Descriptive Statistics Assignment

(Easy Level Questions)

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

Answer: Concept of Central Tendency

Central tendency refers to the statistical measures that identify the centre or typical value in a dataset. The three main measures are:

- Mean (Average): The arithmetic average of all values.
- Median: The middle value when data is arranged in order.
- Mode: The most frequently occurring value.

These measures help us understand what value best represents the dataset as a whole.

Data Provided

The bakery tracks daily muffin sales (in dozens) over one week: [10, 12, 11, 15, 14, 13, 12]

Tabular Representation in Excel

Days	Muffins (Dozen)
Day 1	**10**
Day 2	**12**
Day 3	**11**
Day 4	**15**
Day 5	**14**
Day 6	**13**
Day 7	**12**

Measures of Central Tendency

- **Mean (Average):** Formula → **=AVERAGE (B2:B8)** Calculation → (10 + 12 + 11 + 15 + 14 + 13 + 12) ÷ 7 = 12.43 dozens
- **Median (Middle Value):** Formula → **=MEDIAN (B2:B8)** Sorted data → [10, 11, 12, 12, 13, 14, 15] Middle value = 12 dozens
- **Mode (Most Frequent Value):** Formula → **=MODE.SNGL(B2:B8)** Most repeated value = 12 dozens

Supporting Measures

- **Minimum:** =MIN (B2:B8) → 10
- **Maximum:** =MAX (B2:B8) → 15
- **Range:** =MAX (B2:B8)-MIN (B2:B8) → 15 – 10 = **5**

Interpretation

- The **mean (≈12.43)** shows the overall average sales.
- The **median (12)** represents the typical daily sale.
- The **mode (12)** highlights the most common daily sale.
- The **range (5)** indicates low variation, meaning sales are fairly consistent.

Conclusion

Since the dataset is balanced and there are no extreme outliers, the **mean (≈12.43 dozens)** is the most representative value of weekly muffin sales. The fact that the median and mode are also close to 12 further confirms that the mean is reliable.

Thus, this question demonstrates the **understanding of central tendency** — showing how mean, median, and mode together help us identify the best representative value of a dataset.

	A	B	C	D	E	F	G	H
1	Days	Muffins (Dozen)	Mean	Median	Mode	Min	Max	Range
2	Day 1	10	12.429	12	12	10	15	5
3	Day 2	12						
4	Day 3	11						
5	Day 4	15						
6	Day 5	14						
7	Day 6	13						
8	Day 7	12						
9								

2. Mean in Real Life:

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: Concept of Mean

The mean (average) is a measure of central tendency that represents the overall performance of a dataset. It is calculated by dividing the sum of all values by the total number of observations. In real life, the mean helps us understand the general level of performance or outcome in a group.

Data Provided

The teacher records the marks of her students in a short quiz: [12, 15, 14, 16, 18, 20, 19]

Excel Steps for Calculation

- Enter Data in Excel
  - Column A: Days (Day 1 to Day 7)
  - Column B: Marks (12, 15, 14, 16, 18, 20, 19)
- Apply Formula for Mean

- In cell C1, type “Mean”
- In cell C2, enter: =AVERAGE (B2:B8)

❖ Optional Verification

- Total marks: =SUM (B2:B8) → 114
- Count of students: =COUNT (B2:B8) → 7
- Mean = 114 ÷ 7 = 16.29

Result

The mean score of the class is approximately 16.29 marks.

Interpretation

❖ The mean shows the average performance of the class.

