Chapter 4 Econ 3210 A

End of Chapter Problem Solutions for Selected Problems

4.1 Suppose that a researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression:

$$\widehat{\text{Test Score}} = 520.4 - 5.82 \times CS, \quad R^2 = 0.08$$
$$SER = 11.5$$

(a) A classroom has 22 students. What is the regression's prediction for that classroom's average test score?

Ans: $\widehat{\text{Test Score}} = 520.4 - (5.82)(22)$
$= 392.36$

(b) Last year, a classroom had 19 students and this year it has 23 students. What is the regression's prediction for the change in the classroom average test score?

Ans. When $CS = 23$:
Then $\widehat{\text{Test Score}} = 520.4 - (5.82)(23)$
$= 520.4 - 133.86$
$= 386.54$

When $CS = 19$:
Then $\widehat{\text{Test Score}} = 520.4 - (5.82)(19)$
$= 520.4 - 110.58$
$= 409.82$

Change/Difference in classroom test scores is
$409.82 - 38.54 = 23.28$

(C) The sample average class size across the 100 classrooms is 21.4. What is the sample average of the test scores across the 100 class rooms?

Ans We are given $\overline{CS} = 21.4$.

We need to find $\overline{TestScore}$

Recall: $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

In this example: $520.4 = \overline{TestScore} - (5.82) \times 21.4$

$$\Rightarrow \overline{TestScore} = 520.4 - (5.82)(21.4)$$
$$= 395.85$$

(d) What is the sample standard deviation of test scores across the 100 classrooms?

Ans: We are looking for $S_Y^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} = \frac{TSS}{n-1}$

Now $\underset{\substack{\downarrow \\ \text{Total sum of squares}}}{TSS} = \underset{\substack{\downarrow \\ \text{Explained sum of squares}}}{ESS} + \underset{\substack{\searrow \\ \text{Residual sum of squares.}}}{RSS}$ —— (1)

We also know that $SER = \sqrt{\frac{SSR}{n-2}}$

$$\text{or } (SER)^2 = \frac{SSR}{n-2}$$

$$\text{or } SSR = (SER)^2 (n-2)$$

Substituting $SER = 11.5$:

$$SSR = (11.5)^2 (100-2) = (11.5)^2 (98) = 12961$$

Using the relationship between R^2 and SSR:

$$R^2 = 1 - \frac{SSR}{TSS}$$

$$\Rightarrow (TSS)(R^2) = TSS - SSR$$

$$\Rightarrow TSS(R^2 - 1) = -SSR$$

$$\Rightarrow TSS = \frac{SSR}{1-R^2}$$

$$\Rightarrow TSS = \frac{12961}{1-0.88} = 14088$$

Sample standard deviation $= S_Y = \sqrt{\frac{TSS}{n-1}}$

$$= \sqrt{\frac{14088}{99}} = 11.9$$

<u>4.2</u> Suppose that a random sample of 200 20 year-old men is selected from a population and these men's height and weight are recorded. A regression of weight on height yields:

$$\widehat{Weight} = -99.41 + 3.94 \times Height$$

$$R^2 = 0.81, \; SER = 10.2$$

where weight is measured in pounds and Height is measured in inches

<u>Notice</u> SER = 10.2. This is a measure of average spread of the observed Y_i around the regression line, measured in units of Y_i. So, the difference between the observed weight and average weight is around 10 pounds on average.

(a) What is the regression's weight prediction for someone who is 70 in. tall? 65 in tall? 74 inches tall?

<u>Ans</u>

$$\widehat{Weight} = -99.41 + 3.94 \times (70) = 176.39$$

$$\widehat{Weight} = -99.41 + 3.94 \times (74) = 192.15$$

$$\widehat{Weight} = -99.41 + (3.94) \times (65) = 156.69.$$

(b) A man has a late growth spurt and grows 1.5 in. over the course of a year. What is the regression's prediction for the increase in this man's weight?

<u>Ans</u> we know that the slope parameter β_1 is equal to $\frac{\Delta Y}{\Delta X}$

$$\Rightarrow 3.94 = \frac{\Delta Weight}{\Delta Height}$$

$$\Rightarrow 3.94 = \frac{\Delta Weight}{1.5}$$

$$\Rightarrow \Delta \widehat{Weight} = (3.94)(15) = 59.1$$

C) Suppose that instead of measuring weight and height in pounds and inches, these variables are measured in centimeters and kilograms. What are the regression estimates from this new centimeter-kilogram regression. (Give all results, estimated coefficients, R^2 and SER)

Ans We know that 1 inch = 2.54 cm.
1 lb = 0.4536 kg.

- The intercept of the equation is measured in terms of units of dependent variable. So we will multiply it by 0.4536
  $\Rightarrow \hat{\beta}_0 = (0.4536)(-99.41) = -45.092$ kg

  The slope parameter measures the change in Y for a unit change in X.
  So $\hat{\beta}_1 = \frac{0.4536}{2.54} = 0.7036$ kg per cm.

  So the new equation is:
  $\widehat{Weight} = (-45.092) + 0.7036$ Height
- The R^2 is units free, so it will not change.
- SER is measured in the units of dependent variable, so the new SER = $(10.2)(0.4536)$
  $= 4.6267$ kg.

