MOOC Course- Regression Analysis and Forecasting - January 2017

Assignment 2

[1] Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, 2, \ldots, n,\ \epsilon_i \sim N(0, \sigma^2)$ are identically and independently distributed. The parameters $\beta_0,\ \beta_1$ and σ^2 are estimated by ordinary least squares estimation where $\hat{\beta}_1 = s_{xy}/s_{xx},\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},\ s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$, $s_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2,\ s_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2,\ \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,\ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\hat{y}_i$ is the i^{th} fitted value. The variance of $\hat{\beta}_0$ is specified by

(i) $\frac{1}{n-2}(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}})\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

(ii) $\frac{1}{n-2}(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}})\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$.

(iii) $\frac{1}{n-2}(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}})(s_{yy} - \hat{\beta}_1^2 s_{xx})$.

(iv) $\frac{1}{n-2}(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}})(s_{yy} - \hat{\beta}_1 s_{xy})$.

A. (i) and (ii).

B. (ii) and (iv).

C. (ii), (iii) and (iv).

D. (i), (ii), (iii) and (iv).

[2] If the random errors ϵ_i's having zero mean and unknown variance σ^2 in the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, 2, \ldots, 40$ do not follow the normal distribution, then which of the following test is used to test the significance of null hypothesis $H_0 : \beta_1 = 5$?

A. t test.

B. Z test.

C. Any one of the t test or Z test.

D. None of the t test or Z test.

[3] Let $s^*_{xx} = \sum_{i=1}^n x_i^2$, $s^*_{yy} = \sum_{i=1}^n y_i^2$ and $s^*_{xy} = \sum_{i=1}^n x_i y_i$ The $100(1-\alpha)\%$ confidence interval of β in the simple linear regression model $y_i = \beta x_i + \epsilon_i$ where ϵ_i's are identically and independently distributed following normal distribution with zero mean and unknown variance σ^2 is

A. $\left(\frac{s^*_{xy}}{s^*_{xx}} - \frac{z_{\alpha/2}}{\sqrt{(n-1)s^*_{xx}}}\sqrt{s^*_{yy} - \frac{(s^*_{xy})^2}{s^*_{xx}}}, \frac{s^*_{xy}}{s^*_{xx}} + \frac{z_{\alpha/2}}{\sqrt{(n-1)s^*_{xx}}}\sqrt{s^*_{yy} - s\frac{(s^*_{xy})^2}{s^*_{xx}}}\right).$

B. $\left(\frac{s^*_{xy}}{s^*_{xx}} - \frac{t_{\alpha/2,n-1}}{\sqrt{(n-1)s^*_{xx}}}\sqrt{s^*_{yy} - \frac{(s^*_{xy})^2}{s^*_{xx}}}, \frac{s^*_{xy}}{s^*_{xx}} + \frac{t_{\alpha/2,n-1}}{\sqrt{(n-1)s^*_{xx}}}\sqrt{s^*_{yy} - \frac{(s^*_{xy})^2}{s^*_{xx}}}\right).$

C. $\left(\frac{s^*_{xy}}{s^*_{xx}} - \frac{t_{\alpha/2,n-2}}{\sqrt{(n-2)s^*_{xx}}}\sqrt{s^*_{yy} - \frac{(s^*_{xy})^2}{s^*_{xx}}}, \frac{s^*_{xy}}{s^*_{xx}} + \frac{t_{\alpha/2,n-2}}{\sqrt{(n-2)s^*_{xx}}}\sqrt{s^*_{yy} - \frac{(s^*_{xy})^2}{s^*_{xx}}}\right).$

D. None of these.

[4] The 95% confidence interval of β_1 in the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, 2, \ldots, 7$, $\epsilon_i \sim N(0, \sigma^2)$ where σ^2 is unknown, is obtained for a given set of data as $(10, 20)$. The tabulated value of t statistic at 5% level of significance with 5 degrees of freedom is 2.0. Which of the following statements are correct.

Statement 1 : $\hat{\beta}_1 = 15.0$ and standard error of $\hat{\beta}_1$ is 2.5.

Statement 2 : $\hat{\beta}_1 = 15.0$ and variance of $\hat{\beta}_1$ is 6.25.

Statement 3 : The null hypothesis $H_0 : \beta_1 = 5$ is accepted at 5% level of significance.

Statement 4 : The null hypothesis $H_0 : \beta_1 = 5$ is rejected at 5% level of significance.

A. Statements 1, 2 and 3 are correct.

B. Statements 1, 2 and 4 are correct.

C. Statements 1 and 3 are correct.

D. All the statements 1 and 4 are correct.

[5] A simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, 2, \ldots, n$ is fitted on the basis of given set of data. The fitted model is obtained as $y = 2 + 3x$. The interpretation of the value 3 is

A. when x changes by one unit, then the average value of y changes by 3 units.

B. when x changes by one unit, then the value of y changes by 3 units.

C. when x changes by one unit, then the average value of y lies in the interval (0,3).

D. when x changes by one unit, then the value of y lies in the interval (0,3).

[6] A simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, 2, \ldots, n$ is fitted on the basis of given set of data. The fitted model is obtained as $y = 6 + 7x$. The interpretation of the value 6 is

A. when $x = 0$, then the average value of y is 6 units.

B. when $x = 0$, then the value of y is 6 units.

C. when x changes by 7 units, then the change in the average value of y is 6 units.

D. when x changes by 7 units, then the change in the value of y is 6 units.

Question 7 - 10 are based on the following output of a software which is obtained while fitting a simple linear regression model $y = \beta_0 + \beta_1 x + \epsilon$, $\epsilon \sim N(0, \sigma^2)$ to a given set of data.

[7] The fitted regression model is

A. $y = 44.2 + 2.89x$

B. $y = 2627.8 - 37.15x$

C. $y = -37.15 + 2627.8x$

D. $y = 59.47 - 12.86x$

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1527483	1527483	165.38	0.000
Error	18	166255	9236		
Total	19	1693738			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2627.8	44.2	59.47	0.800	
x	-37.15	2.89	-12.86	0.000	1.00

Figure 1: Software output for Questions 7 - 10

[8] The null hypothesis $H_0 : \beta_1 = 0$ at 5% level of significance is

A. accepted.

B. rejected.

C. nothing can be said.

D. Inadequate data.

[9] The null hypothesis $H_0 : \beta_0 = 0$ at 5% level of significance is

A. accepted.

B. rejected.

C. nothing can be said.

D. Inadequate data.

[10] The least squares estimate of σ^2 is

A. 44.2

B. 2.89

C. 166255

D. 9236

Solution to Assignment 2

Answer of Question 1 – D

Answer of Question 2 – D

Answer of Question 3 – B

Answer of Question 4 – B

Answer of Question 5 – A

Answer of Question 6 – A

Answer of Question 7 – B

Answer of Question 8 – B

Answer of Question 9 – A

Answer of Question 10 – D