Foundation Mathematics Units 3 & 4

Chapter 8 — Probability and long-term data

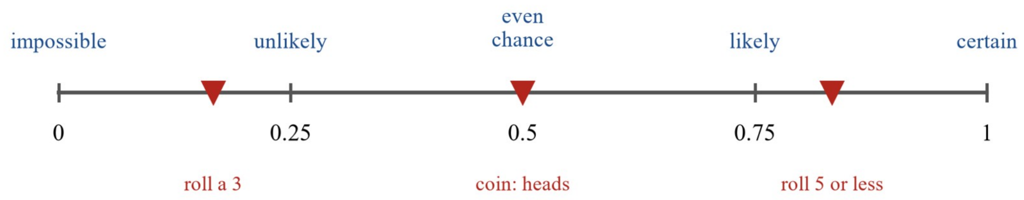
Aboriginal and Torres Strait Islander content has been excluded as accuracy cannot be guaranteed, and it is better to be respectful than cover the entire curriculum inaccurately.

Significant effort has been made to ensure this content is legal. Please send any areas of concern to admin@vwx.au with appropriate document/legal references so errors can be verified and changes be made.

Contents

Theory

Probability is a number that measures how likely an event is to happen. Every probability lies between 0 and 1: an **impossible** event has probability 0, a **certain** event has probability 1, and an event with an **even chance** has probability $1/2$.



The probability scale and describing chance.

We describe everyday chance with words — *impossible*, *unlikely*, *even chance*, *likely*, *certain* — and each corresponds to a position on the scale above. An event with probability less than $1/2$ is *unlikely*; one with probability greater than $1/2$ is *likely*.

Theoretical probability.

When every outcome of a situation is **equally likely**, the **theoretical probability** of an event is

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

For example, a fair six-sided die has six equally likely outcomes $1, 2, 3, 4, 5, 6$. The event "roll an even number" has three favourable outcomes ($2, 4, 6$), so

$$Pr(\text{even}) = \frac{3}{6} = \frac{1}{2}.$$

Writing a probability three ways.

A probability can be written as a **fraction**, a **decimal** or a **percentage** — they all mean the same thing. Here $Pr(\text{even}) = 1/2 = 0.5 = 50\%$. Use whichever form the question asks for; if none is specified, an exact fraction is safest.

The sample space.

The **sample space** is the list of all possible outcomes. Writing it out helps you count correctly. For two coins tossed together the sample space is

$$\{\text{HH}, \text{HT}, \text{TH}, \text{TT}\},$$

four equally likely outcomes, so $Pr(\text{two heads}) = 1/4$.

The complement.

The **complement** of an event A is "A does **not** happen", written not A. Because something either happens or it doesn't,

$$Pr(\text{not } A) = 1 - Pr(A).$$

This is often the quickest way to find a probability — for the die, $Pr(\text{not even}) = 1 - 1/2 = 1/2$.

Expected number of occurrences.

If an event has probability p and a situation is repeated n times, the **expected number** of times it happens is

$$\text{expected number} = p \times n.$$

This is a long-run average, not a guarantee. Rolling a die 60 times, you would *expect* about $1/6 \times 60 = 10$ sixes.

Worked examples

Example 1 — scaffolded

A fair six-sided die is rolled. Find the probability of rolling (a) an even number and (b) a number greater than 4, giving each as a fraction, a decimal and a percentage.

Working	Comment
The equally likely outcomes are $1, 2, 3, 4, 5, 6$ — six in total.	Identify the sample space.
(a) Even numbers: $2, 4, 6$ — three favourable. $Pr(\text{even}) = \frac{3}{6} = \frac{1}{2}$.	Favourable ÷ total, then simplify.
$\frac{1}{2} = 0.5 = 50\%$.	The three equivalent forms.
(b) Numbers greater than 4: $5, 6$ — two favourable. $Pr(>4) = \frac{2}{6} = \frac{1}{3}$.	Two favourable out of six.
$\frac{1}{3} = 0.333\ldots \approx 33.3\%$.	A recurring decimal; round the percentage if asked.

Example 2 — scaffolded

A bag contains 5 red, 3 blue and 2 green balls, all the same size. One ball is drawn at random. Find (a) $Pr(\text{red})$ and (b) $Pr(\text{not red})$. (c) If the ball is returned and this is repeated 30 times, how many reds would you expect?

Working	Comment
Total balls $= 5 + 3 + 2 = 10$.	Count the whole sample space.
(a) $Pr(\text{red}) = \frac{5}{10} = \frac{1}{2}$.	Five reds out of ten.
(b) $Pr(\text{not red}) = 1 - \frac{1}{2} = \frac{1}{2}$.	Use the complement.
(c) Expected reds $= \frac{1}{2} \times 30 = 15$.	Probability × number of trials.

Example 3 — unscaffolded

A spinner is divided into 8 equal sectors: 4 yellow, 3 blue and 1 red. The spinner is spun once. Find $Pr(\text{yellow})$, $Pr(\text{red})$ and $Pr(\text{not blue})$.

The 8 sectors are equal, so each is equally likely. Then

$$Pr(\text{yellow}) = \frac{4}{8} = \frac{1}{2}, \quad Pr(\text{red}) = \frac{1}{8}.$$

For "not blue", there are 3 blue sectors, so $Pr(\text{blue}) = 3/8$ and

$$Pr(\text{not blue}) = 1 - \frac{3}{8} = \frac{5}{8}.$$

$\therefore Pr(\text{yellow}) = 1/2$, $Pr(\text{red}) = 1/8$ and $Pr(\text{not blue}) = 5/8$.

