

Sampling and data

A Level Maths · Knowledge organiser

KEY DEFINITIONS

Population and sample

The population is everything you want to know about; a sample is the part you collect data from.

Random sample

Every member of the population has an equal chance of being chosen.

Outlier

A value far from the rest, often defined as more than $1.5 \times \text{IQR}$ beyond a quartile.

Coding

Changing data with $y = (x - a)/b$ to make the numbers easier, then undoing it.

YOU SHOULD BE ABLE TO...

- Distinguish between a population, a sample and a census, and discuss the advantages and disadvantages of sampling compared with collecting data on everyone.
- Classify data as qualitative or quantitative and discrete or continuous, and become familiar with the variables and structure of the large data set.
- Calculate the variance and standard deviation from raw data and from summary statistics such as Σx and Σx^2 , and interpret them as measures of spread.
- Draw cumulative frequency diagrams from grouped data and use them to estimate the median, quartiles, percentiles and the number of values above a given value.
- Draw and interpret scatter diagrams, describe the type and strength of correlation, and explain why correlation does not show that one variable causes another.
- Use and interpret the product moment correlation coefficient as a measure of linear correlation, and describe the strength and direction it shows.

COMMON MISTAKES

- ✗ Premature rounding and truncating
- ✗ Quota sampling called random
- ✗ Stratified sample sizes not rounded sensibly
- ✗ Census and sample advantages swapped

KEY FORMULAS

B = given in the exam (formula booklet or list); learn the rest.

Mean $\bar{x} = \frac{\Sigma x}{n} = \frac{\Sigma fx}{\Sigma f}$

Standard deviation $\sigma = \sqrt{\frac{S_{xx}}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$

Outliers (common rule) $\backslash (\text{ or } > Q_3 + 1.5(Q_3 - Q_1)$

Coding $y = \frac{x-a}{b}$ $\bar{x} = a + b\bar{y}$, $\sigma_x = |b| \sigma_y$

WORKED EXAMPLE

Data are coded with $y = (x - 100)/5$. The coded mean is 3.2 and the coded standard deviation is 1.4. Find the mean and standard deviation of x .

1. $x = 100 + 5y$, so $\bar{x} = 100 + 5 \times 3.2$
2. Adding 100 does not change spread; multiplying by 5 does: $\sigma_x = 5 \times 1.4$

Answer: Mean 116, standard deviation 7