ≤ Var(β̂ 1). A more direct way to see this is to write 2 1 ()n i i x x∑ − = 221 ()n i i x nx∑ − , whichis less than 2 1 n i i x∑ unless x = 0.(iv) For a given sample size, the bias in β 1 increases as x increases (holding the sum of the 2 xi fixed). But as x increases, the variance of 1 β̂increases relative to Var(β 1). The bias in β 1 is also small when β 0 is small. Therefore, whether we prefer β 1 or β̂ 1 on a mean squared error basis depends on the sizes of β 0 , x, and n (in addition to the size of 2 1 n i i x∑).2.11 (i) When cigs = 0, predicted birth weight is 119.77 ounces. When cigs = 20, bwghtn = 109.49. This is about an 8.6% drop. (ii) Not necessarily. There are many other factors that can affect birth weight, particularly overall health of the mother and quality of prenatal care. These could be correlated with cigarette smoking during birth. Also, something such as caffeine consumption can affect birth weight, and might also be correlated with cigarette smoking. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, 12(iii) If we want a predicted bwght of 125, then cigs = (125 – 119.77)/(–.524) ≈–10.18, or about –10 cigarettes! This is nonsense, of course, and it shows what happens when we are trying to predict something as complicated as birth weight with only a single explanatory variable. The largest predicted birth weight is necessarily 119.77. Yet almost 700 of the births in the sample had a birth weight higher than 119.77. (iv) 1,176 out of 1,388 women did not smoke while pregnant, or about 84.7%. Because we are using only cigs to explain birth weight, we have only one predicted birth weight at cigs = 0. The predicted birth weight is necessarily roughly in the middle of the observed birth weights at cigs = 0, and so we will under predict high birth rates. SOLUTIONS TO COMPUTER EXERCISESC2.1 (i) The average prate is about 87.36 and the average mrate is about .732. (ii) The estimated equation is nprate= 83.05 + 5.86 mrate n = 1,534, R 2 = .075. (iii) The intercept implies that, even if mrate = 0, the predicted participation rate is 83. percent. The coefficient on mrate implies that a one-dollar increase in the match rate – a fairly large increase – is estimated to increase prate by 5.86 percentage points. This assumes, of course, that this change prate is possible (if, say, prate is already at 98, this interpretation makes no sense). (iv) If we plug mrate = 3.5 into the equation we get pratê = 83.05 + 5.86(3.5) = 103.59. This is impossible, as we can have at most a 100 percent participation rate. This illustrates that, especially when dependent variables are bounded, a simple regression model can give strange predictions for extreme values of the independent variable. (In the sample of 1,534 firms, only 34 have mrate ≥ 3.5.) (v) mrate explains about 7.5% of the variation in prate. This is not much, and suggests that many other factors influence 401(k) plan participation rates. C2.2 (i) Average salary is about 865.864, which means $865,864 because salary is in thousands of dollars. Average ceoten is about 7.95. (ii) There are five CEOs with ceoten = 0. The longest tenure is 37 years. (iii) The estimated equation is log(nsalary)= 6.51 + .0097 ceoten n = 177, R 2 = .013. We obtain the approximate percentage change in salary given Δceoten = 1 by multiplying the coefficient on ceoten by 100, 100(.0097) = .97%. Therefore, one more year as CEO is predicted to increase salary by almost 1%. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, 14The estimated elasticity of rd with respect to sales is 1.076, which is just above one. A one percent increase in sales is estimated to increase rd by about 1.08%. C2.6 (i) It seems plausible that another dollar of spending has a larger effect for low-spending schools than for high-spending schools. At low-spending schools, more money can go toward purchasing more books, computers, and for hiring better qualified teachers. At high levels of spending, we would expend little, if any, effect because the high-spending schools already have high-quality teachers, nice facilities, plenty of books, and so on. (ii) If we take changes, as usual, we obtain Δ=math 10 β 11 Δ ≈ Δlog(expend) (β/100)(% expend), just as in the second row of Table 2.3. So, if %1Δexpend= 0,Δmath 10 =β 1 /10. (iii) The regression results are n 2 10 69.34 11.16 log( ) 408,. math expend nR =− +==(iv) If expend increases by 10 percent, mathn 10 increases by about 1.1 percentage points. This is not a huge effect, but it is not trivial for low-spending schools, where a 10 percent increase in