Basic probability

An **event** is an **outcome** or set of outcomes in a probability experiment.

The **probability** of an event is a measure of the likelihood of that event occurring. Probabilities lie between 0 and 1.

- An **impossible event** has probability zero.
- The more unlikely an event is, the closer its probability is to 0.
- A **certain event** has probability 1.
- The more likely an event is, the closer its probability is to 1.

For equally likely outcomes, the probability of an event is equal to the number of outcomes in the event divided by the total number of outcomes.

Example

When a fair die is rolled, the sample space has six equally likely outcomes {1, 2, 3, 4, 5, 6}.

1. Each of the six numbers has a probability of $\frac{1}{6}$ of occurring.
2. There are two numbers which are greater than 4 (i.e. 5 and 6) so the probability of a number greater than 4 is $\frac{2}{6}$ or $\frac{1}{3}$.

Relative frequency

If a selection is made 'at random' from a group, then each member of the group is equally likely to be chosen.

The probability of a particular event, A, is written P(A). This probability is worked out using **relative frequency** (the number of occurrences of the event is divided by the total number of possible occurrences). A relative frequency is also called a **proportion**.

Two-way tables

A **two-way table** has rows which represent one **category variable** (such as male/female), and columns which represent another category variable (such as sports choices).

Example

On an activities day, 670 students signed up for one of three summer sports.

The 2-way (contingency) table shows their choices.

	Tennis	**Swimming**	**Cricket**	**Totals**
Boys	120	90	140	350
Girls	150	80	90	320
Totals	270	170	230	670

A student is chosen at random. Find the probability that the student

1. is a boy who plays tennis **2.** chose tennis or cricket **3.** is not a cricketer.

Solution

1. P(student is a boy who plays tennis) $= \frac{120}{670}$ 120 boys play tennis
2. P(student chose tennis or cricket) $= \frac{500}{670}$ $270 + 230 = 500$ chose tennis or cricket
3. P(student is not a cricketer) $= \frac{440}{670}$ $670 - 230 = 440$ did not choose cricket

ESA PUBLICATIONS ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

Answers p. 67

1. Two fair dice are rolled together and the numbers added.

a. Complete the table showing the set of all 36 possible outcomes.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5			
3	4					
4	5					
5	6					
6						

b. When two dice are rolled and the numbers added, use the table to work out the probability that the sum is:

i. equal to 6

ii. equal to 13

iii. an odd number

c. Find the probability the sum is:

i. less than 5 __________

ii. more than 7 __________

iii. less than 13 __________

iv. a multiple of 3 __________

d. Ben said that getting a sum of 4 has the same chance as getting the sum of another number.
What is this other number?

2. Amy carried out a week-long survey of customers who bought items in her shop. She had a total of 304 customers during the week, of which 115 were male.

a. What proportion of customers were male?

b. Estimate the probability that the next customer in her shop is female.

c. Why is your answer to part **b.** only an estimate? Discuss how confident you are that this is a good estimate of the actual likelihood that the next customer is female.

3. A shop takes orders for dining tables, which come in three sizes and two colours.

	Small	Medium	Large	Total
Black	28	45	37	110
Brown	36	51	48	135
Total	64	96	85	245

An order is selected at random. Find the probability it is for

a. a small black table __________

b. a medium table __________

c. a brown table __________

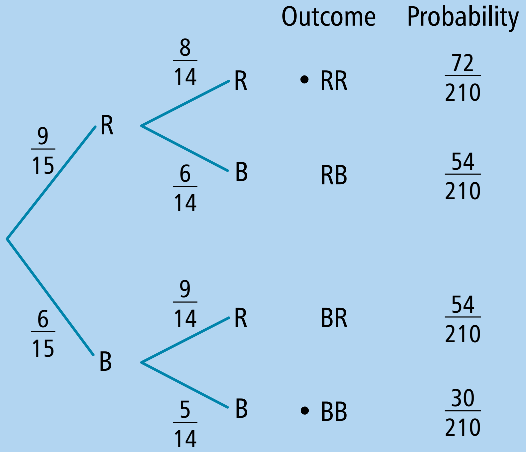
d. a table that is not large or black __________

 ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

The tree diagram for this situation is shown below.

