

Averages, Spread & Cumulative Frequency

Amari Cross & Matthew Williams • Math • May 6, 2026

Averages describe the centre of a dataset, while spread describes how far the values are from each other. Two classes can have the same mean score but very different consistency, so both centre and spread are needed to understand the data properly.

CSEC questions often move from calculation to interpretation: find the mean, median, range, quartiles, or cumulative frequency, then say what the result tells you. Do not treat these as isolated formulas. After calculating, write a short sentence explaining what the number means in the context of the question.

A **measure of central tendency** represents the "typical" or "average" value in a dataset.

Mean (Average)

The mean uses every value, so it is affected by very large or very small outliers. It is useful when the data is fairly balanced.

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Number of values}}$$

For raw data:

Example

Scores: 5, 7, 8, 9, 5

[MathBlock]

$$\text{Mean} = \frac{5 + 7 + 8 + 9 + 5}{5} = \frac{34}{5} = 6.8$$

[/MathBlock]

For ungrouped frequency data:

$$\text{Mean} = \frac{\sum(\text{value} \times \text{frequency})}{\sum \text{frequency}}$$

Example

Score	Frequency
5	2
7	1
8	1
9	1

[MathBlock]

$$\text{Mean} = \frac{(5 \times 2) + (7 \times 1) + (8 \times 1) + (9 \times 1)}{2+1+1+1} =$$

$$\frac{10+7+8+9}{5} = \frac{34}{5} = 6.8$$

[/MathBlock]

For grouped data:

Use class midpoints:

$$\text{Mean} = \frac{\sum(\text{midpoint} \times \text{frequency})}{\sum \text{frequency}}$$

Example

Class	Frequency	Midpoint
10-19	3	14.5
20-29	5	24.5
30-39	2	34.5

[MathBlock]

$$\text{Mean} = \frac{(14.5 \times 3) + (24.5 \times 5) + (34.5 \times 2)}{3+5+2}$$

[/MathBlock]

[MathBlock]

$$= \frac{43.5 + 122.5 + 69}{10} = \frac{235}{10} = 23.5$$

[/MathBlock]

Median

The median is the middle value after ordering the data. It is useful when outliers would distort the mean.

The **median** is the middle value when data is arranged in order.

For raw data:

- 1. Arrange values from smallest to largest
- 2. If odd number of values: median is the middle one
- 3. If even number of values: median is the average of the two middle ones

Example

Scores: 5, 5, 7, 8, 9 (5 values, odd)

Median = 7 (the 3rd value)

Scores: 5, 5, 7, 8, 9, 10 (6 values, even)

Median = $(7 + 8) \div 2 = 7.5$

For grouped data:

Use the cumulative frequency table and interpolation:

$$\text{Median} = L + \frac{\frac{n}{2} - CF}{f} \times w$$

Where:

- [Math: L] = lower boundary of median class
- [Math: n] = total frequency
- [Math: CF] = cumulative frequency before median class
- [Math: f] = frequency of median class
- [Math: w] = class width

Example

Class	Frequency	Cumulative
10-19	3	3
20-29	5	8
30-39	2	10

Total = 10, so median position = $10 \div 2 = 5$

Median class is 20-29 (cumulative frequency reaches 5 here)

[MathBlock]

$$\text{Median} = 19.5 + \frac{5 - 3}{5} \times 10 = 19.5 + \frac{2}{5} \times 10 = 19.5 + 4 = 23.5$$

[/MathBlock]

<JSXGraph id="stat-median-nl" title="Median is the middle value of an ordered list"

boundingbox=[-1, 1.5, 10, -1.5] axis={false} height={150} code="var vals=[2,4,5,7,9];for(var

i=0;i<vals.length;i++){board.create('point',[2*i+1,0],{name:String(vals[i]),size:4,fixed:true,fillColor:i===2?'#dc26

/>

Mode

The mode identifies the most common value. It is especially useful for categorical data, where mean and median may not make sense.