Example 4 — unscaffolded

A factory finds that $2\,\%$ of the items it makes are defective. A quality inspector checks a batch of 500 items. (a) How many defective items would be expected? (b) What is the probability that a randomly chosen item is **not** defective?

(a) The probability an item is defective is $2\,\% = 0.02$, so the expected number in 500 items is

$$0.02 \times 500 = 10\,\text{defective items}.$$

(b) By the complement, $Pr(\text{not defective}) = 1 - 0.02 = 0.98$.

$\therefore$ about 10 items would be expected to be defective, and $Pr(\text{not defective}) = 0.98$ (that is, $98\,\%$).

Practice questions

Scaffolded

Q1. A fair six-sided die is rolled. (a) Find $Pr(3)$. (b) Find $Pr(\text{odd number})$. (c) Find $Pr(\text{a number } 2 \text{ or less})$.

Q2. A bag holds 4 red and 6 blue counters. One counter is drawn at random. (a) Find $Pr(\text{red})$ as a fraction, a decimal and a percentage. (b) Find $Pr(\text{blue})$. (c) Find $Pr(\text{not red})$.

Q3. A prize wheel has 5 equal sections, of which 1 says "WIN" and 4 say "TRY AGAIN". (a) Find $Pr(\text{win})$. (b) Find $Pr(\text{lose})$. (c) If the wheel is spun 20 times, how many wins would you expect?

Q4. One card is drawn from a standard pack of 52 playing cards. (a) Find $Pr(\text{a heart})$. (b) Find $Pr(\text{a king})$. (c) Find $Pr(\text{not a heart})$.

Q5. Two fair coins are tossed together. (a) List the sample space. (b) Find $Pr(\text{two heads})$. (c) Find $Pr(\text{at least one head})$.

Q6. For each event, describe its chance in words (impossible, unlikely, even chance, likely or certain) and estimate its probability. (a) The sun will rise tomorrow. (b) A fair coin lands on heads. (c) You roll a 7 on an ordinary die.

Unscaffolded

Q7. A fair die is rolled. Find the probability of rolling a multiple of 3.

Q8. A jar contains 3 red, 5 yellow and 2 green lollies. One is taken at random. Find $Pr(\text{green})$ and $Pr(\text{not yellow})$.

Q9. Two coins are tossed. Find the probability of getting exactly one head.

Q10. A raffle sells 200 tickets and a person buys 8 of them. Find the probability that this person wins the single prize, as a fraction and as a percentage.

Q11. A spinner has 10 equal sectors numbered 1 to 10. Find (a) $Pr(\text{even number})$ and (b) $Pr(\text{a number greater than } 7)$.

Q12. One card is drawn from a standard pack of 52. Find the probability that it is a face card (a jack, queen or king — 12 cards in all).

Q13. A weather forecast gives the probability of rain on any day in November as 0.3. (a) Find the probability that it does **not** rain on a given day. (b) Over the 30 days of November, how many rainy days would you expect?

Q14. A machine produces components, 4% of which are faulty. In a run of 250 components, (a) how many faulty components would be expected, and (b) what is the probability a component is **not** faulty?

Q15. A bag of red and white marbles is such that $Pr(\text{red}) = 1/4$. There are 3 red marbles. (a) How many marbles are in the bag altogether? (b) How many are white?

Q16. A bag has 6 red and 4 white balls; a fair die is also rolled. Which is more likely: drawing a red ball from the bag, or rolling a 5 or a 6 on the die? Justify your answer with probabilities.

Q17. A fair six-sided die is rolled. Find the probability of rolling a prime number ($2, 3$ or 5), giving your answer as a fraction, a decimal and a percentage.

Q18. A scratch-card game has a probability of 0.05 of winning on each card. A shop sells 60 of these cards in a day. How many winning cards would be expected?

Answers

Q1. (a) $1/6$. (b) $1/2$. (c) $1/3$. **Q2.** (a) $2/5 = 0.4 = 40\%$. (b) $3/5$. (c) $3/5$. **Q3.** (a) $1/5$. (b) $4/5$. (c) 4 wins. **Q4.** (a) $1/4$. (b) $1/13$. (c) $3/4$. **Q5.** (a) $\{HH, HT, TH, TT\}$. (b) $1/4$. (c) $3/4$. **Q6.** (a) certain, probability 1. (b) even chance, probability $1/2$. (c) impossible, probability 0. **Q7.** $2/6 = 1/3$. **Q8.** $Pr(\text{green}) = 2/10 = 1/5$; $Pr(\text{not yellow}) = 5/10 = 1/2$. **Q9.** $2/4 = 1/2$. **Q10.** $8/200 = 1/25 = 4\%$. **Q11.** (a) $1/2$. (b) $3/10$. **Q12.** $12/52 = 3/13$. **Q13.** (a) 0.7. (b) 9 rainy days. **Q14.** (a) 10 faulty. (b) 0.96 (that is, 96%). **Q15.** (a) 12 marbles. (b) 9 white. **Q16.** Red: $6/10 = 0.6$; a 5 or 6: $2/6 = 1/3 \approx 0.33$. Drawing a red ball is more likely. **Q17.** $3/6 = 1/2 = 0.5 = 50\%$. **Q18.** 3 winning cards.

