

Write your name here

Surname

Other names

Pearson
Edexcel GCSE

Centre Number

--	--	--	--	--

Candidate Number

--	--	--	--	--

Statistics
Paper 1H

Higher Tier

Thursday 21 June 2018 – Morning
Time: 2 hours

Paper Reference
5ST1H/01

You must have:

Ruler graduated in centimetres and millimetres, protractor, pen,
HB pencil, eraser, electronic calculator.

Total Marks

--

Instructions

- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions.
- Answer the questions in the spaces provided
– *there may be more space than you need.*

Information

- The total mark for this paper is 100.
- The marks for **each** question are shown in brackets
– *use this as a guide as to how much time to spend on each question.*
- Questions labelled with an **asterisk** (*) are ones where the quality of your written communication will be assessed
– *you should take particular care on these questions with your spelling, punctuation and grammar, as well as the clarity of expression.*

Advice

- Read each question carefully before you start to answer it.
- Keep an eye on the time.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ►

P56825A

©2018 Pearson Education Ltd.

1/6/6/6/8/6/1/



Pearson

Higher Tier Formulae

**You must not write on this page.
Anything you write on this page will gain NO credit.**

Mean of a frequency distribution $= \frac{\sum fx}{\sum f}$

Mean of a grouped frequency distribution $= \frac{\sum fx}{\sum f}$, where x is the mid-interval value.

Variance $= \frac{\sum (x - \bar{x})^2}{n}$

Standard deviation (set of numbers) $\sqrt{\left[\frac{\sum x^2}{n} - \left(\frac{\sum x}{n} \right)^2 \right]}$

or $\sqrt{\left[\frac{\sum (x - \bar{x})^2}{n} \right]}$

where \bar{x} is the mean set of values.

Standard deviation (discrete frequency distribution) $\sqrt{\left[\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f} \right)^2 \right]}$

or $\sqrt{\left[\frac{\sum f(x - \bar{x})^2}{\sum f} \right]}$

Spearman's Rank Correlation Coefficient $1 - \frac{6 \sum d^2}{n(n^2 - 1)}$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Answer ALL the questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 Suha is carrying out market research in the town centre.

She has been asked to interview a total of 60 people of different ages and genders, as shown in the table.

	Age 18–30 years	Age 31–55 years	Age 56 years and over
Male	10	10	10
Female	10	10	10

(a) Write down the statistical name of this sampling method.

.....
(1)

(b) Give one advantage and one disadvantage of this sampling method.

Advantage.....
.....

Disadvantage.....
.....
(2)

(Total for Question 1 is 3 marks)



2 Supul is investigating how long pupils in Year 10 in his school spent on homework.

He asked each pupil to record the time taken, to the nearest minute, to do their homework one night.

*(a) Describe the type of data the pupils recorded.

.....
..... (2)

Supul collected each pupil's recorded time.

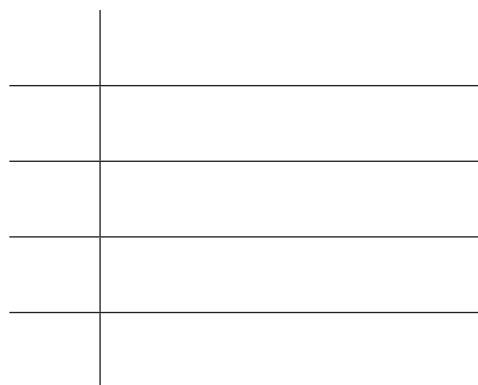
*(b) Discuss how reliable the data might be.

.....
.....
..... (2)

Supul selected a sample of 20 of the pupils.
Here are their recorded times.

55	53	35	31	21	47	64	53	23	37
50	32	58	51	40	45	63	33	41	60

(c) Complete an ordered stem and leaf diagram for these times.
You must include a key.



(3)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(d) Find the median time.

..... minutes
(1)

The mean time is 44.6 minutes.

(e) How many of the 20 pupils took more than the mean time to do their homework?

.....
(1)

The times were recorded to the nearest minute.

(f) Find the **maximum** possible range for the times.

