BINDURA UNIVERSITY OF SCIENCE EDUCATION SFM212

Bachelor Of Science in Financial Mathematics

LINEAR REGRESSION ANALYSIS - DEC 2019

Time : 3 hours

Answer ALL questions in Section A and at most TWO questions in section B.

SECTION A (40 marks)

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

A1. Given that $\hat{\beta} = (X'X)^{-1}X'Y$

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

(b) Show that $E(\hat{\beta}) = \beta$, [4]

(c) Show that $var(\hat{\beta}) = \sigma^2(X'X)^{-1}$. [5]

A2. An investigator interested in the dependents of the speed of sound on temperature obtained the following measurements;

temp	-20	0	20	50	100
speed	323	327	340	364	384

(a) Write down the model in matrix form, [2]

(b) Find the least square estimates of β_0 and β_1, [5]

(c) Calculate the 95% confidence interval for $y_0 = \beta_0 + 80\beta_1$. [5]

A3. (a) State any three assumptions of linear regression model $Y_i = \beta_0 + \beta_1 X_i + e_i$, [3]

(b) Briefly describe how any two(2) of the model assumptions outlined in (a) are checked. [2]

A4. Define the following terms as they are used in statistics;

(a) autocorrelation, [2]

(b) multi-collinearity [2]

(c) heteroscedasticity [2]

A5. Given that $(X'X)^{-1} = c_{jj}$. Show that the confidence interval of $\hat{\beta}_j$ is given by $\hat{\beta}_j \pm st_{(n-p,\frac{\alpha}{2})}\sqrt{c_{jj}}$. [6]

SECTION B (60 marks)

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

B6. 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. [3]

(b) Fit a multiple regression model to this data. [8]

(c) Calculate the fitted values and compute the residuals. [5]

(d) Construct the basic ANOVA table, hence test the significance of regression at $\alpha = 0.01$ [8]

(e) Find 95% confidence intervals for $\hat{\beta}_0$ and $\hat{\beta}_1$. [6].

B7. (a) A study was made to determine the effect of stirring rate on the amount of impurity in paint produced by a chemical process. The study yielded the following data:

Stirring rate(pm)(x)	20	22	24	26	28	30	32	34	36	38	40	42
Impurity (%),(y)	8.4	9.5	11.8	10.4	13.3	14.8	13.2	14.7	16.4	16.5	18.9	18.5

(i) Fit a simple linear model to the data. [7]

(ii) Test for the significance of the regression with the fitted model at $\alpha = 0.01$. [8]

(iii) Obtain a 95% confidence interval for β_0 and β_1. [8]

(b) Outline the effects of multicollinearity. [4]

(c) Outline any three ways that can be used to deal with multicollinearity. [3]

B8. (a) Outline problems associated with auto-correlation. [6]

(b) Suppose we wish to determine whether four different tips produce different readings on a hardness testing machine. There are four tips and four available metal coupons. Each tip is tested once on each coupon resulting in a randomized complete block design. The data obtained is given below:

	Coupon(Block)			
Type of tip	1	2	3	4
1	9.3	9.4	9.6	10.0
2	9.4	9.3	9.8	9.9
3	9.2	9.4	9.5	9.7
4	9.7	9.6	10.0	10.2

(i) Write the appropriate model for this data(define all the terms used). [5]

(ii) Can we conclude at 5% that the type of tip affects mean hardness reading. [10]

(c) Distinguish Latin square from randomized block design. [4]

(d) Explain the procedure for testing and assessing regression model validity. [5]

END OF QUESTION PAPER