The **mode** is the value that appears most often.

Example

Scores: 5, 5, 5, 7, 8, 9, 9

Mode = 5 (appears 3 times)

Data: 2, 5, 5, 7, 7, 9

Two modes: 5 and 7 (both appear twice) = **bimodal**

Data: 2, 5, 7, 9

No mode (all appear once) = **no mode**

Choosing Mean, Median, or Mode

Choosing the average is a reasoning skill. The best measure depends on the shape of the data and what the question is trying to describe.

Use **MEAN** when:

- Data is roughly symmetric
- No extreme outliers
- You want to use all values
- Example: class average on a test

Use **MEDIAN** when:

- Data has outliers or is skewed
- You want the "typical" middle value
- Example: house prices (skewed by luxury homes)

Use **MODE** when:

- Categorical data (colors, preferences)
- Discrete data with clear peaks
- Example: favorite color, most common shoe size

 **Exam Tip****Example: Which average?**

House prices: 120,000, 125,000, 130,000, 140,000, 2,000,000

- Mean = $(120+125+130+140+2000)\div 5 = \mathbf{502,300}$ (way too high!)
- Median = **130,000** (better—the luxury home is an outlier)
- Mode = no mode

Best answer: Median, because the data has an outlier.

Part 5: Measures of Spread (Dispersion)

Spread measures how far apart the data values are from each other.

Range

Range gives a quick sense of spread, but it only uses the smallest and largest values. One unusual value can make the range misleading.

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

 **Example**

Scores: 5, 7, 8, 9, 5

$$\text{Range} = 9 - 5 = 4$$

Problem: Only uses the extreme values. Doesn't show middle spread.

Quartiles and Interquartile Range

Quartiles split ordered data into four parts. The interquartile range focuses on the middle half of the data, so it is less affected by extremes.

Quartiles divide the data into 4 equal parts.

- Q_1 (1st quartile) = 25th percentile
- Q_2 (2nd quartile) = 50th percentile = **median**
- Q_3 (3rd quartile) = 75th percentile

Interquartile Range (IQR):

$$\text{IQR} = Q_3 - Q_1$$

This shows the spread of the middle 50% of data.

Example

Test scores: 5, 6, 7, 7, 8, 8, 8, 9, 9, 10 (10 values)

Arrange in order: 5, 6, 7, 7, 8, 8, 8, 9, 9, 10

Q_1 position = $(10+1) \div 4 = 2.75$ 'between 2nd and 3rd values = $6 + 0.75(7-6) = 6.75$

Q_2 position = $(10+1) \div 2 = 5.5$ 'between 5th and 6th values = 8

Q_3 position = $3(10+1) \div 4 = 8.25$ 'between 8th and 9th values = $9 + 0.25(9-9) = 9$

[MathBlock]

$$\text{IQR} = Q_3 - Q_1 = 9 - 6.75 = 2.25$$

[/MathBlock]

<JSXGraph id="stat-box" title="Box-and-whisker plot showing quartiles" boundingbox=[[0, 3, 22, -2]] axis={false} height={220}

code="board.create('axis',[[0,0],[22,0]],{ticks:{insertTicks:false,ticksDistance:2,minorTicks:0}});board.create('se
/>

Semi-Interquartile Range

The semi-interquartile range is half of the IQR. It gives a compact measure of spread around the middle of the dataset.

$$\text{Semi-IQR} = \frac{\text{IQR}}{2} = \frac{Q_3 - Q_1}{2}$$

Example

Using the example above:

[MathBlock]

$$\text{Semi-IQR} = \frac{2.25}{2} = 1.125$$

[/MathBlock]

Remember

- **Range:** Uses only extremes, sensitive to outliers
- **IQR:** Shows spread of middle 50%, ignores outliers
- **Semi-IQR:** Half of IQR, useful for comparison

Part 6: Cumulative Frequency and Ogives

Cumulative Frequency Table

Cumulative frequency is a running total. It answers questions like "how many values are less than or equal to this point?"

