EE 325 Answer HW 2

1. In Table 1.a. X_i is total microeconomics exam point (total points are 100) and Y_i is GPA of each student.

Table 1.a

Student	Y_i	X_i
1	2.8	63
2	3.4	72
3	3	78
4	3.5	81
5	3.6	87
6	3.0	75
7	2.7	75
8	3.7	90

1.1 Now consider the two-variable model $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID(0,\sigma^2)$ Use OLS to find the estimator of $\beta_1 and \beta_2$. Interpret the regression.

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{17.4375}{511.875} = 0.0341$$

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X} = 3.2125 - 0.0341(77.625) = 0.5681$$

When total microeconomics exam point is equal to zero, student's GPA is 0.5681. When total microeconomics exam point increase 1 point, student's GPA increases approximately 0.0341.

1.2 Find $\hat{Y}_i$ *and* $\hat{u}_i$ and show that $\sum_{i=1}^{n} \hat{u}_i \approx 0$

Construct the table that contain $\hat{Y}_i$ *and* $\hat{u}_i$, you will get $\sum_{i=1}^{n} \hat{u}_i = 8.327 \times 10^{-16}$

1.3 Find $\text{var}(\hat{u}_i), \text{var}(\hat{\beta}_1), and \ \text{var}(\hat{\beta}_2)$

$$\text{var}(\hat{u}_i) = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2} = \frac{0.4347253}{8-2} = 0.0725$$

$$\text{var}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} x_i^2} \hat{\sigma}^2 = \frac{48717(0.0725)}{8(511.875)} = 0.862$$

$$\text{var}\left(\hat{\beta}_2\right) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2} = \frac{0.0725}{511.875} = 0.00014163$$

1.4 Test the hypothesis that total microeconomics exam point has no influence on GPA at $\alpha = 5\%$

$$H_0 : \beta_2 = 0,\ H_1 : \beta_2 \neq 0$$

$$t = \frac{0.0341 - 0}{\sqrt{0.00014163}} = 2.863324$$

$$t_{\frac{0.05}{2},8-2} = 2.447$$

$$t = 2.863324 > t_{\frac{0.05}{2},8-2} = 2.447$$

Reject the null hypothesis. Total microeconomics exam point has influence on GPA at $\alpha = 5\%$

1.5 What percentage of the total variation in Y explained by the regression model?

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{8} \hat{y}_i^2}{\sum_{i=1}^{8} y_i^2} = \frac{0.59402}{1.02875} = 0.5774$$

The total variation in Y explained by the regression model is 57.74%.

1.6 Establish a 95 percent confidence interval for $E(Y \mid X = 77.6)$

$$E\left(Y \middle| X = 77.6\right) = 0.5681 + 0.0341\left(77.6\right) = 3.21$$

$$\Pr\left[\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\frac{\alpha}{2},n-k} se\left(\hat{Y}_0\right) \leq \beta_1 + \beta_2 X_0 \leq \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\frac{\alpha}{2},n-k} se\left(\hat{Y}_0\right)\right] = 1 - \alpha$$

$$\text{var}\left(\hat{Y} \middle| X = 77.6\right) = \sigma^2 \left[\frac{1}{n} + \frac{\left(X - \bar{X}\right)^2}{\sum_{i=1}^{n} x_i^2}\right] = 0.0725\left[\frac{1}{8} + \frac{\left(77.6 - 77.625\right)^2}{511.875}\right] = 0.0091$$

$$se\left(\hat{Y} \middle| X = 77.6\right) = \sqrt{0.0091} = 0.0952$$

$$t_{\frac{0.05}{2},8-2} = 2.447$$

$$\Pr\left[3.21 - 2.447\left(0.0952\right) \leq E\left(Y \middle| X = 77.6\right) \leq 3.21 + 2.447\left(0.0952\right)\right] = 0.95$$

$$\Pr\left[2.98 \leq E\left(Y \middle| X = 77.6\right) \leq 3.44\right] = 0.95$$

1.7 Establish a 95 percent confidence interval for $E(Y \mid X = 100)$. Compare the answer with 1.6 whether the confidence interval is wider or narrower?

