

Statistics

IGCSE Mathematics ■ Topic 9

IGCSE Mathematics



What we will cover

- 1 Averages & range
- 2 Frequency tables
- 3 Quartiles & box plots
- 4 Pie charts
- 5 Cumulative frequency & histograms
- 6 Scatter & correlation

1

Averages

Averages and range



$$\text{median} = 3$$

$$\text{mean} = \frac{2+3+3+5+7}{5} = 4$$

$$\text{mode} = 3 \text{ (most common)}$$

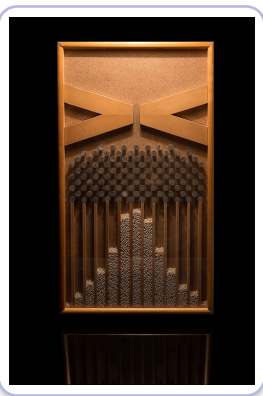
$$\text{range} = 7 - 2 = 5$$

- **mean** 平均数
$$= \frac{\text{sum of all values}}{\text{how many values}}$$
- **median** 中位数 = the middle value when the data is put in order.
- **mode** 众数 = the value that appears most often.
- **range** 极差 = largest value – smallest value (it shows how spread out the data is).

Mean from a frequency table

$$\text{mean} = \frac{\sum(\text{value} \times \text{frequency})}{\sum \text{frequency}}$$

Values 1, 2, 3 with frequencies 4, 5, 1:
 $\frac{1(4) + 2(5) + 3(1)}{10} = 1.7$. For **grouped data** 分组数据 use each group's **midpoint** 中点;
the **modal class** 众数组 has the highest frequency.



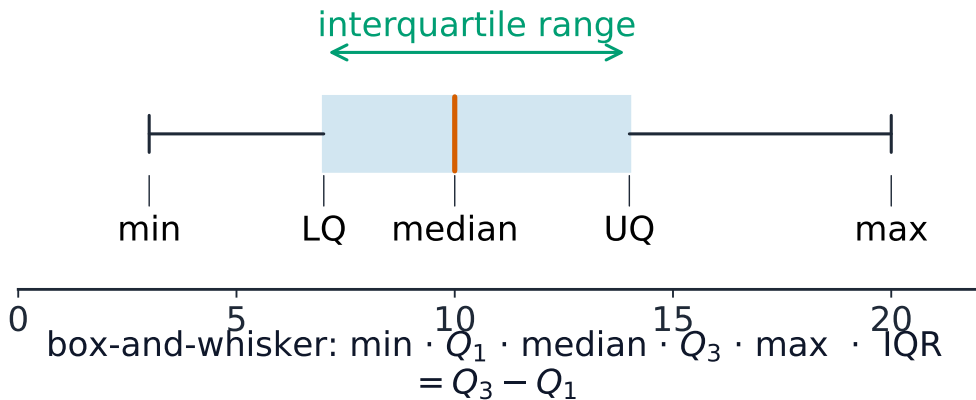
Distribution galton

Quartiles and box plots

Quartiles 四分位数 cut ordered data into four.

Interquartile range 四分位距 = $UQ - LQ$ – the spread of the middle half, ignoring extremes.

Quartiles and box plots, in detail



A box-and-whisker plot shows the five-number summary; the box length is the interquartile range.

2

Charts

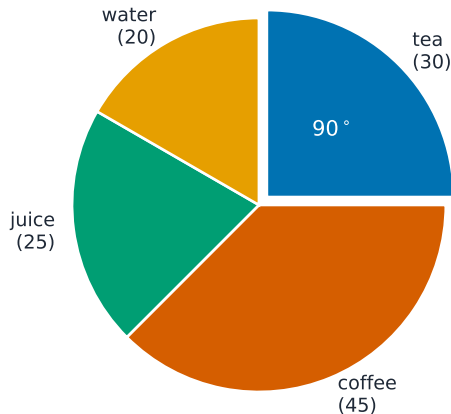
Pie charts

Each slice is its fraction of 360° :

$$\text{angle} = \frac{\text{frequency}}{\text{total}} \times 360^\circ$$

A **bar chart** 条形图 or **pictogram** 象形图 shows each group's frequency.

Pie charts, in detail

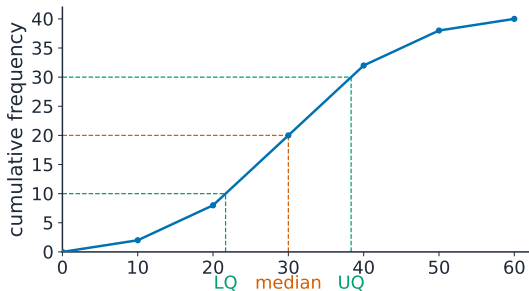


Each slice is its fraction of 360°:

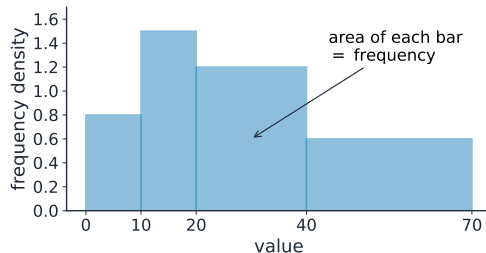
$$\text{angle} = \frac{\text{frequency}}{\text{total}} \times 360^\circ$$

A **bar chart** 条形图 or **pic-togram** 象形图 shows each group's frequency.

