

GCSE MATHS TUTOR

A red abstract background featuring a prominent, swirling pattern that resembles a vortex or a stylized eye. The colors range from deep red to bright red, with some darker, almost black, areas in the shadows of the swirl.

Revision Guide

PART FOUR

INFORMATION

www.gcsemathstutor.com

This book is under copyright to GCSE Maths Tutor. However, it may be distributed freely provided it is not sold for profit.

Contents

time series graphs	3
representing data	5
cumulative frequency	8
histograms	11
mean, mode, median & range	14
probability	18

Time Series Graphs

Time series graphs - this is the collective name for all graphs that have **time** as the x-axis. There are 3 types of graph:

- random/erratic
- repeated/cyclic
- ones with a particular trend/direction

'Seasonality' & Cyclic time series

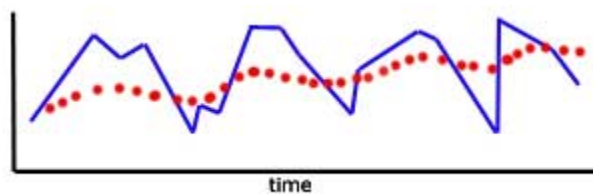
Seasonality is the term for data that has a periodicity of one year. That is, it alters over a period of one year, then repeats to some degree. The highs and lows may alter, but the general shape of the graph is similar year on year.

Similarly a **Cyclic** time series repeats itself. However, this is a more general term. The period may be seconds (like the beat of a heart) or thousands of years (like the coming and going of ice ages).

Trends - the 'Moving Average'

A **trend** is a simplification of chaotic time-related data to show the underlying movement of values.

A **Moving Average** is simply the average of consecutive blocks of data. In this way, fluctuations in a curve are 'ironed out'.



The number of pieces of data in a block is termed the number of 'points'.

- 3 pieces of data in a block is a '3 point' moving average
- 10 pieces of data in a block is a '10 point' moving averageetc.

it is important to remember that the starting point of each block advances by one number each time.

Example - calculate **four point** moving averages for the following results:

1 3 8 4 5 7 3 8 2 13

4 point numbers	4 point average
1 3 8 4	4
3 8 4 5	5
8 4 5 7	6
4 5 7 3	4.75
5 7 3 8	5.75
7 3 8 2	5
3 8 2 1	3.5
8 2 13	3.5

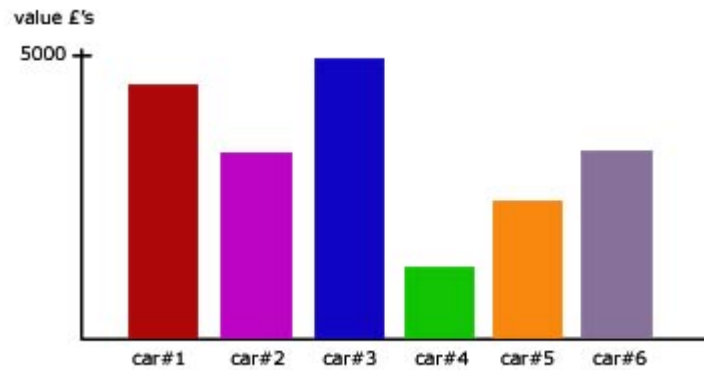
NB plotting of moving average points - the moving average for each block of data should be plotted in the **middle** of each number block

That is, the first moving average should be plotted between the 2nd and 3rd. reading along the x-axis.

The second moving average should be plotted between the 5th and 6th reading, and so on.

Representing Data

Bar/block graph

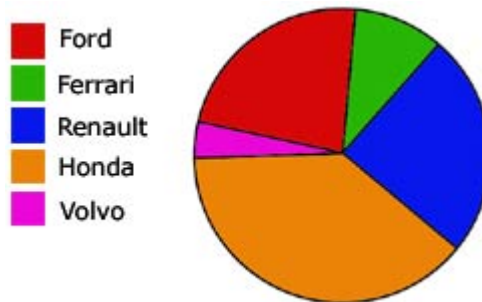


Not to be confused with a histogram, a bar chart/graph has columns of equal width, with height representing some variable, usually a number(eg frequency, % money).

The base of each column can represent anything (eg a car type, a person's name, a company).

Unlike a histogram, in a bar graph the columns need not be adjacent to each other .

