Measures of location and spread

S2.3 · Statistics · revision notes

Calculating and interpreting measures of location (mean, median, mode) and spread (range, interquartile and interpercentile ranges, variance and standard deviation) for raw, grouped and summarised data, including coding. These skills are used throughout the statistics on Paper 3, often in the first part of a longer question.

Key facts

- $\bar{x} = \dfrac{\Sigma x}{n}$ or $\dfrac{\Sigma fx}{\Sigma f}$
- $S_{xx} = \Sigma x^2 - \dfrac{(\Sigma x)^2}{n}$
- Coding $y = \dfrac{x - a}{b}$: $\bar{x} = a + b\bar{y}$, $\sigma_x = b\sigma_y$
- IQR $= Q_3 - Q_1$; interpercentile range, e.g. $P_{90} - P_{10}$
- $\sigma^2 = \dfrac{\Sigma x^2}{n} - \bar{x}^2 = \dfrac{S_{xx}}{n}$
- Interpolation: lower boundary + $\dfrac{\text{position} - \text{cf before}}{\text{class frequency}}$ × class width
- Working backwards: $\Sigma x = n\bar{x}$, $\Sigma x^2 = n(\sigma^2 + \bar{x}^2)$

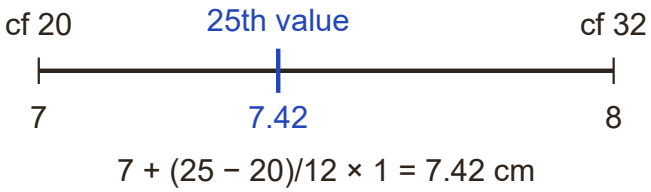
Notes

Measures of location

- Mean $\bar{x} = \dfrac{\Sigma x}{n}$, or $\dfrac{\Sigma fx}{\Sigma f}$ for a frequency table. For grouped data use the class midpoints: the result is an estimate.
- The mean uses all the data but is affected by extreme values; the median is not. The mode (or modal class) suits discrete or non-numerical data.

Quartiles and percentiles

- For a list of n ordered values, to find Q_1 work out $\frac{n}{4}$. If it is a whole number, Q_1 is halfway between that value and the next one; if not, round up and take that value. Use $\frac{n}{2}$ for the median and $\frac{3n}{4}$ for Q_3 in the same way.
- For grouped continuous data use **linear interpolation**: Q_1 is the $\frac{n}{4}$th value, the median the $\frac{n}{2}$th, Q_3 the $\frac{3n}{4}$th and the kth percentile the $\frac{kn}{100}$th.
- Estimate = lower class boundary + $\frac{d}{f}$ × class width, where d is the position minus the cumulative frequency before the class, and f is the frequency of the class.

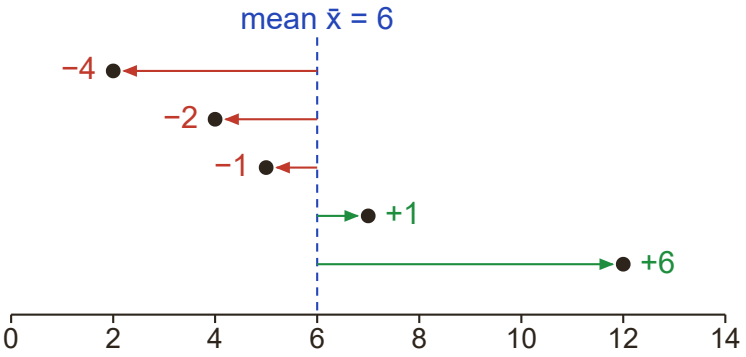


The 12 values in the class are assumed evenly spread: the 25th value is 5 of the 12 steps along

- IQR $= Q_3 - Q_1$; the 10th to 90th interpercentile range is $P_{90} - P_{10}$

Variance and standard deviation

- Variance $\sigma^2 = \frac{S_{xx}}{n} = \frac{\Sigma x^2}{n} - \bar{x}^2$, where $S_{xx} = \Sigma(x - \bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$
- For a frequency table, $\sigma^2 = \frac{\Sigma f x^2}{\Sigma f} - \bar{x}^2$
- Standard deviation $\sigma = \sqrt{\text{variance}}$: the spread about the mean, in the same units as the data.



Deviations x − x̄ always add to 0, so they are squared: σ² = Σ(x − x̄)² ÷ n

- Edexcel's formula uses the divisor n (σ_x on a calculator); answers using $n - 1$ (s_x) are normally accepted too.

Coding

- If $y = \dfrac{x - a}{b}$ (with $b > 0$), then $\bar{y} = \dfrac{\bar{x} - a}{b}$ and $\sigma_y = \dfrac{\sigma_x}{b}$. Decode with $\bar{x} = a + b\bar{y}$ and $\sigma_x = b\sigma_y$
- Adding or subtracting a constant does not change the spread; multiplying or dividing scales it.

How to answer each type of question

M = method mark, A = answer mark, B = a mark on its own.

Find the mean and standard deviation from summary statistics 3 marks **Grade C**

1. Mean $= \Sigma x \div n$
2. Variance $= \Sigma x^2 \div n - \bar{x}^2$, using the unrounded mean.
3. Square root, and give units and 3 significant figures.

Example. For a random sample of 10 days, the daily total sunshine, h hours, gives $\Sigma h = 68.0$ and $\Sigma h^2 = 518.6$. Find the mean and the standard deviation of h

---

Model answer.

Mean $= \dfrac{68.0}{10} = 6.8$ hours (B1)

$\sigma = \sqrt{\dfrac{518.6}{10} - 6.8^2}$ (M1) $= \sqrt{5.62} = 2.37$ hours (A1)

(Using $n - 1$ gives 2.50, which is normally accepted.)