Cumulative frequency = total count up to and including that class.

Example

Class	Frequency	Cumulative Frequency
10-19	3	3
20-29	5	3+5 = 8
30-39	7	8+7 = 15
40-49	4	15+4 = 19
50-59	1	19+1 = 20

```
<JSXGraph id="stat-cf-table" title="Cumulative frequency: running total of frequencies"
boundingbox=[-1, 6, 11, -1] axis={false} height={240} code="var
rows=[['Class','f','cf'],['0-10','3','3'],['10-20','7','10'],['20-30','12','22'],['30-40','5','27']];for(var
r=0;r<rows.length;r++){for(var
c=0;c<3;c++){board.create('text',[1+c3,5-r1.1,rows[r][c]},{fontSize:14,anchorX:'middle',cssStyle:r===0?'font-wei
/>
```

Cumulative Frequency Curve (Ogive)

An ogive turns cumulative totals into a graph. It is useful for estimating medians, quartiles, and percentiles from grouped data.

An **ogive** is an S-shaped curve showing cumulative frequency.

How to draw:

- 1. Use class boundaries on x-axis (not class limits)
- 2. Use cumulative frequency on y-axis
- 3. Plot point at upper boundary of each class
- 4. Connect points with a smooth curve

Example

Using the table above:

Upper Boundary	Cumulative Frequency
19.5	3
29.5	8
39.5	15
49.5	19
59.5	20

```
<JSXGraph id="stat-ogive" title="Ogive (cumulative frequency curve)" boundingbox=[[-2, 35, 55, -5]] height={320} code="var
pts=[[0,0],[10,3],[20,10],[30,22],[40,27],[50,30]];board.create('curve',[function(t){var
p=pts;for(var i=0;i<p.length-1;i++){if(t>=p[i][0]&&t<=p[i+1][0])return
p[i][1]+(p[i+1][1]-p[i][1])*(t-p[i][0])/(p[i+1][0]-p[i][0]);}return
p[p.length-1][1];},function(t){return t;},0,50},{strokeColor:'#2563eb',strokeWidth:2});for(var
i=0;i<pts.length;i++){board.create('point',pts[i],{name:'',size:3,fillColor:c.blue});}"/>
```

Reading from an Ogive

To read from an ogive, move horizontally from the cumulative frequency value to the curve, then down to the data value. This is an estimate, so use the graph carefully.

You can read:

- **Quartiles:** Q_1 at 25% of total, Q_2 at 50%, Q_3 at 75%
- **Percentiles:** any value's percentage position
- **Median:** where cumulative frequency = $n/2$
- **Frequencies above/below** a given value

Example

From the ogive above (n=20):

Q₁ (25% of 20 = 5): Read across from cumulative frequency 5 to curve, then down to x-axis **12**

Median (50% of 20 = 10): Read from cumulative 10 **18**

Q₃ (75% of 20 = 15): Read from cumulative 15 = **39.5**

<JSXGraph id="stat-ogive-q" title="Ogive with quartiles marked" boundingbox=[-2, 35, 55, -5] height={340} code="var

pts=[[0,0],[10,3],[20,10],[30,22],[40,27],[50,30]];board.create('curve',[function(t){var

p=pts;for(var i=0;i<p.length-1;i++){if(t>=p[i][0]&&t<=p[i+1][0])return

p[i][1]+(p[i+1][1]-p[i][1])*(t-p[i][0])/(p[i+1][0]-p[i][0]);}return

p[p.length-1][1];},function(t){return t;},0,50},{strokeColor:'#2563eb',strokeWidth:2});var

qs=[7.5,15,22.5];var labels=['Q1','Q2','Q3'];var colors=['#16a34a','#dc2626','#f59e0b'];for(var

i=0;i<3;i++){board.create('line',[[0,qs[i]],[50,qs[i]]],{strokeColor:colors[i],dash:2,straightFirst:false,straightLast:f