Printed Page:-05 **Subject Code**:- B**CSBS0201**

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: II - THEORY EXAMINATION - (2024- 2025)

Subject: Statistical Methods & Modelling

Time: 3 Hours **Max. Marks: 100**

General Instructions:

IMP: *Verify that you have received the question paper with the correct course, code, branch etc.*

1. *This Question paper comprises of* ***three Sections -A, B, & C.*** *It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.*

2. *Maximum marks for each question are indicated on right -hand side of each question.*

3. *Illustrate your answers with neat sketches wherever necessary.*

4. *Assume suitable data if necessary.*

5. *Preferably, write the answers in sequential order.*

6. *No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.*

SECTION-A 20

1. Attempt all parts:-

1-a. Choose the correct option regarding the sampling method? (CO1, K2) 1

(a) The sample is the population's part

(b) It helps in determining sampling error

(c) Sampling saves money, time, and energy

(d) All these options are correct

1-b. What refers to elements from where you choose the samples for the research?(CO1, K2) 1

(a) Infinite population

(b) Finite population

(c) Sampling population

(d) Target population

1-c. Which of the following is true for the coefficient of correlation? (CO2, K2) 1

(a) The coefficient of correlation is not dependent on the change of scale.

(b) The coefficient of correlation is not dependent on the change of origin.

(c) The coefficient of correlation is not dependent on both the change of scale and change of origin.

(d) None of the above

1-d. Regression modelling is a statistical tool for building a mathematical equation depicting how (CO2, K2) 1

(a) One explanatory and one or above response variables are related

(b) There is a link between one response variable and one or many explanatory variables

(c) Several explanatory and response variables are related

(d) All of the above are correct.

1-e. Which of the following statements are true? (CO3, K2) 1

a) The sum of the deviations from mean (ignoring algebraic signs) is greater than the sum of the deviations from median (ignoring algebraic signs)

b) Standard deviation is independent of change of origin and change of scale

c) In a symmetrical distribution, mean deviation equals 4/5 of standard deviation

d) In a symmetrical and bell-shaped distribution quartile deviation is 1/3 of standard deviation

Choose the correct answer from the options given below:

(a) b) and d)

(b) a) and c)

(c) c) and d)

(d) a) and b)

1-f. Criteria to check a point estimator to be good are (CO3, K2) 1

(a) Degrees of freedom

(b) The t-ratio

(c) Standard Error of the means

(d) All of the above

1-g. What is the meaning of the testing of the hypothesis? (CO4, K2) 1

(a) It is a significant estimation of the problem

(b) It is a rule for acceptance or rejection of the hypothesis of the research problem

(c) It is a method of making a significant statement

(d) None of the above

1-h. While performing Kruskal-Wallis test, the ranks are assigned: (CO4, K2) 1

(a) Independently to the observation for each treatment

(b) for observations in each block independently

(c) by pooling all the observations

(d) None of the above

1-i. Time series does not relate to (CO5) 1

(a) Seasonal variations

(b) Secular trends

(c) Cross sectional variations

(d) Cyclical variations

1-j. For an autoregressive process to be considered stationary (CO5, K2) 1

(a) The roots of the characteristic equation must all lie inside the unit circle

(b) The roots of the characteristic equation must all lie on the unit circle

(c) The roots of the characteristic equation must all lie outside the unit circle

(d) The roots of the characteristic equation must all be less than one in absolute value

2. Attempt all parts:-

2.a. What is population? (CO1, K1) 2

2.b. What is multiple regression? (CO2, K1) 2

2.c. Define statistical inference. (CO3, K1) 2

2.d. What is alternate hypothesis? (CO4, K1) 2

2.e. What is differencing? (CO5, K1) 2

SECTION-B 30

3. Answer any five of the following:-

3-a. Write three merits and demerits of stratified sampling.(CO1, K1) 6

3-b. A class of five pupils took an exam, and the scores were 12, 55, 74, 79, and 90. Determine the standard error. (CO1, K3) 6

3-c. Calculate Karl Pearson's coefficient of correlation from the following data and interpret its value: (CO2, K3) 6

Roll No.:	1	2	3	4	5
Marks in Accountancy:	48	35	17	23	47
Marks in statistics:	45	20	40	25	45

3-d. From the data given below, find the number of items (n): 6

$r = 0.5, \sum xy = 120, \sigma_y = 8, \sum x^2 = 90$

where x and y are deviations from arithmetic mean. (CO2, K3)

3.e. In a hospital, 480 females and 520 male babies were born in a week. Do these figures confirm the hypothesis that males and females are born in equal number? (CO3, K3) 6

3.f. A factory is producing 50,000 pairs of shoes daily. From a sample of 500 pairs, 2% were found to be of sub-standard quality. Estimate the number of pairs that can be reasonably expected to be spoiled in the daily production and assign limits at 95% level of confidence. (CO4, K3) 6

3.g. Explain Dickey-Fuller test with example. (CO5, K2) 6

SECTION-C

4. Answer any one of the following:- 50

4-a. Given Data below, calculate the standard error of the estimate (residual standard error) using the fitted regression line (CO1, K3) 10

X	1	2	3	4
y	2	4	5	7

4-b. Define the following- (CO1, K1) 10

i. Cluster sampling

ii. Snowball sampling

iii. Stratified sampling

iv. Quota sampling

5. Answer any one of the following:-

5-a. The following data, based on 450 students, are given for marks in statistics and economics at a certain examination. 10

Mean marks in statistics 40
Mean marks in Economics 48
S.D. of marks in Statistics 12
The variance of marks in Economics 256
The sum of the products of deviation of marks from their respective mean is 42075.
Give the equations of the two lines of regression and estimate the average marks in Economics of a candidate who obtained 50 marks in statistics. (CO2, K3)

5-b. Calculate the least squares regression line Y = a + b X , for following data:(CO2, K3) 10

X: 1, 2, 3, 4, 5
Y: 2, 4, 5, 4, 5

6. Answer any one of the following:-

6-a. $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution $N(\mu, \sigma^2)$, where both μ and σ^2 are unknown. Using the method of moments, derive the moment estimators for the population mean μ and variance σ^2.(CO3, K3) 10

6-b. Let $x_1, x_2, \ldots, x_n$ be independent and identically distributed observations from a Bernoulli distribution with parameter p, i.e., $P(X = 1) = p$, $P(X = 0) = 1 - p$. Show that the $T = (1 / n(n-1)) \times \Sigma x_i(\Sigma x_i - 1)$ is an unbiased estimator of p^2. (CO3, K3) 10

7. Answer any one of the following:-

7-a. Ten students' scores before and after a new teaching method are as follows. Use the Sign Test at the 5% level of significance to test whether the new teaching method has a significant effect on student performance.(CO4, K3) 10

Before: 70, 68, 75, 80, 65, 72, 78, 66, 74, 69
After: 75, 65, 78, 80, 70, 76, 77, 70, 74, 73.

7-b. A company claims that its LED bulbs last 1000 hours on average. A consumer group tests 50 bulbs and finds the sample mean is 980 hours. The population standard deviation is known to be 60 hours. Test the company's claim at the 5% level of significance using a two-tailed Z-test. (Z-critical value: **±1.96**) (CO4, K3) 10

8. Answer any one of the following:-

8-a. Explain the concept of stationarity. Write applications of time series. (CO5, K2) 10

8-b. In time series forecasting, write the characteristics of a model. Explain the application of forecasting in business. (CO5, K2) 10

.