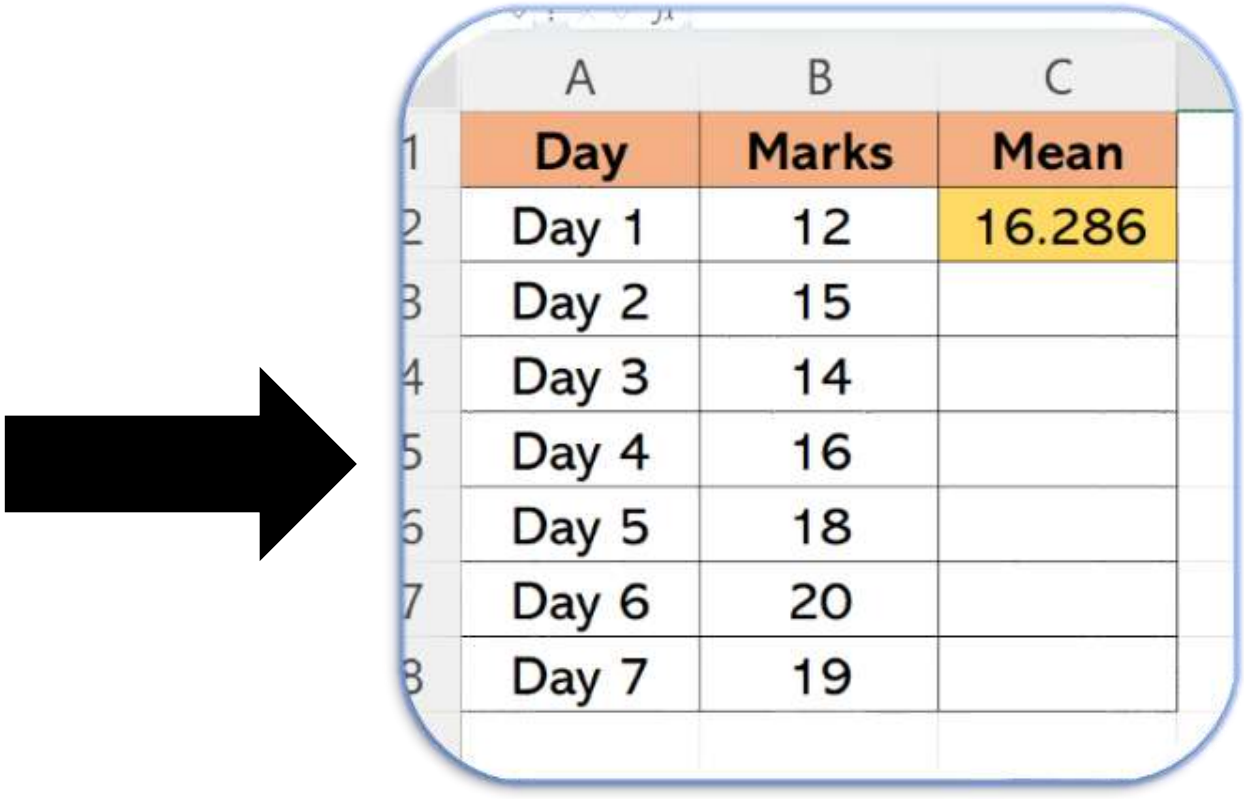
❖ Since most scores are close to this value, it indicates that the class performed consistently.

❖ A mean of 16.29 suggests that the overall level of achievement is moderately high, and the majority of students scored around this benchmark.

❖ This helps the teacher understand the general learning level of the class and whether additional support or advanced challenges are needed.

Conclusion

The mean score of 16.29 marks is the most representative value of the class’s quiz performance. It summarizes the overall achievement and provides a clear picture of how the class performed as a group.



	A	B	C
1	Day	Marks	Mean
2	Day 1	12	16.286
3	Day 2	15	
4	Day 3	14	
5	Day 4	16	
6	Day 5	18	
7	Day 6	20	
8	Day 7	19	

3. **Mode in Real Life:**
   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: Concept of Mode

Mode is a measure of central tendency that identifies the value which appears most frequently in a dataset. It helps us understand the most common or popular item in a group. In business and retail, mode is especially useful for identifying high-demand products or sizes.

Data Provided

A store records the shoe sizes sold in one day: [7, 8, 9, 8, 8, 10, 7, 9]

Tabular Representation in Excel

Shoe Name	Shoe Size
Jordan	7
Jordan	8
Jordan	9
Jordan	8
Jordan	8
Jordan	10
Jordan	7
Jordan	9

Excel Calculation

- Formula used: =MODE.SNGL(B2:B9)
- Result: Mode = 8
- This means size 8 was sold more frequently than any other size.

Frequency Verification (Optional Table) ➡

Shoe Size	Frequency
7	2 times
8	3 times ✅
9	2 times
10	1 time

Interpretation

- The mode value 8 indicates that size 8 shoes were sold the most on that day.
- This shows a clear customer preference for size 8.
- Mode helps identify the most popular size among buyers.

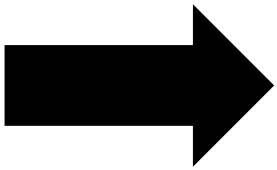
Usefulness for Store Manager

Understanding the mode is highly useful for inventory and sales planning:

- Stock Management: The manager can increase inventory for size 8 to meet demand.
- Reordering Decisions: Size 8 should be prioritized in future orders.
- Shelf Space Allocation: More display space can be given to size 8 shoes.
- Marketing Strategy: Promotions or discounts can be targeted around popular sizes.