$$E\left(Y \middle| X = 100\right) = 0.5681 + 0.0341(100) = 3.98$$

$$\Pr\left[\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\frac{\alpha}{2},n-k} se\left(\hat{Y}_0\right) \le \beta_1 + \beta_2 X_0 \le \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\frac{\alpha}{2},n-k} se\left(\hat{Y}_0\right)\right] = 1 - \alpha$$

$$\operatorname{var}\left(\hat{Y} \middle| X = 100\right) = \sigma^2 \left[\frac{1}{n} + \frac{\left(X - \bar{X}\right)^2}{\sum_{i=1}^{n} x_i^2}\right] = 0.0725\left[\frac{1}{8} + \frac{(100 - 77.625)^2}{511.875}\right] = 0.07997$$

$$se\left(\hat{Y} \middle| X = 100\right) = \sqrt{0.0091} = 0.2828$$

$$t_{\frac{0.05}{2},8-2} = 2.447$$

$$\Pr\left[3.98 - 2.447(0.2828) \le E\left(Y \middle| X = 100\right) \le 3.98 + 2.447(0.2828)\right] = 0.95$$

$$\Pr\left[3.29 \le E\left(Y \middle| X = 77.6\right) \le 4.67\right] = 0.95$$

The confidence interval is wider than 1.6.

2. Data is listed in the table

X_i	Y_i
10	0
12	2
14	5
16	6
18	7
22	10
24	10
26	15
28	16
30	20

2.1 From the simple regression model $Y_i = \beta_1 + \beta_2 X_i + u_i, u_i \sim NIID(0, \sigma^2)$ Find estimators of β_1 and β_2 from the OLS method and interpret the meaning.

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = 0.8955$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = -8.8091$$

2.2 Find the value of $\hat{Y}_i$ and $\hat{u}_i$. Show that $\sum \hat{u}_i \approx 0$

2.3 Plot graph and draw regression line. Does the line pass $(\bar{X}, \bar{Y})$?

$\bar{X} = 20, \ \bar{Y} = 9.1$

2.4 If $X_i = 16$, what is the predicted Y?

$Y_i = -8.8091 + 0.8955(16) = 5.5189$

2.5 Find $\text{var}(\hat{u}_i)$, $\text{var}(\hat{\beta}_1), \text{var}(\hat{\beta}_2)$

$$\text{var}\left(\hat{u}_i\right) = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2} = 1.7614$$

$$\text{var}\left(\hat{\beta}_1\right) = \frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n} x_i^2}\hat{\sigma}^2 = 1.7774$$

$$\text{var}\left(\hat{\beta}_2\right) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} x_i^2} = 0.0040$$

3. Given Y is wages per hour (\$), X is year of schooling (years) from a sample of 528 observations

$$\hat{Y}_t = 0.7437 + 0.6416 X_t$$
$$se = (0.8355)(0.0664)$$
$$r^2 = 0.8944, \hat{\sigma}^2 = 0.8040$$
$$\bar{X} = 12 \ \sum x_i^2 = 2054$$

3.1 Test the hypothesis that year of schooling has a positive influence on wages per hour at $\alpha = 1\%$

$H_0 : \beta_2 \le 0$

$H_1 : \beta_2 > 0$

$$t = \frac{0.6416 - 0}{0.0664} = 9.6625$$

Critical t value is $t_{0.05, df=526} = 2.326$

$$t = \frac{0.6416 - 0}{0.0664} = 9.6625 > t_{0.05, df=526} = 2.326$$

Reject the null hypothesis. Year of schooling has a positive influence on wages per hour at $\alpha = 1\%$

3.2 Interpret the regression.

$\beta_2 = 0.6416$

When year of schooling increases 1 year, wages per hour ($) will increase approximately $ 0.6416.

3.3 If Miss Lily has 8 years of schooling, what is the predicted average on wages per hour ($)?

$$E\left(Y_0 \middle| X_0 = 8\right) = 5.8765$$

$$\text{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right]$$

$$\hat{\sigma}^2 = 0.8040, n = 528, \bar{X} = 12, \sum x_i^2 = 2054, X_0 = 8$$

$$\Pr\left[\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\frac{\alpha}{2},n-k} se\left(\hat{Y}_0\right) \le \beta_1 + \beta_2 X_0 \le \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\frac{\alpha}{2},n-k} se\left(\hat{Y}_0\right) \right] = 1 - \alpha$$

$$\Pr\left(5.6492 \le E\left(Y_0 \middle| X_0 = 8\right) \le 6.1038\right) = 0.95$$

4. Consider the two-variable model $Y_i = \beta_1 + \beta_2 X_i + u_i$ Y is Supply for good (unit: hundred pieces), X is price of good (unit: thousand baht) from a sample of 5 observation

$$\hat{Y}_i = 475.9444 - 0.4579 X_i$$
$$se = (54.3277) \quad (0.0828)$$
$$t = (8.7606) \quad (-5.5294)$$
$$\hat{\sigma}^2 = 2364.7694$$

TSS = 43245.2

RSS= 7094.3080

4.1 Test the hypothesis that β_2 is different from zero at $\alpha = 5\%$

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

$t_{\beta_2} = -5.5294$

$\left| t_{\beta_2} \right| = 5.5294$

$t_{0.025,3} = 3.182$

$\left| t_{\beta_2} \right| > 3.182$ Reject the null hypothesis.

