BINDURA UNIVERSITY OF SCIENCE EDUCATION SFM212

BACHELOR OF STATISTICS AND FINANCIAL MATHEMATICS

LINEAR REGRESSION ANALYSIS

Time : 3 hours MAR 2024

Candidates should attempt 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 A4

A1. (a) Briefly describe the following terms used in regression analysis.

(i) Analysis of variance, [2]

(ii) Lack of fit, [2]

(b) Outline any four assumptions of the least squares technique. [4]

A2. Given that $\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$ is a design matrix,

(a) Show that $\hat{\beta}$ can be expressed as $\hat{\beta} = AY$, where A is to be determined. [3]

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

(c) Show that $\text{Var}(\hat{\beta}) = \sigma^2(X^T X)^{-1}$. [6]

A3. The following is a partial edited multiple regression SPSS output relating price to age, audience size, and attendance.

ANOVA Model 1: Dependent Variable :Price

	Sum of Squares	df	Mean Square	F	Sig.
Regression	5288169.883	b	1762723.294	d	.000
Residual	a	20	c		
Total	5724333.333	23			

Coefficients

	Parameter Estimate	Standard Error	Beta	t-value	Sig.
Constant	20.8	7.046		22.826	.000
Age	6.2	.0	.194	2.472	.000
Audience Size	0.3	.025	.952	2.113	.000
Attendance	9.8	2.999	.678	2.9090	.078

(a) Write down the equation of the fitted regression line. [4]

(b) Find the values of a, b, c and d in the ANOVA table. [7]

(c) Test the significance of the regression model, using $\alpha = 0.05$. [4]

A4. The regression model for parameters X and Y is defined as $Y = \beta_0 + \beta_1 X + \epsilon$. Given that the confidence interval for Y_0 at $X = 80$ is $(364.65 \quad 385.90)$, Test the hypothesis given below:

$$H_0 : Y_0 = 320$$
$$H_1 : Y_0 \neq 320$$

[4]

SECTION B (60 marks)

Candidates may attempt TWO questions being careful to number them B5 to B7.

B5. Consider the following data set, where Y is the dependent variable and X is the explanatory variable

X	4	8	5	9	10
Y	2	3	4	6	8

Suppose the data can be described by model $Y_i = \beta_0 + \beta_1 X_i + e_i$ where $e_i \sim N(0, \sigma^2)$ and $\text{Cov}(e_i, e_j) = 0$ for $i \neq j$.

(a) Express the above model in matrix form. [1]

(b) Obtain the design matrix X, $[X'X]$, $[X'Y]$ and $[X'X]^{-1}$. [1,2,2,3]

(c) Find the least squares estimator of β. [3]

(d) Construct the ANOVA table and test for significance of the regression line using $\alpha = 0.05$. [10]

(e) Test the hypothesis $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$ at $\alpha = 0.05$. [3]

(f) Estimate Y at $X = 7$ and find the 95% confidence interval for this mean response. [2,3]

B6. To study the per capita consumption of chicken in response to income in Zimbabwe, you are given data below, where Y denotes per capita consumption of chickens (Kg) and X denotes real disposable income per capita ($).

Y	36	72	48	51	80	40	55	72	39	47
X	240	450	250	320	450	250	330	430	240	320

(a) Fit the regression model to this data using the method of least squares. [7]

(b) Construct the ANOVA table and test for significance of the regression line. Use $\alpha = 0.05$. [10]

(c) State the conditions required for the lack of fit test. [2]

(d) Test for lack of fit at $\alpha = 0.05$, clearly stating H_0 and H_1. [8]

(e) Calculate the coefficient of variation and interpret your answer. [3]

B7. (a) Outline the 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 randomised block design. [4]

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

END OF QUESTION PAPER