Q1: Understanding Central Tendency (Easy)

Q1: Understanding Central Tendency (Easy)

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

The most representative value of the weekly sales, which is the bakery's daily muffin sales (in dozens) over a week: [10, 12, 11, 15, 14, 13, 12], can be identified using measures of **Central Tendency**. Central tendency helps summarise a dataset with one representative value. The key measures are the Mean, Median, and Mode.

1. **Mean (Arithmetic Average)**:
   - **Calculation:** The mean is the sum of all values divided by the number of values.

     Mean = (10+12+11+15+14+13+12/7) =12.43

   - **Why it's representative:** The mean represents the total sales equally distributed across the days. Since the dataset [10, 12, 11, 15, 14, 13, 12] is continuous and does not have extreme outliers, the mean (12.43 dozens) is an excellent representative value.
2. **Median (Middle Value)**:
   - **Calculation:** First, order the dataset: [10, 11, 12, **12**, 13, 14, 15]. The median is the middle value of the ordered dataset. Since there is an odd number of values, the median is the single middle value, which is **12**.
   - **Why it's representative:** The median is the exact centre of the sales data. The median is often used when a dataset has extreme outliers.
3. **Mode (Most Frequently Occurring Value)**:
   - **Calculation:** The mode is the value that occurs most frequently. In the dataset, the value **12** appears twice (more than any other value).
   - **Why it's representative:** The mode shows the most common daily sales figure.

Conclusion:

Both the **Mean (approx 12.43)** and the **Median (12)** are the most representative values for this dataset because the data is continuous and not skewed by extreme outliers. The Mean gives the average sales, and the Median provides the middle sales value.

Q2: Mean in Real Life (Easy)

Q2: 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?

Solution:

1. Calculating the Mean

To determine the mean score, the sum of all the individual marks is divided by the total number of students.

- **Sum of marks: 12 + 15 + 14 + 16 + 18 + 20 + 19 = 114**
- **Number of students: 7**
- **Calculation:**
  Mean = approx 16.29

2. Interpretation of Performance

The calculated mean score is 16.29.

This value represents the central tendency of the class's results. It tells us that the "typical" student scored approximately 16.3 out of 20. Since this average is significantly higher than the midpoint of the scale, it indicates that the class's overall performance on this quiz was strong.

Q3: Mode in Real Life (Easy)

Q3: 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?

Solution:

1. Finding the Mode

To find the mode, the dataset is analysed to identify the value that appears most frequently.

- **Dataset: 7, 8, 9, 8, 8, 10, 7, 9**
- **Frequency Count:**
  - **Size 7: 2 times**
  - **Size 8: 3 times**
  - **Size 9: 2 times**
  - **Size 10: 1 time**

The mode is 8, as it is the shoe size with the highest frequency of sales.

2. Utility for the Store Manager

The mode is particularly useful for the store manager because it identifies the "most popular" item. Unlike the mean, which might result in a non-existent size (e.g., size 8.25), the mode represents a real product that is in the highest demand.

Knowing that size 8 is the mode allows the manager to:

- **Optimise Inventory: Stock more pairs of size 8 to prevent running out of the most requested size.**
- **Prioritise Display: Place the most popular sizes in easily accessible areas.**
- **Forecast Trends: accurately predict which specific products will drive the majority of sales.**

Q4: Median in Real Life (Medium)

Q4: 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.

Solution:

1. Calculating the Median

To find the median, the data is arranged in ascending order to identify the middle value.

- Sorted Data: $[\$8,000, \$9,500, \mathbf{\$10,200}, \$11,000, \$50,000]$
- Total Observations (n): 5
- Calculation: Since there is an odd number of observations, the median is the value in the exact centre (the 3rd position).

The median price is $10,200.

2. Why the Median is a Better Measure

In this specific case, the median is a better measure of central tendency than the mean because the dataset contains a significant outlier.

- The value $50,000 is much higher than the other four values.
- If the mean were calculated, the outlier would skew the result to approximately $17,740. This average would be misleading, as it is higher than the price of four out of the five cars available.
- The median ($10,200) is robust against outliers, providing a much more accurate representation of the "typical" price a customer should expect to pay for a car at this dealership.

Q5: Dispersion Introduction (Medium)

Q5: 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?

Solution:

1. Calculating the Range

The range is the simplest measure of dispersion and is calculated by subtracting the minimum value in the dataset from the maximum value.

- **Dataset: 25, 30, 27, 35, 40**
- **Maximum Value: 40**
- **Minimum Value: 25**
- **Calculation:**
  Range = 40 - 25 = 15

2. Interpretation of Variation

The calculated range is 15 minutes.

This value indicates the total spread of the student's performance. It tells us that there is a 15-minute gap between the student's fastest and slowest attempts. In terms of variation, this suggests that the puzzle-solving speed is not perfectly consistent; the performance fluctuates significantly from day to day rather than remaining steady around a single time.

Q6: Range in Action (Medium)

Q6: 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?

Solution:

1. Finding the Range

To find the range, the lowest value in the dataset is subtracted from the highest value.

- Dataset: [100, 105, 98, 110, 120]
- Maximum Weight: 120 kg
- Minimum Weight: 98 kg
- Calculation:
  Range = 120 - 98 = 22

The range is **22 kg**.

2. Application for Packaging

Calculating the range helps the farmer understand the fluctuation in harvest volume. A range of 22 kg indicates the spread between the lightest and heaviest harvest weeks.

