Statistics and probability

4MA1 full topic 6 package: Foundation 6.1-6.3 and Higher 6.1-6.3 with June 2024 demand

Statistics and probability

- present data in tables, charts and histograms, and compare distributions;
- find the **mean**, median, **mode** and range, including estimates for grouped data;
- draw **cumulative frequency** graphs and find the median, quartiles and interquartile range (Higher);

Statistics and probability (2)

- find probabilities from sample spaces, Venn diagrams and experiments;
- use tree diagrams for independent and dependent events, and **conditional probability** (Higher).

Collecting and presenting data

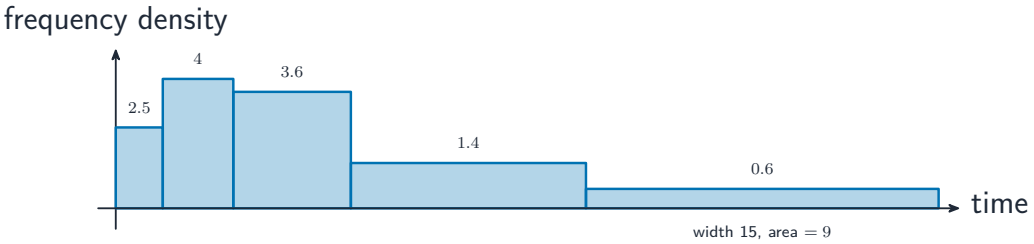
- a bar chart for categories or discrete data;
- a pie chart to show proportions of a whole;
- a pictogram, with a key;

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

Collecting and presenting data (2)

- a frequency polygon for grouped data, plotting each frequency at the class midpoint.
- Frequency densities: $\frac{5}{2} = 2.5$, $\frac{12}{3} = 4$, $\frac{18}{5} = 3.6$, $\frac{14}{10} = 1.4$ and $\frac{9}{15} = 0.6$.
- Check the last bar: area $= 0.6 \times 15 = 9$, its frequency. It is wide, but it is low.

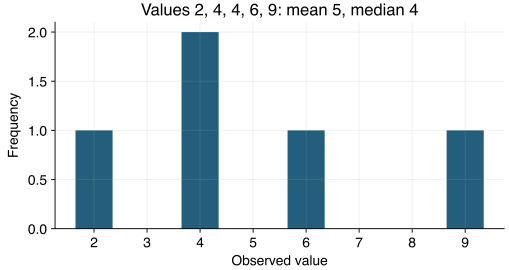
Figure: Collecting and presenting data



area = frequency: height is
density = frequency over width

- Height = **frequency density**
- Area of each bar = frequency

Figure: Collecting and presenting data (2)



- For 2, 4, 4, 6, 9: **mean** $= 25 \div 5 = 5$
- Median $= 4$, **mode** $= 4$

Averages and spread

- the **mean** 平均数: add the values and divide by how many there are
- the **median** 中位数: the middle value when the data are in order
- the **mode** 众数: the most common value.

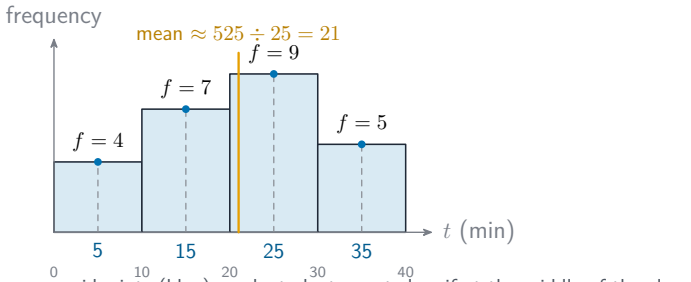
Averages and spread (2)

- $\sum fm = 20 + 105 + 225 + 175 = 525$ and $\sum f = 25$.
- Estimated **mean** $= 525 \div 25 = 21$ minutes. Divide by 25, the number of students, not by 4, the number of classes.
- The **modal class** is $20 < t \leq 30$.

Averages and spread (3)

- The median is $\frac{9+10}{2} = 9.5$.
- Lower half 4, 6, 7, 9: $Q_1 = 6.5$. Upper half 10, 12, 14, 15: $Q_3 = 13$.
- **Interquartile range** $= 13 - 6.5 = 6.5$.

Figure: Averages and spread



- Each student is placed at the class midpoint
- Estimated **mean** $= 525 \div 25 = 21$ minutes

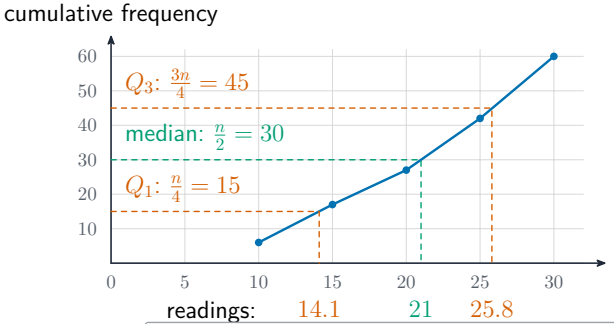
Cumulative frequency diagrams

- the median is at $\frac{n}{2}$ on the **cumulative frequency** axis;
- the lower **quartile** is at $\frac{n}{4}$ and the upper quartile at $\frac{3n}{4}$;
- to find how many values are above a number, read the cumulative frequency at that number and subtract it from n.

