Exercise walk-through

6.1.3 Statistical Samples

AQA · GCSE · Mathematics

Samples and data

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

Histograms, cumulative frequency and box plots (H)

- **Histograms.** In a **histogram** the area of each bar, not its height, shows the frequency.
- When the **class intervals** have different widths, the height is the **frequency density**:
- **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** of the last class is 20, so the frequency densities are $20 \div 10 = 2$, $30 \div 10 = 3$ and $20 \div 20 = 1$.

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

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.

Exercise 2.1

2.1 A factory makes 2000 bulbs a day. 50 of them are tested. Write down the population and the sample. [1]

Exercise 2.1

2.1 A factory makes 2000 bulbs a day. 50 of them are tested. Write down the population and the sample. [1]

2.1

- population: all **2000 bulbs** made that day; sample: the **50 tested**

Setup for exercise 2.2

2.2 Of the 50 bulbs tested in **2.1**, 2 are faulty. Estimate the number of faulty bulbs made each day. [2]

Exercise 2.2

2.3 Explain what is meant by a random sample. [1]



2.4 Ana surveys people waiting at a bus stop to find out how people in her town travel to work. Explain why her sample is biased. [1]



2.5 In a random sample of 60 voters, 33 support a new park. The town has 12 000 voters. Estimate how many support the park. [2]

Exercise 2.2

2.3 Explain what is meant by a random sample. [1]



2.4 Ana surveys people waiting at a bus stop to find out how people in her town travel to work. Explain why her sample is biased. [1]



2.5 In a random sample of 60 voters, 33 support a new park. The town has 12 000 voters. Estimate how many support the park. [2]

2.2

- $\frac{2}{50} = 0.04$ of the sample are faulty

Full solution on next slide

exercises.lol exercises.com.cn

Full solution 2.2

2.2

- $\frac{2}{50} = 0.04$ of the sample are faulty
- $2000 \times 0.04 = \mathbf{80}$ bulbs

2.3

- every member of the population has an **equal chance** of being chosen

2.4

- people at a bus stop are more likely to travel by bus, so other ways of travelling are under-represented

2.5

- $\frac{33}{60} = 0.55$
- $12\,000 \times 0.55 = \mathbf{6600}$ voters

Exercise 3.1

3.1 Explain why a sample of 200 is likely to give a better estimate than a sample of 20. [1]

Exercise 3.1

3.1 Explain why a sample of 200 is likely to give a better estimate than a sample of 20. [1]

3.1

- a larger sample is less affected by chance, so its proportion is likely to be closer to the population's

Setup for exercise 3.2

3.2 A survey about the favourite sport of a town's teenagers only asks members of a football club. Give one criticism of the survey and one way to improve it. [2]

Exercise 3.2

3.3 A questionnaire asks “How many hours of TV do you watch?” with the boxes 0–1, 1–2, 2–3 and more than 4. Give two criticisms of the boxes. [2]



3.4 Amy’s random sample of 80 students has 24 walkers. Ben’s random sample of 40 students from the same school has 15 walkers. Combine the samples to estimate how many of the school’s 600 students walk. [3]



3.5 *(H)* 40 fish in a lake are caught, marked and released. Later 50 fish are caught, and 8 of them are marked. Estimate the number of fish in the lake, and state one assumption you have made. [4]

Exercise 3.2

3.3 A questionnaire asks "How many hours of TV do you watch?" with the boxes 0–1, 1–2, 2–3 and more than 4. Give two criticisms of the boxes. [2]



3.4 Amy's random sample of 80 students has 24 walkers. Ben's random sample of 40 students from the same school has 15 walkers. Combine the samples to estimate how many of the school's 600 students walk. [3]



3.5 *(H)* 40 fish in a lake are caught, marked and released. Later 50 fish are caught, and 8 of them are marked. Estimate the number of fish in the lake, and state one assumption you have made. [4]

3.2

- criticism: football club members are more likely to choose football, so the sample is **biased**

Full solution on next slide

exercises.lol exercises.com.cn

Full solution 3.2 (1/2)

3.2

- criticism: football club members are more likely to choose football, so the sample is **biased**
- improvement: take a **random** sample of all the town's teenagers

3.3

- the boxes **overlap**: 1 and 2 each fit in two boxes
- there is a **gap**: between 3 and 4 hours has no box; and no time frame (per day or per week) is given

3.4

Full solution 3.2 (2/2)

- combined: $24 + 15 = 39$ walkers out of $80 + 40 = 120$
- $\dfrac{39}{120} = 0.325$
- $600 \times 0.325 = \mathbf{195}$ students

3.5

- marked fraction in the second sample: $\dfrac{8}{50}$, which should equal $\dfrac{40}{N}$
- $N = \dfrac{40 \times 50}{8} = \mathbf{250}$ fish

- assumption: the marked fish mix fully with the others, and no fish are born, die or leave between the catches