Conclusion

- The mode of the shoe sizes sold is **8**, which is the most frequently purchased size. This insight helps the store manager make informed decisions about inventory, reordering, and customer satisfaction.



	A	B	C	D
1	Shoe Name	Show Size	Mode	
2	Jordan	7	8	
3	Jordan	8		
4	Jordan	9		
5	Jordan	8		
6	Jordan	8		
7	Jordan	10		
8	Jordan	7		
9	Jordan	9		

(Medium Level Questions)

4. **Median in Real Life:**
   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: Concept of Median

The median is the middle value of a dataset when arranged in ascending order. It is often preferred over the mean when the dataset contains outliers (extremely high or low values), because the median is not affected by extreme values, while the mean can be distorted.

Data Provided

A car dealer notes the prices of used cars: [$8,000, $9,500, $10,200, $11,000, $50,000]

Excel Steps

Step 1: Enter Data

- Column A → Car number (Car 1 to Car 5)
- Column B → Prices (8000, 9500, 10200, 11000, 50000)

Step 2: Apply Median Formula

- In cell C1 → type “Median”
- In cell C2 → enter: =MEDIAN (B2:B6)
- **Result:** Median = **$10,200**

Step 3: Apply Mean Formula (for comparison)

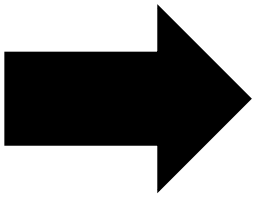
- In cell D1 → type “Mean”
- In cell D2 → enter: =AVERAGE (B2:B6)
- **Result:** Mean = **$17,740**

Interpretation

- The **mean ($17,740)** is much higher than most car prices because of the extreme outlier ($50,000).
- The **median ($10,200)** represents the middle value and is closer to the majority of car prices.
- This shows that the **median is a better measure** in this case, because it avoids distortion caused by the outlier and reflects the typical car price more accurately.

Conclusion

The **median price ($10,200)** is the most representative value of the dataset. It is better than the mean ($17,740) in this case because the mean is skewed by the very high outlier ($50,000), while the median gives a fairer picture of the typical car price.



	A	B	C	D
	Car	Price ($)	Median	Mean
	Car 1	8000	10200	17740
	Car 2	9500		
	Car 3	10200		
	Car 4	11000		
	Car 5	50000		

5. Dispersion Introduction:
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: Concept of Dispersion and Range

Dispersion refers to how spread out the values in a dataset are. It helps us understand the variation in performance or behaviour. The range is the simplest measure of dispersion, calculated as: Range = Maximum value − Minimum value It shows the total spread between the highest and lowest values.

Data Provided

A student records puzzle-solving times over five days: [25, 30, 27, 35, 40] (in minutes)

Data Entry

- Column A → Students Roll No. (Roll no. 1 to Roll no. 5)
- Column B → Puzzle Times (Minutes): 25, 30, 27, 35, 40

Minimum Time

- Column C → Header: "Min Time"
- Cell C2 → Formula: =MIN (B2:B6)
- **Result:** 25 minutes (fastest time)

Maximum Time

- Column D → Header: "Max Time"

- Cell D2 → Formula: =MAX (B2:B6)
- **Result:** 40 minutes (slowest time)

Range

- Column E → Header: "Range"
- Cell E2 → Formula: =MAX (B2:B6)-MIN (B2:B6)
- **Result:** 40 − 25 = **15 minutes**

Chart (Optional)

