

ESAT

Maths 1

Statistics and probability

Free lesson notes

Covers	M6.1 to M6.4, M7.1 to M7.7
Format	Lesson, worked examples, self-check
Price	Free

About this lesson

Statistics and probability questions on the ESAT are short and logical. They test whether you understand what each measure means, rather than whether you can push numbers through a formula. This lesson covers averages and spread, grouped data and histograms, scatter graphs, and probability from relative frequency through tree diagrams, Venn diagrams and conditional probability.

How to use it. Try each worked example first. Then do the self-check, then the practice set. For probability, always ask: *is it AND (multiply) or OR (add)?* and *does the situation change after each pick (without replacement)?*

Heurisco is an independent practice provider and is not affiliated with, endorsed by, or connected to UAT-UK, Pearson VUE, the UCAT Consortium, or any university. All questions are original.

1 Averages and spread

- **Mean** = $\frac{\text{sum of values}}{\text{number of values}}$, so $\text{sum} = \text{mean} \times \text{number}$. This identity solves most puzzle questions.
- **Median**: the middle value of the ordered list (the average of the middle two if the count is even).
- **Mode**: the most frequent value. **Range** = largest – smallest.
- **Quartiles**: split the ordered data in halves. Lower quartile = median of the lower half, upper quartile = median of the upper half. $\text{IQR} = \text{UQ} - \text{LQ}$.

To combine groups, use the sums, not the means: the overall mean is $\frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$.

Worked example: combined mean

A class of 20 students has a mean mark of 60. Another class of 30 has a mean of 70. What is the mean of all 50?
Total marks: $20 \times 60 + 30 \times 70 = 1200 + 2100 = 3300$. Mean = $\frac{3300}{50} = 66$.

Worked example: quartiles

Find the IQR of 3, 5, 6, 8, 9, 11, 12, 14, 16, 17, 20.
The median is the 6th value, 11. Lower half: 3, 5, 6, 8, 9 with median 6. Upper half: 12, 14, 16, 17, 20 with median 16.
 $\text{IQR} = 16 - 6 = 10$.

Common mistakes

- Averaging the two means when the groups are different sizes. Use a weighted mean.
- Including the median in both halves when finding quartiles of an odd-sized list.

2 Grouped data, histograms and cumulative frequency

For grouped data you only know which class each value is in, so:

- To estimate the mean, use the class mid-points: $\bar{x} \approx \frac{\sum f \times \text{midpoint}}{\sum f}$.
- In a **histogram**, the area of a bar is the frequency, so frequency = frequency density $\times$ class width.
- On a **cumulative frequency** graph, the median is read at $\frac{1}{2}n$ and the quartiles at $\frac{1}{4}n$ and $\frac{3}{4}n$. Between plotted points, assume values are spread evenly (linear interpolation).

Worked example: estimating the mean

Classes 0–10, 10–20, 20–30, 30–40 have frequencies 4, 6, 8, 2. Estimate the mean.

Mid-points 5, 15, 25, 35. $\sum fx = 20 + 90 + 200 + 70 = 380$ and $\sum f = 20$, so the mean is $\frac{380}{20} = 19$.

Worked example: a histogram

A bar from 20 to 30 has frequency density 2.0. How many values does it represent?

Frequency = $2.0 \times 10 = 20$. If part of a bar is wanted, take the same fraction of its width: half of that bar represents 10.

Common mistakes

- Reading the height of a histogram bar as the frequency. It is the density.
- Using the class boundary rather than the mid-point when estimating the mean.

3 Scatter graphs

A scatter graph shows whether two variables are related. *Positive correlation*: both increase together. *Negative*: one increases as the other decreases. A line of best fit passes close to the points, with roughly equal numbers on each side, and can be used to estimate one variable from the other. Two cautions:

- Predictions *inside* the data range (interpolation) are reasonable; predictions far *outside* it (extrapolation) are unreliable.
- Correlation does not prove that one variable *causes* the other.

Common mistakes

- Saying correlation proves causation.
- Forcing the line of best fit through the origin or through every point.

4 Probability basics

For equally likely outcomes, $P(\text{event}) = \frac{\text{number of favourable outcomes}}{\text{total number of outcomes}}$.

