STAT 363 Homework 2

Merve Ogretmek

November 29, 2020

1

a. (Original Question) Ordinary least squares estimators are

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \tag{1}$$

where $\hat{\beta}$ is a 3x1 matrix. To find the OLS estimator, we need to calculate the following

$$\hat{\beta} = (X'X)^{-1}X'y \tag{2}$$

by putting the values given in the question, we get

$$\hat{\beta} = \begin{bmatrix} 8.6 & 0.081 & -0.0099 \\ 0.08 & 0.0021 & -0.00013 \\ -0.0099 & -0.00013 & 0.000012 \end{bmatrix} \begin{bmatrix} 924.0 \\ 20299.8 \\ 935906.0 \end{bmatrix} = \begin{bmatrix} 325.2144 \\ -5.1182 \\ -0.5557 \end{bmatrix} \tag{3}$$

Fitted model can be written as

$$\hat{Y} = 325.2144 - 5.1182X_1 - 0.5557X_2 \tag{4}$$

a. (Updated Question) $\hat{\beta}$ matrix is given as

$$\hat{\beta} = \begin{bmatrix} 350.99427 \\ -1.27199 \\ -0.15390 \end{bmatrix} \tag{5}$$

Then, the fitted model is

$$\hat{Y} = 350.99427 - 1.27199X_1 - 0.1539X_2 \tag{6}$$

b. Estimated variance can be found by the formula

$$\hat{\sigma}^2 = MS_{Res} = \frac{SS_{Res}}{n-p} = \frac{y'y - \hat{\beta}'X'y}{n-p} = \frac{SS_{Res}}{n-p} = \frac{\sum_{i=1}^{6} y_i^2 - \hat{\beta}'X'y}{n-p} \tag{7}$$

where n is the number of observations and p is the number of parameters, for this question n = 6 and p = 3. By putting the values according to the data in question

$$\hat{\sigma}^2 = \frac{156408 - \begin{bmatrix} 350.99427 & -1.27199 & -0.15390 \end{bmatrix} \begin{bmatrix} 924.0 \\ 20299.8 \\ 935906.0 \end{bmatrix}}{6 - 3} \tag{8}$$

$$\hat{\sigma}^2 = \frac{156408 - (154461.62)}{3} = 648.8 \tag{9}$$

c. To find the estimated covariance between the estimators of slope terms, we need to estimate the covariance matrix. Covariance matrix calculated as following

$$\hat{Cov}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} \tag{10}$$

We have found that $\hat{\sigma}^2 = 648.8$, so it yields

$$\hat{Cov}(\hat{\beta}) = 648.8 \begin{bmatrix} 8.6 & 0.081 & -0.0099 \\ 0.081 & 0.0021 & -0.00013 \\ -0.0099 & -0.00013 & 0.000012 \end{bmatrix} \tag{11}$$

$$\hat{Cov}(\hat{\beta}) = \begin{bmatrix} \hat{Cov} & \hat{\beta}_0 & \hat{\beta}_1 & \hat{\beta}_2 \\ \hat{\beta}_0 & 5579.68 & 52.5528 & -6.42312 \\ \hat{\beta}_1 & 52.5528 & 1.36248 & -0.084344 \\ \hat{\beta}_2 & -6.42312 & -0.084344 & 0.0077856 \end{bmatrix} \tag{12}$$

Above figure states that

$$
\begin{aligned}
\hat{Cov}(\hat{\beta}_0, \hat{\beta}_0) &= \hat{Var}(\hat{\beta}_0) = 5579.68 \\
\hat{Cov}(\hat{\beta}_1, \hat{\beta}_1) &= \hat{Var}(\hat{\beta}_1) = 1.36248 \\
\hat{Cov}(\hat{\beta}_2, \hat{\beta}_2) &= \hat{Var}(\hat{\beta}_2) = 0.0077856 \\
\hat{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= \hat{Cov}(\hat{\beta}_1, \hat{\beta}_0) = 52.5528 \\
\hat{Cov}(\hat{\beta}_0, \hat{\beta}_2) &= \hat{Cov}(\hat{\beta}_2, \hat{\beta}_0) = -6.42312 \\
\hat{Cov}(\hat{\beta}_1, \hat{\beta}_2) &= \hat{Cov}(\hat{\beta}_2, \hat{\beta}_1) = -0.084344
\end{aligned} \tag{13}
$$

d. First of all, to make inference by using hypothesis testing we need to assume that ϵ(error) has identical and independent normal distribution with 0 mean and σ^2(constant variance) for every observation. For testing the overall significance of regression, we need to test the following hypothesis

$$H_0 : \beta_1 = \beta_2 = 0 \quad \text{against} \quad H_1 : \text{at least one them is not equal to 0} \tag{14}$$

To find the test statistic we need to calculate MS_R and MS_{Res}. We have already found that $MS_{Res} = 648.8$. MS_R is

$$MS_R = \frac{SS_R}{k} = \frac{\hat{\beta}'X'y - \frac{\left(\sum_{i=1}^{6} y_i\right)^2}{n}}{k} = \frac{\begin{bmatrix} 350.99427 & -1.27199 & -0.15390 \end{bmatrix} \begin{bmatrix} 924.0 \\ 20299.8 \\ 935906.0 \end{bmatrix} - \frac{924^2}{6}}{2} \tag{15}$$

$$MS_R = \frac{154461.62 - 142296}{2} = 6082.81 \tag{16}$$

Test statistic is

$$F_0 = \frac{MS_R}{MS_{Res}} = \frac{6082.81}{648.8} = 9.375 \tag{17}$$

and critical value at 0.05 significance level and degrees of freedom (2,29) is

$$F_{2,29} = 3.328 \tag{18}$$

Since $F_0 = 9.375 > 3.328 = F_{2,29}$, we reject the null hypothesis and conclude that at least one of the independent variables has a significant linear relationship with dependent variable.

