Total number of printed pages–7

3 (Sem–6/CBCS) STA HC 1

2024

STATISTICS

(Honours Core)

Paper : STA–HC–6016

(Design of Experiments)

Full Marks : 60

Time : Three hours

The figures in the margin indicate full marks for the questions.

1. Answer the following as directed : 1×7=7

   (a) Name a design where principle of local control is not used.

   (b) The degree of freedom for error in a $m \times m$ LSD is ________ .

   (c) What is a treatment contrast ?

Contd.

(d) For a 2^2 factorial experiment in an r-randomised block, the sum of squares for the main effect in the analysis of variance table is ______ .

(e) Write down the main effects and interaction effects for a 3^2 design with two factors A and B each at three levels 0, 1, 2.

(f) In the linear model considered in analysis of variance the error term is distributed as ______ .

(g) The error df in an RBD with 4 blocks comparing 5 treatments is ______ .

2. Answer the following : 2×4=8

(a) Define the terms 'main effect' and 'interaction effect' in relation to a 2^3 experiment.

(b) Give the layout of a 4 × 4 Latin square design.

(c) Show that for 2^3 factorial experiment the main effect A and interaction effect AB are mutually orthogonal contrasts.

(d) In an RBD with 6 treatments and 5 blocks, the following results were obtained :

Mean S.S due to block = 20

Mean S.S. due to treatment = 25

Total S.S = 245

Construct the ANOVA table.

3. Answer **any three** questions from the following : 5×3=15

(a) Explain the basic principles of experimental design giving brief explanatory note for each.

(b) Explain the concept of partial, total and balanced confoundings in factorial experiments.

(c) What is a split-plot design? Describe the situations in which a split-plot confounding rather than the general type of confounding would be more suitable.

(d) Obtain the estimate of the missing plot in a randomised block design.

(e) Give an idea of 3^2 factorial experiment.

4. Answer the following questions : ***(any three)***

10×3=30

(a) Define a Balanced Incomplete Block Design (BIBD) with parameters v, b, r, k, λ.

When do you call such a design

(i) symmetric

(ii) resolvable

(iii) affine resolvable?

Also write down the efficiency of BIBD relative to RBD. 5+5=10

(b) The elements of control block of each of six replication of a 2^4 design are (1), *ab*, *acd*, *bcd*. Identify the complete subgroup and give an outline of the analysis of the data obtained from the experiment.

(c) In a Latin square design the expectation of the yield X_{ijk} of the plot in the ith row, jth column receiving the kth treatment is $\mu + \alpha_i + \beta_j + \gamma_k$ where $\Sigma\alpha_i = 0$, $\Sigma\beta_j = 0$, $\Sigma\gamma_k = 0$ and variance of X_{ijk} is σ^2. Show how to estimate the parameters μ, α_i, β_j and γ_k and find the variance of the estimate of α_i and β_j.

(d) In testing the value of three fertilisers *N*, *P* and *K* each at two levels, eight pairs of blocks of 4 plots are used. The treatment (*N*, *P*, *K* and *NPK*) are put in one block. What should be the composition of the other block for completely confounding the second order interaction? Give the analysis of variance table for the confounded factorial design. 5+5=10

(e) What is the use of 'missing plot technique'?

Show that in an RBD with r blocks and 't' treatments the analysis can be carried out by substituting the value

$$y = \frac{rB + tT - G}{(r-1)(t-1)}$$

for the missing yield, B = the actual total of the block with the missing unit, T = the total of yield of the treatment with the missing unit and G = the grand total.

Show that by using the above missing value, the treatment sum of squares is overestimated by $\left[B-(t-1)y\right]^2 / t(t-1)$.

2+5+3=10

(f) Find the standard error of difference between two treatment means, one of which has a missing observation in RBD. Compare the efficiency of this design with CRD. 5+5=10

Total number of printed pages–8

3 (Sem–6/CBCS) STA HC 2

2024

STATISTICS

(Honours Core)

Paper : STA–HC–6026

(Multivariate Analysis and Non-parametric Methods)

Full Marks : 60

Time : Three hours

The figures in the margin indicate full marks for the questions.

1. Answer the following questions as directed : $1 \times 7 = 7$

   (a) Let X and Y be two normally correlated variables. In derving the bivariate normal distribution we make the following asumptions :

   (i) Regression of Y on X is linear.

   (ii) The arrays are homoscedastic.

   (iii) The distribution of Y in different arrays is normal.

Contd.

Choose the correct answer from the following options :

(I) Only (i) and (ii) are the basic assumptions required

(II) Only (ii) is the basic assumption deriving bivariate normal distribution

(III) No assumptions are required to derive bivariate normal distribution

(IV) All three assumptions (i), (ii) and (iii) are required to derive bivariate normal distribution

(b) The moment generating function of the bivariate normal distribution is

(I) $M_{XY}(t_1, t_2) = exp\left[\frac{1}{2}\left(t_1^2 + t_2^2 + 2Pt_1t_2\right)\right]$

(II) $M_{XY}(t_1, t_2) = exp\left[\frac{1}{2}\left(t_1 + t_2 + 2Pt_1t_2\right)\right]$

(III) $M_{XY}(t_1, t_2) = exp\left[\left(t_1^2 + t_2^2 + 2Pt_1t_2\right)\right]$

(IV) $M_{XY}(t_1, t_2) = exp\left[2\left(t_1^2 + t_2^2 + \frac{1}{2}Pt_1t_2\right)\right]$

(Choose the correct option)

(c) If the marginal distribution of X and Y are normal (Gaussian)

(I) it does not necessarily imply that the joint distribution of (X, Y) is bivariate normal

(II) it implies that the joint distribution of (X, Y) is bivariate normal

(III) it implies that the variables X and Y are independently distributed

(IV) the joint distribution is bivariate normal, but the random variables X and Y are independently distributed

Choose the appropriate answer from the above options.

