

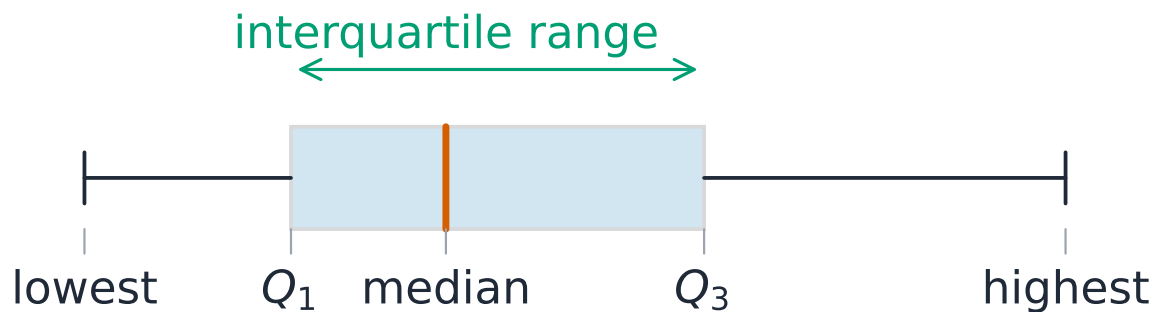
Exercise walk-through

5.1

Representation of data

A-Level Mathematics

Box plot: quartiles



box: Q_1 - Q_3 · median line · whiskers: min/max

box plot: quartiles

box plot: quartiles

Choose the diagram that suits the data

Diagram	What it is for
stem-and-leaf	keeps the original values <i>and</i> shows the shape; a back-to-back version compares two sets round a shared stem
histogram	grouped data, where the area of each bar gives the frequency
box-and-whisker	the five-figure summary: min, Q_1 , median, Q_3 , max
cumulative frequency	read the median across from half the total, quartiles from a quarter and three quarters

In a histogram it is **area**, not height, that means frequency – which is why unequal class widths need frequency *density*.

The diagrams, and what each shows

Stem-and-leaf diagram

A **stem-and-leaf diagram** keeps every original value while showing the shape – so the median and quartiles can be read straight off it.

Box-and-whisker plot

A **box-and-whisker plot** shows the five-number summary and makes two distributions easy to compare side by side.

Histogram

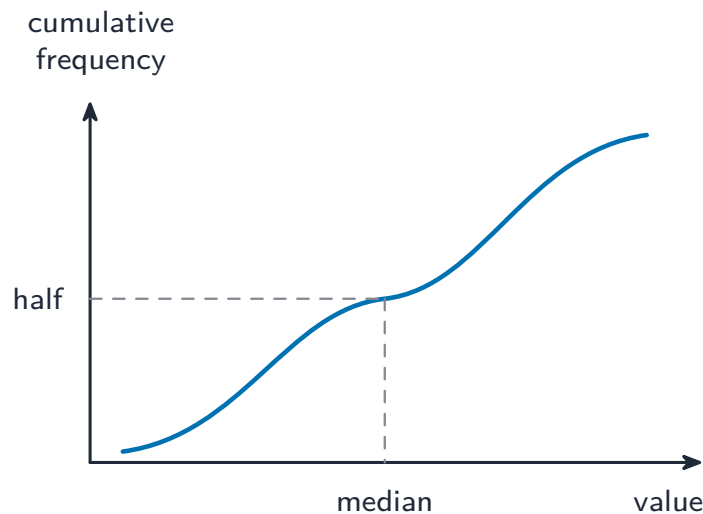
For continuous data in unequal classes: it is the *area* that represents frequency, so the vertical axis is frequency density.

Cumulative frequency

Plot at the upper class boundary and read **percentiles** off the curve.

A **measure of central tendency** says where the data sits; the **range (of data)**, interquartile range and standard deviation say how spread out it is.

Representation of data



A cumulative frequency curve is S-shaped; read the median across from half the total frequency, and the quartiles from one and three quarters

Averages and spread

■ **mean** $\bar{x} = \frac{\sum x}{n}$, **median** ,
mode

■ **standard deviation**

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

■ **interquartile range** $Q_3 - Q_1$

The mean, median and mode are **measures of central tendency** ; the standard deviation and IQR are measures of **variation** . **Coding**

makes big numbers easy: put $t = x - a$ for an **assumed mean** a , then $\bar{x} = a + \bar{t}$ and the standard deviation is **unchanged**.

$$n = 10, \sum x = 50, \sum x^2 = 300$$

$$\bar{x} = 5, \sigma = \sqrt{30 - 25} = 2.24$$

Exercises 2.1·2.2

2.1 Find the interquartile range when $Q_1 = 14$ and $Q_3 = 23$. [1]

2.2 For 8 values, $\sum x = 96$. Find the mean. [1]

Exercises 2.1·2.2

2.1 Find the interquartile range when $Q_1 = 14$ and $Q_3 = 23$. [1]

2.1

- $\text{IQR} = 23 - 14 = 9$
-

2.2 For 8 values, $\sum x = 96$. Find the mean. [1]

2.2

- $\bar{x} = \frac{96}{8} = 12$

Exercise 3.1

3.1 For data with $\sum x = 200$, $\sum x^2 = 4500$ and $n = 10$, the variance is: [1]

☐

A 25

B 50

C 450

D 5

Exercise 3.1

3.1 For data with $\sum x = 200$, $\sum x^2 = 4500$ and $n = 10$, the variance is: [1]

☐

A 25

B 50

C 450

D 5

3.1 B

- $$\text{Variance} = \frac{4500}{10} - \left(\frac{200}{10}\right)^2 = 450 - 400 = 50$$

Exercise 3.2 (a)

3.2 The masses (kg) of 7 parcels are 3, 5, 5, 8, 9, 12, 16.

(a) Find the median and the interquartile range.

[3]

Exercise 3.2 (a)

3.2 The masses (kg) of 7 parcels are 3, 5, 5, 8, 9, 12, 16.

(a) Find the median and the interquartile range.

[3]

(a)

- ordered already; median (4th) = 8; $Q_1 = 5$, $Q_3 = 12$, $\text{IQR} = 7$

Exercise 3.2 (b)

3.2 The masses (kg) of 7 parcels are 3, 5, 5, 8, 9, 12, 16.

(b) Find the mean and the standard deviation.

[4]

Exercise 3.2 (b)

3.2 The masses (kg) of 7 parcels are 3, 5, 5, 8, 9, 12, 16.

(b) Find the mean and the standard deviation.

[4]

(b)

$$\begin{aligned} \bullet \quad \sum x &= 58, \quad \bar{x} = \frac{58}{7} = 8.29; \quad \sum x^2 = 9 + 25 + 25 + 64 + 81 + 144 + 256 = 604; \\ \sigma &= \sqrt{\frac{604}{7} - 8.29^2} = \sqrt{86.29 - 68.7} = \sqrt{17.6} = 4.20 \end{aligned}$$

Exercise 3.3

3.3 A cumulative frequency graph for 200 students rises smoothly. The median is read as 52, $Q_1 = 40$ and $Q_3 = 63$. Estimate how many students scored more than 63. [2]

Exercise 3.3

3.3 A cumulative frequency graph for 200 students rises smoothly. The median is read as 52, $Q_1 = 40$ and $Q_3 = 63$. Estimate how many students scored more than 63. [2]

3.3

- above Q_3 is the top quarter $= \frac{1}{4} \times 200 = 50$ students