

Summary Statistics

AP Statistics | Unit 1 | mean vs median, IQR, standard deviation, boxplots

1. Center: mean vs. median

Mean $\bar{x} = \frac{\sum x_i}{n}$: every value pulls on it. **Median**: the middle value of the ordered data (with an even count, the average of the two middle values).

Nine people in a café earn \$35k–\$55k: mean = median = \$45k. A billionaire walks in: the mean jumps to about **\$100 million**, the median moves only to **\$46.5k**.

Resistance — and which one to report

The median is **resistant** to outliers and skew; the mean is **not**.

Roughly symmetric, no outliers $\Rightarrow$ mean (and standard deviation). Skewed or outliers $\Rightarrow$ median (and IQR). That is why incomes and house prices are reported as medians.

2. Spread

Range = max – min: depends on the two most extreme values only.

Quartiles: Q_1 is the median of the lower half, Q_3 the median of the upper half (leave the median itself out when n is odd). **IQR** = $Q_3 - Q_1$, the spread of the middle half — resistant. For class B (25 test scores, hypothetical): $Q_1 = 75$, median = 81, $Q_3 = 84$, so IQR = 9.

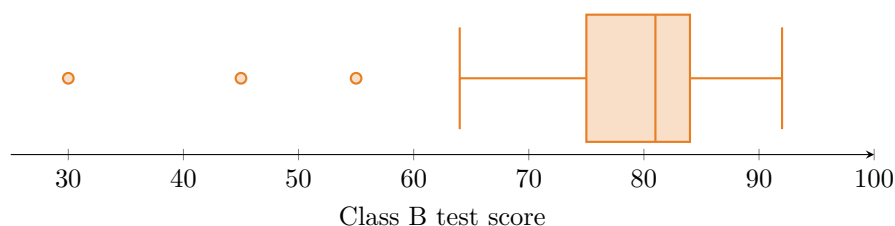
Standard deviation — roughly, the typical distance of a value from the mean:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}.$$

Quiz scores 4, 7, 8, 9, 12: $\bar{x} = 8$, deviations $-4, -1, 0, 1, 4$ (they always sum to 0), squares sum to 34, $s^2 = 34/4 = 8.5$, $s \approx 2.92$ points. Dividing by $n - 1$ keeps the sample s from coming out too small on average. s is **not** resistant.

3. Five-number summary, boxplots, outliers

Min, Q_1 , median, Q_3 , max. **1.5 IQR rule**: a value is an outlier if it is below $Q_1 - 1.5\text{IQR}$ or above $Q_3 + 1.5\text{IQR}$. Class B: fences $75 - 13.5 = 61.5$ and $84 + 13.5 = 97.5$, so 30, 45 and 55 are outliers. Whiskers go to the most extreme values that are *not* outliers (64 and 92); outliers are plotted as separate points.



Long left whisker and outliers on the left $\Rightarrow$ skewed left.

4. Changing every value

	class B	add 5 to every score	double every score
mean	76	81	152
median	81	86	162
IQR	9	9	18
standard deviation	14.1	14.1	28.2
range	62	62	124

Adding a constant shifts the center and leaves the spread alone; multiplying by a constant scales both center and spread (e.g. $F = 1.8C + 32$ multiplies the spread by 1.8).

Where students lose points

1. Reporting the mean for skewed data or data with outliers.
2. Calling the standard deviation resistant (only the median and IQR are).
3. Drawing whiskers all the way to the outliers.
4. Thinking that adding a constant changes the spread.
5. Numbers without context: “the middle half of the scores span 9 points”, not “IQR = 9”.

Remember

- Resistant: median, IQR. Not resistant: mean, s , range.
- Skewed or outliers $\rightarrow$ median & IQR; symmetric $\rightarrow$ mean & s .
- Outlier: below $Q_1 - 1.5\text{IQR}$ or above $Q_3 + 1.5\text{IQR}$.
- Add a constant: center shifts. Multiply: center and spread scale.

Now practise it

Theory sticks once you use it. On **sad-academy.com** you'll find practice questions for this exact subtopic in the AP format, and an AI that grades your free-response answers against the scoring guidelines, the way an AP reader would.

sad-academy.com