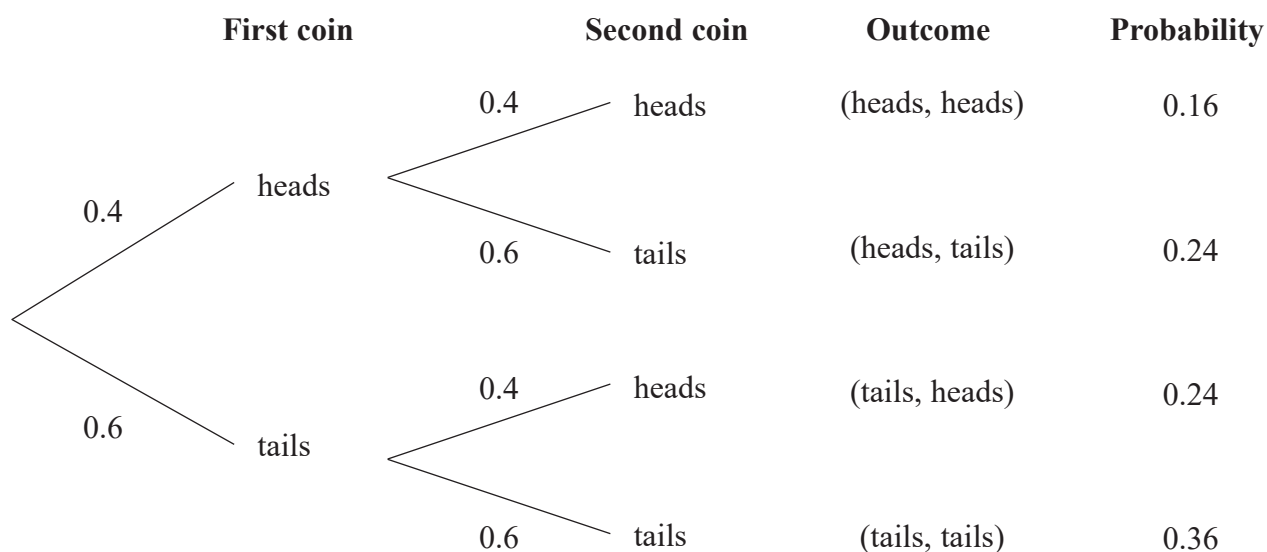
..... minutes
(2)

(Total for Question 2 is 11 marks)



3 Simon spins two identical biased coins.

He drew a tree diagram to help identify the possible outcomes.



(a) Show why the probability of the outcome (heads, heads) is 0.16

(1)

Simon says that the likelihood of getting one head and one tail is nearly evens.

(b) Show why Simon is correct.

(2)

Simon spins the two coins 100 times.

The outcome (tails, tails) happened 25 times.

(c) How does this compare with the expected outcome?

(2)

(Total for Question 3 is 5 marks)

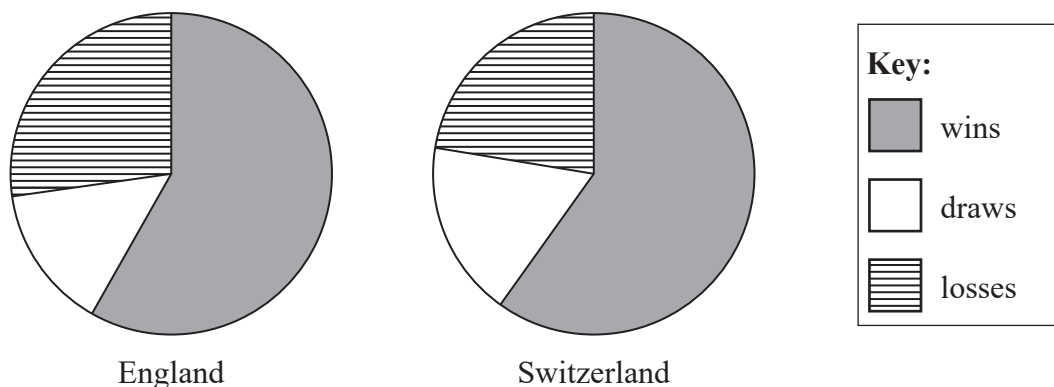


DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

4 The pie charts show information about the proportion of wins, draws and losses in the matches played by the England football team and by the Switzerland football team for the years 2012 to 2015



Source: www.worldfootball.net

(a) Compare the proportion of losses for the two football teams.

