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.	**4+3+3**	**CO3**

Study Hours	8	10	7	9	6	11	5	12
Test Score	75	82	68	78	65	85	60	90

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

Formula

Measures of Central Tendency

Mean

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

W.M= $\bar{x}=\frac{w_1x_1+w_2x_2+\cdots+w_nx_n}{w_1+w_2+\cdots+w_n}$

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

H.M=$\frac{n}{\frac{1}{x1}+\frac{1}{x2}+\cdots+\frac{1}{xn}}$

Median

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

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

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=$\frac{\sigma}{\mu}\times100$

Coefficient of variation for sample, C.V=$\frac{s}{\bar{x}}\times100$

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,

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

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×IQR

&

Q_3 +1.5×IQR

Outer fence:

Q_1- 3×IQR

&

Q_3 + 3×IQR

Shape of the distribution

Coefficient of Sknewness,

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

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