Fully-worked solutions

Q1. The die has six equally likely outcomes $1, 2, 3, 4, 5, 6$. (a) One outcome is a 3, so $Pr(3) = 1/6$. (b) The odd numbers are $1, 3, 5$ (three of them), so $Pr(\text{odd}) = 3/6 = 1/2$. (c) "2 or less" means 1 or 2 (two outcomes), so $Pr(\le 2) = 2/6 = 1/3$.

Q2. There are $4 + 6 = 10$ counters. (a) $Pr(\text{red}) = 4/10 = 2/5 = 0.4 = 40\%$. (b) $Pr(\text{blue}) = 6/10 = 3/5$. (c) $Pr(\text{not red}) = 1 - 2/5 = 3/5$ (the same as blue, since blue is the only other colour).

Q3. The wheel has 5 equal sections. (a) One section wins: $Pr(\text{win}) = 1/5$. (b) Four sections lose: $Pr(\text{lose}) = 4/5$ (or $1 - 1/5$). (c) Expected wins $= 1/5 \times 20 = 4$.

Q4. A standard pack has 52 cards. (a) There are 13 hearts, so $Pr(\text{heart}) = 13/52 = 1/4$. (b) There are 4 kings, so $Pr(\text{king}) = 4/52 = 1/13$. (c) $Pr(\text{not a heart}) = 1 - 1/4 = 3/4$.

Q5. (a) Listing both coins, the sample space is $\{HH, HT, TH, TT\}$ — four equally likely outcomes. (b) Only HH gives two heads, so $Pr(\text{two heads}) = 1/4$. (c) "At least one head" is HH, HT, TH — three outcomes — so $Pr(\ge 1\,\text{head}) = 3/4$ (or $1 - Pr(\text{no heads}) = 1 - 1/4$).

Q6. (a) The sun rising is **certain**, probability 1. (b) A fair coin landing heads is an **even chance**, probability $1/2$. (c) An ordinary die has no 7, so rolling a 7 is **impossible**, probability 0.

Q7. The multiples of 3 on a die are 3 and 6 (two outcomes), so $Pr(\text{multiple of } 3) = 2/6 = 1/3$.

Q8. There are $3 + 5 + 2 = 10$ lollies. $Pr(\text{green}) = 2/10 = 1/5$. "Not yellow" means red or green, $3 + 2 = 5$ lollies, so $Pr(\text{not yellow}) = 5/10 = 1/2$ (or $1 - 5/10$).

Q9. The sample space is $\{HH, HT, TH, TT\}$. Exactly one head occurs for HT and TH — two outcomes — so $Pr(\text{exactly one head}) = 2/4 = 1/2$.

Q10. The person holds 8 of the 200 equally likely tickets, so $Pr(\text{win}) = 8/200 = 1/25 = 0.04 = 4\%$.

Q11. The spinner has 10 equal sectors 1–10. (a) The even numbers are $2, 4, 6, 8, 10$ (five of them), so $Pr(\text{even}) = 5/10 = 1/2$. (b) Numbers greater than 7 are $8, 9, 10$ (three of them), so $Pr(>7) = 3/10$.

Q12. There are 12 face cards out of 52, so $Pr(\text{face card}) = 12/52 = 3/13$.

Q13. (a) $Pr(\text{no rain}) = 1 - 0.3 = 0.7$. (b) Expected rainy days $= 0.3 \times 30 = 9$.

Q14. (a) Expected faulty $= 4\% \times 250 = 0.04 \times 250 = 10$. (b) $Pr(\text{not faulty}) = 1 - 0.04 = 0.96$.

Q15. (a) If $Pr(\text{red}) = 1/4$ and there are 3 red marbles, then 3 is one quarter of the total, so the total is $3 \times 4 = 12$ marbles. (b) White marbles $= 12 - 3 = 9$.

Q16. For the bag, $Pr(\text{red}) = 6/10 = 3/5 = 0.6$. For the die, $Pr(5 \text{ or } 6) = 2/6 = 1/3 \approx 0.33$. Since $0.6 > 0.33$, drawing a red ball is more likely.

Q17. The prime numbers on a die are $2, 3, 5$ (three outcomes), so $Pr(\text{prime}) = 3/6 = 1/2 = 0.5 = 50\%$.

Q18. Expected winning cards $= 0.05 \times 60 = 3$.

Theory

When we cannot list equally-likely outcomes to work out a **theoretical** probability, we instead look at what has actually happened over many trials. This is the idea behind **relative frequency** — using real data to estimate how likely an event is.

Relative frequency.

The **relative frequency** of an event is the proportion of trials in which the event happened:

$$\text{relative frequency} = \frac{\text{number of times the event occurred}}{\text{total number of trials}}.$$