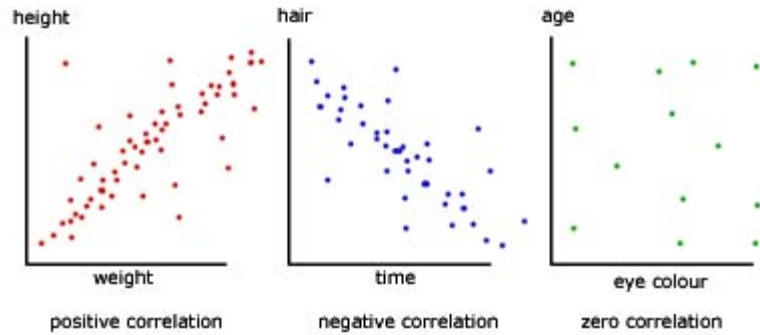
Pie Chart



car make	% imports into country X
Ford	20
Ferrari	12
Renault	26
Honda	35
Volvo	7

The angle in degrees for each car make is given by:

$$\frac{\text{'% import'} \times 360}{100}$$

Scatter diagrams/graphs

When two sets of data are plotted against each other, a scatter of points is produced.

Correlation is the relationship between one set of data and the other.

An exact correlation would be '1' (a straight line graph with all the points on the line), while a zero correlation is '0'.

The first graph would have a correlation between 0 and 1.

The second graph would have a correlation between 0 and -1.

The third graph would have a correlation close to zero.

Stem & Leaf table

Stem & leaf tables are similar to bar charts but differ in two distinct ways:

the pattern is horizontal not vertical

the display is made from numbers, not block colour

The data is placed in number order in groups of tens. It is then displayed horizontally for each tens grouping, starting with 0-9 at the top.

0	2						
1	1 3 5						
2	0 4 4 6 9						
3	4 6 8 8 9 9 9						
4	7 8 9 9 9						
5	1 1 3 4 8 9						
6	0 1 1 2						
7	3 4 5						
8	2						

The numbers displayed are:

2

11 13 15

20 24 24 26 29

34 36 38 38 39 39 39

47 48 49 49 49

51 51 53 54 58 59

60 61 61 62

73 74 75

82

Cumulative Frequency

Introduction

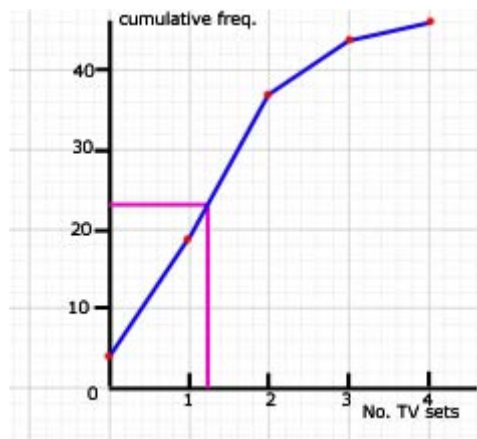
Frequency is used to describe the number of times results occur. On the other hand, **cumulative frequency** is a 'running total'. It is the sum of frequencies moving through the data.

Example - A survey was done to look at how many TV's there were in a household.

no. of TV's	frequency		cumulative frequency
0	4	4	4
1	15	4+15	19
2	18	4+15+18	37
3	7	4+15+18+7	44
4	2	4+15+18+7+2	46

The definition of the **median** is that particular value half way through the data.

If the cumulative total of frequencies is 46, then the median is the 23 rd. value.



So the median is 1 (nearest whole number).

Where there are lots of values, say more than 10, the data is best presented as '**grouped data**'.

Quartiles

The **upper quartile** is the particular **value** $\frac{3}{4}$ through the cumulative frequency.

The **lower quartile** is the particular **value** $\frac{1}{4}$ through the cumulative frequency.

In the example given above:

upper quartile = $0.75 \times 46 = 34.5$ (rounded to 36) - this gives a value close to 2

lower quartile = $0.25 \times 46 = 11.5$ (rounded to 12) - this gives a value close to zero

note: **values** are the readings along the bottom of a cumulative frequency graph