(1)

England played a total of 48 matches.

The angle representing wins in the England pie chart is 210°

(b) Work out how many of these matches were wins.

(2)

Both teams had the same **number** of draws.

(c) Which team played the greater total number of matches?
Give a reason for your answer.

(2)

(Total for Question 4 is 5 marks)



- 5 The table gives information about the numbers of overseas visits by UK residents for the years 2010 to 2014

It shows the type of travel (Air, Sea, or Tunnel) and the reason for the visit.

	Number of overseas visits (thousands)					Average annual growth (%)
	2010	2011	2012	2013	2014	
Air						
Holiday	28 420	28 827	28 573	29 676	31 025	2.2
Business	5 156	5 590	5 569	5 332	5 359	1.0
Visiting friends or relatives	8 799	9 474	9 842	10 151	11 061	5.9
Miscellaneous	865	832	932	866	893	0.8
All visits	43 239	44 723	44 916	46 025	48 337	2.8
Sea						
Holiday	5 435	5 288	4 550	4 526	4 580	-4.2
Business	646	536	562	588	554	-3.8
Visiting friends or relatives	1 435	1 513	1 221	1 354	1 411	-0.4
Miscellaneous	540	521	423	454	464	-3.8
All visits	8 056	7 857	6 755	6 922	7 009	-3.4
Tunnel						
Holiday	2 567	2 705	3 050	2 948	2 915	3.2
Business	837	720	826	830	842	0.1
Visiting friends or relatives	616	608	735	807	800	6.8
Miscellaneous	247	223	257	260	179	-7.8
All visits	4 267	4 255	4 867	4 845	4 735	2.6
Total						
Holiday	36 422	36 819	36 173	37 149	38 519	1.4
Business	6 639	6 846	6 956	6 750	6 756	0.4
Visiting friends or relatives	10 850	11 594	11 797	12 313	13 272	5.2
Miscellaneous	1 652	1 576	1 612	1 580	1 535	-1.8
All visits	55 562	56 836	56 538	57 792	60 082	2.0

Source: ons.gov.uk

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



(a) Find the number of overseas visits made by 'Tunnel' to visit friends or relatives in 2013

..... thousands
(1)

(b) For 'All visits' in 2013, work out how many more people travelled by air than travelled by sea.

..... thousands
(2)

(c) For 2011, work out the percentage of all visits made by air which were for business. Give your answer to 1 decimal place.

..... %
(2)

Angelos makes two conclusions using the data for 2014

- 1. Holiday is the most common reason for travel
- 2. The most common type of travel is by air

(d) Explain how the data in the table supports each of these conclusions.

1.....
.....
.....

2.....
.....
.....
(2)

(Total for Question 5 is 7 marks)



- 6 The table gives information about the race times of the 78 competitors in the first round of the Women's 100m Freestyle event at the British Swimming Championships 2015

Race time (t seconds)	Frequency (f)	
$52 \leq t < 54$	1	
$54 \leq t < 56$	8	
$56 \leq t < 58$	22	
$58 \leq t < 60$	29	
$60 \leq t < 62$	18	

Source: www.swimmingresults.org.uk

- (a) Find an estimate of the mean race time.
Give your answer to 2 decimal places.