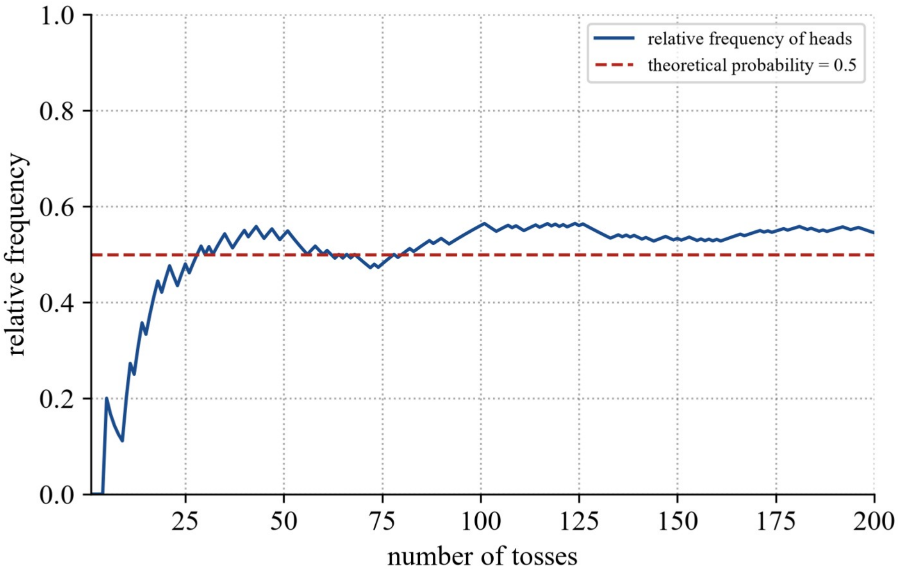
Like any probability, it can be written as a fraction, a decimal or a percentage. For example, if a bus was late on 9 of the last 30 mornings, the relative frequency of a late bus is $\frac{9}{30} = 0.3 = 30\%$.

Experimental versus theoretical probability.

The relative frequency found from data is also called the **experimental probability**. It is an *estimate* of the true probability, and it will not usually equal the theoretical value exactly. Rolling a fair die 60 times will rarely give each face exactly 10 times, even though the theoretical probability of each face is $\frac{1}{6} \approx 0.167$.

The long-run pattern.

The key fact is what happens as the number of trials grows: **the more trials you carry out, the closer the relative frequency tends to get to the theoretical probability.** A handful of coin tosses can easily give 70% heads, but after thousands of tosses the relative frequency of heads settles down close to 0.5. The graph below shows the relative frequency of heads “settling” toward 0.5 as the number of tosses increases.



Estimating a probability from long-term data.

Because relative frequency approaches the true probability, a large set of **long-term data** gives a good estimate of the probability of an event:

$$Pr(A) \approx \frac{\text{number of times } A \text{ occurs}}{\text{total number of trials}}.$$

This is how probabilities are estimated in the real world, where outcomes are not equally likely — for example the chance of rain, of a manufactured part being faulty, of a team winning, or of a person having a particular health condition. The larger the data set, the more reliable the estimate.

Predicting an expected number.

Once you have an estimated probability, you can predict how often an event should happen in a **larger** group:

$$\text{expected number} = \text{estimated probability} \times \text{number of trials}.$$

If 3% of parts are faulty, then in a batch of 2000 you would expect about $0.03 \times 2000 = 60$ faulty parts. This is a prediction, not a guarantee — the actual number will vary — but it is the best estimate available from the data.

Reading long-term data critically.

Long-term data is used to make decisions about epidemics and health, climate and the environment, sport, and quality control. When you use it, check that the sample is **large enough** and **representative** of the situation you are predicting; an estimate from a small or biased sample can be well off the true value.

Worked examples

Example 1 — scaffolded

A die was rolled 60 times. The results are shown below. Find the relative frequency of rolling a 6, and compare it with the theoretical probability.

Face	1	2	3	4	5	6
Frequency	8	11	9	13	10	9

Working	Comment
Total trials $= 8 + 11 + 9 + 13 + 10 + 9 = 60$.	The frequencies add to the number of rolls.
Relative frequency of a $6 = \frac{9}{60} = 0.15$.	Frequency of the event over the total.
As a percentage, $0.15 = 15\%$.	A relative frequency can be written three ways.
Theoretical probability $= \frac{1}{6} \approx 0.167$ (16.7%).	A fair die has 6 equally-likely faces.
The experimental value 0.15 is close to, but not equal to, 0.167.	With only 60 rolls, some variation is expected.

Example 2 — scaffolded

In a quality-control check, 15 of a sample of 500 phone chargers were found to be faulty. Estimate the probability that a charger is faulty, and predict how many faulty chargers would be expected in a production run of 2000.

Working	Comment
Relative frequency of a faulty charger $= \frac{15}{500} = 0.03$.	Number faulty over the sample size.
So $Pr(\text{faulty}) \approx 0.03 = 3\%$.	Long-term data estimates the probability.
Expected faulty in $2000 = 0.03 \times 2000 = 60$.	Expected number = probability × trials.
About 60 chargers in the run of 2000 would be expected to be faulty.	A prediction — the exact number will vary.

Example 3 — unscaffolded

Over a season a basketballer made 68 of her 80 free-throw attempts. Estimate the probability that she makes a free throw, and predict how many she would make in her next 40 attempts.

The relative frequency of a successful free throw is

$$\frac{68}{80} = 0.85 = 85\,\%,$$

so $Pr(\text{made}) \approx 0.85$. In her next 40 attempts the expected number made is

$$0.85 \times 40 = 34.$$

∴ her estimated probability of making a free throw is 0.85, and she would be expected to make about 34 of her next 40 attempts.

Example 4 — unscaffolded

In a random sample of 250 people in a town, 40 were found to be immune to a particular virus. Use this long-term data to estimate the probability that a resident is immune, and predict how many of the town's 5000 residents are immune.

The relative frequency of immunity in the sample is

$$\frac{40}{250} = 0.16 = 16\,\%,$$

so $Pr(\text{immune}) \approx 0.16$. For the whole town of 5000 residents the expected number immune is

$$0.16 \times 5000 = 800.$$

∴ the estimated probability of immunity is 0.16, and about 800 of the 5000 residents are expected to be immune (assuming the sample is representative of the town).