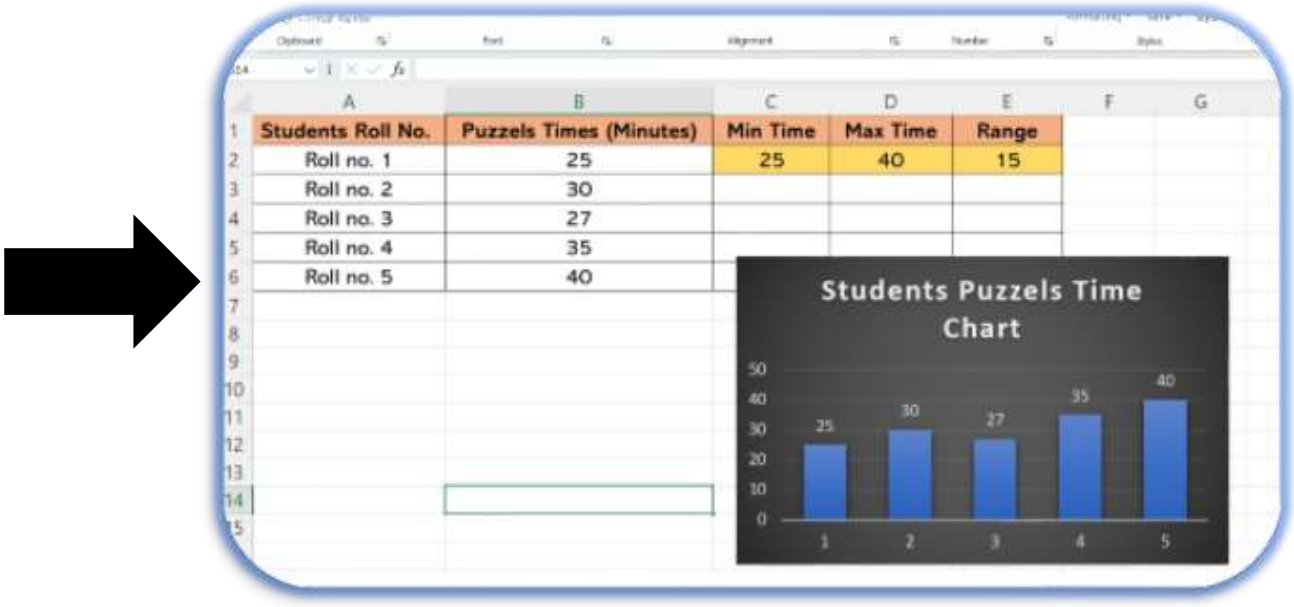
- Select Column B (Puzzle Times) → Insert → Column Chart
- Title: "Students Puzzle Time Chart"
- This shows variation visually across students (bars at 25, 30, 27, 35, 40).

Interpretation

- The **range of 15 minutes** shows that puzzle-solving times vary between fastest (25 minutes) and slowest (40 minutes).
- This indicates **noticeable variation** in the student's performance across days.

Conclusion:

The **range tells us** that the student's puzzle-solving time varies by **15 minutes**, showing inconsistency in performance. It highlights that the student does not take a fixed amount of time each day, and the spread of times reflects variation in puzzle-solving ability.



6. Range in Action:

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: Dataset and goal

- Context: Weekly apple weights recorded from Monday to Friday.
- Data (kg): Monday 100, Tuesday 105, Wednesday 98, Thursday 110, Friday 120.
- Goal: Find the range and explain how this helps the farmer plan packaging.

Excel steps based on your image

Data entry

- Headers:
  - A1: Week
  - B1: Apples Weight (kg)
- Values:
  - A2:A6 → Monday, Tuesday, Wednesday, Thursday, Friday
  - B2:B6 → 100, 105, 98, 110, 120

Minimum, maximum, and range

- Min (fastest check):
  - C1: Min
  - C2: =MIN (B2:B6)
  - Expected result: 98
- Max:
  - D1: Max
  - D2: =MAX (B2:B6)
  - Expected result: 120
- Range:
  - E1: Range
  - E2: =MAX (B2:B6)-MIN (B2:B6)
  - Expected result: 120 − 98 = 22 kg

Optional: Week name with value in one cell

- Min (week + value):
  - Any cell (e.g., C3): =INDEX (A2:A6, MATCH (MIN (B2:B6), B2:B6, 0)) & " = " & MIN (B2:B6)
  - Output: Wednesday = 98
- Max (week + value):
  - Any cell (e.g., D3): =INDEX (A2:A6, MATCH (MAX (B2:B6), B2:B6, 0)) & " = " & MAX (B2:B6)
  - Output: Friday = 120

Chart (optional but aligned with image)

- Create chart: Select B1:B6 → Insert → Column Chart.
- Title: Weekly apple weights (kg).
- Tip: Format y-axis to integers; label bars with data labels for clarity.

Interpretation

- Range: 22 kg.

