

1 CHAPTER OVERVIEW

Your visual roadmap through the chapter

Measures of Central Tendency

Mean, median, mode, geometric & harmonic mean

1

Measures of Dispersion

Range, IQR, variance, standard deviation, coefficient of variation

2

Measures Compared

Choosing the right average and dispersion measure; skewness; box plots

3

WHAT IT'S ABOUT

Statistical measures that summarize data's central value (averages) and its spread or variability (dispersion), supporting descriptive and predictive analysis.

WHY IT MATTERS FOR QAB

Comparing datasets fairly requires knowing not just their typical value but also how consistent or spread out they are.

EXAM SKILLS REQUIRED

Calculate mean/median/mode for grouped and ungrouped data; compute variance, SD and coefficient of variation; construct and read box plots and ogives.

2 KEY DEFINITIONS

Source-based terminology, exam-ready

Descriptive Statistics

The part of statistics that describes or summarizes information from data, e.g. via central value or dispersion.

Key word: describe/summarize

Arithmetic Mean

The sum of the values in a series divided by the count of values — a mathematical representation of the typical value.

Key word: sum ÷ count

Median

The middle value when data is arranged in ascending or descending order.

Key word: middle value

Mode

The most frequently occurring value (or the highest frequency in grouped data). There may be more than one mode, or none at all.

Key word: most frequent

Geometric Mean

Useful when several quantities are multiplied together to arrive at a product, e.g. average growth rate over several years.

Key word: average growth rate

Harmonic Mean

The number of observations divided by the sum of the reciprocals of each value; useful for average speeds.

Key word: average speed

Variance & Standard Deviation

Variance is the average of the squared deviations from the mean. Standard deviation is its square root.

Key word: $SD = \sqrt{\text{variance}}$

Coefficient of Variation

A relative measure of dispersion expressing standard deviation as a percentage of the mean, used to compare variability between datasets.

Key word: relative spread

3 CORE CONCEPT

Explained in exam-oriented language

CONCEPT → Central Tendency

MEANING

A single value that represents the "centre" or typical value of a dataset.

PURPOSE

Compare two or more datasets using the same units.

APPLICATION

Choosing mean, median, mode, geometric or harmonic mean depending on the data's nature.

CONCEPT → Dispersion

MEANING

How spread out or clustered the data is around its central value.

PURPOSE

Reveal consistency that measures of central tendency alone cannot show (e.g. two batsmen with the same average).

APPLICATION

Absolute measures (range, IQR, variance, SD) vs relative measures (coefficient of variation).

4 KEY RULES & PRINCIPLES

Every source rule, exam-ready

RULE 1 Properties of an ideal average

- Well defined; based on all values; capable of further algebraic calculation
- Not affected by extreme values; stable in repeated sampling

RULE 2 Properties of the arithmetic mean

- Sum of deviations from the mean is always zero; sum of squared deviations is always least
- Combining group means requires the weighted combined-mean formula, NOT a simple average of the means
- Mean of constant values is that constant; mean is affected by a change of origin or scale ($\text{Mean}(x \pm c) = \text{Mean}(x) \pm c$; $\text{Mean}(cx) = c \times \text{Mean}(x)$)

RULE 3 Properties of variance & standard deviation

- SD and variance of a constant value are always zero
- Independent of a change of origin: $\text{SD}(x \pm c) = \text{SD}(x)$
- Affected by a change of scale: $\text{SD}(Cx) = |C| \times \text{SD}(x)$; $\text{Var}(Cx) = C^2 \times \text{Var}(x)$
- If symmetrical: 68.27% of values lie within ± 1 SD, 95.45% within ± 2 SD, 99.73% within ± 3 SD of the mean

RULE 4 Empirical relationship & skewness

- $(\text{Mean} - \text{Mode}) = 3 \times (\text{Mean} - \text{Median})$
- Perfectly symmetric: mean = median
- Skewed left/negative: mean < median | Skewed right/positive: mean > median

5 FORMULA ZONE — CENTRAL TENDENCY

Every formula supported by the source text

MEDIAN (UNGROUPED) & MEAN

$$\text{Position} = (n+1) \div 2 \quad | \quad \text{Mean}(\bar{x}) = \Sigma x \div n \text{ or } \Sigma fx \div \Sigma f$$

Mean formula on the right applies to grouped data using class mid-points (x).

MEDIAN (GROUPED)

$$\text{Median} = L + [(\Sigma f \div 2 - cf) \div f] \times h$$

L=lower boundary of median class, **cf**=cumulative frequency before it, **f**=its frequency, **h**=its width.

MODE (GROUPED)

$$\text{Mode} = L + [(f_m - f_1) \div ((f_m - f_1) + (f_m - f_2))] \times h$$

f_m=frequency of modal class, **f₁**=class before, **f₂**=class after.

