Statistics

AQA GCSE 8300 topic 3.6 full package: three-tier S1-S6 with June 2024 Higher anchors

Statistics

- understand samples and populations, and the limits of what a **sample** can tell you;
- draw and read tables, bar charts, pie charts and time-series graphs;
- calculate and compare averages and measures of spread, including from frequency tables;

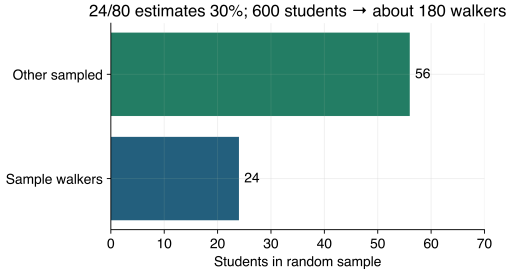
Statistics (2)

- (higher tier) draw and read histograms, **cumulative frequency** graphs and box plots;
- use scatter graphs, **correlation** and lines of best fit.

Samples and data

- **Primary data 原始数据** is collected by you; **secondary data 二手数据** was collected by someone else.
- **Categorical data 分类数据** is in words, such as favourite colour.
- **Discrete 离散** data can take only certain values, such as the number of brothers and sisters.
- **Continuous data 连续数据** can take any value in a **range**, such as a time or a height, even if it is rounded when it is recorded.

Figure: Samples and data

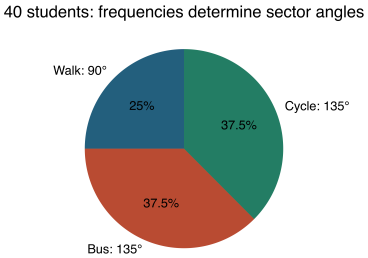


- 24 out of 80 is 30%, so about $0.3 \times 600 = 180$ walkers in the school
- Read the axes and units first.

Tables, charts and time series

- Choose a chart that suits the data.
- Bar charts, pictograms and **pie charts** 饼图 suit **categorical data**.
- Vertical line charts suit **discrete** numbers.
- A **time series** 时间序列 shows how something changes over time, such as monthly rainfall, and is drawn as a line graph with the points joined in time order.

Figure: Tables, charts and time series



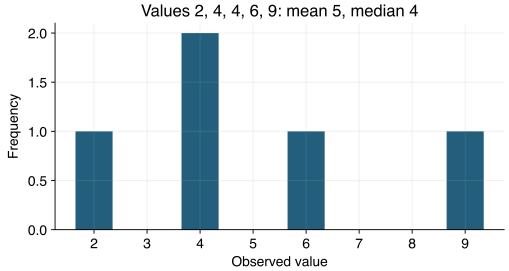
- Each student is $360° \div 40 = 9°$: 10 walkers make 90°, 15 cyclists 135°
- Read the axes and units first.

Averages and spread

- The **mean** 平均数 is the total divided by the number of values.
- The **median** 中位数 is the middle value when the data is in order.
- The **mode** 众数 is the most common value; for grouped data, the **modal class** 众数组 is the group with the highest frequency.

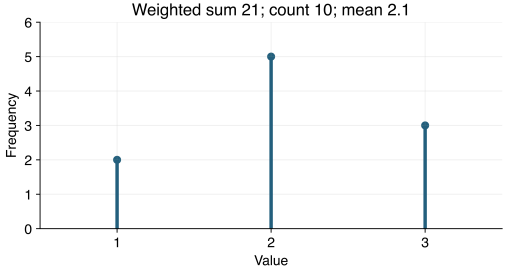
$$\text{mean} = \frac{\sum fx}{\sum f}$$

Figure: Averages and spread



- For 2, 4, 4, 6, 9: **mean** $= 25 \div 5 = 5$, **median** $= 4$
- Read the axes and units first.

Figure: Averages and spread (2)



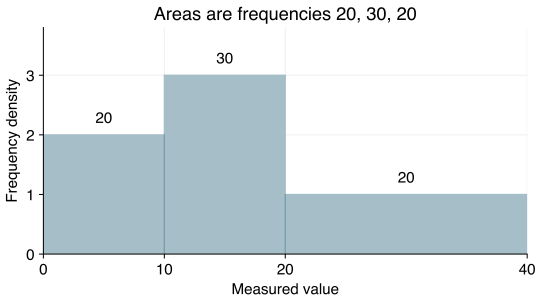
- **Mean** $= (1\times2 + 2\times5 + 3\times3) \div 10 = 21 \div 10 = 2.1$
- Match each part of the diagram to the text.