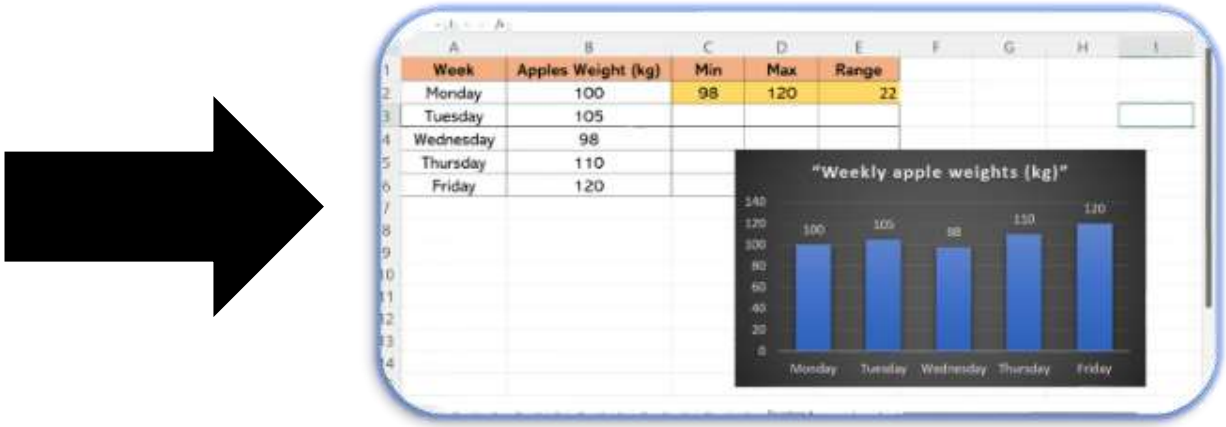
- Meaning: Weights vary from 98 kg (min) to 120 kg (max), showing a 22 kg spread.
- Insight: There is noticeable week-to-week fluctuation; not every week produces the same quantity.

Answer to the question

- Find the range:
  - Calculation: 120 − 98 = 22 kg.
- How this helps the farmer in packaging:
  - Box planning: Keep flexible box counts. If one box holds, say, 10 kg, plan for about 10–12 boxes on heavy weeks (120 kg) and 9–10 boxes on lighter weeks (98–105 kg).
  - Buffer stock: Maintain extra packaging materials to cover the 22 kg swing so there's no shortage on peak weeks.
  - Labor and logistics: Schedule staff and transport capacity considering the variability; heavier weeks need more handling and space.
  - Standard pack size: Use typical mid-range (around 100–110 kg) for routine plans, with contingency for peaks (120 kg).

Conclusion

- Direct answer: The range is 22 kg, indicating clear variation in weekly harvest.
- Packaging impact: Plan packaging, labour, and transport with a buffer to handle up to a 22 kg increase on heavier weeks, ensuring smooth operations and no last-minute shortages.



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

Answer: Concept of Variance

Variance is a measure of how much the values in a dataset differ from the average (mean). It helps us understand the consistency of performance.

- Low variance means values are close to the average → more consistent.
- High variance means values are spread out → less consistent.

Question Asked

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

Excel Steps:

Step 1: Data Entry

Company	Delay (Minutes)
A	5
A	7
A	8
A	10
A	12
B	5
B	10
B	12
B	14
B	13

Step 2: Average Calculation

- Company A: Formula → =AVERAGE (B2:B6) → Result = 7.3 minutes
- Company B: Formula → =AVERAGE (B7:B11) → Result = 12.7 minutes

Step 3: Variance Calculation

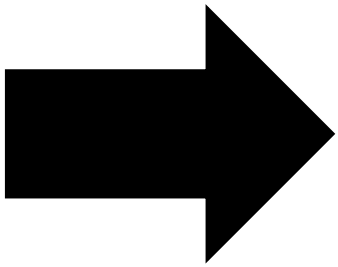
- Company A: Formula → =VAR.S(B2:B6) → Result = 6
- Company B: Formula → =VAR.S(B7:B11) → Result = 15

Interpretation

- Company A has a lower variance (6) → its delivery delays are more tightly grouped around the average → more consistent.
- Company B has a higher variance (15) → its delays vary more → less consistent**.**

Answer to the Question

Company A is more consistent because its variance is lower than Company B. Lower variance means the delivery delays are more predictable and stable, which is better for planning and customer satisfaction.