Ranges

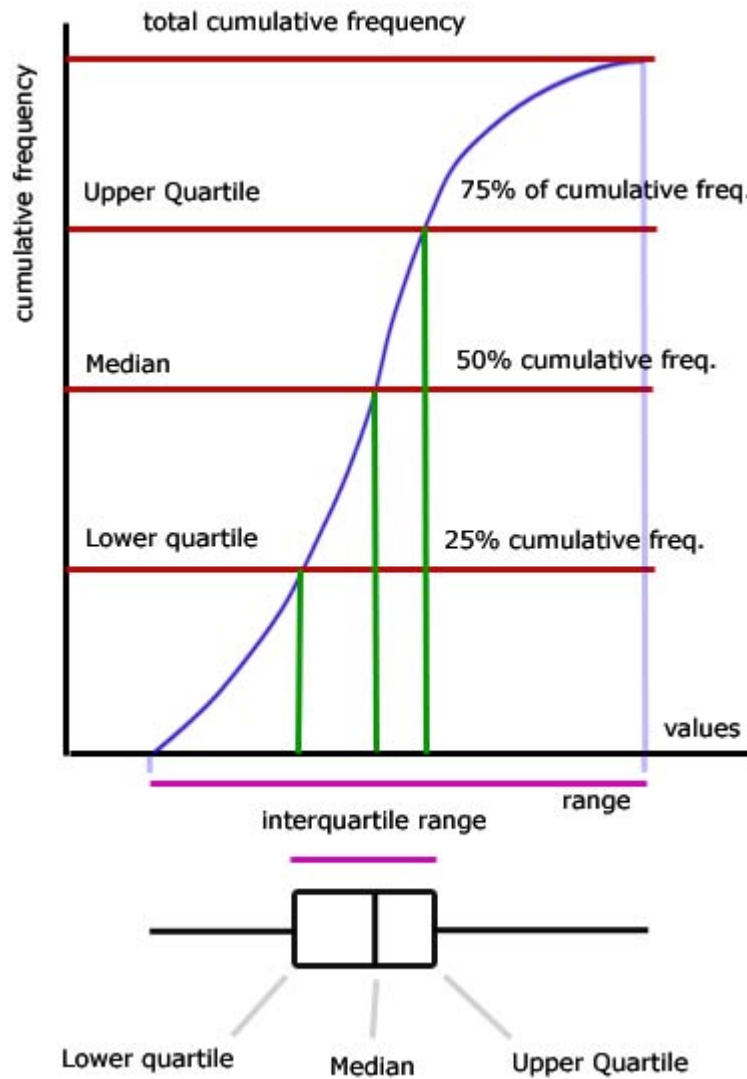
The **interquartile range** is the difference between the lower and upper quartiles.

$$\text{interquartile range} = 34.5 - 11.5 = 23$$

The interquartile range is a measure of how spread out data is. With reference to products (eg the shelf-life of foods) a small value for the interquartile range means a more accurate result.

Box & Whisker Plot (Box Plot)

The plot is derived from a cumulative frequency graph and shows the range of data, the interquartile range, and where the quartiles are in relation to the median.



Histograms

Block graphs

As a comparison you may wish to revise 'block graphs'(in 'representing data') before going on to study histograms.

Grouped Data (Grouped Frequencies)

To understand frequency density and it's role in histograms, you need first to appreciate the meaning of a number of terms relating to grouped data:

Example

height(cm)	71-75	76-80	81-85	86-90	91-95
frequency	5	17	21	15	2

class boundaries are the exact values(for each set of grouped data) where one set of values ends and the other begins.

In our example, the class boundaries are: **75.5 __ 80.5__ 85.5__ 90.5**

You must appreciate that these numbers are the 'deciders' as to which group data is placed.

For example 75.45 would be rounded down to 75 and be placed in the first group, but 75.55 would be rounded up and placed in the second.

class width is the width of each block of values(not frequencies).

In our example(above) height 71-75 gives a class width of **5** and NOT 4 (75-71). There are 5 numbers in the group 71 72 73 74 75 .

grouping using inequalities The values are grouped according to an inequality rule.

mid-interval values are useful in estimating the 'mean' of a set of grouped data. This is dealt with in detail in the topic 'mean, mode and median' [here](#).

Frequency Density

The frequency density is a the **frequency** of values **divided by** the **class width** of values.

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

Histograms

The area of each block/bar represents the total of frequencies for a particular class width.

The width of the block/bar (along the x-axis) relates to the **size** of the class width. So the width of a block/bar can vary within a histogram.