This information assists in planning packaging by:

- **Inventory Management:** It ensures that enough packaging materials (crates or boxes) are always available to accommodate the maximum expected weight of 120 kg, rather than planning based solely on the average.
- **Standardisation:** It helps determine if a single standard box size is sufficient or if variable packaging options are needed to handle the volume shifts efficiently.

Q7: Variance for Decision-Making (Medium)

Q7: Variance for Decision-Making (Medium)

Two delivery companies track delivery delays (in minutes).

- **Company A: variance = 6**
- **Company B: variance = 15**

Solution:

1. Analysing the Variance

Variance is a statistical measure that quantifies the dispersion or spread of a set of data points around their mean value. A lower variance indicates that the data points are clustered closely around the average, while a higher variance indicates that the data points are more spread out.

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

2. Determining Consistency

Company A is more consistent.

This conclusion is based on the fact that Company A has a lower variance (6) compared to Company B (15). In the context of delivery times, a lower variance means that the delivery delays for Company A are relatively stable and predictable. Company B, with a higher variance, experiences wider fluctuations in delay times, making its service less reliable and less consistent.

Q8: Standard Deviation in Context (Hard)

Q8: 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?

Solution:

1. Analysing Standard Deviation Standard deviation is a statistical measure that quantifies the amount of variation or dispersion in a set of values. In the context of finance and investment, standard deviation is the most common metric used to measure **volatility**.

- **Coin A:** Standard Deviation = $30
- **Coin B:** Standard Deviation = $120

2. Determining Investment Risk Coin B is riskier to invest in.

This conclusion is drawn because Coin B has a significantly higher standard deviation ($120) compared to Coin A ($30). A higher standard deviation indicates that the price of Coin B fluctuates much more dramatically from its average price. While this high volatility suggests the potential for higher returns, it implies a much greater level of uncertainty and the possibility of substantial losses in a short period. Coin A, with a lower standard deviation, is more stable and predictable.

Q9: Combining Measures (Hard)

Q9: 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?

1. Calculating the Mean The mean represents the average monthly electricity consumption.

- **Dataset:** [400, 420, 390, 450, 410]
- **Sum of values:** 400 + 420 + 390 + 450 + 410 = 2070
- **Count (n):** 5
- **Calculation:** Mean = 2070 / 5 = 414 kWh

2. Calculating the Standard Deviation The standard deviation measures how much the monthly usage varies from the average. (Using the sample standard deviation formula, n-1).

- **Step A: Find squared deviations from the Mean (414):**
  - (400 - 414)^2 = (-14)^2 = 196
  - (420 - 414)^2 = (6)^2 = 36
  - (390 - 414)^2 = (-24)^2 = 576
  - (450 - 414)^2 = (36)^2 = 1296
  - (410 - 414)^2 = (-4)^2 = 16
- **Step B: Sum of squared deviations:** 196 + 36 + 576 + 1296 + 16 = 2120
- **Step C: Calculate Variance and Standard Deviation:**
  - Variance = 2120 / (5 - 1) = 2120 / 4 = 530
  - Standard Deviation = square root of 530 = 23.02 kWh (approx)

3. Interpretation of Energy Use Pattern Together, these values describe the family's consumption habits:

- **The Mean (414 kWh)** establishes the baseline, predicting that the family will typically use around 414 units of electricity per month.
- **The Standard Deviation (23.02 kWh)** provides context on reliability. Since 23.02 is relatively small compared to the mean (only about 5.5% fluctuation), it indicates that the family's energy usage is **consistent**. They do not experience extreme spikes or drops in consumption; their electricity needs are stable and predictable month-to-month.

Q10: Practical Application (Hard)

Q10: 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?

Solution:

1. Calculations

- **Dataset:** [15, 18, 20, 22, 25, 17, 19, 21]
- **Count (n):** 8

Mean:

- Sum = 15 + 18 + 20 + 22 + 25 + 17 + 19 + 21 = 157
- Mean = 157 / 8 = **19.625**

Median:

- Sorted Data: 15, 17, 18, 19, 20, 21, 22, 25
- Since there is an even number of values, the median is the average of the two middle values (19 and 20).
- Median = (19 + 20) / 2 = **19.5**

Mode:

- All values appear exactly once.
- Result: **No Mode** (or the dataset is uniform).

Range:

- Maximum: 25
- Minimum: 15
- Range = 25 - 15 = **10**

Standard Deviation:

- Step A: Calculate squared deviations from the Mean (19.625).
  - (15 - 19.625)^2 = 21.39
  - (18 - 19.625)^2 = 2.64
  - (20 - 19.625)^2 = 0.14
  - (22 - 19.625)^2 = 5.64
  - (25 - 19.625)^2 = 28.89
  - (17 - 19.625)^2 = 6.89
  - (19 - 19.625)^2 = 0.39
  - (21 - 19.625)^2 = 1.89
- Step B: Sum of squared deviations = 67.87

- Step C: Divide by (n-1) for sample variance = 67.87 / 7 = 9.70
- Step D: Square root of 9.70 = **3.11** (approx)

2. Insights on Performance

These measures provide a comprehensive view of the player's scoring ability:

- **Consistency:** The Standard Deviation (3.11) is relatively low compared to the Mean (19.6). This indicates the player is consistent; they rarely have "bad" games or erratic scoring spikes.
- **Predictability:** The Mean (19.625) and Median (19.5) are almost identical. This symmetry suggests a stable performance where the player reliably scores around 19-20 points per game.
- **Versatility:** The Range (10) shows that while the player is consistent, they are capable of reaching higher scores (25) when needed, but their "floor" (15) is still respectable.