PRACTICE PAPER SET 3

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]

Explain percentiles and boxplots with example. Draw a neat diagram.

b. [5 Marks]

Illustrate central limit theorem with a neat diagram.

c. [5 Marks]

Define standard deviation and interquartile range with examples.

d. [5 Marks]

What is the empirical CDF function? Explain with example.

e. [5 Marks]

Write a short note on Random sampling methods.

f. [5 Marks]

Discuss Bootstrapping vs. re-sampling techniques.

Q2. [20 Marks]

a. [10 Marks]

Find the correlation coefficient from the following data and interpret the result:

Product	Price (X) in ₹100	Units Sold (Y)
1	10	95
2	15	85
3	20	75
4	25	70
5	30	60
6	35	55

Explain the relationship between price and units sold.

b. [10 Marks]

What is the primary purpose of conducting one-way ANOVA? Explain key components. A researcher wants to compare productivity scores of employees under three different management styles:

Style A	Style B	Style C
92	85	78
88	90	82
95	82	85
90	88	80
93	85	83

Test at $\alpha = 0.05$ if there's significant difference.

Q3. [20 Marks]

a. [10 Marks]

Find standard error of estimate for average monthly expenses in households. Find 95% confidence interval:

Household	Monthly Expense (₹1000)
1	15
2	18
3	12
4	20
5	16
6	14
7	19
8	17
9	13
10	21

b. [10 Marks]

What is the concept of correlation in statistics? How is it different from regression? Give practical examples from AI/Data Science applications.

Q4. [20 Marks]

a. [10 Marks]

A quality control manager monitors defect rates. The speeds are normally distributed with mean = 100 units/hour and standard deviation = 15 units/hour. Find:

1. Probability of production rate more than 120 units/hour
2. Probability of production rate between 90 and 110 units/hour
3. Production rate that 90% of machines exceed

b. [10 Marks]

Explain Numerical and Categorical data types with appropriate examples from real-world datasets. How do statistical methods differ for these data types?

Q5. [20 Marks]

a. [10 Marks]

A manufacturing company claims average product lifetime is 400 days. A researcher randomly selected 25 products and tested them. Mean lifetime was 380 days with SD = 60 days. Test at $\alpha = 0.05$:

1. State null and alternative hypotheses
2. Calculate test statistic
3. Determine if claim is supported

b. [10 Marks]

Explain linear least square regression (LLSR) along with its advantages and disadvantages. Find regression equation for:

Input Feature (X)	Output (Y)
2	15
4	22
6	28
8	35
10	42
12	48

Calculate R^2 and interpret.

Q6. [20 Marks]

a. [10 Marks]

Explain the concept of two-way ANOVA. How does it differ from one-way ANOVA? Describe assumptions of two-way ANOVA and how to check them. Briefly explain Friedman's test as non-parametric alternative.

b. [10 Marks]

A fitness trainer is testing a new workout technique. Average weight loss with standard method is 5 kg with SD = 2 kg. With new technique, 30 participants showed mean weight loss of 6 kg. Test at $\alpha = 0.05$ if new technique is more effective:

1. State hypotheses (one-tailed test)
2. Calculate z-score

3. Make decision using critical value method
4. Make decision using p-value method
5. Interpret results

END OF PAPER

Best of luck!

Compiled by: Nitin Gupta