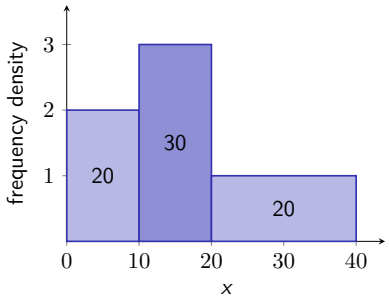
Data presentation and interpretation

AQA A-level Mathematics 7357, section L: histograms, measures of location and spread, box plots, outliers, correlation

Centre and spread

Measure	Kind	Note
Mean $\bar{x}$	centre	uses every value; pulled by outliers
Median	centre	middle value; resistant to outliers
Standard deviation σ	spread	typical distance from the mean
IQR $= Q_3 - Q_1$	spread	spread of the middle half

Histograms show frequency as area



- Classes 0–10, 10–20, 20–40: frequencies 20, 30, 20.
- Densities $\frac{20}{10} = 2$, $\frac{30}{10} = 3$, $\frac{20}{20} = 1$.
- The wide bar is shortest, yet holds 20.

$$\text{frequency density} = \frac{\text{frequency}}{\text{class width}}$$

Reading frequencies from a histogram

1. **Question** how many values lie between 15 and 30?
2. **15 to 20** $3 \times 5 = 15$
3. **20 to 30** $1 \times 10 = 10$
4. **Total** $15 + 10 = 25$ (an estimate)

Decision

An estimate, because values inside a class are spread evenly only by assumption.

Estimating a grouped mean

Class	**Midpoint** x	f	fx
0–10	5	20	100
10–20	15	30	450
20–40	30	20	600
Total		70	1150

$$\bar{x} \approx \frac{\sum fx}{\sum f} = \frac{1150}{70} \approx 16.4$$

The median by interpolation

1. **Given** 70 values: 20 below 10, 50 below 20
2. **Locate** the 35th value lies in the class 10–20
3. **Into the class** $35 - 20 = 15$ of its 30 values
4. **Interpolate** $10 + \frac{15}{30} \times 10 = 15$

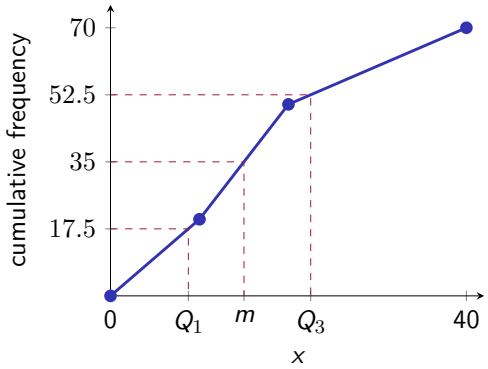
Decision

Median ≈ 15, a little below the mean of 16.4: a slight **positive skew**.

$$\text{median} \approx L + \frac{\frac{n}{2} - F}{f} \times w$$

L lower bound, F frequency before the class, f its frequency, w its width

A cumulative frequency polygon



- Plot at the upper class bounds.
- Read Q_1, median, Q_3 at 17.5, 35, 52.5.
- $Q_1 \approx 8.75$, median 15, $Q_3 \approx 22.5$.

Standard deviation from summary statistics

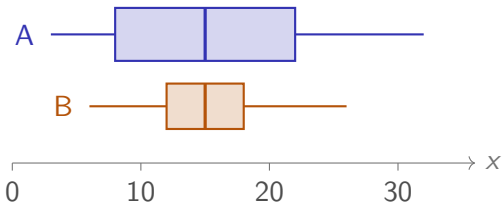
1. **Given** $n = 10$, $\sum x = 50$, $\sum x^2 = 310$
2. **Mean** $\bar{x} = \frac{50}{10} = 5$
3. **Variance** $\frac{310}{10} - 5^2 = 31 - 25 = 6$
4. **SD** $\sigma = \sqrt{6} \approx 2.45$

Decision

Use the calculator's statistics mode on raw data; use the formula on summaries.

$$\sigma^2 = \frac{\sum x^2}{n} - \bar{x}^2$$

Comparing box plots



- Same median, 15.
- IQR: A is 14, B is 6.
- B is more consistent; A is more spread.

Compare in context: "B's middle half is far less spread, though both have median 15."

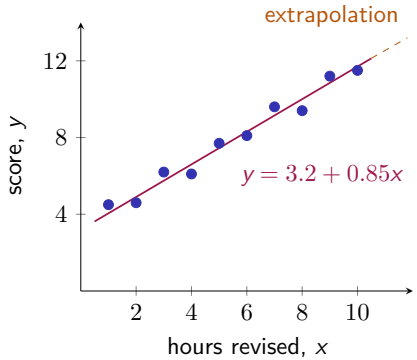
Is it an outlier?

1. **Quartiles** $Q_1 = 12$, $Q_3 = 20$, IQR $= 8$
2. **Fences** $12 - 1.5 \times 8 = 0$ and $20 + 1.5 \times 8 = 32$
3. **Test** 35 lies beyond 32: an **outlier** by this rule
4. **Alternative** or beyond $\bar{x} \pm 2\sigma$, if the question says so

Decision

Remove it only if it is an error; a genuine extreme value stays in the data.

Scatter diagrams and correlation



- Positive, negative or no **correlation**.
- Correlation is not causation.
- Predict only inside the data range.

Interpreting a regression line

1. **Given** $y = 3.2 + 0.85x$ for 1 to 10 hours of revision
2. **Gradient** each extra hour adds about 0.85 marks
3. **Intercept** 3.2 marks with no revision, outside the data: treat with care
4. **Interpolate** $x = 6.5$: $y = 3.2 + 0.85 \times 6.5 = 8.7$, reliable
5. **Extrapolate** $x = 40$: $y = 37.2$, beyond the data and the maximum mark

Traps in presenting data

Misleading

A **histogram**'s tallest bar called the biggest class.

A truncated axis that exaggerates a change.

Equal medians called equal distributions.

Sound

Area is frequency; compare areas.

Read the scale and its starting value.

Compare a spread as well as a centre.

Check yourself

Question	**Answer**
Class 30–50 holds 40 values: density?	2
$n = 5$, $\sum x = 20$, $\sum x^2 = 100$: σ^2?	$20 - 16 = 4$
IQR of $Q_1 = 6$, $Q_3 = 15$?	9
Upper fence for that IQR?	$15 + 13.5 = 28.5$
Predict outside the data range?	**extrapolation**: unreliable