Practice questions

Scaffolded

Q1. A coloured spinner was spun 50 times, landing on red 18 times, blue 20 times and green 12 times. (a) Find the relative frequency of red, as a fraction and a decimal. (b) Find the relative frequency of blue, as a percentage. (c) Find the relative frequency of green.

Q2. In a batch of 200 light globes, 6 were faulty. (a) Find the relative frequency of a faulty globe. (b) Estimate the probability that a globe is faulty. (c) Predict the number of faulty globes in a batch of 1500.

Q3. It rained on 45 of the last 90 days at a weather station. (a) Find the relative frequency of a rainy day. (b) Write this as a percentage. (c) Estimate the probability that it rains on a given day.

Q4. A die was rolled 120 times and showed a 4 on 18 rolls. (a) Find the relative frequency of rolling a 4. (b) State the theoretical probability of rolling a 4. (c) Comment on how the experimental and theoretical values compare.

Q5. A goalkeeper saved 34 of the last 40 penalties faced. (a) Find the relative frequency of a save. (b) Estimate the probability of a save. (c) Predict how many saves would be expected from the next 25 penalties.

Q6. A student tosses a coin 20 times and gets 13 heads, a relative frequency of 0.65. (a) State the theoretical probability of a head. (b) Explain why the student's relative frequency is not 0.5. (c) Explain what would happen to the relative frequency if the student tossed the coin 1000 times instead.

Unscaffolded

Q7. In a sample of 800 machine parts, 24 did not meet the required standard. Find the relative frequency of a substandard part, as a fraction, a decimal and a percentage.

Q8. A basketballer made 45 of 120 three-point attempts. Estimate the probability that she makes a three-pointer, and predict how many she would make in 200 attempts.

Q9. At a weather station there was no rain on 219 of the 365 days in a year. Find the relative frequency of a dry day, as a decimal and a percentage.

Q10. A coin was tossed 500 times and landed heads 260 times. Find the relative frequency of heads, and comment on how it compares with the theoretical probability.

Q11. In a survey of 2000 people, 320 were left-handed. Estimate the probability that a person is left-handed, and predict the number of left-handed people in a city of $25\,000$.

Q12. In a vaccine trial, 4800 of the 6000 participants developed immunity. Estimate the probability of developing immunity, and predict how many of $15\,000$ vaccinated people would develop immunity.

Q13. A die was rolled 300 times and showed a 6 on 55 rolls. Find the relative frequency of a 6, compare it with the theoretical probability, and state which estimate — this one or the 120-roll experiment in Q4 — you would trust more, and why.

Q14. A factory's records show that 45 of every 1500 items produced are rejected. Estimate the probability that an item is rejected, and predict the number of rejects in a production run of $10\,000$.

Q15. A cricketer hit a boundary from 60 of the 400 balls she faced last season. Estimate the probability that she hits a boundary from a ball, and predict how many boundaries she would hit from 1000 balls.

Q16. Over 200 summer days, the temperature exceeded 30°C on 73 days. Find the relative frequency of a day over 30°C, as a decimal and a percentage.

Q17. To estimate the probability of an event whose true probability is 0.5, one student did 50 trials and got a relative frequency of 0.40, while another did 500 trials and got 0.48. (a) Show how each relative frequency was obtained (20 of 50, and 240 of 500). (b) State which student's estimate is likely to be more reliable, and why.

Q18. A health study found that 1200 of 8000 people surveyed had a particular condition. Estimate the probability that a person has the condition, and predict how many people in a town of $20\,000$ would be expected to have it. State one assumption you are making.

Answers

Q1. (a) $\frac{18}{50}=\frac{9}{25}=0.36$. (b) 40%. (c) $\frac{12}{50}=0.24=24\%$. **Q2.** (a) 0.03. (b) $Pr(\text{faulty})\approx 0.03$. (c) 45 faulty globes. **Q3.** (a) $\frac{45}{90}=0.5$. (b) 50%. (c) $Pr(\text{rain})\approx 0.5$. **Q4.** (a) $\frac{18}{120}=0.15$. (b) $\frac{1}{6}\approx 0.167$. (c) 0.15 is close to 0.167; the small difference is normal variation. **Q5.** (a) $\frac{34}{40}=0.85$. (b) ≈ 0.85. (c) $0.85\times 25=21.25$, so about 21 saves. **Q6.** (a) 0.5. (b) 20 tosses is too few for the relative frequency to settle; short runs vary a lot. (c) It would be expected to get much closer to 0.5. **Q7.** $\frac{24}{800}=\frac{3}{100}=0.03=3\%$. **Q8.** $Pr\approx\frac{45}{120}=0.375$; expected in $200=0.375\times 200=75$. **Q9.** $\frac{219}{365}=0.6=60\%$. **Q10.** $\frac{260}{500}=0.52$; close to the theoretical 0.5, a little above. **Q11.** $Pr\approx 0.16$; expected $=0.16\times 25\,000=4000$. **Q12.** $Pr\approx\frac{4800}{6000}=0.8$; expected $=0.8\times 15\,000=12\,000$. **Q13.** $\frac{55}{300}\approx 0.183$ vs theoretical 0.167; trust the 300-roll estimate more (more trials → closer to the true value). **Q14.** $Pr\approx 0.03$; expected $=0.03\times 10\,000=300$ rejects. **Q15.** $Pr\approx\frac{60}{400}=0.15$; expected $=0.15\times 1000=150$ boundaries. **Q16.** $\frac{73}{200}=0.365=36.5\%$. **Q17.** (a) $\frac{20}{50}=0.40$ and $\frac{240}{500}=0.48$. (b) the 500-trial estimate (0.48), because a larger number of trials gives a relative frequency closer to the true probability. **Q18.** $Pr\approx\frac{1200}{8000}=0.15$; expected $=0.15\times 20\,000=3000$. Assumption: the town is similar to the surveyed group.

