

2.5 Collect, classify and display data without overstating evidence

Name: _____ Class: _____ Date: _____

Total: 24 marks

KEY WORDS population 总体 • sample 样本 • bias 偏差

Objective

Collect, classify and display data without overstating evidence. Original GAC004-aligned practice, not an official assessment. Show the method as well as the result. All applied scenarios are invented classroom problems.

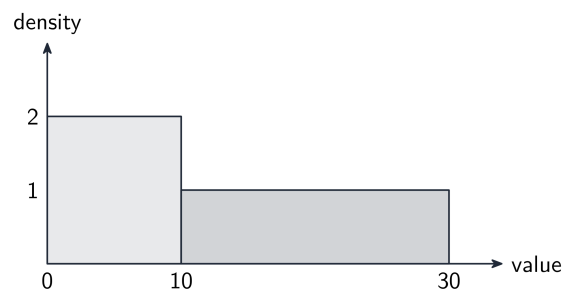
1 Key knowledge and worked examples

■ Key knowledge: a sample and a graph both need a defined scope

The **population** 总体 is the group a question concerns; a **sample** 样本 is the observed subset. A **bias** 偏差 can systematically favour some members or responses. A large convenience sample does not automatically represent the population. Record units, collection method, nonresponse and the question asked.

Categorical data use labels; quantitative data use numbers with a meaningful measurement scale. Discrete data are counts and continuous data are measured over an interval. Use separated bars for categories. For unequal-width continuous histogram classes, frequency density is frequency divided by class width; area, not height alone, represents frequency. Association in observational data does not establish causation.

■ Worked example: equal bar heights can conceal unequal frequencies



Known: class 0 to 10 has frequency 20; class 10 to 30 has frequency 20. Widths differ, so calculate frequency density.

$$D = \frac{f}{w}$$

$$D_1 = \frac{20}{10} = 2$$

$$D_2 = \frac{20}{20} = 1$$

The second bar is half as high but twice as wide. Both areas represent 20 observations.

2 Practice

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]

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]

2.4 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]

3 Exam-style questions

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]

3.2 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]

Solutions

2.1

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

2.2

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

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.

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.

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.

3.2

(a) Divide replies by invited students.

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

Response rate.

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

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.