Exercise walk-through

2.5 Statistics

GAC Mathematics

Statistics: collecting and displaying data

- A **population** is everyone you want to describe; a **sample** is who you actually measure.
- To generalise, the **sample** must be **representative**: chosen so that every part of the **population** has a fair chance of being included.
- **Qualitative data** are categories; **quantitative data** measure amounts, either **discrete** (counted) or **continuous** (measured).
- **Bias** enters through who you ask, who answers and how the question is worded. The *Literary Digest* poll had all three problems: its lists favoured the wealthy, and the 24% who replied were a self-selected group.

$$D_1 = \frac{20}{10} = 2 \qquad D_2 = \frac{20}{20} = 1$$

Statistics: measures of variability

- The **range** is the largest value minus the smallest, so one extreme value can change it a lot.
- The **interquartile range (IQR)** is $Q_3 - Q_1$, the spread of the middle half. A **box plot** shows the minimum, quartiles, **median** and maximum.
- The **standard deviation** measures spread using squared distances from the **mean**. For a whole **population**, $\sigma = \sqrt{\sum (x-\mu)^2/n}$; for a sample estimate, $s = \sqrt{\sum (x-\bar{x})^2/(n-1)}$. It has the same units as the data.
- Two data sets can share a mean and be completely different: a city with every day at 20 °C and one swinging between 0 °C and 40 °C have the same mean temperature.

$$\text{range} = 12 - 2 = 10 \qquad IQR = 8 - 4 = 4$$

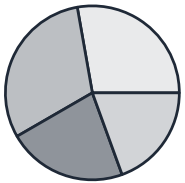
Sources for this lesson

- GAC010 Mathematics II course outline (sequences, finance, probability and statistics).
- *Labour Law of the People's Republic of China* (1994), Article 44, on overtime pay.
- National Bureau of Statistics of China, *Statistical Communiqué on the 2023 National Economic and Social Development* (per capita disposable income, **mean** and **median**).

Exercises 2.1·2.2

2.1 A school studies all 600 students but surveys 60 randomly selected students. Identify its population and sample. [2]

2.2 Classify each variable as categorical or quantitative: (a) preferred bus route label, (b) travel time in minutes. [2]



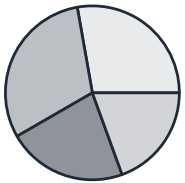
Exercises 2.1·2.2

2.1 A school studies all 600 students but surveys 60 randomly selected students. Identify its population and sample. [2]

2.1

Population: all 600 students. Sample: the 60 selected and surveyed students.

2.2 Classify each variable as categorical or quantitative: (a) preferred bus route label, (b) travel time in minutes. [2]

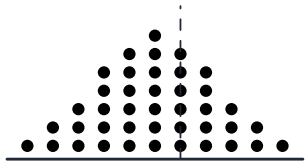


2.2

Route label is categorical, even if routes use numeral names. Travel time is quantitative.

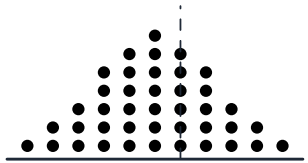
Exercise 2.3

2.3 A school asks only members of its sports club whether more money should go to sports facilities. Identify a sampling limitation, explain its effect, and suggest a better selection method. [3]



Exercise 2.3

2.3 A school asks only members of its sports club whether more money should go to sports facilities. Identify a sampling limitation, explain its effect, and suggest a better selection method. [3]

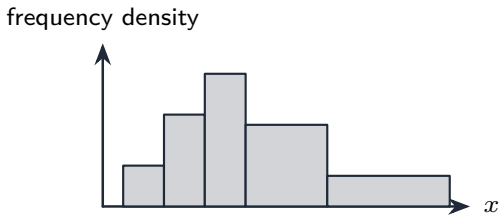


2.3

The sample excludes students outside the club. Club members may have different preferences, so results need not represent the entire school. Select randomly from the full student list, or use a specified representative stratified design including club and non-club students. A better design still needs honest handling of nonresponse.