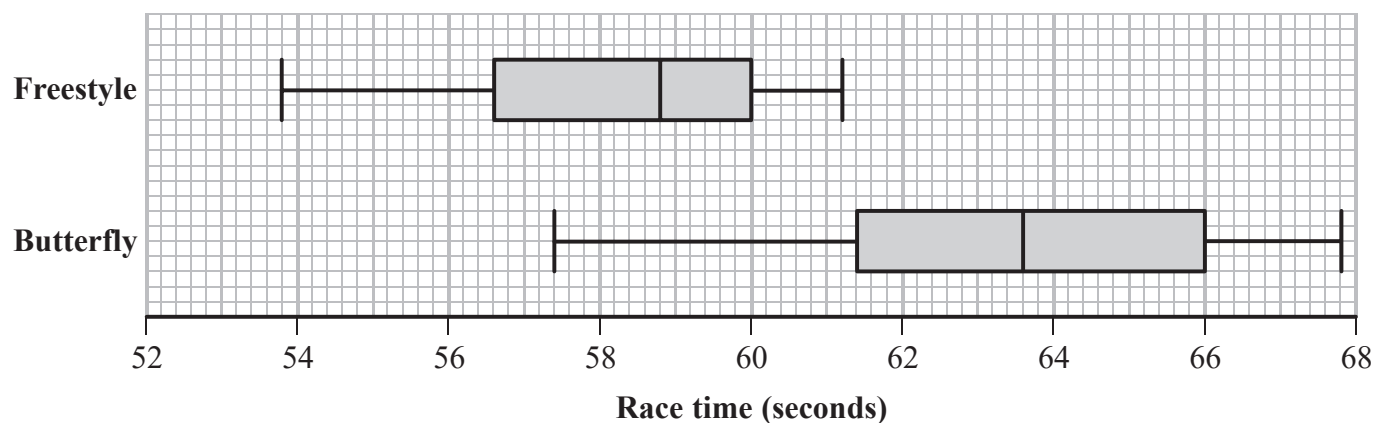
.....seconds
(3)

The true value for the mean race time is 58.37 seconds.

- (b) Explain why this is different from your estimate in part (a).

.....
.....
(1)

The comparative box plots show information about the race times in the first rounds of the Women's 100m Freestyle event and the Women's 100m Butterfly event.



Source: www.swimmingresults.org.uk



(c) (i) Describe the skew shown by the box plot for the Freestyle event.

.....

(ii) Explain how the mean race time for the Freestyle event confirms your answer to part (i).

.....

.....

(2)

(d) Calculate and compare the interquartile ranges.
Interpret your comparison.

.....

.....

.....

(4)

*(e) Given a competitor's race time for only one of the two events, explain how the box plots could be used to help identify which of the two events the time was for.

.....

.....

.....

.....

.....

.....

.....

.....

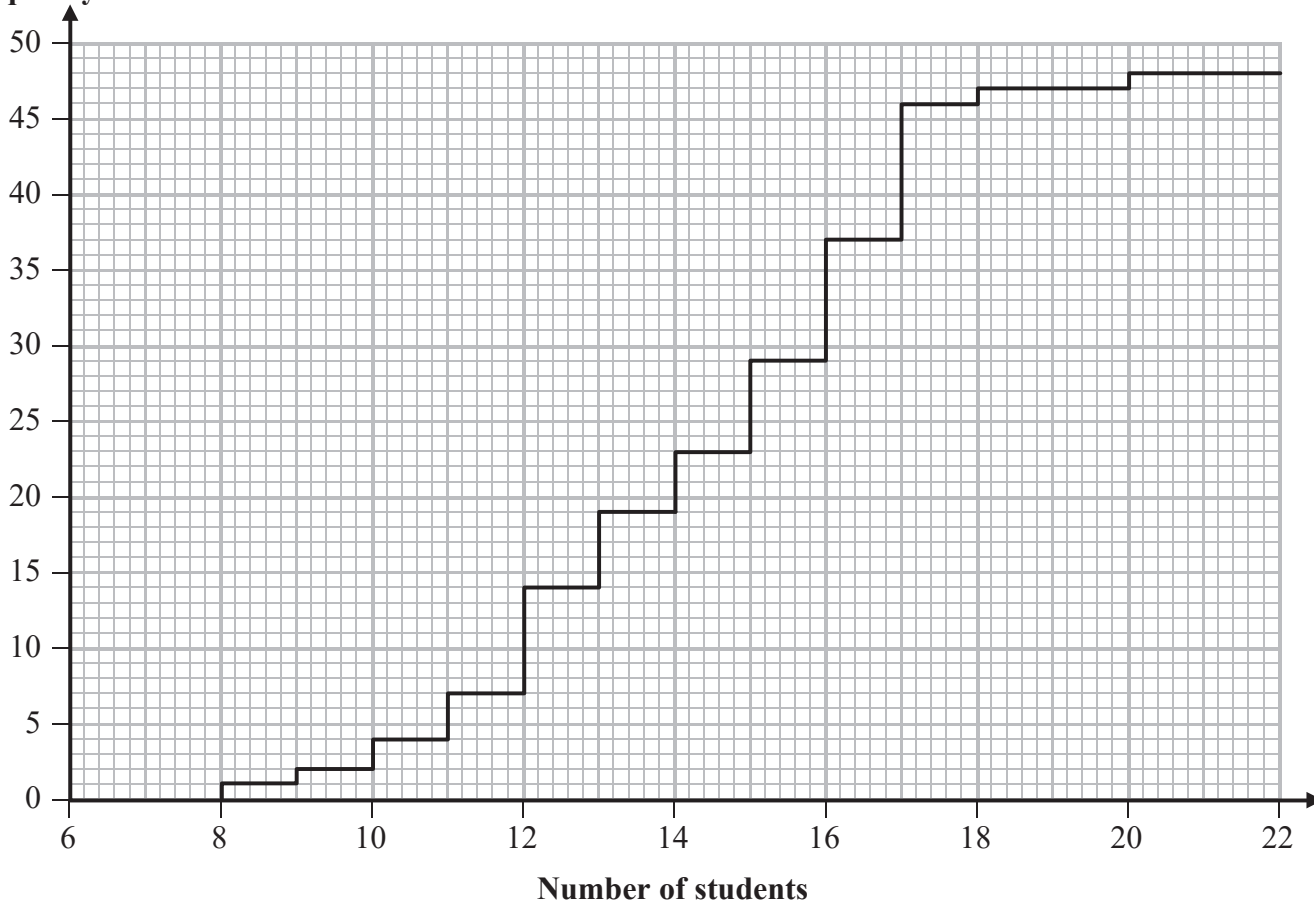
(3)

