Chapter 2: The Simple Regression Model Problems
Book Title: Introductory Econometrics
Printed By: Wanwiphang Manachotipong (wanwiphang@econ.tu.ac.th)

Chapter Review

Problems

1. Let *kids* denote the number of children ever born to a woman, and let *educ* denote years of education for the woman. A simple model relating fertility to years of education is

$$kids = \beta_0 + \beta_1 educ + u,$$

where *u* is the unobserved error.

   i. What kinds of factors are contained in *u*? Are these likely to be correlated with level of education?

   ii. Will a simple regression analysis uncover the ceteris paribus effect of education on fertility? Explain.

2. In the simple linear regression model $y = \beta_0 + \beta_1 x + u$, suppose that $\mathrm{E}(u) \neq 0$. Letting $\alpha_0 = \mathrm{E}(u)$, show that the model can always be rewritten with the same slope, but a new intercept and error, where the new error has a zero expected value.

3. The following table contains the *ACT* scores and the *GPA* (grade point average) for eight college students. Grade point average is based on a four-point scale and has been rounded to one digit after the decimal.

Student	*GPA*	*ACT*
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25

Student	*GPA*	*ACT*
7	2.7	25
8	3.7	30

i. Estimate the relationship between *GPA* and *ACT* using OLS; that is, obtain the intercept and slope estimates in the equation

$$\widehat{GPA} = \widehat{\beta}_0 + \widehat{\beta}_1 ACT.$$

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much higher is the *GPA* predicted to be if the *ACT* score is increased by five points?

ii. Compute the fitted values and residuals for each observation, and verify that the residuals (approximately) sum to zero.

iii. What is the predicted value of *GPA* when $ACT = 20$?

iv. How much of the variation in *GPA* for these eight students is explained by *ACT*? Explain.

4. The data set BWGHT contains data on births to women in the United States. Two variables of interest are the dependent variable, infant birth weight in ounces (*bwght*), and an explanatory variable, average number of cigarettes the mother smoked per day during pregnancy (*cigs*). The following simple regression was estimated using data on $n = 1{,}388\ births$:

$$\widehat{bwght} = 119.77 - 0.514\ cigs$$

i. What is the predicted birth weight when $cigs = 0$? What about when $cigs = 20$ (one pack per day)? Comment on the difference.

ii. Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain.

iii. To predict a birth weight of 125 ounces, what would *cigs* have to be? Comment.

iv. The proportion of women in the sample who do not smoke while pregnant is about .85. Does this help reconcile your finding from part (iii)?

1)

i) factor that contained in u is error term, other factor that may effect kids. u is not correlated with education since SLR 4 : $E(u_i | x_i) = 0$ to hold the OLS estimator to be blue

ii) No, Simple Linear Regression only have 1 variable that explain y, in this case kids.

4)

i) predicted $\widehat{bwght}$ when cigs = 0

$\widehat{bwght} = 119.77 - 0.514(0) = 119.77$ ounces

predicted $\widehat{bwght}$ whe cigs = 20

$\widehat{bwght} = 119.77 - 0.514(20) = 109.49$ ounces

In conclusion, when amount of cigarette increase by 1, the birth weight of baby ($\widehat{bwght}$) will decrease by $\hat{\beta}_1$, that is 0.514 ounce

ii) $\frac{\partial \widehat{bwght}}{\partial cigs} = -0.514 = \hat{\beta}_1$

It implies that, when amount of cigarette increases by 1, the birth weight will change by $\hat{\beta}_1$, -0.514 ounce

iii) when $\widehat{bwght}$ = 125 ounces

$125 = 119.77 - 0.514(cigs)$

$cigs = \frac{119.77 - 125}{0.514} = -10.175$ cigarettes

if we want $\widehat{bwght} = 125$, mother have to consume cigarette -10.175 units but it is impossible, so the max $\widehat{bwght}$ is 119.77 ounces

iv) from part iii the answer is impossible, it may because of proportion of sample who do not smoke is too small and the regression is not accurate, we can increase sample who smoke during pregnancy to make regression be more accurate.

Chapter 3: Multiple Regression Analysis: Estimation Problems
Book Title: Introductory Econometrics
Printed By: Wanwiphang Manachotipong (wanwiphang@econ.tu.ac.th)

Chapter Review

Problems

1. Using the data in GPA2 on 4,137 college students, the following equation was estimated by OLS:

$$\widehat{colgpa} = 1.392 - .0135\, hsperc + .00148\, sat$$
$$n = 4{,}137,\ R^2 = .273,$$

where *colgpa* is measured on a four-point scale, *hsperc* is the percentile in the high school graduating class (defined so that, for example, $hsperc = 5$ means the *top* 5% of the class), and *sat* is the combined math and verbal scores on the student achievement test.

   i. Why does it make sense for the coefficient on *hsperc* to be negative?

   ii. What is the predicted college GPA when $hsperc = 20$ and $sat = 1{,}050$?

   iii. Suppose that two high school graduates, A and B, graduated in the same percentile from high school, but Student A's SAT score was 140 points higher (about one standard deviation in the sample). What is the predicted difference in college GPA for these two students? Is the difference large?

   iv. Holding *hsperc* fixed, what difference in SAT scores leads to a predicted *colgpa* difference of .50, or one-half of a grade point? Comment on your answer.