Estimate the median and quartiles by interpolation

3 to 4 marks | **Grade C**

1. Make a cumulative frequency column.
2. Find the position ($\frac{n}{2}$, $\frac{n}{4}$ or $\frac{3n}{4}$) and the class it falls in.
3. Interpolate using that class's lower boundary, frequency and width.

Example. The table shows the lengths, l cm, of 50 leaves.

Length, l cm	Frequency
$3 \leq l < 5$	6
$5 \leq l < 7$	14
$7 \leq l < 8$	12
$8 \leq l < 10$	13
$10 \leq l < 14$	5

(a) Use linear interpolation to estimate the median.
(b) Estimate the interquartile range.

Model answer.

Cumulative frequencies:

Upper class boundary	5	7	8	10	14
Cumulative frequency	6	20	32	45	50

(a) The 25th value is in $7 \leq l < 8$: $7 + \frac{25 - 20}{12} \times 1$ (M1) $= 7.42$ cm (A1)

(b) Q_1 (12.5th value) $= 5 + \frac{12.5 - 6}{14} \times 2 = 5.929$ and Q_3 (37.5th value)

$= 8 + \frac{37.5 - 32}{13} \times 2 = 8.846$ (M1)

IQR $= 8.846 - 5.929$

$= 2.92$ cm (A1)

Use coding

3 to 4 marks **Grade B**

1. Write the coding with x as the subject.
2. Decode the mean: apply both the multiplying and the adding.
3. Decode the standard deviation: apply only the multiplying.

Example. The masses, m grams, of 40 apples are coded using $y = \dfrac{m - 150}{5}$. For the coded data, $\bar{y} = 2.4$ and $\sigma_y = 1.8$. Find the mean and the standard deviation of the masses.

Model answer.

$m = 5y + 150$, so $\bar{m} = 5 \times 2.4 + 150$ (M1) $= 162$ g (A1)

$\sigma_m = 5 \times 1.8 = 9$ g (B1)

Combine two sets of data

5 to 6 marks **Grade A**

1. For each set find n, Σx and Σx^2, using $\Sigma x = n\bar{x}$ and $\Sigma x^2 = n(\sigma^2 + \bar{x}^2)$ if needed.
2. Add the values of n, Σx and Σx^2
3. Calculate the combined mean and standard deviation.

Example. Sample A has 12 values with mean 20 and standard deviation 3. Sample B has 8 values with $\Sigma x = 176$ and $\Sigma x^2 = 3960$. Find the mean and the standard deviation of all 20 values.

Model answer.

Sample A: $\Sigma x = 12 \times 20 = 240$ (M1) and $\Sigma x^2 = 12(3^2 + 20^2) = 4908$ (M1)

Combined: $\Sigma x = 416$, $\Sigma x^2 = 8868$, so the mean is $\dfrac{416}{20} = 20.8$ (A1)

$\sigma = \sqrt{\dfrac{8868}{20} - 20.8^2}$ (M1) $= \sqrt{10.76} = 3.28$ (A1)

Don't lose marks

- Forgetting to square root the variance, or subtracting $\bar{x}$ instead of $\bar{x}^2$
- Rounding the mean before using it in the variance formula.
- Using class limits instead of class boundaries, or the wrong class width, when interpolating.
- Applying the added constant to the standard deviation when decoding.
- Dividing by the number of classes instead of Σf

More tips

Memory tricks

- Variance is ‘the mean of the squares minus the square of the mean’.
- Use your calculator's statistics mode (with a frequency column) to check $\bar{x}$, σ_x, Σx and Σx^2
- Coding: an added or subtracted constant moves the mean but never changes the standard deviation.
- Working backwards from a mean and standard deviation: $\Sigma x = n\bar{x}$ and $\Sigma x^2 = n(\sigma^2 + \bar{x}^2)$

Exam technique

- Give answers to 3 significant figures unless told otherwise, and keep full calculator values between steps.
- When asked to choose a measure, justify it: the median and IQR are not affected by outliers or skew; the mean and standard deviation use all the data.
- Show the formula with the numbers substituted: the method mark is for that line, even if you then use your calculator.

What each grade needs

What you need to be able to do, from the first marks up to the top grade.

Grade E — **Find the mean, median and mode of data**

Mean $= \dfrac{\Sigma x}{n}$; the median is the middle value of the ordered data; the mode is the most common value.

Grade E — **Find quartiles and the interquartile range**

Use the rule for positions in an ordered list, then IQR $= Q_3 - Q_1$

Grade D — **Estimate the mean from grouped data**

Use the class midpoints: $\bar{x} \approx \dfrac{\Sigma fx}{\Sigma f}$

Grade C — **Calculate standard deviation from summary statistics**

Use $\sigma = \sqrt{\dfrac{\Sigma x^2}{n} - \bar{x}^2}$, or $\sqrt{\dfrac{S_{xx}}{n}}$

Grade C — **Interpolate medians and percentiles from grouped data**

Find the position, locate its class, then go the right fraction of the way through that class.

Grade B — **Use coding to find mean and standard deviation**

Reverse the coding for the mean; for the standard deviation only the scaling matters.

Grade A — **Combine data sets using Σx and Σx^2**

Rebuild $\Sigma x = n\bar{x}$ and $\Sigma x^2 = n(\sigma^2 + \bar{x}^2)$ for each set, add them, then recalculate.

Grade A* — **Find unknown data values from summary statistics**

Form and solve equations (sometimes quadratics) using Σx, Σx^2, the mean or the variance.

Free notes, Full packs, practice papers and mock exams at chhetriacademy.com. Written for this site; not taken from past papers. Stuck? 1-to-1 tutoring with a free 30-minute trial lesson.