(Total for Question 6 is 13 marks)



7 The cumulative frequency step polygon shows information about the numbers of students in the 48 teaching sets in Year 12 at Pearson Academy.

Cumulative frequency



(a) For the numbers of students in the teaching sets, find

(i) the median,

.....

(ii) the interquartile range,

.....

(iii) the mode.

.....

(4)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(b) Write down the number of students in the largest teaching set.

.....
(1)

(c) Find the number of teaching sets which have more than 16 students.

.....
(2)

(Total for Question 7 is 7 marks)



P 5 6 8 2 5 A 0 1 3 2 4

8 Paul is investigating dog and cat ownership in the UK.

The table gives the numbers of dogs owned and numbers of cats owned, in millions, for different years.

Year	1975	1980	1985	1990	1994	1997	2000	2002
Dogs owned (millions)	5.7	5.6	6.3	7.4	6.9	6.6	6.5	6.1
Cats owned (millions)	4.5	4.9	6.1	6.8	7.2	7.7	8.0	7.5

Source: pmfa.org.uk

- (a) Suggest a suitable diagram to show the changes in dog ownership and cat ownership between 1975 and 2002

(1)

The difference between the numbers of cats owned and dogs owned for each year changed over these years.

- (b) Find the biggest difference shown in the table.

..... millions
(1)

Paul uses the information in the table to predict the numbers of dogs owned and cats owned in 2018

- (c) Explain whether or not this is a suitable thing to do.

(2)

- (d) What could Paul do to improve his prediction?

Write down a problem he might have.

(2)

(Total for Question 8 is 6 marks)