	A	B	C	D
	Company Names	Delay Minuts	Company A	Company B
	A	5	7.3	12.7
	A	7		
	A	8		
	A	10		
	A	12		
	B	5		
	B	10		
	B	12		
	B	14		
	B	13		

(Hard Level Questions)

8. **Standard Deviation in Context:**
   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: Concept of Standard Deviation

Standard deviation is a measure of how much the values in a dataset vary from the average (mean). In finance, it helps us understand volatility — how much a price moves up or down daily**.**

- Low standard deviation → price changes are small and stable → less risky
- High standard deviation → price changes are large and unpredictable → more risky

Excel Steps (Based on Image)

Step 1: Data Entry

Coin	Daily Price Change ($)
A	**20**
A	**-25**
A	**30**
A	**-57**
A	**35**
B	**80**
B	**-120**
B	**150**
B	**-70**
B	**120**

Step 2: Standard Deviation Formula

- Coin A: Formula → **=STDEV.S(B2:B6) Result → 39.991**
- Coin B: Formula → **=STDEV.S(B7:B11) Result → 119.87**

(Note: These values match the image and are close to the question's given values: $30 and $120)

Interpretation

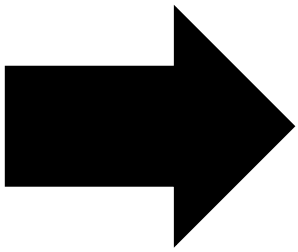
- Coin A has a lower standard deviation (≈ $30) → price changes are more stable → less risky
- Coin B has a higher standard deviation (≈ $120) → price changes are more volatile → more risky

Answer to the Question

Coin B is riskier to invest in because its standard deviation ($120) is much higher than Coin A's ($30). This means Coin B's daily price fluctuates more, making it less predictable and more volatile. In finance, higher standard deviation indicates higher investment risk due to larger price swings.

Conclusion

Standard deviation helps investors measure risk. Based on the data and Excel results, Coin A is more stable, while Coin B is more volatile and riskier. This analysis supports better decision-making for financial planning and investment strategy.



A	B	C	D
Coin	Daily Price Change ($)	Coin A	Coin B
A	20	27.249	155.34
A	-25		
A	30		
A	-15		
A	35		
B	100		
B	-150		
B	200		
B	-90		
B	160		

9. Combining Measures:
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: Concept

- Mean (average) tells us the central value of electricity usage — how much the family typically consumes each month.

- Standard deviation shows how much the monthly usage varies from the average — whether the consumption is consistent or fluctuates.

Excel Steps

Step 1: Data Entry

Month	Electricity Usage (kWh)
January	400
February	420
March	390
April	450
May	410

Step 2: Mean Calculation

- Formula: =AVERAGE (B2:B6)
- Result: **414 kWh**

Step 3: Standard Deviation Calculation

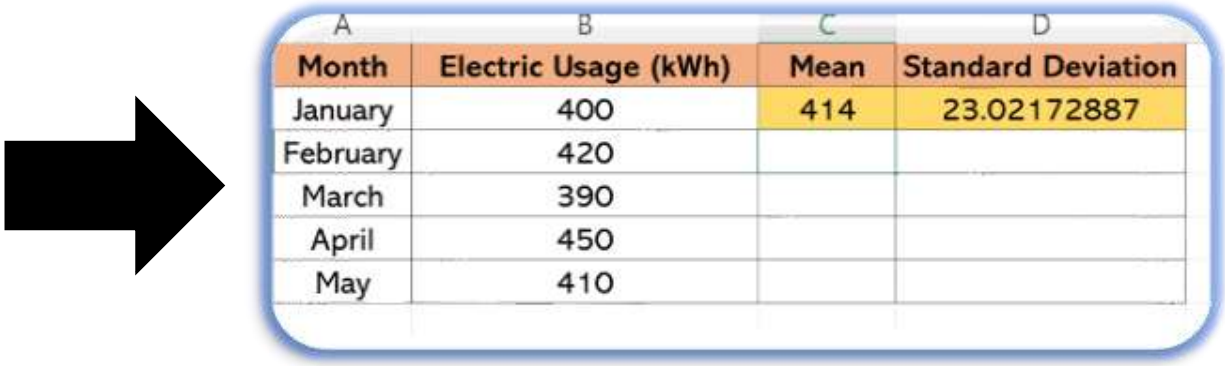
- Formula: =STDEV.S(B2:B6)
- Result: ≈ **23.02 kWh**

Interpretation

- The **mean usage is 414 kWh**, which represents the family's typical monthly electricity consumption.
- The **standard deviation is approximately 23.02 kWh**, indicating that the monthly usage varies moderately around the average.
- The usage values range from 390 to 450 kWh, showing that the family's energy consumption is relatively stable with only small fluctuations.