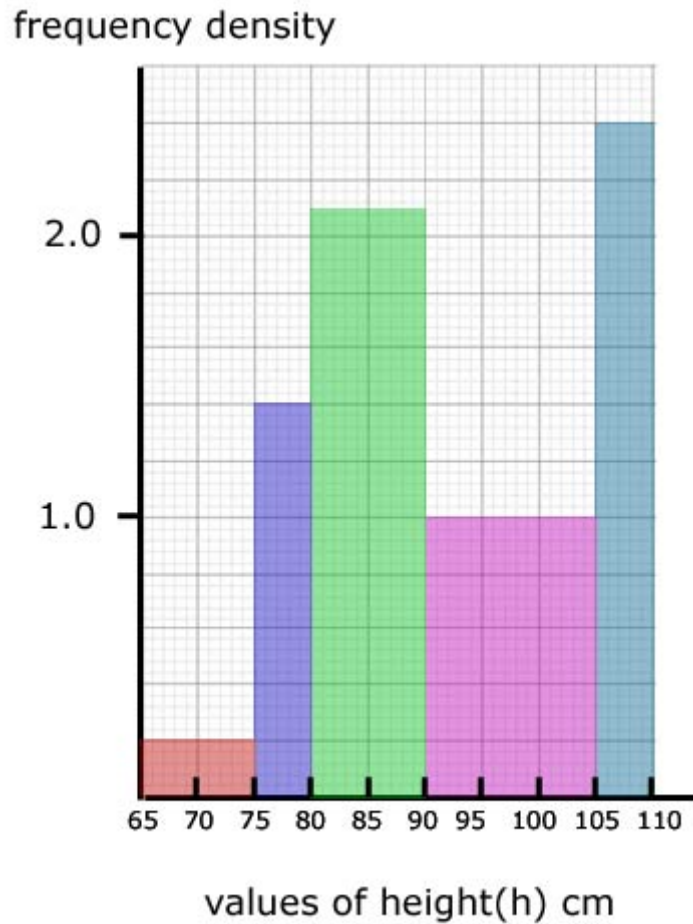
The frequency density is always the y-axis of a histogram.

Histograms are only used for numerical continuous data that is grouped.

Example Here is a table of data similar to the last one but with values of height grouped differently using inequalities.

note: because the class is grouped using inequalities, one 'equal to and greater' and the other 'less than' , the class width is a straight subtraction of the two numbers making up the class group.

class (height - h) cm	class width	frequency (f)	frequency density
$65 \leq h < 75$	10	2	$2/10 = 0.2$
$75 \leq h < 80$	5	7	$7/5 = 1.4$
$80 \leq h < 90$	10	21	$21/10 = 2.1$
$90 \leq h < 105$	15	15	$15/15 = 1.0$
$105 \leq h < 110$	5	12	$12/5 = 2.4$



Significance of area

The area on a histogram is important in being able to find the total number of values/individual results in the data.

In our histogram(from the table), the 65 to 75 block represents 2 children, the 75 to 80 block represents 7 children, and so on.

So **one block square** represents **one** child.

If we count the square blocks in the whole sample we get 57 - the sum of all the frequencies i.e. the total number of children taking part - the number of individual results.

Mean, Mode, Median & Range

The Mean

This is the average value for a set of single data values. It is calculated by adding up all the values and dividing by the number of values.

The Mean for grouped data

This is similar to some extent with the median for grouped data. No exact value can be found. It can only be estimated.

method:

find the mid-value for each group of data (call these values $m_1 \dots m_2 \dots m_3$ etc.)

in turn, multiply the mid-value for each group by its frequency ($f_1 m_1 \dots f_2 m_2 \dots$ etc.)

sum the (product of mid-value x frequency) $f_1 m_1 + f_2 m_2 + f_3 m_3 \dots$ etc.

divide the sum by the sum of all frequencies

$$\text{estimated mean} = \frac{\sum fm}{\sum f}$$

Example

The table represents the heights of pupils within a school.

Using the data calculate the mean value.

class	mid value	frequency	mid value x frequency
(height - h) cm	(m)	(f)	(fm)
$70 \leq h < 75$	72.5	2	145
$75 \leq h < 80$	77.5	7	542.5
$85 \leq h < 90$	87.5	21	1837.5
$95 \leq h < 100$	97.5	15	1462.5
$105 \leq h < 110$	107.5	12	1290
	total no. results	57	5277.5

$$\begin{aligned} \text{estimated mean} &= \frac{5277.5}{57} \\ &= 92.59 \\ &= 93 \text{ (rounded)} \end{aligned}$$

The Mode & Modal Class

For single values in a set of data, the mode is simply the value with the **highest frequency**.