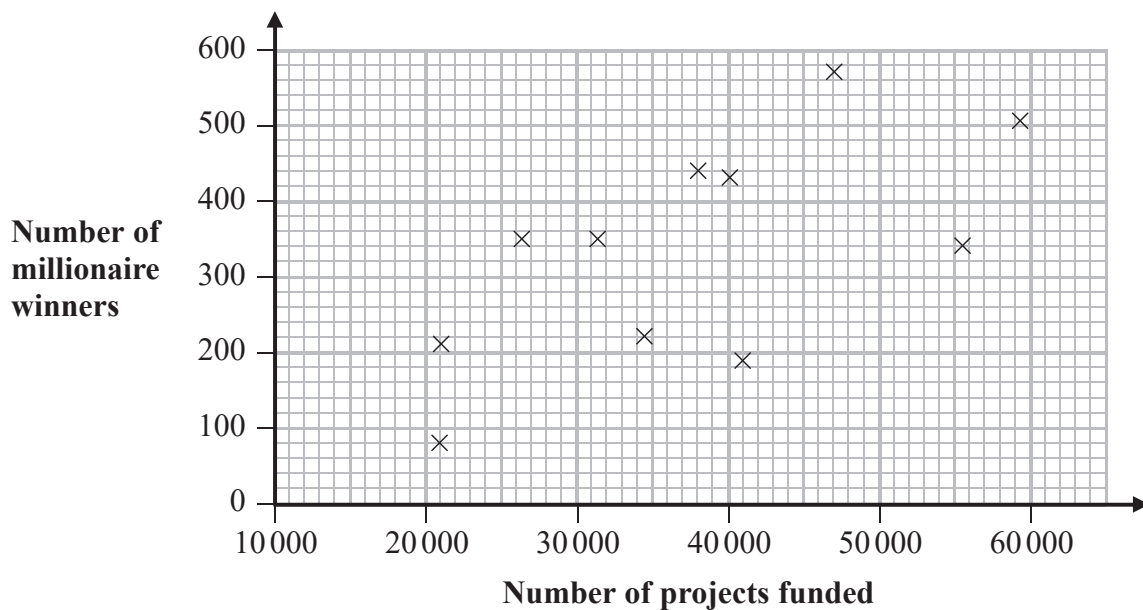


DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

9 The scatter diagram shows information about the number of projects funded by the National Lottery and the number of millionaire winners of the National Lottery, for 11 different regions of the UK up until 2015



Source: The National Lottery

(a) Describe and interpret any correlation shown by the scatter diagram.

.....

.....

.....

(2)

Dale is going to work out Spearman's rank correlation coefficient for the data in the scatter diagram.

He finds that $\sum d^2 = 96$

(b) Determine whether or not Spearman's rank correlation coefficient for the data supports your answer to part (a). Explain your answer.

.....

.....

(3)

(Total for Question 9 is 5 marks)



10 A librarian wants to investigate, for books in her library, if there is a relationship between how old a book is and for how long it is borrowed.

(a) Suggest a hypothesis that the librarian could use.

(1)

(b) Write down the two variables the librarian should use for her investigation.

and

(1)

(c) For the librarian's investigation write down

(i) the population,

(ii) a suitable sampling frame.

(2)

The table shows the number of books of each type in the library.

Book type	Fiction	Non-fiction	Children's
Number of books	23 000	12 000	8000

A stratified sample of 60 books is taken.

(d) (i) Explain why a stratified sample is appropriate in this case.

(ii) Work out how many children's books should be in the sample.

(3)



*(e) Describe how to take a stratified sample of 60 books from the library.
Do **not** do any further calculations.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(3)

(Total for Question 10 is 10 marks)



11 A restaurant offers a meat dish, a fish dish and a vegetarian dish each day.

The table shows the percentages of customers who, in the past, have ordered each dish.

Meat	Fish	Vegetarian
50%	30%	20%

On Saturday, the restaurant manager expects 80 customers.

He is going to use random numbers to simulate the ordering of the meat dish, the fish dish and the vegetarian dish.

The table below shows the numbers he intends to use to represent each dish.

Dish	Meat	Fish	Vegetarian
Numbers	0, 1, 2, 3, 4	5, 6, 7	8, 9

(a) Comment on whether or not his choice of numbers is sensible.

(1)

The frequency table below shows the first 70 results from the simulation.

Dish	Tally	Frequency								
Meat	<table style="display: inline-table; border: none;"> <tr><td> </td><td> </td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td> </td><td></td></tr> </table>									
Fish	<table style="display: inline-table; border: none;"> <tr><td> </td><td> </td><td> </td></tr> <tr><td> </td><td> </td><td></td></tr> </table>									
Vegetarian	<table style="display: inline-table; border: none;"> <tr><td> </td><td> </td></tr> <tr><td> </td><td></td></tr> </table>									

The final ten random numbers are

0 9 1 1 7 0 3 4 7 2

(b) Using these numbers, complete the frequency table.

(2)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

*(c) Explain why the manager needs to repeat the simulation many more times.

.....

.....

.....

.....

.....

(2)

(Total for Question 11 is 5 marks)



P 5 6 8 2 5 A 0 1 9 2 4

- 12 The table gives information about the number of patients admitted to hospital in England in 2014 for different age groups.

