BINDURA UNIVERSITY OF SCIENCE EDUCATION SFM212

BSc. IN STATISTICS AND FINANCIAL MATHEMATICS

LINEAR REGRESSION

Time : 3 hours

Candidates may attempt ALL questions in Section A and at most TWO questions in Section B. Each question should start on a fresh page.

SECTION A (40 marks)

Candidates may attempt ALL questions being careful to number them A1 to A5.

A1. Define the following terms as they are used in statistics:

(a) multi-collinearity, [2]

(b) leverage, [2]

(c) autocorrelation. [2]

A2. (a) State the assumptions underlying the use of the least squares technique. [2]

(b) Given a simple linear model $Y_i = \beta_0 + \beta_1 x_i + e_i$, derive the least squares estimators for β_0 and β_1. [6]

(c) Prove that

$$\widehat{\beta_0} \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)\right).$$

[4]

A3. (a) What do you understand by Latin Square Design? [2]

(b) Give a statistical equation for the Latin Square Design. [2]

(c) Distinguish between a Latin Square Design from a Randomised Block Design. [2]

A4. A certain experiment was conducted and the data below was collected. The regression model assumed is

$$Y_i = \beta_0 + \beta_1 x_i + e_i$$

Observation	X_i	Y_i	Observation	X_i	Y_i
1	125	160	7	75	42
2	100	112	8	175	124
3	200	124	9	125	150
4	75	28	10	200	104
5	150	143	11	100	136
6	175	156	12	150	161

(a) Construct an analysis of variance table. [4]

(b) Test at the 5% level of significance, the lack of fit of the model. [4]

A5. Given that $\hat{\beta} = (X^T X)^{-1} X^T Y$.

(a) Show that $\hat{\beta}$ can be expressed as $\hat{\beta} = AY$, where A is to be determined. [2]

(b) Show that $E(\hat{\beta}) = \beta$. [3]

(c) Show that $Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$. [3]

SECTION B (60 marks)

Candidates may attempt TWO questions being careful to number them B6 to B8.

B6. (a) The response time in milliseconds was determined for different types of circuits that could be used in an automatic value shut-off mechanism. The results were:

Circuit Type	Response Time				
1	09	12	10	08	15
2	20	21	23	17	30
3	06	05	08	16	07

(i) Suggest an appropriate model for the data. [4]

(ii) Estimate the model parameters μ and τ_i. [4]

(iii) Test the hypothesis that the three circuit types have the same response time. (Use $\alpha = 0.05$) [10]

(b) Assume the following table came from the analysis of a randomised block design. Given the following data set

Source	df	SS	MS	F
Treatments	4	b	e	5
Blocks	a	c	48	6
Error	20	d	f	

(i) Determine the values of a through to f. Show all working. [6]

(ii) Provide tests of treatments and blocks with $\alpha = 0.05$. What are your conclusions? [6]

B7. The following data shows a segment of the series on percentage wage change for Y, unemployment X_1 and percentage price changes, X_2 inflation.

Y	X_1	X_2
3	3	5
1	1	4
8	5	6
3	2	4
5	4	6

(a) Briefly discuss the underlying assumptions of multiple regression. [4]

(b) Fit a multiple regression model to this data. [7]

(c) Calculate the fitted values and compute the residuals. [4]

(d) Construct the basic ANOVA table, hence test the significance of regression at $\alpha = 0.10$. [6]

(e) Prove that $\hat{\beta}_j \pm t_{(\frac{\alpha}{2}, n-p)} S\sqrt{C_{jj}}$, where $C_{jj} = X^T X)^{-1}$. [5]

(f) Find 95% confidence interval for $\hat{\beta}_1$. [4]

B8. A test is to be run on a given process for the purpose of determining the effect of an independent variable X (such as process temperature) on a certain characteristic property of finished product Y (such as density). Four observations are to be taken at each of the five settings of the independent variable X. When the test was actually run, the following results were obtained.

$$\overline{x} = 5.0; \qquad \sum_{i=1}^{n}(x_i - \overline{x})^2 = 160.0; \qquad \overline{y} = 3.0$$

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 80; \qquad \sum_{i=1}^{n}(y_i - \overline{y})^2 = 83.2$$

Assume that a model of the type $Y_i = \beta_0 + \beta_1 x_i + e_i$

(a) Calculate the fitted regression equation. [6]

(b) Prepare the ANOVA table. [8]

(c) Determine 95% confidence limits for the true mean value of y when $x = 5.0$. [6]

(d) Test for the significance of the regression model. [6]

(e) Determine a 95% confidence interval for β_1. [4]

END OF QUESTION PAPER