As before, probabilities of outcomes are calculated by multiplying along the branches.

For example P(RR) $= \frac{9}{15} \times \frac{8}{14} = \frac{72}{210}$



1. P(marbles same colour) = P(RR or BB)

$$= \frac{72}{210} + \frac{30}{210}$$

$$= \frac{102}{210}$$

$$= \frac{17}{35} \text{ or } 0.4857$$

2. Require the conditional probability

P(both marbles red given that both same colour).

This is given by:

$$\frac{P(RR)}{P(\text{same colour})} = \frac{\frac{72}{210}}{\frac{102}{210}}$$

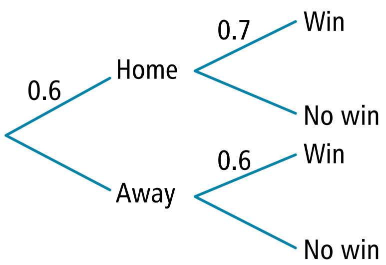
$$= \frac{72}{102} \text{ or } 0.7059$$

Note: If this experiment were run 210 times, then the expected number of times that you would get ‘two marbles the same colour’ is 102, and the expected number of times that you would get ‘two red marbles’ is 72.

 ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

Answers p. 69

1. 60% of Tom's hockey games are 'home' games. Tom's team wins 70% of 'home' games and 60% of 'away' games. This information is on the tree diagram alongside.



Find the following probabilities:

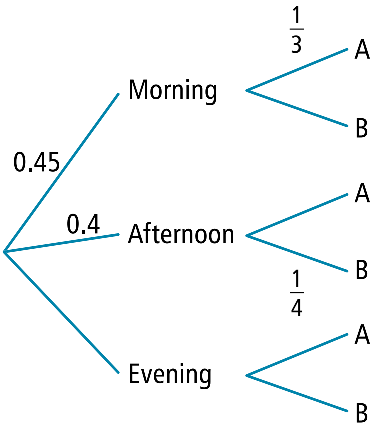
a. P(no win given that game is away)

b. P(no win given that game is at home)

c. P(no win)

d. P(home game given that there was no win)

2. In a tournament, tennis games are played during the day at two locations: A and B.
   - 45% of games are played in the morning and a third of these are at location A.
   - 40% of games are played in the afternoon and a quarter of these are at location B.
   - four fifths of the evening games are played at location A



Complete the tree diagram and use it to find the following probabilities.

a. P(game is at B given that it is a morning game)

b. P(game is at B given that it is an evening game)

c. P(game is at B)

d. P(morning game given that game is at B)

 ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

2. Use Standard normal tables to work out the following probabilities.
Draw a diagram for each.

a. $P(Z \geq 0.67)$

b. $P(Z \geq 2.08)$

c. $P(Z \geq -1)$

d. $P(Z \leq 2.11)$

e. $P(Z \leq 0.95)$

f. $P(Z \leq -0.95)$

g. $P(Z \geq -0.234)$

h. $P(Z < -2 \text{ or } Z > 2)$

ESA PUBLICATIONS ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

Probabilities using the normal curve

Normal probability problems usually involve variables X (such as the weights of school bags or the lifetime of watch batteries) where the mean is not 0 and/or the standard deviation is not 1.

To solve these problems the normal distribution (with mean μ and standard deviation σ) must be converted to a standardised normal distribution (with mean 0, standard deviation 1).

The conversion is done using the formula:

$$Z = \frac{X - \mu}{\sigma}$$

where Z = standard normal variable

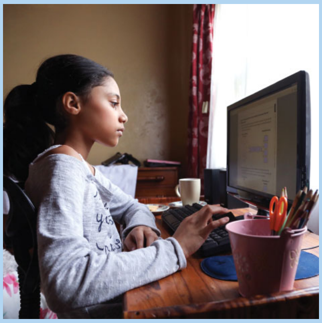
X = normal variable

μ = mean of X

σ = standard deviation of X

Example

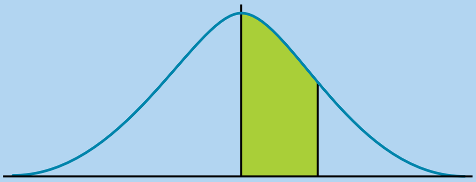
The time Ramani spends each night on the internet is Normally distributed with a mean of 70 minutes and a standard deviation of 8 minutes. Calculate the probability that the length of time Ramani spends on the internet is:



1. between 70 and 80 minutes
2. between 64 and 74 minutes
3. more than 82 minutes
4. more than 64 minutes.