Age (x years)	Number of patients admitted (10 000s)
$0 \leq x < 5$	190
$5 \leq x < 15$	200
$15 \leq x < 25$	280
$25 \leq x < 45$	500
$45 \leq x < 65$	360
$65 \leq x < 100$	385

Source: Adapted from Health & Social Care Information Centre

- (a) Show that the class interval $0 \leq x < 5$ has a frequency density of 38

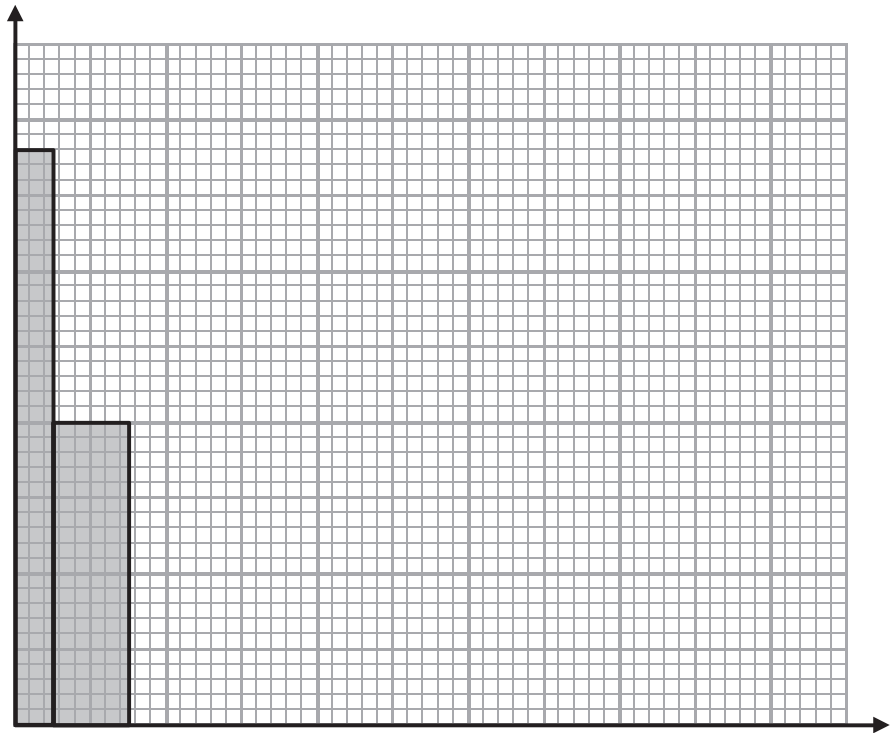
(1)



(b) On the grid below, complete the histogram for the information in the table.

The bars for the first two class intervals have already been drawn.

Label the axes.



(4)

(c) Find an estimate for the number of patients admitted to hospital who were aged 60 or over.

(2)

(Total for Question 12 is 7 marks)



13 In 2010, a factory had the following monthly running costs.

Category	Staff wages	Premises	Raw material
Monthly running cost (£)	52 000	8000	20 000

*(a) Show why appropriate weightings for the running cost categories are

Category	Staff wages	Premises	Raw material
Weighting	65	10	25

(2)

Using 2010 as the base year, the index numbers for the running cost categories in 2015 were

Category	Staff wages	Premises	Raw material
Index number	107.2	114.6	112.0

(b) Find the monthly cost of staff wages in 2015

£.....

(2)

The factory manager says that the total monthly running costs have risen by less than 10% since 2010

(c) By finding the weighted index number for the total running costs of the factory in 2015, show that she is correct.

(3)

(Total for Question 13 is 7 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

14 The list below gives three probability distributions which could be used to model real life data.

- A Discrete uniform distribution
- B Binomial distribution
- C Normal distribution

From this list, choose the best distribution which could be used to model each of the variables below.

You must give a reason for each answer.

(i) The time taken to travel the same journey to work each day.

Distribution.....

Reason.....

.....

(ii) The next digit selected from a random number table.

Distribution.....

Reason.....

.....

(Total for Question 14 is 4 marks)



15 Koharu has a bag containing 2 red balls, 3 white balls and 4 blue balls only.

She takes out two balls at random, **without** replacing them.

Find the probability that the two balls are

(a) both red,

.....
(2)

(b) different colours.

.....
(3)

(Total for Question 15 is 5 marks)

TOTAL FOR PAPER IS 100 MARKS

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

