Problem-Solving and Data Analysis

Original skill preparation; versioned source scope

Problem-Solving and Data Analysis on the SAT

- Work with ratios, a **rate** 比率, percentages and unit conversions.
- Describe data with the mean, **median** 中位数 and spread.
- Read scatterplots and lines of best fit.

Problem-Solving and Data Analysis on the SAT (2)

- Find probabilities from tables.
- Judge what a **random sample** 随机样本 or an experiment allows you to conclude.

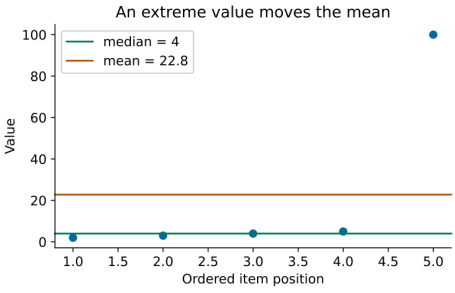
Ratios, rates and percentages

- A **rate** compares two quantities with different units: 60 km per hour, 3 dollars per kg.
- Convert units by multiplying by fractions equal to 1: $\dfrac{1000\text{ m}}{1\text{ km}}$.
- Percentage change $= \dfrac{\text{new} - \text{old}}{\text{old}} \times 100\%$.
- Successive changes multiply: up 20% then down 20% is $1.2 \times 0.8 = 0.96$, a 4% fall overall.

Centre and spread

- The mean is pulled toward extreme values; the median is not.
- The range and standard deviation measure spread. Adding the same number to every value changes the mean but not the spread.
- In a skewed distribution with a long right tail, the mean is greater than the median.

Figure: Centre and spread



- One extreme value pulls the mean up to 22.8 but leaves the median at 4
- Match each part of the diagram to the text.

Scatterplots and models

- A line of best fit summarises a linear trend; its slope is the predicted change in y per unit of x.
- A residual is actual minus predicted.
- Predicting far outside the data's range (extrapolation) is risky.

Probability and tables

- Probability = favourable outcomes ÷ possible outcomes.
- In a two-way table, "given" restricts you to one row or column: divide by that row's or column's total.

Samples and experiments

- A **random sample** lets you generalise to the population it was drawn from, with a margin of error.
- Random assignment to treatments lets you conclude cause and effect.
- Without random assignment, an association does not show cause.

Worked example: mean and median

- Median: the middle value, 4.
- Mean: $114/5 = 22.8$.
- Removing the 100 changes the mean to 3.5 and the median to 3.5. The median was the better summary of a typical value.

Worked example: a table

- $P(\text{instrument}) = (45 + 35)/200 = 0.40$.
- $P(\text{instrument} \mid \text{sport}) = 45/120 = 0.375$.
- $P(\text{sport} \mid \text{instrument}) = 45/80 = 0.5625$. The "given" group sets the denominator.

Common errors

- Adding successive percentages instead of multiplying.
- Using the mean to describe a skewed set of data.
- Dividing by the whole table when the question says "given".

Common errors (2)

- Concluding cause from a survey.
- Generalising from a sample to a population it was not drawn from.

Key terms

- **rate**: A ratio comparing quantities with different units.
- **median**: The middle value of an ordered list of data.
- **random sample**: A sample in which every member of the population has a known chance of being chosen.

Check yourself

- A price rises 10% and then 10% again. What is the total percentage increase?
- A data set has mean 50 and SD 6. Every value is increased by 4. What are the new mean and SD?
- A study randomly assigns volunteers to two diets. Can it show the diet caused a difference?