Fully-worked solutions

Q1. Total spins $=50$. (a) Red: $\frac{18}{50}=\frac{9}{25}=0.36$. (b) Blue: $\frac{20}{50}=0.4=40\%$. (c) Green: $\frac{12}{50}=0.24=24\%$. (Check: $0.36+0.40+0.24=1$.)

Q2. (a) Relative frequency of a faulty globe $=\frac{6}{200}=0.03$. (b) So $Pr(\text{faulty})\approx 0.03$. (c) Expected faulty in $1500=0.03\times 1500=45$.

Q3. (a) $\frac{45}{90}=\frac{1}{2}=0.5$. (b) $0.5=50\%$. (c) The long-term data gives $Pr(\text{rain})\approx 0.5$.

Q4. (a) $\frac{18}{120}=0.15$. (b) A die has six equally-likely faces, so the theoretical probability of a 4 is $\frac{1}{6}\approx 0.167$. (c) The experimental value 0.15 is close to the theoretical 0.167; with 120 rolls a small difference is expected.

Q5. (a) $\frac{34}{40}=0.85$. (b) $Pr(\text{save})\approx 0.85$. (c) Expected saves in $25=0.85\times 25=21.25$, so about 21 saves.

Q6. (a) A fair coin has two equally-likely sides, so $Pr(\text{head})=0.5$. (b) Only 20 tosses is a very short run, and relative frequency varies a lot over short runs, so 0.65 instead of 0.5 is not surprising. (c) Over 1000 tosses the relative frequency would be expected to settle much closer to 0.5.

Q7. Relative frequency $=\frac{24}{800}=\frac{3}{100}=0.03=3\%$.

Q8. $Pr(\text{three})\approx\frac{45}{120}=0.375$. Expected in 200 attempts $=0.375\times 200=75$.

Q9. Dry days $=219$ out of 365, so the relative frequency is $\frac{219}{365}=0.6=60\,\%$.

Q10. Relative frequency of heads $=\frac{260}{500}=0.52$. This is close to the theoretical probability 0.5, just slightly above — consistent with a fair coin over many tosses.

Q11. $Pr\left(\text{left-handed}\right)\approx\frac{320}{2000}=0.16$. In a city of $25\,000$, expected number $=0.16\times 25\,000=4000$.

Q12. $Pr\left(\text{immune}\right)\approx\frac{4800}{6000}=0.8$. For $15\,000$ people, expected number $=0.8\times 15\,000=12\,000$.

Q13. $\frac{55}{300}\approx 0.183$, compared with the theoretical $\frac{1}{6}\approx 0.167$. The 300-roll estimate is more trustworthy than the 120-roll estimate in Q4, because a larger number of trials gives a relative frequency closer to the true probability.

Q14. $Pr\left(\text{reject}\right)\approx\frac{45}{1500}=0.03$. Expected rejects in $10\,000=0.03\times 10\,000=300$.

Q15. $Pr\left(\text{boundary}\right)\approx\frac{60}{400}=0.15$. Expected boundaries from 1000 balls $=0.15\times 1000=150$.

Q16. Relative frequency of a day over 30°C $=\frac{73}{200}=0.365=36.5\,\%$.

Q17. (a) The first student: $\frac{20}{50}=0.40$. The second student: $\frac{240}{500}=0.48$. (b) The second student's estimate (0.48) is more reliable, because it is based on far more trials (500 versus 50), and more trials give a relative frequency closer to the true probability of 0.5.

Q18. $Pr\left(\text{condition}\right)\approx\frac{1200}{8000}=0.15$. Expected number in a town of $20\,000=0.15\times 20\,000=3000$. Assumption: the town's population is similar to (representative of) the group that was surveyed.

Topic 2 Test A — Multiple-choice

Data analysis, probability and statistics (Chapters 5–8). Approved materials: one bound reference that may be annotated; one scientific calculator. Section A: 15 multiple-choice questions (15 marks). Time: about 25 minutes. Unless otherwise indicated, diagrams are not drawn to scale.

Section A — Multiple-choice questions

Choose the response that is correct for the question.

Question 1

The number of text messages a person sends in a day is an example of

A. categorical (nominal) data

B. categorical (ordinal) data

C. discrete numerical data

D. continuous numerical data

Question 2

A factory checks the quality of every 20th item coming off a production line. This is an example of

A. systematic sampling

B. stratified sampling

C. convenience sampling

D. a census

Question 3

A website invites visitors to click to vote in an opinion poll. The main problem with the resulting sample is that

A. the sample is too large

B. the sample is stratified

C. it is actually a census

D. the sample is self-selected and not representative

Question 4

In a pie chart, a category that makes up 25% of the total is shown by a sector with an angle of

A. 25°

B. 45°

C. 120°

D. 90°

Question 5

The most appropriate graph to show how a household budget is divided among spending categories is a

A. line graph

B. pie chart

C. scatterplot

D. stem-and-leaf plot

Question 6

The median of $4, 7, 7, 9, 12, 15, 15, 15, 20$ is

A. 9

B. 12

C. 13

D. 15

Question 7

A data set has the five-number summary $8, 14, 20, 27, 35$. Its interquartile range is

A. 13

B. 6

C. 21

D. 27

Question 8

The mean of $5, 8, 10, 12, 15$ is

A. 9

B. 10

C. 12

D. 50

Question 9

The daily maximum temperatures (°C) over a week were $3, 9, 12, 15, 18, 22, 28$. The range is

A. 15

B. 22

C. 25

D. 28

Question 10

Which of the following is a measure of spread?

