



FORMULAS YOU MUST KNOW

$$\bar{x} = \frac{\sum f_i x_i}{n}$$

Sample mean:

$$x_{\max} - x_{\min}$$

Range:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

Population SD:

$$\bar{x} + k$$

$$k\bar{x}$$

$$\bar{x}$$

$$\mu = \frac{\sum f_i x_i}{n}$$

Population mean:

$$Q_3 - Q_1$$

IQR:

$$\sigma^2$$

Variance:

$$\bar{x} - k$$

$$\bar{x}/k$$

$$\sigma_x$$

KEY METHODS

- In a 1-mark sampling question, name the method and its defining feature: "systematic — every k -th member". Both are needed for the mark.
- If the question gives " n values, sum = S , mean = μ ", use $n = S/\mu$ to find n in one step.
- "Each value multiplied by 10" $\Rightarrow$ new $\bar{x} = 10\bar{x}$, new $\sigma = 10\sigma$, new variance = $100\sigma^2$. Apply the scaling factor squared to the variance.
- 2-Variable Stats: Menu $\rightarrow 4 \rightarrow 1 \rightarrow 2$. Gives $\bar{x}$, $\bar{y}$, σ_x , σ_y simultaneously — useful for confirming the mean point $(\bar{x}, \bar{y})$.
- To estimate the number of values above a threshold t : read off the cf at t , then subtract from n . Works for "more than" questions in reverse.
- Box plots are compared using: median (centre), IQR (middle spread), range (total spread), and symmetry / skewness. Always address all four when comparing two distributions.
- r^2 (the coefficient of determination) is given by the GDC alongside r . It tells you what fraction of the

EXAM TRAPS TO AVOID

- Quota $\neq$ Stratified. Quota is proportional to stratum size; stratified uses any fixed number from each stratum.
- For grouped data you cannot find the exact mean — write "estimate" or use $\approx$. Also, the modal class is not the same as the mode.
- Adding a constant changes the mean but never changes the standard deviation or variance. Students routinely add k to σ — do not.
- The GDC gives two SDs: σ_x (population) and s_x (sample). IB Math AA HL defaults to population — read σ_x unless the question says "sample standard deviation".
- Always plot against the upper boundary — not the midpoint, not the lower boundary. Using the wrong boundary shifts the entire curve and every quartile will be wrong.
- Each quarter of a box plot contains exactly 25% of the data — but can span very different widths on the axis. A wide box half does not mean more data; it

variation in y is explained by x . Not required for IB but useful for interpreting strength.

means more spread. A narrow box half means data is tightly clustered, not that fewer values are there.

- r close to ± 1 does NOT mean causation. Never write " x causes y " based on correlation alone — always say "there is a (strong / positive) correlation between x and y ".
-