COMBINED MEAN & WEIGHTED MEAN

$$\text{Combined} = (\text{Mean}_A n_A + \text{Mean}_B n_B) \div (n_A + n_B) \quad | \quad \text{Weighted} = \frac{\Sigma wx}{\Sigma w}$$

Never average two group means directly — weight by each group's size (n) or importance (w).

GEOMETRIC MEAN

$$\text{G.M.} = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} \quad \text{or} \quad \sqrt[n]{\text{end value} \div \text{start value}}$$

Use (1+growth rate) as each x for compounding growth rates over n periods.

HARMONIC MEAN

$$\text{H.M.} = n \div \Sigma (1 \div x)$$

Number of items divided by the sum of the reciprocals of each value. Useful for average speed/rate problems.

FORMULA ZONE —

5 DISPERSION

Every formula supported by the source text

RANGE & INTER-QUARTILE RANGE

$$\text{Range} = \text{max} - \text{min} \quad | \quad \text{IQR} = Q_3 - Q_1 \quad | \quad \text{Semi-IQR} = (Q_3 - Q_1) \div 2$$

For grouped data, range = upper limit of last class — lower limit of first class.

VARIANCE & STANDARD DEVIATION

$$s^2 = \Sigma (x - \bar{x})^2 \div n \quad \text{or} \quad \Sigma fx^2 \div \Sigma f - (\Sigma fx \div \Sigma f)^2 \div \Sigma f \quad | \quad s = \sqrt{s^2}$$

Second variance formula (grouped) avoids computing each deviation individually.

COEFFICIENT OF VARIATION

$$\text{CV} = (s \div \bar{x}) \times 100$$

Lower CV = greater precision/consistency. Used to compare datasets with different units or scales.

PEARSON'S COEFFICIENT OF SKEWNESS

$$\text{Sk} = 3(\text{Mean} - \text{Median}) \div \text{Standard Deviation}$$

Sk=0: symmetric. Positive Sk: right-skewed. Negative Sk: left-skewed.

6 CALCULATION ROADMAP

General path for any statistical measure question

1- Read Question

Grouped or ungrouped? What measure is needed?

2- Organize Data

Sort, tabulate class/frequency/mid-point

3- Select Formula

Match the measure and data type

4- Substitute

Put known values into the formula

5- Calculate & Check

Solve, recheck, state final answer

7 STEP-BY-STEP METHODS

Kept separate — do not mix methods

METHOD A · Median (Grouped Data)

- 1 Build the cumulative frequency column; find $\Sigma f \div 2$
- 2 Identify the median class (where the cf first reaches/exceeds $\Sigma f \div 2$)
- 3 Note L, cf (before median class), f and h for that class
- 4 Substitute into $\text{Median} = L + [(\Sigma f / 2 - cf) / f] \times h$

METHOD D · Geometric Mean

- 1 Express each period's growth as (1+rate), or note start/end values
- 2 Multiply all growth factors together (or divide end value by start value)
- 3 Take the n-th root, where n=number of periods
- 4 Subtract 1 and convert to a percentage for the average growth rate

METHOD B · Mode (Grouped Data)

- 1 Identify the modal class (highest frequency)
- 2 Note L, f_m (modal class), f_1 (class before), f_2 (class after), and h
- 3 Substitute into $\text{Mode} = L + [(f_m - f_1) / ((f_m - f_1) + (f_m - f_2))] \times h$
- 4 Alternative: draw the histogram and read off where the two diagonal lines cross

METHOD E · Variance & Standard Deviation

- 1 Ungrouped: find $\bar{x}$, then each $(x - \bar{x})$ and square it
- 2 Sum the squared deviations and divide by n for variance
- 3 Grouped: use $\Sigma fx^2 / \Sigma f - (\Sigma fx / \Sigma f)^2$ instead
- 4 Take the square root of variance for standard deviation

METHOD C · Mean (Grouped / Weighted / Combined)

- 1 Grouped: find each class mid-point (x), multiply by frequency (f)
- 2 Sum fx and f , then divide $(\Sigma fx \div \Sigma f)$
- 3 Weighted: multiply each value by its weight, sum, divide by Σw
- 4 Combined: use each group's mean AND size in the combined-mean formula

METHOD F · Coefficient of Variation & Skewness

- 1 Calculate the mean and SD for each dataset being compared
- 2 $CV = (SD \div \text{mean}) \times 100$ for each; lower CV = more consistent
- 3 Skewness: find mean, median and SD
- 4 Substitute into $Sk = 3(\text{Mean} - \text{Median}) / SD$; interpret the sign

8 DIFFERENT CASES / SCENARIOS

Recognise the situation, pick the right measure

IF you want the most commonly-used, calculation-friendly average

→ USE the **arithmetic mean**

IF the distribution is skewed / has extreme values

→ USE the **median** (not affected by extremes)

IF you want the single most frequently occurring value

→ USE the **mode**

IF averaging growth/decline rates or ratios over time

→ USE the **geometric mean**

IF averaging a rate like speed (distance per constant time)

→ USE the **harmonic mean**

IF some values matter more than others

→ USE the **weighted mean**

IF combining the means of two different-sized groups

→ USE the **combined mean formula** — never just average the two means

IF comparing variability between datasets with different units/scales

→ USE the **coefficient of variation** (relative), not raw SD

9 COMPARISON TABLES

Don't confuse these!