A. the interquartile range

B. the median

C. the mean

D. the mode

Question 11

A data set of weekly wages contains one very high value. Compared with the median, the mean will be

A. lower than the median

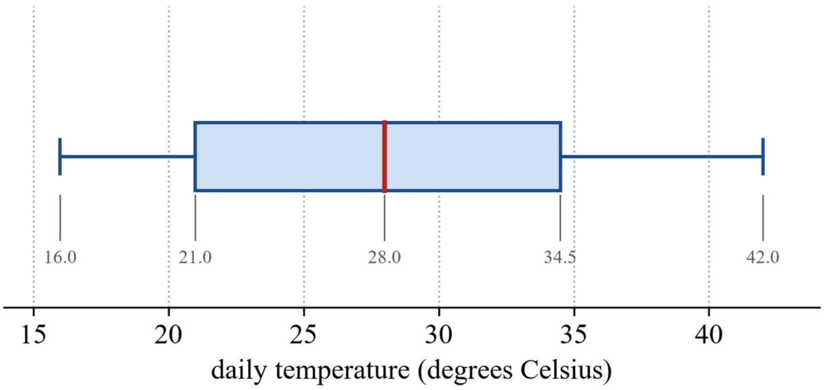
B. the same as the median

C. zero

D. higher than the median

Question 12

The boxplot below shows the daily maximum temperatures recorded at a town.



The median daily maximum temperature is

A. 22

B. 25

C. 28

D. 30

Question 13

A bag contains 3 red, 5 blue and 4 green marbles. A marble is drawn at random. The probability that it is red is

A. $\frac{1}{4}$

B. $\frac{1}{12}$

C. $\frac{1}{3}$

D. $\frac{3}{5}$

Question 14

The probability that it rains tomorrow is 0.35. The probability that it does **not** rain is

A. 0.35

B. 0.65

C. 0.7

D. 1.35

Question 15

In quality testing, 40 out of 250 items were found to be faulty. The relative frequency of a faulty item is

A. 0.016

B. 0.4

C. 0.16

D. 1.6

End of Section A

Topic 2 Test A — Solutions

Question	Answer	Working
1	**C**	A count of messages takes whole-number values, so it is discrete numerical data.
2	**A**	Choosing every 20th item is systematic sampling.
3	**D**	Visitors choose whether to vote, so the sample is self-selected and unrepresentative.
4	**D**	$0.25\times 360^\circ = 90^\circ$.
5	**B**	A pie chart shows parts of a whole.
6	**B**	With $n=9$, the median is the 5th value, 12.
7	**A**	$\text{IQR} = Q_3 - Q_1 = 27 - 14 = 13$.
8	**B**	$\dfrac{5+8+10+12+15}{5} = \dfrac{50}{5} = 10$.
9	**C**	$28 - 3 = 25$.
10	**A**	The interquartile range measures spread; the median, mean and mode are measures of centre.
11	**D**	A single very high value pulls the mean up but not the median.
12	**C**	The median is the line inside the box, at 28.
13	**A**	$\dfrac{3}{3+5+4} = \dfrac{3}{12} = \dfrac{1}{4}$.
14	**B**	$1 - 0.35 = 0.65$.
15	**C**	$\dfrac{40}{250} = 0.16$.

Topic 2 Test B — Short answer

Data analysis, probability and statistics (Chapters 5–8). Approved materials: one bound reference; one scientific calculator. Section B: 6 questions (30 marks). Time: about 40 minutes. In questions where more than one mark is available, appropriate working must be shown. Only round when instructed.

Section B

Question 1 (5 marks)

The waiting times (in minutes) of 11 patients at a medical clinic were

$$8, 12, 15, 18, 20, 23, 25, 28, 31, 35, 58.$$

a. Find the mean waiting time, correct to one decimal place. 1 mark

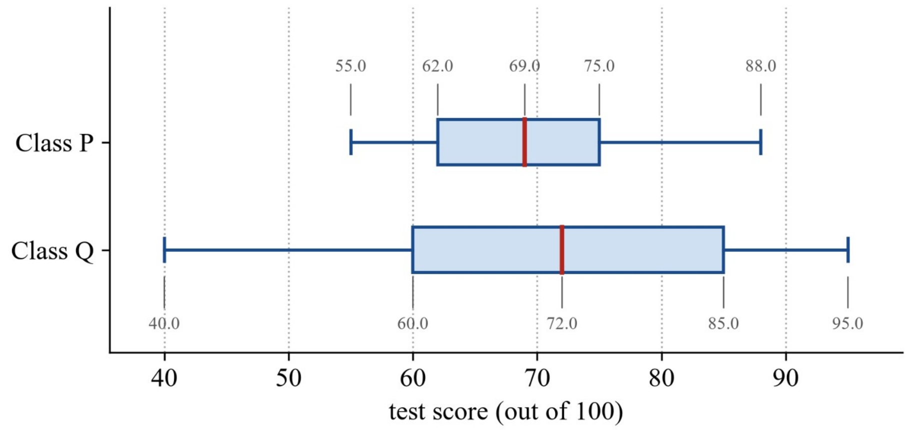
b. Find the median, the lower quartile Q_1 and the upper quartile Q_3. 2 marks

c. Find the interquartile range. 1 mark

d. Show that the 58-minute waiting time is an outlier. 1 mark

Question 2 (5 marks)

The test scores (out of 100) of two classes are shown in the parallel boxplots below.



a. Write down the median score of each class. 2 marks

b. Find the interquartile range of each class. 2 marks

c. The two classes have a similar median. State which class was more consistent, and give a reason. 1 mark

Question 3 (5 marks)

A survey of 40 households recorded the number of pets in each household.

