2021

STATISTICS — HONOURS

Paper : CC-12

(Linear Models and Regression)

Full Marks : 50

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

1. Answer ***any five*** questions : 2×5

(a) State a linear hypothesis in the context of a linear model. How do you test such hypothesis?

(b) Explain why adjusted R^2 is a better tool to measure the goodness of fit of a model?

(c) Define error space in a linear model.

(d) Discuss the use of concominant variable in ANCOVA.

(e) Explain why the standard linear model cannot be used to model count data.

(f) Cite a hypothetical situation where one way ANOVA can be suggested to analyse the dataset.

(g) What do you mean by a set of orthogonal treatment contrasts? What is the maximum number of possible orthogonal treatment contrasts if K group means are available?

(h) Define odds ratio.

2. Answer ***any two*** questions : 5×2

(a) Write down a two-way ANOVA model with more than one observation per cell stating clearly the underlying assumptions. Assuming a balanced design derive the orthogonal splitting of the sum of squares necessary for the validity of F statistics.

(b) Consider a simple linear regression with a continuous response variable (y) and a continuous independent variable (x). Discuss how to test that the response variable is parallel to x-axis if plotted against the independent variables.

(c) Define a linear model. Discuss how an ANOVA model, a regression model and an ANCOVA model can be described through a linear model.

3. Answer ***any three*** questions : 10×3

(a) (i) Consider the following one-way fixed-effects model :

$$y_{ij} = \mu + \alpha_i + e_{ij};\ i = 1,2,...,k \text{ and } j = 1,2,...,r$$

Find an unbiased estimator of the average variance of the estimator of all possible elementary treatment contrasts.

(ii) What do you mean by valid error in the context of ANOVA? Explain briefly the role of valid error in hypothesis testing in ANOVA. 5+5

(b) For an independent variable x, assume that, the response variable y is dichotomous. If $\pi(x)$ is the average proportion of success for y

(i) define the logit function based on $\pi(x)$.

(ii) Discuss a regression technique to regress the logit function on x

(iii) Interpret the regression coefficients. 2+4+4

(c) Consider the following analysis of covariance model :

$$y_{ij} = \mu + b_i + t_j + rx_{ij} + e_{ij} \; i = 1,2,..,r; \; j = 1,2,...,k$$

where e_{ij}'s are random errors which follow normal distribution with mean zero and variance σ^2.

(i) Obtain a test for $H_0 : t_1 = t_2 = ... = t_k$

(ii) Obtain an unbiased estimates for $t_j - t_{j'} \quad j \neq j'$ 7+3

(d) (i) "On the linear model $E(\underset{\sim}{Y}) = A\underset{\sim}{\beta}$, if a parametric function $\underset{\sim}{\lambda}'\underset{\sim}{\beta}$ is estimable then best estimator of $\underset{\sim}{\lambda}'\underset{\sim}{\beta}$ is $\underset{\sim}{\lambda}'\underset{\sim}{\hat{\beta}}$, where $\underset{\sim}{\hat{\beta}}$ is any solution of the equation $A'A\underset{\sim}{\beta} = A'\underset{\sim}{y}$" — Prove the statement.

(ii) Let the model equation be

$$y_1 = 2\alpha_1 + 3\alpha_2 + e_1$$
$$y_2 = 3\alpha_1 + 4\alpha_2 + e_2$$
$$y_3 = 4\alpha_1 + 5\alpha_2 + e_3$$

Obtain the expression for the estimable parametric function. Hence check whether α_1 is estimable. If yes, find the best estimate of α_1. 5+5

(e) Consider a multiple regression model. Suggest a test for checking significance of the regression. Derive simultaneous confidence interval for regression coefficients. 7+3