2. The data in WAGE2 on working men was used to estimate the following equation:

$$\widehat{educ} = 10.36 - .094\, sibs + .131\, meduc + .210\, feduc$$
$$n = 722,\ R^2 = .214,$$

where *educ* is years of schooling, *sibs* is number of siblings, *meduc* is mother's years of schooling, and *feduc* is father's years of schooling.

i. Does *sibs* have the expected effect? Explain. Holding *meduc* and *feduc* fixed, by how much does *sibs* have to increase to reduce predicted years of education by one year? (A noninteger answer is acceptable here.)

ii. Discuss the interpretation of the coefficient on *meduc*.

iii. Suppose that Man A has no siblings, and his mother and father each have 12 years of education. Man B has no siblings, and his mother and father each have 16 years of education. What is the predicted difference in years of education between B and A?

3. The following model is a simplified version of the multiple regression model used by Biddle and Hamermesh (1990) to study the tradeoff between time spent sleeping and working and to look at other factors affecting sleep:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + u,$$

where *sleep* and *totwrk* (total work) are measured in minutes per week and *educ* and *age* are measured in years. (See also Computer Exercise C3 in Chapter 2.)

i. If adults trade off sleep for work, what is the sign of β_1?

ii. What signs do you think β_2 and β_3 will have?

iii. Using the data in SLEEP75, the estimated equation is

$$\begin{aligned} \widehat{sleep} &= 3,638.25 - .148\, totwrk - 11.13\, educ + 2.20\, age \\ n &= 706,\ R^2 = .113. \end{aligned}$$

If someone works five more hours per week, by how many minutes is *sleep* predicted to fall? Is this a large tradeoff?

iv. Discuss the sign and magnitude of the estimated coefficient on *educ*.

v. Would you say *totwrk*, *educ*, and *age* explain much of the variation in *sleep*? What other factors might affect the time spent sleeping? Are these likely to be correlated with *totwrk*?

4. The median starting salary for new law school graduates is determined by

$$\log(salary) = \beta_0 + \beta_1 LSAT + \beta_2 GPA + \beta_3 \log(libvol) + \beta_4 \log(cost) + \beta_5 rank + u,$$

where *LSAT* is the median LSAT score for the graduating class, *GPA* is the median college GPA for the class, *libvol* is the number of volumes in the law

1)

i) because hsperc is the percentage position in class, if hsperc = 5 means in a top 5 % of the class, when the percentage is larger means position is lower so, the $\widehat{colgpa}$ will be lower too.

ii) hsperc = 20 → 20% → 0.2 , sat = 1050

$$\widehat{colgpa} = 1.392 - .0135(hsperc) + 0.00148(sat)$$

$$\widehat{colgpa} = 1.392 - 0.135(20) + 0.00148(1050)$$

$$= 1.392 - 0.27 + 1.554 = 2.676$$

iii) $\widehat{colgpa}_A = 1.392 - 0.135(X) + 0.00148(n + 140)$

$\widehat{colgpa}_B = 1.392 - 0.135(X) + 0.00148(n)$

$\Delta \widehat{colgpa}_{A,B} =$ 0.00~~148~~n + 0.00148(140) − 0.~~00~~148n = 0.2072

the different is not large since the coefficient of SAT is small, 0.00148.

constant

iv) $\widehat{colgpa} =$ ~~1.392~~ − .0135(~~hsperc~~) + 0.00148(sat)

$$0.5 = 0.00148(sat)$$

$$sat = \frac{0.5}{0.00148} = 337.8378 \text{ points}$$

2)

i) yes, because it might effect budget

holding feduc, meduc

$-1 = -0.094\ sibs$

$sibs = 10.6383 \approx 11$

to reduce 1 year of education, they might have to have 11 siblings

ii) as 1 year more of mother education (meduc), 0.131 years of child's education increase

iii) A: sibs = 0, meduc = 12, feduc = 12

B: sibs = 0, meduc = 16, feduc = 16

$\widehat{educ}_A = 10.36 + 0.131\ meduc + 0.210\ feduc$

$= 10.36 + 0.131(12) + 0.210(12) = 14.452$

$\widehat{educ}_B = 10.36 + 0.131\ meduc + 0.210\ feduc$

$= 10.36 + 0.131(16) + 0.210(16) = 15.816$

$14.452 - 15.816 = -1.364$

A have less education than B for about 1.4 years