Number of pets	0	1	2	3	4
Frequency	6	10	14	7	3

a. Complete a relative frequency column for this table, giving each value as a percentage. 2 marks

b. What percentage of households have 2 or more pets? 1 mark

c. State the most appropriate type of graph to display this data, and write down the modal number of pets. 2 marks

Question 4 (5 marks)

A gym wants to find out how satisfied its 2000 members are. It surveys the 50 members who attend a Saturday-morning class.

a. Identify the population and the sample in this survey. 2 marks

b. Name the sampling method used, and give one reason the sample may be biased. 2 marks

c. Describe one way the gym could obtain a more representative sample. 1 mark

Question 5 (5 marks)

A spinner has 8 equal sectors, 3 of which win a prize.

a. Find the probability of winning a prize on one spin. 1 mark

b. Find the probability of **not** winning on one spin. 1 mark

c. In 200 spins, a prize was won 54 times. Find the relative frequency of winning. 1 mark

d. Using the theoretical probability, how many wins would you expect in 400 spins? 2 marks

Question 6 (5 marks)

At a kiosk, the number of ice-creams sold, y, is modelled against the day's maximum temperature, x (°C), by the line of best fit $y = 2x + 10$. The data used ranged from 10 °C to 35 °C.

a. Predict the number of ice-creams sold on a 25 °C day, and state whether this is interpolation or extrapolation. 2 marks

b. Predict the number sold on a 45 °C day, and give one reason this prediction may be unreliable. 2 marks

c. The owner says "for every 5 °C hotter, we sell about 10 more ice-creams." Explain how the line of best fit supports this. 1 mark

End of Section B

Topic 2 Test B — Solutions

Question 1.

a. $\bar{x} = \dfrac{8+12+15+18+20+23+25+28+31+35+58}{11} = \dfrac{273}{11} = 24.8$ minutes (to 1 d.p.).

b. With $n = 11$, the median is the 6th value, 23 minutes. The lower half $8, 12, 15, 18, 20$ gives $Q_1 = 15$ minutes; the upper half $25, 28, 31, 35, 58$ gives $Q_3 = 31$ minutes.

c. $\text{IQR} = Q_3 - Q_1 = 31 - 15 = 16$ minutes.

d. The upper fence is $Q_3 + 1.5 \times \text{IQR} = 31 + 1.5 \times 16 = 31 + 24 = 55$ minutes. Since $58 > 55$, the 58-minute wait lies beyond the upper fence and is therefore an outlier.

Question 2.

a. Reading the value printed at the median line inside each box: Class P median $= 69$; Class Q median $= 72$.

b. For Class P, $Q_1 = 62$ and $Q_3 = 75$, so $\text{IQR} = 75 - 62 = 13$. For Class Q, $Q_1 = 60$ and $Q_3 = 85$, so $\text{IQR} = 85 - 60 = 25$.

c. Class P was more consistent: its box and whiskers are much narrower (IQR 13 versus 25), so its scores are less spread out even though the medians are similar.

Question 3.

a. Each relative frequency is $\dfrac{\text{frequency}}{40} \times 100\,\%$:

Number of pets	0	1	2	3	4
Relative frequency	$15\,\%$	$25\,\%$	$35\,\%$	$17.5\,\%$	$7.5\,\%$

b. Households with 2 or more pets: $14 + 7 + 3 = 24$, which is $\dfrac{24}{40} \times 100\,\% = 60\,\%$.

c. The data is discrete numerical (whole-number counts), so a **column (bar) graph** is most appropriate. The modal number of pets is 2 (the highest frequency, 14).

Question 4.

a. The population is all 2000 gym members; the sample is the 50 members who were surveyed at the Saturday-morning class.

b. The method is convenience sampling. It may be biased because members who attend a Saturday-morning class are likely to be more active and more satisfied than the membership as a whole, so the sample is not representative.

c. The gym could take a simple random sample of members from its full membership list (or survey members across a range of days and times), so that every member has an equal chance of being chosen.

Question 5.

a. $Pr(\text{win}) = \dfrac{3}{8}$.

b. $Pr(\text{not win}) = 1 - \dfrac{3}{8} = \dfrac{5}{8}$.

c. Relative frequency $= \dfrac{54}{200} = 0.27$.

d. Expected wins $= \dfrac{3}{8} \times 400 = 150$.

Question 6.

a. At $x = 25$: $y = 2(25) + 10 = 60$ ice-creams. Since 25 °C lies within the data range 10–35 °C, this is **interpolation**.

b. At $x = 45$: $y = 2(45) + 10 = 100$ ice-creams. This is extrapolation beyond the data, so it may be unreliable — the linear trend may not continue at very high temperatures (for example, sales could level off or the kiosk could sell out).

c. The gradient of the line is 2, meaning about 2 more ice-creams are sold for each extra 1 °C. Over 5 °C that is $2 \times 5 = 10$ more ice-creams, which is exactly what the owner claims.