

~~Comp I IT~~
~~lib~~
~~3~~

**University of Mumbai
Examinations Summer 2022**

Max. Marks: 80

Time: 2 hour 30 minutes

| Q1 (20 Marks) | Choose the correct option for following questions. All the Questions are compulsory and carry equal marks |
|---------------------|---|
| 1. | Inspectors for a hospital chain with multiple locations randomly select some of their locations for a cleanliness check of their operating rooms. |
| Option A: | Cluster sampling |
| Option B: | Stratified Sampling |
| Option C: | Quota Sampling |
| Option D: | Snowball Sampling |
| 2. | In MLR, the square of the multiple correlation coefficient or R^2 is called the |
| Option A: | Coefficient of determination |
| Option B: | Variance |
| Option C: | Covariance |
| Option D: | Cross-product |
| 3. | The mode of the calls received on 7 consecutive days 11,13,13,17,19,23,25 is |
| Option A: | 11 |
| Option B: | 13 |
| Option C: | 17 |
| Option D: | 23 |
| 4. | "More than type Ogive" and "less than type Ogive" for a distribution intersect at |
| Option A: | Mean |
| Option B: | Median |
| Option C: | Mode |
| Option D: | Origin |
| 5. | In _____ method, the upper limit of one class is the lower limit of the next class. |
| Option A: | Inclusive |
| Option B: | Exclusive |
| Option C: | Inter |
| Option D: | Intra |
| 6. | If the regression coefficients are $b_{yx} = 0.5$ and $b_{xy} = 0.46$, then the value of coefficient of correlation (r) is |
| Option A: | 0.39 |
| Option B: | 0.48 |
| Option C: | 0.23 |
| Option D: | 0.25 |
| 7. | In regression analysis, if the independent variable is measured in Kilometers, the dependent variable |
| Option A: | Must also be in Kilometers |
| Option B: | Must be in some unit of Distance |
| Option C: | Cannot be in Kilometers |

| | |
|-----------|--|
| Option D: | Can be any units |
| 8. | A linear regression (LR) analysis produces the equation $Y = 0.4X + 3$. This indicates that: Option A: When $Y = 0.4$, $X = 3$ Option B: When $Y = 0$, $X = 3$ Option C: When $X = 3$, $Y = 0.4$ Option D: When $X = 0$, $Y = 3$ |
| 9. | If all the dots of a scatter diagram lie on a straight line falling from left bottom corner to the right upper corner, the correlation is called..... Option A: Zero correlation Option B: High degree of positive correlation Option C: Perfect negative correlation Option D: Perfect positive correlation |
| 10. | A point estimator is defined as _____ Option A: A single value from the sample Option B: Average of all sample values Option C: Average of all population values Option D: A single value that is best estimate of unknown population parameter |

| Q2 (20 Marks) | Solve any Two Questions out of Three | 10 marks each |
|----------------------|---|---------------|
| A | What do you mean by a questionnaire? What is the difference between a questionnaire and a schedule? State the essential points to be remembered in drafting a questionnaire. | |
| B | In a simple study about coffee habits in two Towns A and B the following information is given Town A: Females were 40%, total coffee drinkers were 45% and female non coffee drinkers were 20%. Town B: Males were 55%, male non coffee drinkers were 30% and female coffee drinkers were 15% Present the data into a table format | |
| C | Explain the following Point Estimation Properties with example i) Consistency ii) Unbiasedness | |

| Q3 (20 Marks) | Solve any Two Questions out of Three | 10 marks each |
|----------------------|---|---------------|
| A | What is Hypothesis testing? Explain i) Z-Test for single mean ii) Z-Test for Difference of Mean | |
| B | Perform simple linear regression , Determine slope and intercept | |

The data with regard to the output of gram and cost of seed and labour per hectare at eight farmers' fields, are as given below:

| Sr. No. | Cost of produce (Y) (Rs./hectare) | Cost of Seed (X1) (Rs./hectare) | Cost of Labour (X2) (Rs./hectare) |
|---------|--------------------------------------|------------------------------------|--------------------------------------|
| 1 | 190 | 50 | 10 |
| 2 | 50 | 30 | 10 |
| 3 | 300 | 150 | 15 |
| 4 | 100 | 50 | 15 |
| 5 | 150 | 40 | 20 |
| 6 | 90 | 40 | 10 |
| 7 | 300 | 100 | 35 |
| 8 | 120 | 60 | 14 |

- a) Fit a regression $\hat{y} = a + b_1x_1 + b_2x_2$
- b) Find the coefficient of multiple determination (R^2).
- c) Also test the significance of regression (Given the appropriate Table value, $F = 13.27$, for a significance level of $\alpha = 0.01$)

| Q4 (20 Marks) | Solve any Four Questions out of Six | 05 marks each |
|----------------------|---|---------------|
| A | What is Stratified sampling? Explain the merits and limitations of Stratified sampling. | |
| B | Explain the following methods to check the performance of Regression Model i) MAE ii) MAPE | |
| C | In a trivariate distribution, the simple coefficients of correlation are as follows: If $r_{12} = 0.86$, $r_{13} = 0.65$ and $r_{23} = 0.72$, calculate the coefficient of partial correlation $r_{12.3}$. | |
| D | What is diagrammatic representation of data? Explain its advantages. | |
| E | The manufacturer of a certain make of electric bulbs claims that his bulbs have a mean life of 25 months with standard deviation of 5 months. A random sample of 6 such bulbs gave the following values Life of bulb in months 24,26,30,20,20,18 Is the manufacturer's claim valid at 1% level of significance?(Given that the table values of the appropriate test statistics at said level are 4.032,3.707 and 3.499 for 5, 6 and 7 degree of freedom respectively) | |
| F | Explain method of maximum likelihood with its advantages and disadvantages | |