2

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	4120.59225	2060.29612	**25.9170**	< .0001
Error	**29**	**2305.3765**	**79.4957**		
Total	31	6425.96875			

Variable	Estimate	Standard Error	t Value	Pr > $\lvert t \rvert$
(Intercept)	55.32344	12.53475	**4.41360**	0.0001
AGE	1.04516	**0.38566**	2.71	**0.0112**
QUET	**9.72442**	5.40246	1.80	0.0815

a. Fitted model is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 \tag{19}$$

where $\hat{Y}$ is the predicted value of systolic blood pressure, $\hat{\beta}_0$ is the estimated intercept, $\hat{\beta}_1$ is the estimated coefficient of X_1, X_1 is the observed age, $\hat{\beta}_2$ is the estimated coefficient of X_2 and X_2 is the observed QUET score. By looking at the estimates from table, we can write

$$\hat{Y} = 55.32 + 1.04 X_1 + 9.72442 X_2 \tag{20}$$

b. $\hat{\beta}_1 = 1.04$ is corresponding to the expected change in y when we increase X_1 by one unit while holding the X_2 same. $\hat{\beta}_2 = 9.72442$ is corresponding to the expected change in y when we increase X_2 by one unit while holding the X_1 same.

c. To conduct a hypothesis test, we need to assume that error term has independent and identical normal distribution with mean zero and constant variance σ^2. The hypothesis is

$$H_0 : \beta_1 = \beta_2 = 0 \quad \text{against} \quad H_1 : \text{at least one them is not equal to 0} \tag{21}$$

The test statistic is

$$F_0 = \frac{MS_R}{MS_{Res}} = \frac{2060.29612}{79.4957} = 25.9170 \tag{22}$$

and critical value at 0.05 significance level and degrees of freedom (2,29) is

$$F_{2,29} = 3.328 \tag{23}$$

Since $F_0 = 25.9170 > 3.328 = F_{2,29}$, we reject the null hypothesis and conclude that at least one of the independent variables has a significant linear relationship with dependent variable.
From the p-value approach

$$p - value = 2 * P(t_{29} > 25.9170)$$
$$p - value < 0.0001 < \alpha = 0.05 \tag{24}$$

Either way, we can reject the null hypothesis and say that at least one of the coefficients is not equal to zero.

d. Again, we need to assume that ϵ is distributed iid with normal distribution where mean is zero and variance is σ^2. For β_1,

-Hypothesis:

$$H_0 : \beta_1 = 0 \quad \text{against} \quad H_1 : \beta_1 \neq 0 \tag{25}$$

-Test Statistic:

$$t_0 = \frac{\hat{\beta}_1 - \beta_{10}}{se(\hat{\beta}_1)} = \frac{1.04}{0.38566} = 2.71 \tag{26}$$

-Critical Value:

$$t_{29,0.025} = 2.045 \tag{27}$$

-Conclusion: Since $t_0 = 2.71 > 2.045 = t_{29,0.025}$, we can reject the null hypothesis and say that there is a significant association between age and systolic blood pressure while QUET score is still in the model.

For β_2,

-Hypothesis:

$$H_0 : \beta_2 = 0 \quad \text{against} \quad H_1 : \beta_2 \neq 0 \tag{28}$$

-Test Statistic:

$$t_0 = \frac{\hat{\beta}_2 - \beta_{20}}{se(\hat{\beta}_2)} = \frac{9.72442}{5.40246} = 1.80 \tag{29}$$

-Critical Value:

$$t_{29,0.025} = 2.045 \tag{30}$$

-Conclusion: Since $t_0 = 1.80 < 2.045 = t_{29,0.025}$, we fail to reject the null hypothesis and say that there is no significant linear association between QUET score and systolic blood pressure while age is still in the model.

The above conclusions can also be made by looking at the p-values. P-value for $\hat{\beta}_1$ is 0.0109 which is smaller than $\alpha = 0.05$, so we can again say there is a significant relationship. However, p-value for $\hat{\beta}_1$ is 0.0815 which is larger than $\alpha = 0.05$, so we say there is no significant linear relationship.

e. Fitted equation is

$$\hat{Y} = 55.32 + 1.04X_1 + 9.72442X_2 \tag{31}$$

$$\hat{Y} = 55.32 + 1.04X_1 + 9.72442X_2$$
$$55.32 + (1.04 * 38) + (9.72442 * 2.52) = 119.34 \tag{32}$$

Predicted y for given x values is equal to 119.34.

f. For the given question we need to calculate the R^2:

$$R^2 = \frac{SS_R}{SS_T} = \frac{4120.59225}{6425.96875} = 0.64 \tag{33}$$

From above calculation, we can say that 0.64 percent of the variability in SBP can be explained by the model.