4.3 A regression of average weekly earnings (AWE) measured in dollars on age (measured in years) using a random sample of college educated full time workers aged 25–26 yields the following:

$$\widehat{AWE} = 696.7 + (9.6) \times Age, \quad R^2 = 0.023$$
$$SER = 624.1$$

(a) Explain what the co-efficient values of 696.7 and 9.6 mean?

Ans 696.7 is the intercept of the regression line. It determines the overall level of the line.

Chapter 4 Econ 3210 A

End of Chapter Problem Solutions for Selected Problems

4.1 Suppose that a researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression:

$$\widehat{Test\ Score} = 520.4 - 5.82 \times CS, \quad R^2 = 0.08$$
$$SER = 11.5$$

(a) A classroom has 22 students. What is the regression's prediction for that classroom's average test score?

Ans: $\widehat{Test\ Score} = 520.4 - (5.82)(22)$
$= 392.36$

(b) Last year, a classroom had 19 students and this year it has 23 students. What is the regression's prediction for the change in the classroom average test score?

Ans. When $CS = 23$:

Then $\widehat{Test\ Score} = 520.4 - (5.82)(23)$
$= 520.4 - 133.86$
$= 386.54$

When $CS = 19$:

Then $\widehat{Test\ Score} = 520.4 - (5.82)(19)$
$= 520.4 - 110.58$
$= 409.82$

Change/Difference in classroom test scores is

$409.82 - 38.54 = 23.28$

9.6 shows the marginal effect of age on AWE. More precisely, it shows that each additional year of age, is expected to bring about an increase of \$9.6 in average weekly earnings.

(b) What are the units of measurement for the SER?

Ans SER is measured in units of average hourly earnings which are dollars.

(c) The regression R^2 is 0.023. What are the units of measurement for the R^2?

Ans R^2 is a units free measure.

(d) What does the regression predict will be the earnings for a 25 year old worker? For a 45 year old worker?

Ans.
$$\widehat{AWE} = 696.7 + (9.6)(25)$$
$$= \$936.7.$$

$$\widehat{AWE} = 696.7 + (9.6)(45)$$
$$= \$1128.7.$$

(e) Will the regression give reliable predictions for a 99 year old worker? why or why not?

Ans No. 99 year is an outlier. It is far outside the range of sample data (25−65)

f. No. The distribution of earnings is not normal so the errors are also not normally distributed.

g. The average age in this sample is 41.6 years. What is the average value of AWE in the sample.

Ans. We can use the following identity to answer this question:

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

$$\Rightarrow \quad 696.7 = \overline{AWE} - (9.6)(41.6)$$

$$\Rightarrow \quad \overline{AWE} = 696.7 - 399.36$$

$$= 297.34.$$

4.5. A professor decides to run an experiment to measure the effect of time pressure on final exams scores. He gives each of the 400 students in his course, the same final exam, but some students have 90 minutes to complete the exam, while others have 120 minutes. Each student is randomly assigned one of the examination times, based on the flip of a coin. Let Y_i denote the number of points scored on the exam by the ith student ($0 \le Y_i \le 100$), let X_i denote the amount of time that the student has to complete the exam ($X_i = 90$ or 120) and consider the regression model: $Y_i = \beta_0 + \beta_1 X_i + u_i$

(a) Explain what the error term u_i represents. Why will different students have different values of u_i?

Ans u_i represents factors, other than time pressure, that may affect students' performance on the exam such as the time devoted to study, aptitude for the material and so forth. Different students will have different values of u_i. Some will be below $\hat{Y}_i$ and the others below it.

(b) Explain why $E(u_i/X_i) = 0$ for this regression model?

Ans. Because the X_i (the amount of time allocated to complete the exam) has been assigned randomly. Since $E(u_i) = 0$ and since u_i and X_i are independent, $E(u_i/X_i) = E(u_i) = 0$

(C) Are other assumptions in Key concepts 4.3 Satisfied?

Ans. The assumption # 2 in Key concept (X_i, Y_i) are i.i.d is met if this year's students are similar to students of a typical class and they can be viewed as random draws from the population of students that enrol in the class.
Assumption # 3 (Large outliers are unlikely) is also satisfied as $0 \leq Y_i \leq 100$ and $X_i = 90$ or 120.

$$R^2 = 1 - \frac{SSR}{TSS}$$

$$\Rightarrow (TSS)(R^2) = TSS - SSR$$

$$\Rightarrow TSS(R^2 - 1) = -SSR$$

$$\Rightarrow TSS = \frac{SSR}{1 - R^2}$$

$$\Rightarrow TSS = \frac{12961}{1 - 0.88} = 14088$$

Sample standard deviation $= S_Y = \sqrt{\frac{TSS}{n-1}}$

$$= \sqrt{\frac{14088}{99}} = 11.9$$

4.2 Suppose that a random sample of 200 20 year-old men is selected from a population and these men's height and weight are recorded. A regression of weight on height yields:

$$\widehat{Weight} = -99.41 + 3.94 \times Height$$

$$R^2 = 0.81 \; ; \; SER = 10.2$$

where weight is measured in pounds and Height is measured in inches

Notice $SER = 10.2$. This is a measure of average spread of the observed Y_i around the regression line measured in units of Y_i. So, the difference between the observed weight and average weight is around 10 pounds on average.