Solution

1. Require $P(70 \le X \le 80)$ where X is the time spent on the internet.



Using the formula, the x-values are converted into z-values using the formula $Z = \frac{X - \mu}{\sigma}$

$z_1 = \frac{70 - 70}{8} = 0$ and $z_2 = \frac{80 - 70}{8} = 1.25$

$P(70 \le X \le 80) = P(0 \le Z \le 1.25)$

$= 0.3944$ from tables

The probability that Ramani spends between 70 and 80 minutes on the internet is 0.3944.

2. Require $P(64 \le X \le 74)$

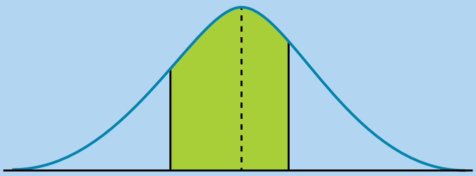
$z_1 = \frac{64 - 70}{8} = -0.75$ and $z_2 = \frac{74 - 70}{8} = 0.5$

Therefore:

$P(64 \le X \le 74) = P(-0.75 \le Z \le 0.5)$

$= 0.2734 + 0.1915$ tables

$= 0.4649$



The probability Ramani spends between 64 and 74 minutes on the internet is 0.6649.

 ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

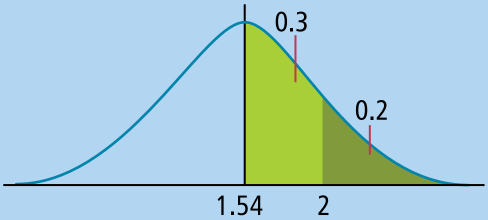
Harder inverse normal problems

At Excellence level you may need to find an unknown **mean** or **standard deviation**.

Example

Q. The depth of a river is normally distributed with mean 1.54 m. 20% of the river is more than 2 metres in depth. Find the standard deviation of the depth of the river.

A. Let X be the depth (in metres) of the river, and s be the standard deviation of X.
The diagram represents the situation.



If $P(X > 2) = 0.2$ then

$P\left(Z > \frac{2 - 1.54}{s}\right) = 0.2$ standardising

$P\left(0 < Z < \frac{0.46}{s}\right) = 0.3$ by symmetry of normal curve

From tables, $P(0 < Z < 0.841) = 0.3$ so,

$\frac{0.46}{s} = 0.841$

$s = \frac{0.46}{0.841}$ rearranging

$s = 0.547$

The standard deviation of the depth of the river is 0.55 m (2 d.p.).

Alternatively, the **inverse normal function** on a graphics calculator may be used, using the standard normal distribution with mean = 0 and standard deviation =1. This will give the required z-value for the given area. Convert to the x-variable by using the formula $Z = \frac{X - \mu}{\sigma}$.

Example

In the above example, the z-value corresponding to an area of 0.2 on the right of the curve is found by entering (right tail = 0.2, $\sigma = 1$, $\mu = 0$) into the inverse normal function.

This gives a z-value of 0.8416

Substituting $x = 2$, $\mu = 1.54$, $z = 0.8416$ into $Z = \frac{X - \mu}{\sigma}$ and solving gives standard deviation as above.

 ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

Answers p. 71

1. Weights of a certain type of carrot are normally distributed with standard deviation 5 g. 3% are packed as 'baby carrots' because they are below 30 g in weight.
   What is the mean weight of this type of carrot?

2. The time taken for solving a simple puzzle is normally distributed with a mean of 2 minutes and 35 seconds. If 10% of the participants take more than 3 minutes, what is the standard deviation of the time taken?

3. Distances travelled daily by an Uber driver are normally distributed with a standard deviation of 23 km. If the driver travels more than 200 km on 12% of days, what is the mean daily distance that he drives?