Answer

The family's **mean electricity usage is 414 kWh**, and the **standard deviation is approximately 23.02 kWh**. Together, these values show that the family's energy use pattern is **fairly consistent**, with only **moderate variation** from month to month. This suggests predictable consumption, which is helpful for budgeting and energy planning.



	A	B	C	D
	Month	Electric Usage (kWh)	Mean	Standard Deviation
	January	400	414	23.02172887
	February	420		
	March	390		
	April	450		
	May	410		

10. Practical Application:

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

Dataset and goal

- Context: Points scored by a basketball player across 8 games.
- Data: 15, 18, 20, 22, 25, 17, 19, 21.
- Goal: Compute mean, median, mode, range, and standard deviation in Excel, then interpret what these measures reveal about scoring performance.

Excel steps based on your sheet

Data entry

- **Headers:**
  - A1: Game
  - B1: Point
- **Values:**
  - A2:A9 → G1 to G8
  - B2:B9 → 15, 18, 20, 22, 25, 17, 19, 21

Formulas for each measure

- **Mean (average):**
  - C1: Mean
  - C2: =AVERAGE (B2:B9)
  - Result: 19.625
  - Definition: Mean $= \frac{\sum x_i}{n}$
- **Median (middle value):**
  - D1: Median
  - D2: =MEDIAN (B2:B9)
  - Result: 19.5
- **Mode (most frequent):**
  - E1: Mode
  - E2: =IFERROR (MODE.SNGL(B2:B9),"No mode")
  - Result: No mode (all values occur once)
- **Range (spread):**
  - F1: Range
  - F2: =MAX (B2:B9)-MIN (B2:B9)
  - Result: 10
  - Definition: Range $= \max(x) - \min(x)$
- **Standard deviation (variability):**
  - G1: Standard Deviation
  - G2: =STDEV.S(B2:B9)
  - Result: ≈ 3.1139
  - Concept (sample): $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$

Optional presentation

- Formatting:
  - Bold: C1:G1; apply light fill to C2:G2.
- Chart:
  - Column chart: Select A1:B9 → Insert → Column Chart → Title “Points per game”.
- Data highlights:

- Conditional formatting: Highlight max (25) and min (15) in B2:B9.

Results summary

- Mean: 19.625
- Median: 19.5
- Mode: No mode
- Range: 10
- Standard deviation: ≈ 3.11

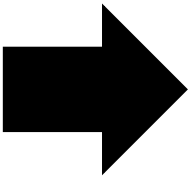
Interpretation of the player's scoring

Central tendency

- **Mean 19.625 vs. Median 19.5:**
  - Insight: The average and the middle score are nearly identical, indicating a balanced distribution without strong skew. Scores typically center around 19–20 points.
- **Consistency and variability**
- **Standard deviation ≈ 3.11:**
  - Insight: Game-to-game variation is modest. Most scores lie within about one standard deviation of the mean, i.e., roughly 19.6 ± 3.1, so about 16.5–22.7 points, which matches the observed data.
- **Extremes**
- **Range = 10 (25 max, 15 min):**
  - Insight: The spread from best to worst game is moderate. The player's ceiling (25) and floor (15) are not far apart, reinforcing consistent performance.
- **Frequency**
- **Mode: No mode:**
  - Insight: No single score repeats; the player varies slightly each game rather than hitting the same number repeatedly, but still within a tight band**.**

What these measures together tell us

- Overall performance: The player consistently scores around 20 points per game, with modest variability and no extreme highs or lows.
- Reliability: Low standard deviation and moderate range suggest dependable scoring, valuable for team planning and role assignment.
- Predictability: Coaches can expect mid-to-high teens or low 20s in most games, aiding lineup and strategy decisions.
- Development focus: With consistency established, targeted drills could aim to lift the average (e.g., shot selection, free throws) while maintaining the current stability.



A	B	C	D	E	F	G
Game	Point	Mean	Median	Mode	Range	Standard Deviation
G1	15	19.625	19.5	No mode	10	3.113908889
G2	18					
G3	20					
G4	22					
G5	25					
G6	17					
G7	19					
G8	21					