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

   ผล เหตุ

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

   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? income, family income

   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)?

$\bar{y} = 3.2125$ from (i) from table!

$\bar{x} = 25.875$ $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

$n = 8$ $\hat{u} = Y_i - \hat{Y}$

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	y GPA	x ACT	① $x-\bar{x}$	② $y-\bar{y}$	① × ②	$①^2$	$\hat{Y}$	$\hat{u}$
1	2.8	21	-4.875	-0.4125	2.0109375	23.76563	2.7143	0.0857
2	3.4	24	-1.875	0.1875	-0.3515625	3.51625	3.0209	0.3791
3	3.0	26	0.125	-0.2125	-0.0265625	0.01563	3.2253	-0.2253
4	3.5	27	1.125	0.2875	0.3234375	1.26563	3.3275	0.1725
5	3.6	29	3.125	0.3875	1.2109375	9.76563	3.5319	0.0681
6	3.0	25	-0.875	-0.2125	0.1859375	0.76563	3.1231	-0.1231
7	2.7	25	-0.875	-0.5125	0.4484375	0.76563	3.1231	-0.4231
8	3.7	30	4.125	0.4875	2.0109375	17.01563	3.6341	0.0659
	sum				5.8125	56.87566		

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

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

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.

(i) $\hat{\beta}_1 = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$ $\beta_0 = \bar{Y} - \beta_1\bar{X}$

$\beta_1 = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$ $\beta_0 = \bar{y} - \beta_1\bar{x}$

$= \dfrac{5.8125}{56.87566}$ $= 3.2125 - (0.1022)(25.875)$

$= 0.1021966$ $= 0.568075$

≈ 0.1022 ≈ 0.5681

>> regression model can be written as

$\hat{Y} = \beta_0 + \beta_1 X_1$

$\widehat{GPA} = 0.5681 + 0.1022(ACT)$

the intercept does not have a useful interpretion because ACT is not close to zero for the population of interest

if ACT is 5 point higher, Gpa increases by 0.1022 (5) = .511

(ii) fitted values = $\hat{Y}$

and the residual

$\sum_{i=1}^{n} \hat{u}_i = 0.0857 + 0.3791 - 0.2253 \ldots + 0.0659$

$= -0.002$

≈ 0

(iii) Predicted value of GPA when ACT is 20

>> from $\widehat{GPA} = 0.5681 + 0.1022(ACT)$

sub ACT = 20

results in $0.5681 + 0.1022(20) = 2.6121$

R-square! (R^2)

: also known as coefficient determination

: interpret as percentage

$$\boxed{R^2 = 1 - \frac{\text{Explained Variation}}{\text{Total Variation}}}$$

$\dfrac{\text{SS regression}}{\text{SS total}}$

$$R^2 = \left(\frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x^2)][n\Sigma y^2 - (\Sigma y^2)]}} \right)^2$$