Cumulative frequency diagrams (2)

- The median is the 30th value. 27 workers are at or below 20 and 42 at or below 25, so the median is $20 + 5 \times \frac{30-27}{42-27} = 21$.
- Q_1 is the 15th value: $10 + 5 \times \frac{15-6}{17-6} = 14.1$.
- Q_3 is the 45th value: $25 + 5 \times \frac{45-42}{60-42} = 25.8$.
- **Interquartile range** $= 25.8 - 14.1 = 11.7$.

Figure: Cumulative frequency diagrams



$n = 60$ workers
median $= 21$ hours
IQR $= 25.8 - 14.1$
$= 11.7$ hours

- Median 21; quartiles 14.1 and 25.8
- IQR $= 25.8 - 14.1 = 11.7$

Probability foundations

- Cricket only: $18 - 7 = 11$. Tennis only: $15 - 7 = 8$. Both: 7. So $11 + 8 + 7 = 26$ like at least one.
- $P(\text{neither}) = \frac{30-26}{30} = \frac{4}{30} = \frac{2}{15}$.

$$P(\text{event}) = \frac{\text{number of favourable outcomes}}{\text{total number of outcomes}}$$

Tree diagrams and combined events

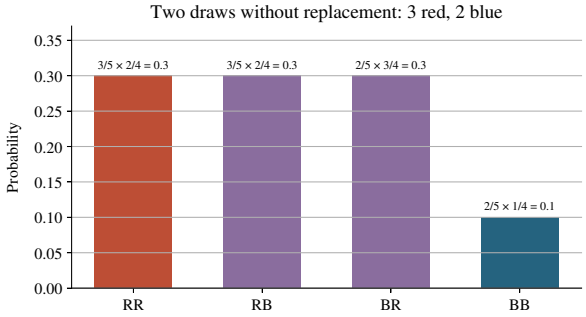
- Red then blue: $\frac{5}{20} \times \frac{6}{19} = \frac{30}{380}$. Blue then red: $\frac{6}{20} \times \frac{5}{19} = \frac{30}{380}$.
- Add the two routes: $\frac{60}{380} = \frac{3}{19}$.
- $\frac{6}{n} \times \frac{5}{n-1} = \frac{1}{3}$, so $90 = n(n-1)$ and $n^2 - n - 90 = 0$.

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

Tree diagrams and combined events (2)

- $(n-10)(n+9)=0$. The number of sweets cannot be negative, so $n=10$.
- Check: $\frac{6}{10}\times\frac{5}{9}=\frac{30}{90}=\frac{1}{3}$.

Figure: Tree diagrams and combined events



- Without replacement: $P(RR) = 3/5 \times 2/4 = 0.3$
- $P(BB) = 2/5 \times 1/4 = 0.1$

Sources for this lesson

- Pearson Edexcel, *International GCSE (9–1) Mathematics A specification* (Issue 2, 2017), topic 6.
- Pearson Edexcel, 4MA1 Papers 1H and 2H, June 2024 (adapted).
- J. Graunt, *Natural and Political Observations upon the Bills of Mortality* (London, 1662).

Common errors

- Drawing **histogram** bars with frequency as the height when the class widths differ.
- Dividing $\sum fm$ by the number of classes instead of the total frequency.
- Plotting **cumulative frequency** at the midpoint of the class instead of the upper end.

Common errors (2)

- Finding the median of grouped data at $\frac{n+1}{2}$ on a **cumulative frequency** graph; use $\frac{n}{2}$.
- Adding probabilities along a branch instead of multiplying.
- Keeping the same denominator on the second branch when items are not replaced.
- Counting the overlap twice in a **Venn diagram**.

Key terms

- **Frequency density**: frequency divided by class width; the height of a **histogram** bar.
- **Median**: the middle value of ordered data.
- **Interquartile range**: the upper **quartile** minus the lower quartile.

Key terms (2)

- **Cumulative frequency**: the running total of the frequencies.
- **Complement**: everything that is not in the event; $P(\text{not } A) = 1 - P(A)$.
- **Independent events**: events where one does not change the probability of the other.

Check yourself

- 1.
- Find the **mean** and median of 5, 9, 2, 8, 6. 2.
- A class $10 < x \leq 30$ has frequency 12.
- Find its **frequency density**. 3.