Histograms, cumulative frequency and box plots (H)

- **Histograms.** In a **histogram 直方图** the area of each bar, not its height, shows the frequency.
- When the **class intervals 组距** have different widths, the height is the **frequency density 频率密度**:
- **Worked example 6 (H).** Classes $0 \leq x < 10$, $10 \leq x < 20$ and $20 \leq x < 40$ have frequencies 20, 30 and 20.
- The **class width 组宽** of the last class is 20, so the frequency densities are $20 \div 10 = 2$, $30 \div 10 = 3$ and $20 \div 20 = 1$.

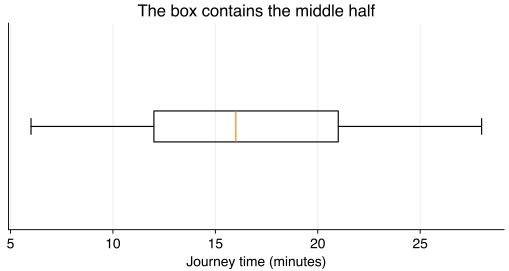
$$\text{frequency density} = \frac{\text{frequency}}{\text{class width}}$$

Figure: Histograms, cumulative frequency and box plots (H)



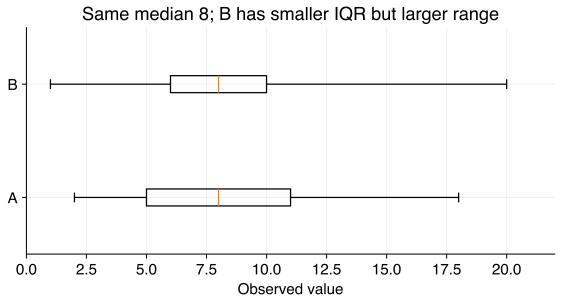
- Area = frequency: $10 \times 2 = 20$, $10 \times 3 = 30$, $20 \times 1 = 20$
- Read the axes and units first.

Figure: Histograms, cumulative frequency and box plots (H) (2)



- The box holds the middle half: quartiles 12 and 21 minutes, IQR 9, **median** 16
- Read the axes and units first.

Figure: Histograms, cumulative frequency and box plots (H) (3)

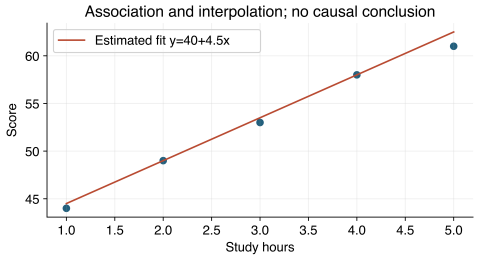


- Same **median** 8; B has the smaller IQR (4 against 6) but the larger **range**
- Read the axes and units first.

Scatter graphs and correlation

- positive **correlation**: as one value increases, the other tends to increase;
- negative correlation: as one value increases, the other tends to decrease;
- no correlation: no clear pattern.

Figure: Scatter graphs and correlation



- Use the line to **interpolate** inside the data, for example $4.5 \rightarrow 60$; it shows association, not cause
- Read the axes and units first.

Sources for this lesson

- AQA, *GCSE Mathematics (8300) specification*, subject content section 3.6 (Statistics).
- "Landon, 1,293,669; Roosevelt, 972,897", *The Literary Digest* (31 October 1936).
- P. Squire, "Why the 1936 Literary Digest poll failed", *Public Opinion Quarterly* 52 (1988), pages 125 to 133.
- AQA, *GCSE Mathematics* Higher Papers 1, 2 and 3 (June 2024), the **pie chart**, frequency table and **box plot** questions, adapted.

Common errors

- Thinking a larger **sample** removes bias.
- Using the **class width** instead of the midpoint to estimate a **mean**.
- Dividing $\sum fx$ by the number of classes instead of the total frequency.

Common errors (2)

- (H) Drawing **histogram** bars with height equal to frequency when the class widths differ.
- (H) Plotting **cumulative frequency** at the midpoint instead of the upper end of the class.
- Comparing only averages; a comparison also needs a measure of spread.
- Saying that **correlation** proves cause.

Key terms

- **Population**: the whole group being studied.
- **Sample**: part of a **population**, used to learn about the whole.
- **Mean**: the total of the values divided by the number of values.

Key terms (2)

- **Median**: the middle value when the data is in order.
- **Range**: the largest value minus the smallest.
- **Interquartile range**: the upper **quartile** minus the lower quartile.

Key terms (3)

- **Frequency density**: frequency divided by **class width**.
- **Correlation**: a relationship between two variables shown on a **scatter graph**.

Check yourself

- 1.
- Find the **mean**, **median**, **mode** and range of 3, 7, 7, 8, 10. 2.
- A pie chart shows 72 people.
- How many degrees stand for each person, and what angle stands for 20 people? 3.