4. The diameter of apples produced by an orchard is normally distributed with mean 72 mm. The supermarket requires apples with a diameter greater than 65 mm, so 8% of the apples from this orchard were too small. What is the standard deviation of the diameters of the apples from this orchard?

 ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

Practice assessment task

Answers p. 72

1. Chocco bars are produced by a machine in a factory. The weights of the bars are normally distributed with a mean of 46 grams and a standard deviation of 0.7 grams.

   a. What is the probability that a Chocco bar weighs between 45 and 47.3 grams?

   b. What percentage of Chocco bars weigh more than 46.5 grams?

   c. i. In one production run, 24 000 Chocco bars are produced. What is the expected number of Chocco bars in this production run that weigh less than 45 grams? Give your answer to 2 s.f.

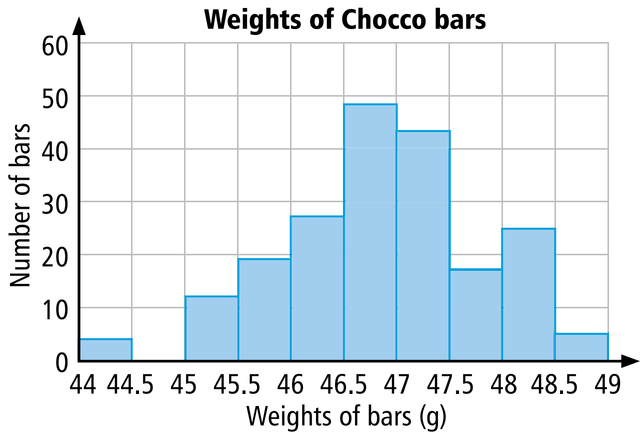
   ii. In another production run, 943 bars are found to be more than 47 g. How big was the production run? Give answer to the nearest hundred bars.

   d. i. 5% of the Chocco bars weigh more than what weight (to 3 s.f.)?

   ii. Between what weights will the middle 90% of Chocco bars lie?

 ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

e. Jamie, the production control manager, wonders if the machine is working properly and producing Chocco bars whose weights are normally distributed with mean 46 kg and standard deviation 0.7 g. Jamie takes a random sample of 200 bars from the production line, weighs each bar and draws a graph of the results.



Compare the distribution of the weights of Jamie's sample of Chocco bars with the expected distribution of the weights if the machine is working properly.

Use statistical terms to explain your answer.

f. Jamie decides to adjust the mean weight of bars produced by this machine so that less than 1% of bars are below 44.5 g. If the settings are calibrated to the nearest tenth of a gram, what should the new mean weight setting be?

 ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

Answers

Basic probability (page 2)

1. a.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

   b. i. $\frac{5}{36}$ ii. 0 iii. $\frac{1}{2}$

   c. i. $\frac{1}{6}$ ii. $\frac{5}{12}$ iii. 1 iv. $\frac{1}{3}$

   d. The other number is 10

2. a. $\frac{115}{304}$ or 0.3783 b. $\frac{189}{304}$ or 0.6217

   c. The probability is an estimate as it is a relative frequency based on the data from a single week of customers. It is unlikely to be a good estimate of the probability that the next customer is female, as it is based on only one week of data, and there is no guarantee that the week was in any way a 'typical' one (nor is it known if the current week (when the probability is being calculated) is in any way 'typical'). There will be considerable variability expected from week to week.

3. a. $\frac{28}{245}$ or $\frac{4}{35}$ b. $\frac{96}{245}$ c. $\frac{135}{245}$ or $\frac{27}{49}$

   d. $\frac{87}{245}$ or 0.3551

Conditional probability (page 4)

1. a. $\frac{140}{350}$ or $\frac{2}{5}$ b. $\frac{150}{270}$ or $\frac{5}{9}$ c. $\frac{140}{230}$ or $\frac{14}{23}$

   d. $\frac{230}{500}$ or $\frac{23}{50}$ e. $\frac{80}{230}$ or $\frac{8}{23}$ f. $\frac{230}{400}$ or $\frac{23}{40}$

2. a. $\frac{3}{8}$ b. $\frac{66}{106}$ or $\frac{33}{53}$ c. $\frac{80}{184}$ or $\frac{10}{23}$

   d. $\frac{50}{174}$ or $\frac{25}{87}$ e. $\frac{124}{200}$ or $\frac{31}{50}$ f. $\frac{140}{176}$ or $\frac{35}{44}$

