University of Ruhuna

Bachelor of Science Special Degree Level I (Semester II) Examination - January-2018

Subject: Statistics
Course Unit: MSP3263 (Regression Analysis)

Time :Two (02) Hours

Answer ALL Questions.
A calculator will be provided.

---

1. Consider the simple linear regression model $y = \beta_0 + \beta_1\, x + \epsilon$, with $E(\epsilon) = 0$, $var(\epsilon) = \sigma^2$, and ϵ are uncorrelated.

   (a) Show that $E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least squares estimates for β_0 and β_1, respectively.

   (b) Show that $Cov(\bar{y}, \hat{\beta}_1) = 0$.

   (c) In the usual notation, derive the formula $Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$.

   (d) Show that $Cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\bar{x}\,\sigma^2}{S_{xx}}$.

   (e) In the usual notation, derive the formula $Var(\hat{\beta}_0) = \sigma^2\,(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}})$.

---

2. The article "Evaluation of the Expansion Attained to Date by Concrete Affected by Alkali-silica Reaction. Part III: Application to Existing Structures" (M. Berube, N. Smaoui, et. al., Canadian Journal of Civil Engineering, 2005: 463-479) reports measurements of expansion for several concrete bridges in the area of Quebec City. Following are measurements of horizontal and vertical expansion (in part per hundred thousand) for several locations on the Pere-Lelivevre bridge.

x	20	15	43	5	18	24	32	10	21
y	58	58	55	80	58	68	57	69	63

Fit a simple linear regression model to this data by performing the following steps:

(a) Write ot the design matrix **X** for this data and the vector **y** of responses.

(b) Compute $\boldsymbol{X'X}$.

(c) Compute $(\boldsymbol{X'X})^{-1}$.

(d) Compute the least squares estimates of the y-intercept β_0 and slope β_1, using suitable matrices.

(e) Compute the mean squared error for the least-squares regression line:

$$\hat{\sigma}^2 = ME_{Res} = (\mathbf{y} - \mathbf{y})'(\mathbf{y} - \mathbf{y}')/(n-2).$$

(f) Compute the estimated covariance matrix for the estimated regresion coefficients, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)'$.

(g) Find a 95% confidence interval for the slope of the regression line.

---

3. (a) Consider the simple linear regression model $y = \beta_0 + \beta_1 x + \epsilon$, with $E(\epsilon) = 0$, $Var(\epsilon) = \sigma^2$, and ϵ uncorrelated. Show that

$$E(MS_{Res}) = \sigma^2.$$

(b) The following data were collected in an experiment to study the relationship between the number of pounds of fertilizer (x) and the yeild of tomatoes in bushels y.

The least squares quadratic model is $y = 4.8000 + 2.508571\ x - 0.07428571\ x^2$.

x	5	10	15	20	25
y	16	21	27	25	21

(i) Using this equation, compute the residuals.

(ii) Compute the error sum of squares SS_{Res} and the total sum of squares SS_T.

(iii) Compute the coefficient of determination R^2 and adjusted R^2.

Continued.

(iv) Compute the value of the F_0 statistic for the hypothesis $H_0 : \beta_1 = \beta_2 = 0$. How many degrees of freedom does this statistic have?

(v) Can the hypotheis $H_0 : \beta_1 = \beta_2 = 0$ be rejected at the 5% level? Explain.

---

4. In an experiment to determine the factors affecting fuel economy in trucks, the fuel consumption (in mi/gal), the weight (in tons), and the odometer reading (in 1000s of miles) were measured for 15 trucks. The following R output presents the results of fitting the model

$$\text{Miles per gallon(MPG)} = \beta_0 + \beta_1 \text{ Weight} + \beta_2 \text{ Odometer}.$$

```
The regrssion equation is
MPG = 8.24 - 0.108 weight - 0.00392 Odometer
```

Predictor	Coef	SE Coef	T	P-Value
Constant	8.2407	0.2871	28.70	0.000
Weight	-0.10826	0.01194	-9.06	0.000
Odometer	-0.0039249	0.001406	-2.79	0.016

```
S=0.3182        R-Sq = 87.8%        R-Sq( adj) = 85.7%
Analysis of Variance
```

Source	DF	Sum of Square	Mean Squares	F	P-Value
Regression	2	8.720	4.360	43.05	0.000
REsidual Error	12	1.215	0.101		
Total	14	9.935			

(a) Predict the miles per gallon for a truck that weighs 10 tons and has an odometer reading of 50,000 miles.

(b) If two trucks have the same weight, and one has 10,000 more miles on the odometer, by how much would you predict their miles per gallon would differ?

(c) If two trucks have the same odometer, and one weighs 5 tons more than the other, by how much would you predict their miles per gallon would differ?

(d) Find a 95% confidence interval for the coefficient of weight.

(e) Can you conclude that $\beta_1 < -0.05$? Perform the appropriate hypothesis test with 95% confidence level.