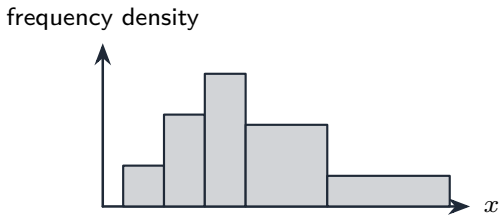
Exercise 2.4

2.4 A histogram class from 20 to 30 minutes has frequency 15. Find its width and frequency density, and state what the area of its bar represents. [3]



Exercise 2.4

2.4 A histogram class from 20 to 30 minutes has frequency 15. Find its width and frequency density, and state what the area of its bar represents. [3]



2.4

Width is 10 minutes.

$$D = f/w = 15/10 = 1.5 \text{ observations per minute}$$

Density 1.5. The bar area represents 15 observations.

Exercise 3.1

3.1 A supplied travel-time table uses classes 0 to less than 10 minutes: 12 students; 10 to less than 20: 18; 20 to less than 40: 20. (a) Find the total count. (b) Find each frequency density for a histogram. (c) Find the proportion taking less than 20 minutes. (d) Explain why using frequencies directly as bar heights would distort the unequal-width histogram. [6]



Exercise 3.1

3.1 A supplied travel-time table uses classes 0 to less than 10 minutes: 12 students; 10 to less than 20: 18; 20 to less than 40: 20. (a) Find the total count. (b) Find each frequency density for a histogram. (c) Find the proportion taking less than 20 minutes. (d) Explain why using frequencies directly as bar heights would distort the unequal-width histogram. [6]



3.1

(a) Total count $N = 12 + 18 + 20 = 50$.

(b) Divide frequency by the corresponding class width.

$$D_1 = f_1/w_1 = 12/10 = 1.2$$

Full solution on next slide

Full solution 3.1

3.1

(a) Total count $N = 12 + 18 + 20 = 50$.

(b) Divide frequency by the corresponding class width.

$$D_1 = f_1/w_1 = 12/10 = 1.2$$

$$D_2 = f_2/w_2 = 18/10 = 1.8$$

$$D_3 = f_3/w_3 = 20/20 = 1.0$$

All three densities correct.

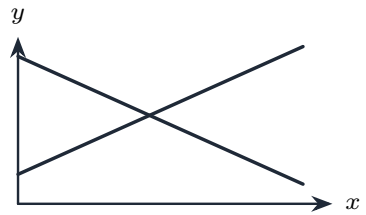
(c) Combine the first two counts.

$$p = \frac{12 + 18}{50} = 0.60$$

Proportion 0.60, or 60%. Frequency-height bars would give the wider class too much area; density ensures area represents frequency.

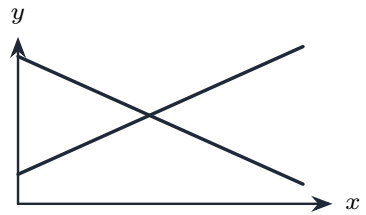
Exercise 3.2

3.2 An optional online survey was sent to 200 students. Sixty responded; 45 of those preferred later club meetings. A report says, “75% of all students want later meetings, so later meetings will improve attendance.” (a) Calculate the response rate and the preference percentage among respondents. (b) Explain two distinct unsupported claims in the report. (c) Propose a better way to study the preference and a separate way to evaluate attendance after a permitted schedule change. [8]



Exercise 3.2

3.2 An optional online survey was sent to 200 students. Sixty responded; 45 of those preferred later club meetings. A report says, “75% of all students want later meetings, so later meetings will improve attendance.” (a) Calculate the response rate and the preference percentage among respondents. (b) Explain two distinct unsupported claims in the report. (c) Propose a better way to study the preference and a separate way to evaluate attendance after a permitted schedule change. [8]



3.2

(a) Divide replies by invited students.

Full solution on next slide

Full solution 3.2

3.2

(a) Divide replies by invited students.

$$r = \frac{60}{200} \times 100\% = 30\%$$

Response rate.

$$p_{respondents} = \frac{45}{60} \times 100\% = 75\%$$

Respondent preference percentage.

(b) The 140 nonrespondents' preferences are unknown, and voluntary responders may differ. The 75% describes respondents rather than all students. A stated preference does not measure a causal effect on attendance.

(c) Use an accessible, clearly worded survey with a specified selection method and follow-up on nonresponse. For a permitted change, record attendance using the same definition over comparable periods and account for competing events or group differences. Describe remaining limitations or, where feasible and approved, use a suitable comparison design rather than announcing causality from a simple before-after total.