For grouped values in a set of data, the modal class is simply that class/group with the highest frequency.

The Median

The median is the **middle value** of a set of data.

1__3__4__5__**9**__10__21__32__45

The median is therefore **9**

The Median for 'grouped values'

For grouped values finding the median is more difficult. It cannot be found exactly, but is estimated using '**interpolation**', which is essentially an educated guess.

The method is:

find which result is in the middle by dividing the total number of results in half. For example if there are 91 results , the median is result number 46(45.5 rounded up).

identify the modal class(the group of values/results) that contains the median

count how many numbers there are in this group

identify the frequency for the group

find the order of the median within the group

the median can be calculated as being the first number in the group **plus** a proportion of the number of results in the group

$$\text{median} = (\text{first no. in group}) + \left(\frac{\text{median position within group}}{\text{no. results in group (ie freq.)}} \right) \times (\text{no. numbers in group})$$

Example

The table represents the heights of pupils within a school.

Using the data calculate the median value.

class (height - h) cm	class width	frequency (f)
$70 \leq h < 75$	5	2
$75 \leq h < 80$	5	7
$85 \leq h < 90$	5	21
$95 \leq h < 100$	5	15
$105 \leq h < 110$	5	12
	total no. results	57

Since there are 57 values/results, the median occurs at position 28.5, ie. 29th to the nearest whole number.

This means that the median is the 29th value.

There are 2 values in the region $70 \leq h < 75$.

and a further 7 values in the region $75 \leq h < 80$.

So the median lies at the 20th value within the range $85 \leq h < 90$. This group has 21 values in it.

note: $(9 + 20 = 29)$

There are 5 numbers in the group.

The first number in the group is 85.

Using our equation:

$$\text{median} = (\text{first no. in group}) + \left(\frac{\text{median position within group}}{\text{no. results in group (ie freq.)}} \right) \times (\text{no. numbers in group})$$

$$\text{median} = (85) + \left(\frac{20}{21} \right) \times (5)$$

$$\text{median} = 85 + 4.76 = 89.76 \sim 90(\text{rounded})$$

Range of values

This is the spread of values, from the lowest to the highest.

Probability

The meaning of probability

Probability is the chance of a particular outcome happening.

Probability is a number between 0 and 1.

$$\text{Probability} = \frac{\text{no. of actual outcomes}}{\text{no. of possible outcomes}}$$

The probability of an event '**C**' occurring when the outcome is certain (ie there is no other outcome) is 1.

$$\text{written } P(\mathbf{C})=1$$

An event '**I**' which is impossible has a probability of 0.

$$\text{written } P(\mathbf{I})=0$$

The probability of an event '**N**' **not** happening = 1 - (probability of it happening '**H**')

$$\text{written } P(\mathbf{N})= 1 - P(\mathbf{H})$$

The sum of probabilities is always 1

$$P(\mathbf{N}) + P(\mathbf{H})= 1$$

Independent events

These are events/outcomes that are independent of each other. This means that one has no effect on the other (eg rolling two dice).

The **AND** rule is used to relate the probability of two independent events occurring.

If $P(M)$ is the probability of an event M happening and $P(N)$ for an event N . Then the probability of both events occurring is given by:

$$P(M + N) = P(M) \times P(N)$$

John picks a card at random from a pack of 52 cards and throws a dice. What is the probability that he will pick the ace of spades and throw a six?

$$\begin{aligned} P(\text{ace of spades} + \text{six}) &= P(\text{ace of spades}) \times P(\text{six}) \\ &= \left(\frac{1}{52}\right) \times \left(\frac{1}{6}\right) \\ &= \frac{1}{312} \\ &= \underline{0.003} \end{aligned}$$

Independent events can be tabulated against each other in a **possibility space** table.

Example #2

There are two spinners, one with four different colours and the other with the first five letters of the alphabet. If both are spun independently, the table below shows all the possible outcomes.

		number spinner				
		1	2	3	4	5
colour spinner	red	1,red	2,,red	3,red	4,red	5,red
	green	1,green	2,,green	3,green	4,green	5,green
	blue	1,,blue	2,blue	3,blue	4,blue	5,blue
	orange	1,,orange	2,orange	3,orange	4,orange	5,orange
	yellow	1,,yellow	2,yellow	3,yellow	4,yellow	5,yellow

There are $5 \times 5 (=25)$ possible outcomes.

