Data analysis: spread, weighted averages and conditional probability

Original classroom teaching; not an official form or score predictor

Data Analysis: Spread, Weighted Averages and Conditional Probability

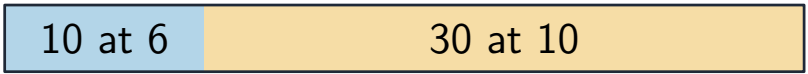
- Combine groups using a **weighted average** 加权平均.
- Compare mean, median, range and **standard deviation** 标准差.
- Find a **conditional probability** 条件概率, including draws without replacement.
- Use "or" correctly, including for **mutually exclusive** 互斥的 events.

Weighted averages

- Combine totals, not means. The larger group counts more.
- The combined mean always lies between the group means, nearer the larger group.

$$\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$$

Figure: Weighted averages



totals: $60 + 300 = 360$

combined mean $= 360 \div 40 = 9$

not $(6 + 10) \div 2 = 8$: the larger group counts more

- Combine totals, not means: the larger group counts more.
- Match each part of the diagram to the text.

Centre and spread

- An outlier pulls the mean but hardly moves the median.
- Adding a constant to every value shifts the mean and median but leaves the range and standard deviation unchanged.
- Multiplying every value by $k > 0$ multiplies the mean, median, range and standard deviation by k.

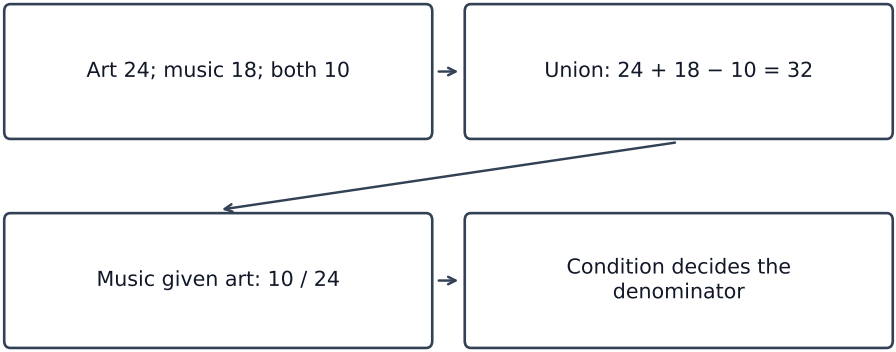
Conditional probability and draws

- "Given that" restricts the group: divide by the size of that group.
- Without replacement, the second probability uses the new totals: $\frac{4}{10} \cdot \frac{3}{9}$, not $\left(\frac{4}{10}\right)^2$.

"Or" and overlaps

- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.
- If A and B are **mutually exclusive**, they cannot both happen, so $P(A \text{ or } B) = P(A) + P(B)$.
- For counts: union = group 1 + group 2 − both.

Figure: "Or" and overlaps



- For music given art, divide by the 24 art students, not by the union of 32

exercises.iol exercises.com.cn

Worked example: a weighted mean and an outlier

- 10 people with mean 6 and 30 people with mean 10: totals $60 + 300 = 360$, so the mean is $\frac{360}{40} = 9$, not the unweighted 8.
- Data 1, 2, 3, 4, 20: mean $\frac{30}{5} = 6$, median 3, range 19. The outlier 20 raises the mean above every other value.
- Adding 5 to each value: mean 11, median 8, standard deviation unchanged.

Worked example: an original application

- Group A has 12 observations with mean 5; group B has eight with mean 10.
- Find the combined mean.
- For the data 1,2,3,4,20 find mean, median and range.
- Describe how adding 7 to every value changes the standard deviation.

exercises.lol exercises.com.cn

Independent transfer

- Of 40 students, 24 study art, 18 study music and 10 study both.
- Find the probability of music given art and the probability of at least one subject.
- Are the events independent?
- Then explain what happens to standard deviation if all measurements are multiplied by -3.

Common errors

- Averaging the group means without weighting.
- Saying that adding a constant changes the spread.
- Dividing by the whole population in a "given" question.

Common errors (2)

- Squaring the first probability for draws without replacement.
- Counting the overlap twice in a union.

Key terms

- **weighted average**: A mean in which each group counts in proportion to its size.
- **standard deviation**: A measure of how far values typically lie from the mean.
- **conditional probability**: The probability of an event within a restricted group.
- **mutually exclusive**: Unable to happen at the same time.

Check yourself

- Group A has 5 people with mean 8; group B has 15 with mean 4. Find the combined mean.
- 24 students take art, 18 take music and 10 take both. Find $P(\text{music} \mid \text{art})$.
- A and B are **mutually exclusive**, with $P(A) = 0.3$ and $P(B) = 0.4$. Find $P(A \text{ or } B)$.