	A.M.	G.M.	H.M.	Median	Mode
Uses all values?	Yes	Yes	Yes	Middle value only	Highest-frequency value only
Effect of extremes	Greatest effect	Less than A.M.	Least effect	No effect	No effect
Zero/negative value	Can calculate	Cannot	Cannot	Can calculate	Can calculate
Typical usage	Default choice	Growth/decline rates	Rates like speed	Skewed/asymmetric data	Most frequent value needed

Measure	Advantage	Disadvantage
Mean	Uses every value; used extensively elsewhere	Distorted by extreme values in small samples
Median	Not influenced by extreme values; an actual value (odd n)	Little use in other statistical applications
Mode	Is an actual value	Little use in other applications
Range	Easy to calculate and understand	Distorted by extreme figures
Semi-IQR	Easy to calculate; ignores extremes	Doesn't use all figures; no insight into clustering
Standard Deviation	Uses all values; useful in further applications	Harder to understand/calculate by hand

10 READING OGIVES & BOX PLOTS

How to interpret the source's graphical tools

Reading an ogive for median/quartiles

The median is the value at $\Sigma f \div 2$ on the cumulative-frequency axis; Q1 and Q3 are read at $\Sigma f \times 25\%$ and $\Sigma f \times 75\%$ respectively, tracing across to the curve then down to the value.

Reading a box-and-whisker plot

The box spans Q1 to Q3 (the middle 50%); the line inside is the median. Whiskers extend from the box to the minimum and maximum values in the data.

11 WORKED CALCULATION LOGIC

Method students must reproduce in the exam

GIVEN → FORMULA → SUBSTITUTION → CALCULATION → ANSWER

Group A: 1,200 students, mean score 60. Group B: 400 students, mean score 70. Find the combined mean.

GIVEN: $\text{Mean}_A = 60$, $n_A = 1,200$; $\text{Mean}_B = 70$, $n_B = 400$

FORMULA: Combined Mean = $(\text{Mean}_A n_A + \text{Mean}_B n_B) \div (n_A + n_B)$

SUBSTITUTION: $= (60 \times 1,200 + 70 \times 400) \div (1,200 + 400) = (72,000 + 28,000) \div 1,600$

FINAL ANSWER: Combined mean = $100,000 \div 1,600 = 62.5$ (NOT the simple average of 65)

12 EXAM QUESTION PATTERNS

- **Direct calculation** — mean/median/mode, grouped or ungrouped
- **Geometric/harmonic mean** — growth rates or average speed
- **Combined/weighted mean** — merging groups correctly
- **Dispersion calculation** — range, IQR, variance, SD, CV
- **Comparison** — which dataset is more consistent (CV)
- Graph/box-plot reading and skewness identification

13 EXAM TRAPS

⚠ WATCH OUT

✗ **Averaging two group means directly** — always use the combined-mean formula weighted by group size.

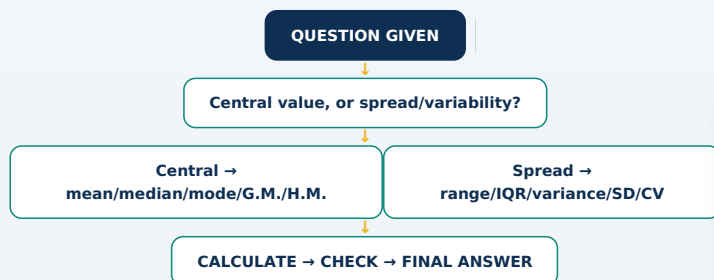
✗ **Variance vs SD mix-up** — forgetting to take the square root of variance to get SD.

✗ **Using arithmetic mean for growth rates** instead of the geometric mean.

✗ **Wrong median class** — picking the class where cf first reaches, not exceeds, $\Sigma f/2$ incorrectly.

✗ **Misreading CV direction** — a LOWER coefficient of variation means MORE consistency, not less.

14 QUICK DECISION TREE



CHAPTER 8 IN 60 SECONDS

Median pos: $(n+1)/2$

Mean: $\Sigma x/n$ or $\Sigma fx/\Sigma f$

Mode: $L + \{[(fm - f1) / ((fm - f1) + (fm - f2))]\} \times h$

G.M.: $\sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$

H.M.: $n / \Sigma (1/x)$

Variance: $\Sigma (x - \bar{x})^2 / n$; **SD** = $\sqrt{\text{variance}}$

CV = $(\text{SD} / \text{mean}) \times 100$

Trap: combined mean $\neq$ simple average