Cumulative frequency and histograms



median at $n/2$ value = UQ - LQ



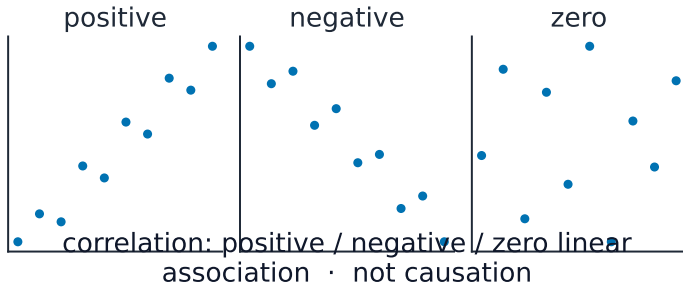
density = freq / width · area = frequency

Read the **median** 中位数 at half the total on a **cumulative frequency** 累积频数 curve. On a **histogram** 直方图 with unequal widths, the **area** (frequency density) shows frequency.

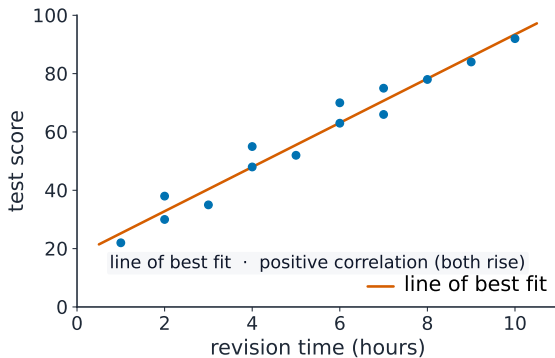
Scatter and correlation

A **scatter diagram** 散点图 shows **correlation** 相关性:

- positive – both rise; negative – one falls
- draw a **line of best fit** 最佳拟合线 to predict



Scatter diagrams and correlation



A line of best fit is one straight ruled line through the middle of the points; use it to predict.

Reading a scatter diagram

Positive correlation

As one goes up, so does the other – the points rise from left to right.

Negative correlation

Negative correlation 负相关: as one goes up, the other falls.

Zero correlation

Zero correlation 零相关: no pattern, and no line of best fit worth drawing.

Collecting the data

A **tally table** 计数表 records the raw counts before any **statistical** 统计的 summary is made.

Correlation is not cause. Say what the connection might be – or say that a third factor could explain both.

Collecting and showing data



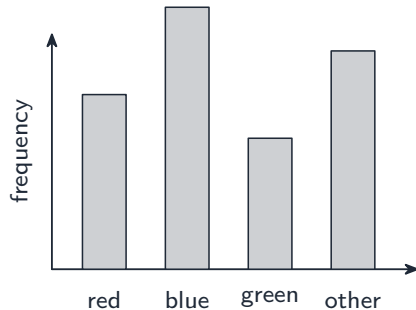
Data is collected from people – a sample drawn from a population.

Measures of spread (Extended)

When data is in order, the **quartiles** 四分位数 cut it into four equal parts.

The lower quartile (LQ) is one quarter of the way up; the upper quartile (UQ) is three quarters of the way up.

Charts and diagrams



A bar chart shows the frequency of each group

Choosing how to show data, and the correlation words

a frequency distribution 频数分布

a table of values against how often each occurs

a two-way table 双向表

sorts data by two features at once – boys/girls against a choice

a bar chart or pictogram

for categories; a histogram for grouped continuous data

a stem-and-leaf diagram 茎叶图

shows the shape and keeps every original value

positive correlation 正相关

as one rises so does the other; negative 负相关 means one rises as the other falls

a percentile 百分位数

read from a cumulative frequency curve – the median is the 50th

Correlation is not causation, and the exam says so. A strong line of best fit shows association only.

Charts and diagrams, continued

Mon



Tue



each ● = 1 item

Wed



A pictogram uses a symbol to stand for a number of items

Exam tips

- Know the three averages: **mean** (add up $\div$ how many), **median** (the middle value when in order), **mode** (the most common). The **range** is highest — lowest.
- From a frequency table the mean is $\frac{\sum fx}{\sum f}$ — divide by the *total frequency*, not the number of rows.
- On a **scatter diagram**, describe the correlation as positive, negative or none, and draw the line of best fit through the mean point.
- For a **histogram** with unequal class widths, the height is the *frequency density* (frequency $\div$ class width), not the frequency.

3

Recap

Worked example – the four averages

Find the mean, median, mode and range of 4, 7, 7, 2, 5.

- Mean = $\frac{4 + 7 + 7 + 2 + 5}{5} = \frac{25}{5} = 5$
- Order first: 2, 4, 5, 7, 7.
- Median = the middle value = 5
- Mode = most common = 7; range = $7 - 2 = 5$

Always order the list before taking the median – that is the single most common lost mark. With an even count, take the mean of the middle two.

Topic 9 – key takeaways

1

Averages

mean, median, mode; range is the spread

2

Spread

$IQR = UQ - LQ$ ignores extremes

3

Pie

slice angle = fraction of 360°

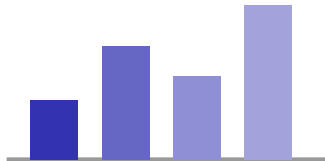
4

Scatter

correlation; a line of best fit predicts

Check yourself

- 1 Which average is the middle value?
- 2 What is the range of 2, 4, 5, 7, 7?
- 3 What is the interquartile range?
- 4 What does a line of best fit help you do?



Answers 1. the median 2. 5 3. $UQ - LQ$ 4. predict / estimate values