Averages & Measures of Spread

Prepared by Nikolas Agisilaou

Mean

The mean is the sum of all values divided by the number of values.

Raw data: $\bar{x} = \dfrac{\sum x}{n}$

Freq. table: $\bar{x} = \dfrac{\sum fx}{\sum f}$

Grouped (est.): $\bar{x} \approx \dfrac{\sum fm}{\sum f}$, m = midpoint

Always add an fx or fm column in working. For grouped data, always state the answer is an **estimate**.

Median

The median is the **middle value** once data is ordered.

Odd n: $\dfrac{n+1}{2}$th value

Even n: mean of $\dfrac{n}{2}$th and $\left(\dfrac{n}{2}+1\right)$th values

For a **frequency table**: add a cumulative frequency column, then locate which group contains the median position.

Not affected by outliers — use when data is skewed.

Mode, Modal Class and Choosing the Right Average

Mode: the value that occurs **most often**.

- A data set can have **no mode**, one, or several modes
- For grouped data: **modal class** = class with the **highest frequency**
- Cannot give an exact value for grouped data

Choosing the right average:

Average	Use when...	Key note
Mean	Symmetric data, no outliers	Uses all values; pulled by extremes
Median	Outliers present / skewed data	Not affected by extremes
Mode	Categorical / discrete data	May not exist

Rule of thumb: outlier present? → prefer **median**. Shoe size or category? → use **mode**.

Measures of Spread and Finding Quartiles

Range: Range = Max − Min
Simple but **sensitive to outliers**.

Interquartile Range (IQR):

$$IQR = Q_3 - Q_1$$

Spread of the **middle 50%**. Not affected by outliers.

Smaller IQR = more consistent data.

Finding quartiles from a list:

1. Order data ascending
2. Q_2 = median of the full set
3. Q_1 = median of the **lower half**
4. Q_3 = median of the **upper half**

(Exclude Q_2 itself when n is odd)

Cumulative frequency graph (ogive):

- Add a cumulative frequency column to the table
- Plot each point at the **upper class boundary**
- Start the curve at (lower boundary of 1st class, 0)
- Join the points with a smooth **S-shaped curve**

Reading off quartiles (where total frequency = n):

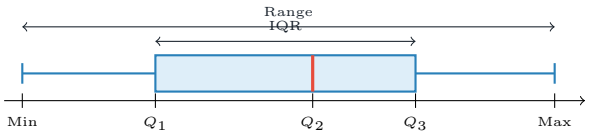
Value	Q_1	Median (Q_2)	Q_3
Read across at cf =	$n/4$	$n/2$	$3n/4$

*All values read from the curve are **estimates**.*

The Panda Math Tutor

Box-and-Whisker Plots and Comparing Two Data Sets

Five-number summary: Min, Q_1, Q_2, Q_3, Max



Skewness from the box:

- Q_2 centred → **symmetric**
- Q_2 near Q_1 → **positive skew** (tail right)
- Q_2 near Q_3 → **negative skew** (tail left)

Always make TWO statements when comparing:

1. **Average** (mean or median): who scored higher / lower on average?
2. **Spread** (range or IQR): who was more / less consistent?

Model sentence starters:

- "On average, Group A scored *higher* because its median is greater…"
- "Group B is *more consistent* as its IQR is smaller…"
- "Group A shows more variability, with a wider range…"

*Always compare **in context**. Numbers alone are not enough.*

Key comparison words: **higher, lower, greater, smaller, more consistent, more variable**.

Common Traps and Pitfalls

1. **Grouped mean trap:** Use **midpoints**, not class boundaries. Always state *"estimate"*.
2. **Cumulative freq. trap:** Plot at **upper boundaries**, not midpoints or lower bounds.
3. **Quartile graph trap:** Q_1 is at $n/4$, **not** $(n+1)/4$. Use n not $n+1$.
4. **Odd n quartile trap:** Exclude the median itself when finding Q_1 and Q_3.
5. **Comparison trap:** Listing two numbers is **not** a comparison. Use comparison words.
6. **Outlier trap:** Extreme values inflate both **mean** and **range** — switch to median and IQR.

Quick Reference: Formulae and Exam Keywords

Key Formulae	Exam Keywords
$\bar{x} = \frac{\sum x}{n}$ *(raw)*	**"estimate mean"** → midpoints
$\bar{x} = \frac{\sum fx}{\sum f}$ *(freq.)*	**"modal class"** → highest freq.
$\bar{x} \approx \frac{\sum fm}{\sum f}$ *(grouped)*	**"use your graph"** → read ogive
Range = Max − Min	**"find the IQR"** → $Q_3 - Q_1$
$IQR = Q_3 - Q_1$	**"compare"** → avg + spread
	"best average?" → outlier/skew
	"draw box plot" → 5-num. summary

The Panda Math Tutor