spending might be a fairly small dollar amount. (v) In this data set, the largest value of math10 is 66.7, which is not especially close to 100. In fact, the largest fitted values is only about 30.2. C2.7 (i) The average gift is about 7.44 Dutch guilders. Out of 4,268 respondents, 2,561 did not give a gift, or about 60 percent. (ii) The average mailings per year is about 2.05. The minimum value is .25 (which presumably means that someone has been on the mailing list for at least four years) and the maximum value is 3.5. (iii) The estimated equation is m 2 2.01 2.4,268,.gift mailsyear nR =+==(iv) The slope coefficient from part (iii) means that each mailing per year is associated with – perhaps even "causes" – an estimated 2.65 additional guilders, on average. Therefore, if each mailing costs one guilder, the expected profit from each mailing is estimated to be 1.65 guilders. This is only the average, however. Some mailings generate no contributions, or a contribution less than the mailing cost; other mailings generated much more than the mailing cost. (v) Because the smallest mailsyear in the sample is .25, the smallest predicted value of gifts is 2.01 + 2.65(.25) ≈ 2.67. Even if we look at the overall population, where some people have received no mailings, the smallest predicted value is about two. So, with this estimated equation, we never predict zero charitable gifts. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. This may not be resold, copied, 15CHAPTER 3 TEACHING NOTESFor undergraduates, I do not work through most of the derivations in this chapter, at least not in detail. Rather, I focus on interpreting the assumptions, which mostly concern the population. Other than random sampling, the only assumption that involves more than population considerations is the assumption about no perfect collinearity, where the possibility of perfect collinearity in the sample (even if it does not occur in the population) should be touched on. The more important issue is perfect collinearity in the population, but this is fairly easy to dispense with via examples. These come from my experiences with the kinds of model specification issues that beginners have trouble with. The comparison of simple and multiple regression estimates – based on the particular sample at hand, as opposed to their statistical properties – usually makes a strong impression. Sometimes I do not bother with the "partialling out" interpretation of multiple regression. As far as statistical properties, notice how I treat the problem of including an irrelevant variable: no separate derivation is needed, as the result follows form Theorem 3.1. I do like to derive the omitted variable bias in the simple case. This is not much more difficult than showing unbiasedness of OLS in the simple regression case under the first four Gauss- Markov assumptions. It is important to get the students thinking about this problem early on, and before too many additional (unnecessary) assumptions have been introduced. I have intentionally kept the discussion of multicollinearity to a minimum. This partly indicates my bias, but it also reflects reality. It is, of course, very important for students to understand the potential consequences of having highly correlated independent variables. But this is often beyond our control, except that we can ask less of our multiple regression analysis. If two or more explanatory variables are highly correlated in the sample, we should not expect to precisely estimate their ceteris paribus effects in the population. I find extensive treatments of multicollinearity, where one "tests" or somehow "solves" the multicollinearity problem, to be misleading, at best. Even the organization of some texts gives the impression that imperfect multicollinearity is somehow a violation of the Gauss-Markov assumptions: they include multicollinearity in a chapter or part of the book devoted to "violation of the basic assumptions," or something like that. I have noticed that master's students who have had some undergraduate econometrics are often confused on the multicollinearity issue. It is very important that students not confuse multicollinearity among the included explanatory variables in a regression model with the bias caused by omitting an important variable. I do not prove the Gauss-Markov theorem. Instead, I emphasize its implications. Sometimes, and certainly for advanced beginners, I put a special case of Problem 3.12 on a midterm exam, where I make a particular choice for the function g(x). Rather than have the students directly