Price of good has influence on supply for good at $\alpha = 5\%$

4.2 Interpret the regression

When price of good is zero, supply for good are 475.9444 hundred pieces. When price of good increases one thousand baht, supply for good decreases 0.4579 hundred pieces.

4.3 Establish a 95 percent confidence interval for β_2 and Test the hypothesis that $\beta_2 = -0.6$ or not.

$H_0 : \beta = -0.6,\ H_1 : \beta \neq -0.6$

$\Pr[\hat{\beta}_2 - t_{\alpha/2} se(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2} se(\hat{\beta}_2)] = 1 - \alpha$

$\Pr[-0.4579 - 3.182(0.0828) \leq \beta_2 \leq -0.4579 + 3.182(0.0828)] = 95$

$\Pr\left(-0.7214 \leq \beta \leq -0.1944\right) = 0.95$

The confidence interval does contain -0.6. We cannot reject the null hypothesis. When price of good increases one thousand baht, supply for good decreases 0.6 hundred pieces.

4.4 What percentage of the total variation in Y explained by the regression model?

R-squared = ESS/TSS or 1-(RSS/TSS)
$R^2 = 0.8360$

4.5 If the unit Supply for good changes from hundred pieces to piece. What is the estimator for β_1 *and* β_2? Interpret the regression

$Y_i^* = w_1 Y_i,\ w_1 = 100$

$\hat{\beta}_1^* = w_1 \hat{\beta}_1 = 47594.44, \hat{\beta}_2^* = \left(\frac{w_1}{1}\right)\hat{\beta}_2 = -45.79$

$\hat{\beta}_1^* = 47594.44$ When price of good is zero, supply for good is 47594.44 pieces.
$\hat{\beta}_2^* = -45.79$ When price of good increases one thousand baht, supply for good decreases 45.79 pieces

4.6 If the unit price of good changes from thousand baht to hundred baht. What is the estimator for β_1 *and* β_2? Interpret the regression

$X_i^* = w_2 X_i,\ w_2 = 10$

$\hat{\beta}_1^* = \hat{\beta}_1 = 475.9444, \hat{\beta}_2^* = \left(\frac{1}{w_2}\right)\hat{\beta}_2 = \frac{-0.4579}{10} = -0.04579$

$\hat{\beta}_1^* = 475.9444$ When price of good is zero, supply for good is 475.9444 hundred pieces.
$\hat{\beta}_2^* = -0.04579$ When price of good increases one hundred baht, supply for good decreases 0.04579 hundred pieces

4.7 If the unit Supply for good changes from hundred pieces to piece and the unit price of good changes from thousand baht to hundred baht. What is the estimator for β_1 *and* β_2? Interpret the regression

$Y_i^* = w_1 Y_i,\ w_1 = 100\ \ X_i^* = w_2 X_i,\ w_2 = 10$

$$\hat{\beta}_1^* = w_1\hat{\beta}_1 = 47594.44, \hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right)\hat{\beta}_2 = \left(\frac{100}{10}\right)0.4579 = -4.579$$

$\hat{\beta}_1^* = 47594.44$ When price of good is zero, supply for good is 47594.44 pieces.

$\hat{\beta}_2^* = -4.579$ When price of good increases one hundred baht, supply for good decreases 4.579 pieces

5. Consider the following regression output:

$$\hat{Y}_i = 0.2033 + 0.6560 X_t$$
$$se = (0.0976) \quad (0.1961)$$
$$r^2 = 0.397$$
$$RSS = 0.0544$$
$$ESS = 0.0358$$

Where Y = labor force participation rate (LFPR) of women in 1972 and X = LFPR of women in 1968. The regression results were obtained from a sample of 19 cities in the United States.

5.1 There is a positive association in the LFPR in 1972 and 1968, which is not surprising in view of the fact since WW II there has been a steady increase in the LFPR of women.

5.2 One tail t-test

$H_0 : \beta_2 \le 1$

$H_1 : \beta_2 > 1$

$$t = \frac{0.6560 - 1}{0.1961} = -1.7542$$

Degree of freedom =19-2 =17

Critical t value at 5 percent significance level is 1.740.

Since the estimated t value is not significant, at this level of significance, we cannot reject the hypothesis that the true slope coefficient is 1 or lesser.