2023

STATISTICS — MINOR

Paper : MN-1

(Descriptive Statistics I and Probability I)

Full Marks : 75

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

1. Answer ***any five*** questions : 2×5

(a) Define statistics and state briefly its primary objective.

(b) Differentiate between ordinal and nominal data.

(c) Provide an example of how Bayes' theorem can be applied in real-life scenarios.

(d) A bag contains 4 red marbles and 3 blue marbles. If one marble is randomly selected from the bag, what is the probability that it is either red or blue?

(e) Name two advantages of presenting data in a tabular format.

(f) Define statistical population.

(g) A fair six-sided die is rolled twice. What is the probability that the first roll is odd and the second roll is even?

(h) State the theorem of total probability.

2. Answer ***any four*** questions : 5×4

(a) Suppose you have three events, A, B, and C with probabilities P(A) = 0.4, P(B) = 0.3, P(C) = 0.5 and P (A ∩ B) = 0.1, P(A ∩ C) = 0.2, P (B ∩ C) = 0.2 and P (A ∩ B ∩ C) = 0.05. Calculate :

(i) P (A ∪ B)

(ii) P ($A^C \cap B^C \cap C^C$)

(iii) P (A ∪ (B ∩ C)).

(b) Interpret the stem-and-leaf plot provided below, which represents the test scores of students in a class :

Stem	Leaf
2	1, 4
3	0, 2, 4, 5, 6
4	0, 2, 3, 4, 5, 7

Please Turn Over

(c) A factory produces light bulbs, and it is known that 5% of the bulbs are defective. If a box contains 10 bulbs and 2 of them are randomly selected for inspection, what is the probability that both bulbs are defective given that the first bulb inspected was defective?

(d) Discuss the limitations of using the mean as a measure of central tendency in skewed distributions. Suggest some modifications over mean to tackle the situations.

(e) Give the classical definition of probability and provide an example to illustrate its application.

(f) A researcher collected data on the heights (in inches) of 50 individuals. The data is as follows :

{64, 67, 70, 68, 65, 63, 71, 69, 72, 66, 68, 64, 70, 67, 69, 65, 71, 68, 66, 63, 70, 67, 69, 65, 68, 64, 71, 69, 67, 69, 66, 63, 70, 68, 72, 67, 69, 65, 63, 71, 68, 66, 64, 70, 67, 69, 66, 63, 70, 68}. Construct a frequency distribution table for this data taking at least five class intervals and determine the mode.

3. Answer ***any three*** questions :

(a) (i) Define a random experiment and provide an example. Discuss its sample space.

(ii) Two six-sided dice are rolled. Determine the sample space and find the probability of obtaining a sum of 7.

(iii) Consider a spinner divided into 8 equal sectors labeled with the numbers 1 through 8. If the spinner is spun once, what is the probability of landing on a multiple of 3?

(iv) In a class of 30 students, 18 study math and 15 study science. If a student is randomly selected, find the lower limit of the probability that the student studies both math and science. (3+2)+4+3+3

(b) (i) Define the Gini coefficient and explain its significance in measuring income inequality.

(ii) Sketch a Lorenz curve and explain how it is used to interpret the distribution of income.

(iii) Calculate the Gini coefficient for a hypothetical population, where 20% of the population earns 10% of the total income, 40% earns 30% of the total income, and 40% earns 60% of the total income. 5+5+5

(c) (i) Define moments in statistics. Discuss the importance of moments in describing the characteristics of a frequency distribution.

(ii) Calculate the skewness and kurtosis for the following data set : 2, 4, 6, 8, 10. Interpret the results.

(iii) Discuss the advantages and limitations of using the median as a measure of central tendency compared to the mean. (2+3)+(3+2)+5

(d) (i) A company conducted a survey among its employees and found that 60% of them prefer working on weekdays and 40% prefer working on weekends. Of those, who prefer working on weekdays, 75% also prefer working in the mornings. What is the probability that an employee chosen at random prefers working on weekdays and in the mornings?

(ii) State and prove the theorem of total probability.

(iii) In a medical diagnosis, if a test for a rare disease has a false positive rate of 5% and a false negative rate of 1%, and the disease affects 1 in 1000 people, calculate the probability that a person with a positive test result actually has the disease. 5+5+5

(e) (i) Differentiate between range (R) and standard deviation (s) as measures of dispersion and illustrate with an example. Also show that $s^2 \leq R^2/4$.

(ii) A group of students took two exams. The mean score for Exam - 1 was 70 with a standard deviation of 5, and the mean score for Exam - 2 was 75 with a standard deviation of 8. Which exam had greater variability and why?

(iii) Explain how box plots can be used to visualize measures of dispersion and skewness.

(3+4)+3+5