(d) There are certain assumptions associated with non-parametric tests, which are

(i) sample observations are independent

(ii) the variables under study is continuous

(iii) the distributions are continuous

(iv) lower order moments exists

Choose the appropriate answer from the following options :

(I) No assumptions are required for non-parametric tests

(II) The non-parametric tests are valid for the discrete variables

(III) All above assumptions *(i), (ii), (iii)* and *(iv)* are required for non-parametric tests

(IV) The non-parametric tests can be done only for discrete distributions.

State which of the following statements is True **or** *False* :

(e) To study the correlation coefficient of two variables in the qualitative data, Pearson correlation being used to find the correlation coefficient of the two varriables.

(I) True

(II) False

Choose the correct answer.

(f) If $R_{1\cdot 23} = 1$, the multiple linear regression of X_1 on X_2 and X_3 is considered as perfect for all predictions.

(I) True

(II) False

Choose the correct answer.

(g) Let $\underset{\sim}{X} = \left(X_1, X_2, \ldots, X_p\right)'$ be the data vector be divided as $\underset{\sim}{X}^{(1)} = \left(X_1, \ldots, X_q\right)'$ and $\underset{\sim}{X}^{(2)} = \left(X_{q+1}, \ldots, X_p\right)'$, then for $i = 1, 2, \ldots, q$, the residual variate $X_{i\cdot q+1}, \ldots, p$ of X_1 is independent of the regressor $\underset{\sim}{X}^{(2)}$.

(I) True

(II) False

Choose the correct answer.

2. Answer the following questions : 2×4=8

(a) State *two* assumptions required for a non-parametric test.

(b) Let X be a p-dimensional random vector and if A is an $(n \times p)$ matrix of known constants, then AX is the product of a vector X by a matrix A, then find $V(AX)$.

(c) Let $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$. If X_1 and X_2 are independent and

$g\left(\underset{\sim}{x}\right) = g^{(1)}(x_1)\, g^{(2)}(x_2)$, prove that

$$E(g(X)) = E\left[g^{(1)}\left(X_{(1)}\right)\right] E\left[g^{(1)}\left(X_{(2)}\right)\right]$$

(d) Write a note on test for randomness.

3. Answer ***any three*** questions from the following : 5×3=15

(a) Write a note on ordinary sign test.

(b) Let $\underset{\sim}{X}$ (with p-components) be distributed according to $N\left(\underset{\sim}{\mu}, \Sigma\right)$, prove that $\underset{\sim}{Y} = C\underset{\sim}{X}$ is distributed according to $N\left(C\underset{\sim}{\mu}, C\Sigma C'\right)$ for C non-singular.

(c) If the bivariate random variables $X = (X_1, X_2)'$ follow bivariate normal distribution with mean vector 0 and dispersion matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Find the pdf of X_1/X_2.

(d) If (X, Y) is $\text{BVN}\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$. Compute the correlation coefficient of e^X and e^Y.

(e) If X and Y are standard normal variates with coefficient of correlation ρ, show that regression of Y on X is linear.

4. Answer ***any three*** questions from the following : 10×3=30

(a) If $X_1, \ldots, X_p$ have a joint normal distribution, a necessary and sufficient condition for one subset of random variables and the subset consisting of the remaining variables to be independent is that each covariance of a variable from one set and a variable from the other set is 0. Prove it.

(b) Let $(X, Y) \sim \text{BVN}$, prove that if and only if every linear combination of X and Y is a normal variate.

(c) Discuss Mann-Whitney test. What is the mean and variance of the test statistic ? How is the test carried out for large sample ? 6+2+2=10

(d) Describe the Kolmogorov-Smirnov test for one sample, explaining assumptions and hypothesis.

(e) (i) Write a note on principal component analysis. 5

(ii) Write a note on Hotelling's T^2. 5

(f) Suppose $\underset{\sim}{X} \sim N_p\left(\underset{\sim}{\mu}, \Sigma\right)$ and let $\underset{\sim}{X}$ be partitioned as $\underset{\sim}{X}' = \left(\underset{\sim}{X}^{(1)'}, \underset{\sim}{X}^{(2)'}\right)$, where $\underset{\sim}{X}^{(1)}$ has k-components and $\underset{\sim}{X}^{(2)} (p-k)$ components. Prove that the regression for the mean of $\underset{\sim}{X}^{(1)}$ on $\underset{\sim}{X}^{(2)}$ coincides with the linear mean square regression of $\underset{\sim}{X}^{(1)}$ on $\underset{\sim}{X}^{(2)}$.