- Probabilities of all the outcomes of an experiment add to 1. $P(\text{not } A) = 1 - P(A)$.
- **Relative frequency** = $\frac{\text{number of successes}}{\text{number of trials}}$ estimates a probability. The *expected number* in N trials is $N \times P$.
- For two dice or two spinners, list a *possibility space* (a 6×6 grid has 36 outcomes).

“At least one” is usually easiest as $1 - P(\text{none})$.

Worked example: at least one

A fair die is rolled 3 times. What is the probability of at least one six?

$P(\text{no sixes}) = \left(\frac{5}{6}\right)^3 = \frac{125}{216}$, so $P(\text{at least one}) = 1 - \frac{125}{216} = \frac{91}{216}$.

Worked example: a possibility space

Two fair dice are rolled. What is the probability that their scores differ by exactly 2?

The pairs are (1, 3), (2, 4), (3, 5), (4, 6) and the four reversed: 8 outcomes out of 36, so $\frac{8}{36} = \frac{2}{9}$.

Common mistakes

- Forgetting that order matters on a possibility space: (1, 3) and (3, 1) are different outcomes.
- Adding when you should multiply for independent events, or the reverse.

5 Combined events, trees, Venn diagrams and conditional probability

- **AND** (independent events): multiply, $P(A \text{ and } B) = P(A) \times P(B)$.
- **OR** (mutually exclusive): add. In general, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.
- **Without replacement**: the second probability uses the *reduced* numbers.
- **Conditional probability**: $P(A | B) = \frac{P(A \text{ and } B)}{P(B)}$. On a tree diagram, multiply along branches and add the branches that give the outcome.

In a Venn diagram, fill in the overlap first, then the “only” regions, then check the total.

Worked example: without replacement

A bag has 5 red and 3 blue counters. Two are taken without replacement. What is the probability that they are the same colour?

$$P(RR) = \frac{5}{8} \times \frac{4}{7} = \frac{20}{56} \text{ and } P(BB) = \frac{3}{8} \times \frac{2}{7} = \frac{6}{56}. \text{ Total} = \frac{26}{56} = \frac{13}{28}.$$

Worked example: conditional probability

It rains with probability 0.3. If it rains, a bus is late with probability 0.6; if not, with probability 0.1. Given that the bus is late, what is the probability that it rained?

$$P(\text{rain and late}) = 0.3 \times 0.6 = 0.18. \quad P(\text{late}) = 0.18 + 0.7 \times 0.1 = 0.25.$$

$$P(\text{rain} | \text{late}) = \frac{0.18}{0.25} = 0.72.$$

Common mistakes

- Using the same denominator for the second pick when counters are not replaced.
- Confusing $P(A | B)$ with $P(B | A)$: the denominator is the *given* event.

Check yourself

Try these without looking back, then compare with the answers on the next page.

1. The mean of 4 numbers is 9. A fifth number is added and the mean becomes 10. What is the fifth number?
2. Find the IQR of 2, 4, 7, 9, 10, 13, 15, 18.
3. A bar in a histogram has class width 5 and frequency density 3.2. Find the frequency.
4. A spinner lands on blue with relative frequency 0.35 in 200 trials. Estimate the number of blues in 600 trials.
5. Two fair coins are tossed. Find $P(\text{at least one head})$.
6. $P(A) = 0.6$, $P(B) = 0.5$, $P(A \text{ and } B) = 0.3$. Find $P(A \text{ or } B)$.

Answers to the self-check

1. New total = 50, old total = 36, so the number is 14.
2. Lower half 2, 4, 7, 9: median 5.5. Upper half 10, 13, 15, 18: median 14. IQR = 8.5.
3. $5 \times 3.2 = 16$.
4. $0.35 \times 600 = 210$.
5. $1 - \frac{1}{4} = \frac{3}{4}$.
6. $0.6 + 0.5 - 0.3 = 0.8$.

Next step

Heurisco's *ESAT Maths 1: Statistics and Probability Practice Set* has twenty original exam-style questions on this lesson, in order of difficulty, with a full worked solution for each.