

same mean, three very different teams: the average never tells the whole story

Why the mean is not enough

- dispersion is the extent to which observations deviate from their average value
- three batsmen X, Y, Z: same size ($n=5$), same mean (50), very different swings
- an average that ignores this spread cannot say who is the more dependable player
- range, variance and standard deviation are the three measures of dispersion used here

Same mean, different spread: runs scored in 5 ODIs

*

Batsman	Runs scored	Total	Mean
X	90, 17, 104, 33, 6	250	50
Y	40, 60, 55, 50, 45	250	50
Z	112, 8, 96, 29, 5	250	50

X and Z scatter far more than Y even though all three average exactly 50. Y is the more consistent series

Range

definition

$$\text{Range} = L - S$$



A, the smallest value, and B, the largest value, mark the two ends of the data on the real line: range is the gap between them

L is the largest value in the data, S the smallest. Range is the simplest measure of dispersion: it uses only these two values and ignores every value in between

Range of raw data

Q. weights (in kg) of 10 students are 70, 62, 38, 55, 43, 73, 36, 58, 65, 47. Find the range

- 1 smallest value $S=36$, largest value $L=73$
- 2 $\text{Range} = L - S = 73 - 36$



A is the smallest weight and B the largest: the arrow between them is the range

Answer 37 kg

for grouped data, L is the upper limit of the highest class and S the lower limit of the lowest class

Variance and Standard Deviation

- range only looks at two values; variance and SD use every observation's own deviation
- variance of X: mean of the squared deviations from its own mean, written $\text{Var}(X)$ or σ^2
- standard deviation is the positive square root of the variance, written σ

Variance and SD for raw data *

variance, by definition $\text{Var}(X) = \sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$

variance, computing form $\sigma^2 = \frac{1}{n} \sum x_i^2 - \bar{x}^2$

standard deviation $\sigma = \sqrt{\text{Var}(X)}$

the two variance formulas are the same number. The second, expanding $(x_i - \bar{x})^2$ and simplifying, is faster whenever you already have $\sum x_i^2$