PRACTICE PAPER SET 2

Statistics for AI & Data Science (Paper Code: 48895)

T.E. Computer Engineering & AI-DS, Semester V

Duration: 3 Hours | Total Marks: 80

By: Nitin Gupta

Instructions to Candidates:

1. Question No. 1 is compulsory
2. Attempt any three questions from remaining five questions
3. All questions carry equal marks
4. Assume suitable data, if required and state it clearly

Q1. Attempt any FOUR: [20 Marks]

a. [5 Marks]

Calculate the z-score if sample mean = 180, population mean = 200, standard deviation = 25, and sample size n = 30.

b. [5 Marks]

Explain null and alternative hypothesis with an example.

c. [5 Marks]

What is the use of scatter plot and box plot? Explain with examples.

d. [5 Marks]

Write a short note on Simple Linear Regression.

e. [5 Marks]

What is a histogram? Give its applications.

f. [5 Marks]

How to detect outliers? Explain any one method.

Q2. [20 Marks]

a. [10 Marks]

Explain Binomial distribution in detail. An insurance agent sells life insurance policies to five equally aged healthy people. The probability of a person living for 30+ years is 3/4. Calculate the probability that after 30 years:

1. All five people are still living
2. At least three people are still living
3. Exactly two people are still living

b. [10 Marks]

Explain Normal and Poisson Distribution. Calls arrive at a customer service center at an average rate of 240 per hour.

1. Find probability that no calls arrive in a given minute
2. What is expected number of calls in 2 minutes?
3. Find probability that exactly 8 calls arrive in 2 minutes

Q3. [20 Marks]

a. [10 Marks]

For a certain computer battery, the time between charges is normally distributed with mean = 60 hours and standard deviation = 12 hours. A user wants to know:

1. Probability that battery lasts between 60 and 75 hours
2. Probability that battery lasts more than 70 hours
3. Battery life value that 95% of batteries exceed

b. [10 Marks]

A researcher claims that average time students spend on social media is 5 hours per day with SD = 1.5 hours. A sample of 50 students showed mean of 5.5 hours. Test at $\alpha = 0.05$:

1. State hypotheses
2. Calculate test statistic
3. Make decision and interpret

Q4. [20 Marks]

a. [10 Marks]

Explain briefly why ANOVA is used. Give difference between one-way and two-way ANOVA. Solve the following using one-way ANOVA:

Group 1	Group 2	Group 3
65	72	80
70	68	85
68	75	78
72	80	88
67	70	82

b. [10 Marks]

Explain Mean, Standard Deviation, Standard Error and Confidence Interval. Find these for the following data:

Data Point	Value
1	6
2	4
3	5
4	7
5	8
6	5
7	9
8	6
9	7
10	8

Calculate 95% confidence interval.

Q5. [20 Marks]

a. [10 Marks]

What is F-Test? A researcher conducted experiments on two different fertilizers. The F-statistic obtained is 3.45 with degrees of freedom (4, 20). Determine whether to reject null hypothesis at:

1. 5% significance level
2. 1% significance level

Use F-table and explain decision-making process.

b. [10 Marks]

Find the simple linear regression equation for the data below. Also calculate coefficient of determination:

Experience (X) years	Salary (Y) in ₹1000
1	25
3	35
5	45
7	55
9	65
11	72

Predict salary for 6 years of experience.

Q6. [20 Marks]

a. [10 Marks]

Explain how Nonparametric methods are different from parametric methods. A department monitored water quality before cleanup, 1 week later, and 2 weeks after cleanup at 8 locations. Use Friedman Test to determine if cleanup was effective:

Location	Before	Week 1	Week 2
1	25.5	22.3	20.1
2	28.2	26.1	24.5
3	30.5	28.2	26.8
4	22.8	21.5	20.2
5	26.3	24.1	22.9
6	24.7	22.9	21.5
7	29.1	27.3	25.8

Location	Before	Week 1	Week 2
8	27.5	25.8	24.2

b. [10 Marks]

Write short notes on (any two):

1. QQ Plots and their interpretation
2. Type I and Type II errors with examples
3. Stem & Leaf Plot
4. Least Square Regression

END OF PAPER

Best of luck!

Compiled by: Nitin Gupta