The chances on getting blue **and** a number more than 3 are shown in blue. In other words, 2 chances in 25 (ie **0.08**)

This can be verified using the AND rule.

probability of getting blue is $P(\text{blue}) = 0.2$

probability of getting a number > 3 is $2/5 (0.4)$

probability of both events happening is $0.2 \times 0.4 = \mathbf{0.08}$

Mutually exclusive events

Events are said to be mutually exclusive if they cannot happen at the same time.

The **OR** rule is used to relate the probability of one independent event happening or another independent event happening at the same time.

The probability of event M or another event N, when the two are mutually exclusive, occurring is given by:

$$P(M \text{ or } N) = P(M) + P(N)$$

Example

What is the probability of picking two doubles from a pack of dominoes?

There are 50 dominoes in a pack.

7 of them are doubles.

The probability of picking one double is therefore $7/50$

The probability of picking a second double is $6/49$

(there are 49 dominoes now and 6 doubles)

Applying the AND rule:

$$\begin{aligned} \text{probability} &= \frac{7}{50} \times \frac{6}{49} \\ &= \frac{42}{2450} \\ &= \underline{0.017} \end{aligned}$$

Tree diagrams

A tree diagram is a way of visualising a probability problem.

rules:

show the event/question above each branch group

write the probabilities of the event/question close to each branch

write the outcome of each event/question at the end of each branch

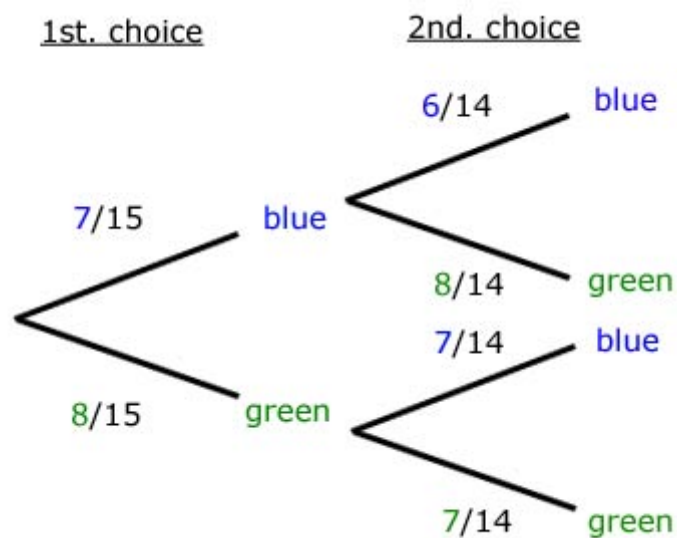
the sum of the probabilities on each branch add up to 1

Example

A black bag contains 15 marbles. 7 of the marbles are blue while the remainder(8) are green.

One marble is chosen from the bag. Then another is chosen, without the first marble being returned.

- i) what is the probability that both choices are blue?
- ii) what is the probability of choosing two marbles of the same colour(in any order)?
- iii) what is the probability of choosing two marbles of different colours(in any order)?



i)

$$\begin{aligned}
 P(\text{both choices are blue}) &= P(\text{first choice blue}) \times P(\text{second choice blue}) \\
 &= \frac{7}{15} \times \frac{6}{14} \\
 &= \frac{42}{210} \\
 &= \underline{0.024}
 \end{aligned}$$

ii)

$$\begin{aligned}
 P(\text{both choices are same}) &= P(\text{1st \& 2nd choice blue}) \text{ or } P(\text{1st \& 2nd choice green}) \\
 &= \left(\frac{7}{15} \times \frac{6}{14} \right) + \left(\frac{8}{15} \times \frac{7}{14} \right) \\
 &= \frac{42}{210} + \frac{56}{210} \\
 &= \frac{98}{210} \\
 &= \underline{0.467}
 \end{aligned}$$

iii)

$$\begin{aligned}
 P(\text{one blue + one green}) &= P(\text{1st blue \& 2nd green}) \text{ or } P(\text{1st green \& 2nd blue}) \\
 &= \left(\frac{7}{15} \times \frac{8}{14} \right) + \left(\frac{8}{15} \times \frac{7}{14} \right) \\
 &= \frac{56}{210} + \frac{56}{210} \\
 &= \frac{112}{210} \\
 &= \underline{0.533}
 \end{aligned}$$

note: probability same colour + probability different colour = 1

notes

This book is under copyright to GCSE Maths Tutor. However, it may be distributed freely provided it is not sold for profit.