<u>**Descriptive Statistics Assignment**</u>

Easy Level

1. Understanding Central Tendency (Easy)

A bakery tracks the daily sales of muffins (in dozens) over a week: [10, 12, 11, 15, 14, 13, 12]. What is the most representative value of their weekly sales, and why?

Answer: To find the most representative value of the bakery's weekly muffin sales, we look at the measures of central tendency: mean, median, and mode.

Given data (in dozens):

$10, 12, 11, 15, 14, 13, 12$

1. Mean (Average)

$$\textbf{Mean} = \frac{\mathbf{10 + 12 + 11 + 15 + 14 + 13 + 12}}{\mathbf{7}} = \frac{\mathbf{87}}{\mathbf{7}} \approx \mathbf{12.43}$$

2. Median (Middle value)

First, arrange the data in ascending order:

$\mathbf{10, 11, 12, 12, 13, 14, 15}$

The middle (4th) value is:

$\textbf{Median} = \mathbf{12}$

3. Mode (Most frequent value)

- The value 12 appears twice
- All other values appear once

$\textbf{Mode} = \mathbf{12}$

Most Representative Value:

12 dozens

Why?

- The data has no extreme outliers, but there is a slightly higher value (15).

- Both the median and mode are 12, indicating that this value best reflects a *typical day's sales*.
- The mean (12.43) is slightly influenced by higher sales days.

Conclusion:
The median (and mode) = 12 dozen is the most representative value of the bakery's weekly sales because it best reflects the typical daily sales without being influenced by higher values.

2. Mean in Real Life (Easy)

A teacher records the marks of her students in a short quiz: [12, 15, 14, 16, 18, 20, 19]. What is the mean score, and what does it tell us about the class's performance?

Answer: Mean in Real Life

Given quiz marks:

$$[12, 15, 14, 16, 18, 20, 19]$$

1. Calculation of Mean

The mean (average) is calculated as:

$$\text{Mean} = \frac{12 + 15 + 14 + 16 + 18 + 20 + 19}{7}$$

$$= \frac{114}{7} \approx 16.29$$

2. Interpretation (What it tells us)

- The mean score is approximately 16.3 marks.
- This represents the average performance of the class in the quiz.
- It suggests that, on average, students scored around 16 out of the total marks.
- Since most scores are clustered between 14 and 20, the class performance is generally good, with no extremely low scores pulling the average down.

Conclusion:

The mean score (≈ 16.3) indicates that the overall class performance in the quiz was above average and fairly consistent.

3. Mode in Real Life (Easy)

A store records the shoe sizes sold in one day: [7, 8, 9, 8, 8, 10, 7, 9]. What is the mode, and why is this information useful for the store manager?

Answer: Mode in Real Life

Given shoe sizes sold:

$[7, 8, 9, 8, 8, 10, 7, 9]$

1. Finding the Mode

Count the frequency of each shoe size:

- **Size 7 → 2 times**
- **Size 8 → 3 times**
- **Size 9 → 2 times**
- **Size 10 → 1 time**

Mode = 8

2. Why this is Useful for the Store Manager

- **The mode (8) represents the most commonly sold shoe size.**
- **It helps the store manager:**
  - Stock more shoes of size 8 to meet customer demand.
  - Reduce the risk of stock-outs for popular sizes.
  - Optimize inventory planning and ordering decisions.

Conclusion:

The mode is 8, and it is useful because it tells the store manager which shoe size is most in demand, helping improve sales and inventory management.

Medium Level

4. Median in Real Life (Medium)

A car dealer notes the prices of used cars: [$8,000, $9,500, $10,200, $11,000, $50,000]. Why is the median a better measure than the mean in this case? Calculate the median.

Answer: Median in Real Life

Given used car prices:

[8,000, 9,500, 10,200, 11,000, 50,000]

1. Calculating the Median

The prices are already in ascending order.

- Number of observations = **5** (odd number)
- The median is the **middle (3rd) value**

$\text{Median} = \$10{,}200$

2. Why Median Is Better Than Mean Here

- The price **$50,000** is an **extreme outlier** compared to the other car prices.
- The **mean** would be pulled **upward** by this unusually high value.
- The **median** is **not affected by extreme values**, so it better represents the **typical price** of a used car.

Conclusion:

Median = $10,200

The median is a better measure than the mean in this case because it gives a more realistic and representative value of typical used car prices when outliers are present.

5. Dispersion Introduction (Medium)

A student times how long it takes to finish a puzzle each day: [25, 30, 27, 35, 40]. What does the range tell us about the variation in the student's puzzle-solving time?

Answer:

6. Range in Action (Medium)

A farmer records the weekly weight of harvested apples (kg): [100, 105, 98, 110, 120]. Find the range. How can this help the farmer in planning his packaging?

Answer: Dispersion Introduction

Given weekly apple weights (kg):

[100, 105, 98, 110, 120]

1. Finding the Range

The **range** is calculated as:

$$\text{Range} = \text{Maximum value} - \text{Minimum value}$$

- Maximum = **120 kg**
- Minimum = **98 kg**

$$\text{Range} = 120 - 98 = 22\ \text{kg}$$

2. How This Helps the Farmer in Planning Packaging

- The range shows the **variation in harvest size** from week to week.
- A range of **22 kg** tells the farmer that weekly harvests can **fluctuate significantly**.
- This helps the farmer:
  - Plan **flexible packaging quantities**.
  - Keep extra boxes or crates ready for **larger harvest weeks**.
  - Avoid shortages or wastage due to **under- or over-packaging**.

