

**Semester II 2024/2025**

|  |  |  |
| --- | --- | --- |
| Subject | : | PROBABILITY & STATISTICAL DATA ANALYSIS  (SECI 1143/SCST 1223) |
| Task | : | Chapter 3 & Chapter 4 |
| Due Date | : | **Week 9 (12 – 16 May 2025)** |

**INSTRUCTION:**

1. This is a **GROUP** assignment. Please clearly write the group members **NAME & MATRIC NUMBER** in the front page of the submission.
2. This assignment contributes to 5% of overall course marks.
3. Only **HANDWRITTEN** submission is accepted:
   1. Submissions using any reporting or statistical tools (e.g.: MS Word, MS Excel, etc.,) will be **REJECTED**.
   2. Make sure the submission is neatly written. Any submission with handwriting that is unreadable, will be **REJECTED**.
   3. For answer that need to draw graphs, using graph paper is optional. You can use plain paper.
   4. Round your answers to **THREE** decimal places.
   5. Please scan/snapshot your work and save as a PDF file.
4. Submission via eLearning – only **ONE** group member needs to submit on behalf of the group.

**QUESTION 1[12 MARKS]**

Table 1 shows data collected on the heights (in cm) of 60 participants from a Youth Basketball Training Camp.

**Table 1**

|  |  |
| --- | --- |
| **Class Interval** | **Frequency** |
|  | 12 |
|  | 20 |
|  | 5 |
|  | 3 |

Based on the above value, calculate the:

1. Mean

(4 Marks)

1. Median

(4 Marks)

1. Mode

(3 Marks)

1. Modal class

(1 Mark)

**QUESTION 2 [15 MARKS]**

You are a science club coordinator tracking the attendance of 10 club members at a series of hands-on lab sessions. Each student could earn more participation points based on their involvement, and the points collected are as follows:

**85, 90, 75, 88, 92, 80, 85, 82, 90, 85.**

You aim to analyze the central tendencies of the scores to better understand student engagement and identify areas for improvement.

1. Calculate the mean, median, and mode of the participation scores.

(5 Marks)

1. Interpret what these values ***(mean, median, and mode)*** indicate about the participation performance of the science club members. Based on these values, determine which statistic that more appropriate represent the overall summary of the participation scores.

(4 Marks)

1. After reviewing attendance logs, you discover that some participation scores were recorded incorrectly. The corrected scores are:

**55, 65, 65, 70, 85, 95, 95, 95, 100, 100**

You decide to reassess the data to ensure your conclusions reflect the true performance and involvement levels of the students.

1. Calculate new mean, median and mode for the updated dataset.
2. Plot the graph of central tendency.
3. Evaluate the overall participation performance of the science club members based on the **corrected scores**, and compare it with the **original data**.

(6 Marks)

**QUESTION 3 [13 MARKS]**

Suppose you are analyzing the monthly sales data for a small business over the past year. You have collected the total sales figures for each month from January to December. Table 2 shows the sales figures (in thousands of Ringgit Malaysia) are as follows:

**Table 2**

|  |  |  |  |
| --- | --- | --- | --- |
| **Months** | **Sales Figure** | **Months** | **Sales Figure** |
| January | 25 | July | 40 |
| February | 30 | August | 38 |
| March | 28 | September | 33 |
| April | 35 | October | 36 |
| May | 32 | November | 31 |
| June | 27 | December | 29 |

1. Calculate the monthly sales figures for:
   1. Range

(2 Marks)

* 1. Variance

(5 Marks)

* 1. Standard deviation

(1 Mark)

1. Interpret the above calculation to understand the variability in the business's monthly sales performance throughout the year.

(3 Marks)

1. Discuss how this understanding can inform strategic decisions for managing the business's finances and operations.

(2 Marks)

**QUESTION 4 [10 MARKS]**

A company wants to assess the level of effectiveness of a new training program designed to improve employees' productivity. The productivity scores are normally distributed with a mean of 50 units and a standard deviation of 10 units.

1. What percentage of employees in the sample showed an increase in productivity after the training program?
2. What is the probability of employees having a productivity score between 37 and 65 after the training?
3. Suppose that the company has 1000 number of employees and plan to conduct a team building for those who not achieved minimum of score of 20 units. The cost for team building is RM200 per head. How much the company need to estimate the budget for the team building cost.
4. If the company wants to identify the top 5% of employees with the highest productivity scores after the training, what minimum score would an employee need to achieve?

**QUESTION 5 [15 MARKS]**

In the Information Technology course, students take an exam with 6 multiple-choice questions. Each question has 4 options, with only one correct answer. The number of correct answers a student gets on the exam represents their performance in the course.

1. Identify the random variable *X*.
2. Construct up the table for the probability of *X*.
3. What is the mean of this distribution?
4. Determine the probability of a student getting at least 3 correct answers on the exam.
5. What is the probability that the first incorrect answer occurs in the fourth question?

**QUESTION 6 [10 MARKS]**

In a coffee shop, customers arrive at random intervals throughout the day. The shop owner is interested in understanding the arrival patterns to better manage staffing levels. The probability of a customer ordering a cappuccino is 0.70. *X* represents the number of customers that will arrive until the 4th customer who orders a cappuccino.

1. What is the probability distribution of *X*?
2. What is the standard deviation of this distribution?
3. Determine the probability that exactly 6 customers will arrive until the 4th customer who orders a cappuccino.
4. Determine the probability that exactly 7 cappuccinos will be ordered within the next 12 customer arrivals.