# Distributions, transformations and spread

Original skill preparation; versioned source scope

# Distributions, Transformations and Spread

- Find the mean and the **median 中位数**, and choose between them.
- Read a **frequency 频数** table, dot plot or histogram.
- Compare spread using range, the **interquartile range 四分位距** and the **standard deviation 标准差**.
- Predict how an **outlier 离群值** or a change to every value affects each measure.

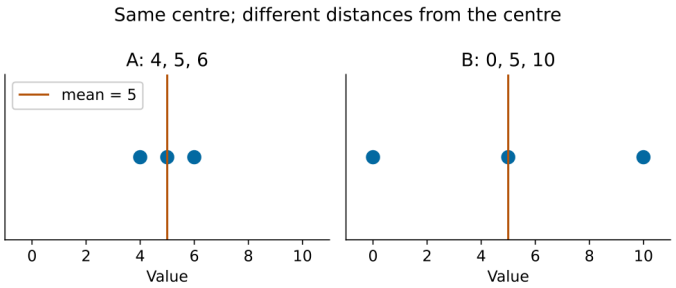
# Centre: mean and median

- The mean is the sum divided by the number of values; the median is the middle value in order (the mean of the middle two if there is an even number).
- An **outlier** pulls the mean towards it but hardly moves the median.
- In a right-skewed distribution (a long tail of high values), the mean is greater than the median.

# Spread

- Range = largest − smallest. It depends only on the two extremes.
- IQR = upper quartile − lower quartile: the spread of the middle half.
- Standard deviation measures the typical distance of values from the mean. The SAT does not ask you to calculate it, only to compare.
- Data bunched near the mean have a smaller standard deviation.

# Figure: Spread




- Data set A (4, 5, 6) and data set B (0, 5, 10), both with mean 5
- Match each part of the diagram to the text.

# Frequency tables and graphs

- In a frequency table, each value counts as many times as its frequency.
- Total $n$ is the sum of the frequencies; the median is at position $(n+1)/2$.
- On a histogram, read the bar heights as counts for each interval, not single values.

# Changing every value

- Adding $k$ to every value adds $k$ to the mean and median; the range, IQR and standard deviation do not change.
- Multiplying every value by $k$ (positive) multiplies every measure of centre and spread by $k$.
- For $z = 3x - 4$: centre $\times 3$ then $-4$; spread $\times 3$ only.

# Worked example: a frequency table

- The data are 2, 2, 2, 5, 5, 8, so $n = 6$.
- Mean $= \dfrac{3(2) + 2(5) + 1(8)}{6} = \dfrac{24}{6} = 4.$
- Median = mean of the 3rd and 4th values $= \dfrac{2 + 5}{2} = 3.5$. Range $= 8 - 2 = 6.$

# Worked example: transforming the data

- New mean $= 3(4) - 4 = 8$; new median $= 3(3.5) - 4 = 6.5$.
- New range $= 3 \times 6 = 18$; the "$-4$" moves every value equally and does not change the spread.
- The standard deviation also triples.

# Common errors

- Forgetting to put the data in order before finding the median.
- Using the number of rows instead of the total frequency as $n$.
- Saying that adding a constant changes the standard deviation.

# Common errors (2)

- Comparing spread with range alone when one value is an **outlier**.
- Choosing the mean for strongly skewed data, such as incomes.

# Key terms

- **median**: The middle value of ordered data.
- **frequency**: The number of times a value occurs.
- **interquartile range**: The upper quartile minus the lower quartile.
- **standard deviation**: A measure of how far values typically lie from the mean.

# Check yourself

- For 3, 4, 4, 5, 24, find the mean and median. Which better describes a typical value?
- Adding 10 to every value: what happens to the standard deviation?
- Which has the larger standard deviation: $\{4, 5, 6\}$ or $\{0, 5, 10\}$?