2506E777

Candidate's Seat No:____

M.C.A. Sem.-2 Examination

Data Analytics
June-2025

Date ; 25/6/2025, Wednesday
Time : 3.00 Hours]

[Max.Marks : 50

Instructions:

- **Write both the Sections in the separate answer book.**
- **Both Sections having equal weightage.**
- **Draw Diagrams wherever necessary.**
- **Make Assumptions wherever necessary.**

SECTION – I

Q-1 A department of transportation's study on driving speed (miles per hour) and miles per gallon for midsize automobiles resulted in the below data: [10]

Speed	30	50	40	55	30	25	60	25	50	55
Miles	28	25	25	23	30	32	21	35	26	25

a. Develop a scatter diagram with x on the horizontal axis.
b. What does the scatter diagram indicate about the relationship between the two variables?
c. Compute and interpret the sample covariance.
d. Compute and interpret the correlation coefficient.

Q-2 Answer the following: **(Any Two)** [10]

(a) What is normalization? Explain its importance in data preprocessing and discuss any two normalization techniques.

(b) Write a Python program to create a Histogram using data stored in lists and array.

(c) Given the following dataset: 46 44 42 46 32
a. Compute the Z-scores for each data point.
b. Identify any outliers in the dataset based on the Z-scores.

Q-3 Answer the following: [5]

1. Which Python library is commonly used for statistical operations?
   a. Matplotlib b. Scipy
   c. Seaborn d. Open CV
2. Convert categorical data into numerical form using one-hot encoding technique.
3. Some hotels ask their guests to rate the hotel's services as excellent, very good, good, and poor. This is an example of the
   a. ordinal scale b. ratio scale
   c. nominal scale d. interval scale
4. Define: Quartiles
5. Mention one function to create a line chart in Python.

P.T.O

E-777-2

SECTION – II

Q-4 Given are five observations for two variables, x and y. [10]

x	3	12	6	20	14
y	55	40	55	10	15

a. Develop a scatter diagram for these data. What is the relationship between the two variables?
b. Develop the estimated regression equation by computing the values of b_0 and b_1.
c. Use the equation to predict the value of y when $x = 10$.
d. Compute SST, SSR and SSE.
e. Compute the coefficient of determination r^2.

Q-5 Answer the following: **(Any Two)** [10]

(a) Airline passengers arrive randomly and independently at the passenger-screening facility at a major international airport. The mean arrival rate is 10 passengers per minute.
a. Compute the probability of no arrivals in a one-minute period.
b. Compute the probability that three or fewer passengers arrive in a one-minute period.
c. Compute the probability of at least one arrival in a 15-second period.

(b) In a survey of MBA students, the data were obtained on "students' first reason for application to the school in which they matriculated."

Enrolment Status		Reason for Application		
		School Quality	Cost or Convenience	Other
	Full Time	421	393	76
	Part Time	400	593	46

a. If a student goes full time, what is the probability that school quality is the first reason for choosing a school?
b. If a student goes part time, what is the probability that cost and convenience is the first reason for choosing a school?

(c) Consider the following hypothesis test:

$$H_0: \mu \geq 80$$
$$H_a: \mu < 80$$

A sample of 100 provided a sample mean of 77 and $\sigma = 12$.
a. Compute the value of the test statistic.
b. What is the p-value and at $\alpha = 0.05$ what is your conclusion?

Q-6 Answer the following: [5]

1. The expected value for a binomial distribution is _____.
2. Which library in Python is used for regression modeling?
   a. Sklearn b. pandas
   c. matplotlib d. scipy
3. Which method is not a smoothing technique?
   a. Moving Average b. Weighted Average
   c. Regression d. Exponential Smoothing
4. Name a Python function for performing a Z-test.
5. What is the difference between discrete and continuous probability distributions?