Conclusion:

Range = 22 kg

Understanding the range helps the farmer anticipate variability in harvest size and plan packaging more efficiently.

7. Variance for Decision-Making (Medium)

Two delivery companies track delivery delays (in minutes). Company A: variance = 6 Company B: variance = 15 Which company is more consistent, and why?

Answer: Variance for Decision-Making

Given:

- Company A: Variance = 6
- Company B: Variance = 15

Which company is more consistent?

Company A is more consistent.

Why?

- Variance measures how spread out the data is around the mean.
- A lower variance means delivery delays are more tightly clustered and more predictable.
- Company A's variance (6) is much lower than Company B's variance (15), indicating:

    - Less fluctuation in delivery delays
    - More reliable and consistent service

Conclusion:

Company A is more consistent because its delivery times vary less, making its performance more predictable and dependable.

Hard Level

8. Standard Deviation in Context (Hard)

A finance student compares the daily price fluctuations of two cryptocurrencies. Coin A: standard deviation = $30 Coin B: standard deviation = $120 Which coin is riskier to invest in, and why?

Answer: Standard Deviation in Context

Given:

- Coin A: SD = $30
- Coin B: SD = $120

Which coin is riskier?

Coin B is riskier.

Why?

- Standard deviation measures how much values deviate from the mean.
- A higher standard deviation means larger fluctuations in price.
- Coin B's SD ($120) is much higher than Coin A's ($30), indicating:
    - Its price changes more drastically day to day
    - Greater potential gain or loss → higher risk for investors

Conclusion:

Coin B is riskier to invest in because its daily price is much more volatile, making it less predictable.

9. Combining Measures (Hard)

A family records their monthly electricity usage (in kWh): [400, 420, 390, 450, 410]. Find the mean and standard deviation. What do these values together tell you about the family's energy use pattern?

Answer: Combining Measures: Mean and Standard Deviation

Given monthly electricity usage (kWh):

$$[400, 420, 390, 450, 410]$$

1. Calculating the Mean

$$\text{Mean} = \frac{400 + 420 + 390 + 450 + 410}{5} = \frac{2070}{5} = 414 \text{ kWh}$$

2. Calculating the Standard Deviation

Step 1: Find deviations from the mean and square them

Value	Deviation (x - mean)	Squared Deviation
400	400 - 414 = -14	196
420	420 - 414 = 6	36
390	390 - 414 = -24	576
450	450 - 414 = 36	1296
410	410 - 414 = -4	16

Step 2: Sum of squared deviations

$$196 + 36 + 576 + 1296 + 16 = 2120$$

Step 3: Divide by n (population SD) or n-1 (sample SD). Assuming sample SD:

$$s^2 = \frac{2120}{5-1} = \frac{2120}{4} = 530$$

Step 4: Standard deviation:

$$s = \sqrt{530} \approx 23.02 \text{ kWh}$$

3. Interpretation

- Mean = 414 kWh → average monthly electricity usage.
- SD ≈ 23 kWh → usage fluctuates moderately around the mean.

Together, they tell us that the family's energy consumption is relatively consistent, with occasional small variations from month to month.

10. Practical Application (Hard)

A basketball player's points in 8 games are recorded: [15, 18, 20, 22, 25, 17, 19, 21]. Find the mean, median, mode, range, and standard deviation. What insights can these measures provide about the player's scoring performance?

Answer: Practical Application: Player's Scoring Performance

Given points per game:

$$[15, 18, 20, 22, 25, 17, 19, 21]$$

1. Mean (Average)

$$\mathbf{Mean} = \frac{15+18+20+22+25+17+19+21}{8} = \frac{157}{8} = 19.625 \approx 19.63$$

2. Median (Middle Value)

Arrange in ascending order:

$15,17,18,19,20,21,22,25$

- **Number of observations = 8 (even)**

$$\text{Median } = \frac{19+20}{2} = 19.$$

3. Mode (Most Frequent Value)

- All values appear once, so no mode.

4. Range

$$\text{Range } = \text{Maximum } - \text{Minimum } = 25 - 15 = 10$$

5. Standard Deviation

Step 1: Find deviations from the mean and square them

Value	Deviation (x - mean)	Squared Deviation
15	15 - 19.625 = -4.625	21.39

17	17 - 19.625 = -2.625	6.89
18	-1.625	2.64
19	-0.625	0.39
20	0.375	0.14
21	1.375	1.89
22	2.375	5.64
25	5.375	28.89

Step 2: Sum of squared deviations

$$21.39 + 6.89 + 2.64 + 0.39 + 0.14 + 1.89 + 5.64 + 28.89 = 67.87$$

Step 3: Sample SD (divide by n-1 = 7)

$$s^2 = \frac{67.87}{7} \approx 9.695$$

$$s = \sqrt{9.695} \approx 3.11$$

6. Insights

- Mean ≈ 19.63 → The player averages ~20 points per game.
- Median = 19.5 → Most typical performance aligns closely with the mean.
- Mode: None → Scores are fairly varied, no single dominant score.
- Range = 10 → Moderate variation between lowest (15) and highest (25) scores.
- SD ≈ 3.11 → Scores are relatively consistent around the average, with small fluctuations from game to game.

Conclusion:
The player is reliably scoring around 20 points per game with moderate consistency and no extreme highs or lows.