Daffodil International University
Faculty of Science & Information Technology
Department of Software Engineering
Midterm Examination, Spring 2026
Course Code: DS411; Course Title: Statistical Data Analysis
Sections & Teachers: A (FA), B (NAJ), C (SAM)

Time: 1:30 Hours | 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.	The Election Commission of Bangladesh plans to conduct a nationwide survey to assess voter awareness and satisfaction with the electoral process. The voting population is large and geographically distributed across divisions, districts, upazilas, and polling areas. Due to constraints related to time, cost, and manpower, it is not feasible to survey voters from every polling center in the country. As a data analyst, you are required to recommend the most suitable sampling technique for conducting this survey. Identify the most suitable sampling method for this scenario and **justify** your choice by explaining how the sampling process would be conducted. Also, demonstrate the advantages and disadvantages of different sampling techniques in this context to justify your answer.	Marks-10	CO-1 *Level-5*
2.	A public health institute is conducting a study to examine the relationship between daily screen time and sleep quality scores among university students. The students are categorized into two groups based on their primary device usage: Mobile Users and Laptop Users. The dataset below shows average daily screen time (in hours) and corresponding sleep quality scores (higher score indicates better sleep quality).	[Marks- 5 +5 = 10	CO-2 *Level-5*

Student ID	Device Type	Screen Time (Hours)	Sleep Quality Score
S1	Mobile	6	62
S2	Mobile	5	65
S3	Mobile	7	58
S4	Mobile	4	68
S5	Laptop	3	72
S6	Laptop	4	70
S7	Laptop	5	66
S8	Laptop	2	75

	The institute wants to analyze both the distribution of sleep quality scores and the relationship between screen time and sleep quality across different device types. **a) Evaluate** the sleep quality scores using appropriate descriptive statistics (Standard deviation) separately for each category; and analyze the relationship between daily screen time and sleep quality score by calculating and interpreting the correlation coefficient. **b)** Draw and **interpret** the most appropriate data visualization techniques (e.g., box plots and scatter plots) to effectively compare the distribution of sleep quality scores between Mobile and Laptop users and to illustrate the relationship between daily screen time and sleep quality score.		
3.	A government agency conducts a study to evaluate public satisfaction with a newly introduced digital service. Data is collected through an online questionnaire distributed via the agency's official website and ~~social media~~ platforms. The results of the study are intended to guide future policy decisions. **a) Explain** the ethical or social concerns associated with the data collection method used in this study. **b) Identify** and explain one type of bias that may arise from this data collection approach and briefly suggest how it could be reduced.	[Marks-3+2=5]	CO-3 *Level-4*