3. a. i. $\frac{26}{80}$ or 0.325 ii. $\frac{25}{80}$ or 0.3125 iii. $\frac{12}{80}$ or 0.15

   b. i. $\frac{9}{26}$ or 0.3462 ii. $\frac{14}{37}$ or 0.3784 iii. $\frac{39}{54}$ or 0.7222

4. a. i. $\frac{746}{5\,536}$ or 0.1348 ii. $\frac{2\,700}{5\,536}$ or 0.4877

   iii. $\frac{1\,080}{5\,536}$ or 0.1951 iv. $\frac{839}{5\,536}$ or 0.1516

   b. i. $\frac{191}{2\,700}$ or 0.0707 ii. $\frac{542}{2\,700}$ or 0.2007

   iii. $\frac{1\,374}{2\,700}$ or 0.5089

   c. i. $\frac{193}{2\,836}$ or 0.0681 ii. $\frac{593}{2\,836}$ or 0.2091

   iii. $\frac{1\,442}{2\,836}$ or 0.5085

   d. $\frac{162}{321}$ or 0.5047 e. $\frac{906}{1\,881}$ or 48.2% (1 d.p.)

5. a. 0.62 b. 0.97

   c. $\frac{92}{900}$ or 0.0356 d. $\frac{170}{338}$ or 0.5030

6. a.

	Factory	Office	Totals
Male	144	30	174
Female	96	90	186
Totals	240	120	360

   b. i. $\frac{30}{360}$ or $\frac{1}{12}$ ii. $\frac{96}{360}$ or $\frac{4}{15}$

   iii. $\frac{174}{360}$ or $\frac{29}{60}$ iv. $\frac{120}{360}$ or $\frac{1}{3}$

   c. $\frac{96}{186}$ or $\frac{16}{31}$ d. $\frac{96}{240}$ or $\frac{2}{5}$

7. $\frac{11}{60}$

Expected numbers (page 9)

1. 640 2. 29 or 30 3. 326 or 327
4. 52 approx. 5. 9 6. 48%
7. Expected number is $\frac{2}{3} \times 80 = 53\frac{1}{3}$ (or 53 goals)
   Amy's claim may be true, and this difference in success rates may be due to the variation that arises from a smallish sample of 80 throws. More trials would be needed before deciding that Amy's claim is incorrect.
8. 60
9. a. 155 b. 267 or 268

   c. *Answers will vary, an example follows*

   Reasonably confident since the survey was of a large sample which was randomly selected, so the sample proportions are likely to be good estimates of the population probabilities. However, the sample may be unrepresentative depending on the day of the week or the time of the day in which the sample was taken, when people of different viewpoints to those selected may not have been available for sampling. Also it is difficult to know how the 9 abstainers would have affected sample proportions.
10. a. 19 or 20 games

   b. It is likely that Paul has been improving as he continues to train, so his success rate at first may have been lower. If so it is likely that it took more than 20 games to reach his 10th strike.
11. 0.15 (2 d.p.)
12. $\frac{Mn}{N}$

Risk and relative risk (page 12)

1. a. $\frac{389}{484}$ or 0.8037 b. $\frac{51}{95}$ or 0.5368

   c. i. $\frac{44}{295}$ or 0.1492 ii. $\frac{51}{189}$ or 0.2698

   d. Risk is $\frac{95}{484} = 0.196$ or 0.2 (1 d.p.) which is $\frac{1}{5}$

   e. Approximately 600 (using $\frac{1}{5}$) or 588 (using 0.196)

   f. 0.553
2. a. 0.2961 b. 0.2834 c. 1.1251

   d. The pass rate for students who attend lectures at the college is 12.5% higher than the pass rate for students who study remotely.
3. a. 0.0567 b. 0.0387 c. 0.5565

   d. The risk of a traveller to Europe making an insurance claim is 55.7% of the risk of a traveller to Asia making an insurance claim.

 ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.

Index

ESA PUBLICATIONS ISBN 978-1-988586-67-0 © Copying or scanning from ESA books is subject to the provisions of the Copyright Act 1994.