



Daffodil International University

Faculty of Science & Information Technology

Department of Computer Science & Engineering

Mid-Term Examination, Fall 2025

Course Code: STA 101, Course Title: Statistics and Probability

Level: 3 Term: 1 Batch: 65

Time: 01:30 Hrs

Marks: 25

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1.	a)	List all level of measurement with appropriate example.	2	CO1																		
	b)	A data scientist wanted to study the income distribution among families in a metropolitan city. To collect the data, the researcher personally surveyed 200 households using a structured questionnaire, asking about their monthly income. Most families earned between \$2,000 and \$5,000 per month, but a few business tycoons reported incomes above \$100,000 per month. Clarify is the data collected source primary or secondary? Explain which measure of central tendency is most appropriate in this case? Interpret if the researcher surveyed only the 200 households, would the calculated mean or median be a statistic or a parameter?	3																			
2.	a)	Customer satisfaction ratings collected from a survey on a 5-point Likert scale. Like 1 for Very Dissatisfied, 2 for Dissatisfied, 3 for Neutral, 4 for Satisfied and 5 for Very Satisfied. The observations are 3, 4, 4, 2, 5, 4, 3, 3, 5, 4, 2, 3, 4, 5, 5, 1, 3, 4, 2, 3, 5, 4, 3, 4, 2, 3, 5, 4, 3, 2, 5, 3, 4, 5, 1. i. Draw an appropriate graph from the above data. ii. Calculate Median from the above data if possible, if not then why?	3+2	CO2																		
	b)	A researcher wanted to study the monthly expenditure of students in a local school to understand their spending patterns. The researcher surveyed students and recorded the following amounts (in Taka): 7,500, 5,000, 12,000, 8,500, 30,000, 5,500, 9,500, 6,500, 4,000, 11,000, 7,000, 12,500, 10,000, 4,500, 11,500, 40,000, 8,000, 9,000, 10,500, 6,000. i. Draw Boxplot from the above data and detect outliers if any. ii. Explore how much the mean differ after trimming outliers.	4+1																			
3.	a)	A school researcher wanted to examine whether the number of hours students' study hours per day affects their test scores (out of 100). The researcher collected data from 8 students in a class. <table border="1"><tr><td>Study Hours</td><td>8</td><td>10</td><td>7</td><td>9</td><td>6</td><td>11</td><td>5</td><td>12</td></tr><tr><td>Test Score</td><td>75</td><td>82</td><td>68</td><td>78</td><td>65</td><td>85</td><td>60</td><td>90</td></tr></table> i. Compare which variable is more consistent? ii. Examine the shape of study hour by calculating skewness and kurtosis. iii. Corelate the two variables and identify the direction and strength of the relationship	Study Hours	8	10	7	9	6	11	5	12	Test Score	75	82	68	78	65	85	60	90	4+3+3	CO3
Study Hours	8	10	7	9	6	11	5	12														
Test Score	75	82	68	78	65	85	60	90														

Formula

Measures of Central Tendency

Mean

$$A.M = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$W.M = \bar{x} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{w_1 + w_2 + \dots + w_n}$$

$$G.M = \bar{x} = ((x_1 \cdot x_2 \cdot x_3 \dots x_n))^{1/n}$$

$$H.M = \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

Median

If "n" is odd, $M_e = X_{(n+1)/2}$

If "n" is even, $M_e = \frac{1}{2} (X_{n/2} + X_{\frac{n}{2}+1})$

Measure of Dispersion

Range = $X_{max} - X_{min}$

Mean Deviation, M.D = $\frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$

Population variance

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

Population standard deviation,

$$\sqrt{\sigma^2}$$

Sample variance

$$s^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{n-1}$$

Population standard deviation,

$$\sqrt{s^2}$$

Coefficient of variation for

population, C.V = $\frac{\sigma}{\mu} \times 100$

Coefficient of variation for

sample, C.V = $\frac{s}{\bar{x}} \times 100$

Measures of Location

Quartile

$$Q_i = \frac{i \times n}{4}$$

Deciles

$$D_i = \frac{i \times n}{10}$$

Percentile

$$P_i = \frac{i \times n}{100}$$

Inner fence:

$$Q_1 - 1.5 \times IQR$$

&

$$Q_3 + 1.5 \times IQR$$

Outer fence:

$$Q_1 - 3 \times IQR$$

&

$$Q_3 + 3 \times IQR$$

Shape of the distribution

Coefficient of Skewness,

$$Sk = \frac{3 \times (\text{Mean} - \text{Median})}{\text{Standard deviation}}$$

$$\text{Kurtosis } \beta_2 = \frac{\mu_4}{\mu_2^2}$$

Correlation Analysis

Correlation Coefficient:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

or,

$$r = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$