

Please check the examination details below before entering your candidate information

Candidate surname

Other names

Centre Number

Candidate Number

Pearson Edexcel
Level 1/Level 2 GCSE (9–1)

--	--	--	--	--

--	--	--	--	--

Specimen Paper (Set 2)

(Time: 1 hour 30 minutes)

Paper Reference **1ST0/2H**

Statistics

Paper 2
Higher Tier

You must have:

Ruler graduated in centimetres and millimetres, protractor,
pair of compasses, pen, HB pencil, eraser, scientific 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.*
- Scientific calculators may be used.
- You must **show all your working out** with **your answer clearly identified** at the **end of your solution**.



Information

- The total mark for this paper is 80
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

Advice

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

Turn over ►

S61447A

©2018 Pearson Education Ltd.

1/1/1




Pearson

Higher Tier Formulae

You must not write on this page.

Anything you write on this page will gain NO credit.

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n} \sum (x - \bar{x})^2}$$

An alternative formula for standard deviation is

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Rates of change (e.g. Crude birth rate = $\frac{\text{number of births} \times 1000}{\text{total population}}$)

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 questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 Chirag wants to investigate patient satisfaction at the hospital where he works.

Here are the first three questions on the questionnaire he has produced.

I would like to find out if you were satisfied with the service provided by certain departments in this hospital during your visit. Please complete this questionnaire.

1. Do you agree that service in this hospital is very good? Please tick (✓)

Yes No Don't know

2. How often do you come into hospital? Please tick (✓)

0 – 1 times 2 – 3 times 4 or more times

3. In which department of the hospital did you receive treatment? Please tick (✓)

A & E Maternity Physiotherapy Plastic surgery Other

None of these questions is appropriate.

In each case, explain why.

1

2

3

(Total for Question 1 is 3 marks)



S 6 1 4 4 7 A 0 3 2 0

- 2 The table shows information about the number of car journeys per person in a year that are a distance of between 0 and 50 miles.

The information is based on a sample of 382 people from the 2016 National Travel Survey.

Distance (x miles)	$0 < x \leq 1$	$1 < x \leq 2$	$2 < x \leq 5$	$5 < x \leq 10$	$10 < x \leq 25$	$25 < x \leq 50$
Frequency	24	65	131	83	62	17

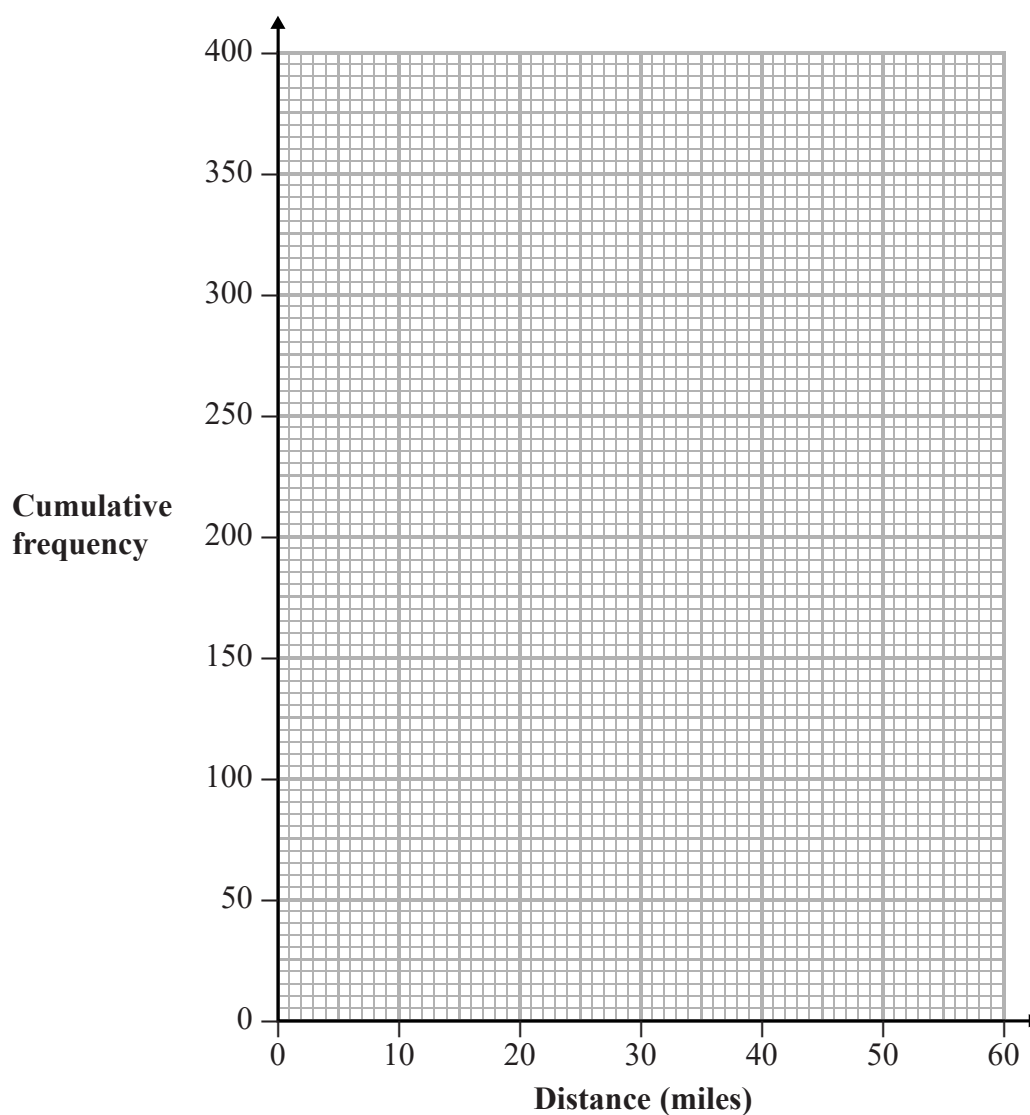
(Source: www.gov.uk)

- (a) Complete the cumulative frequency table for the information above.

Distance (x miles)	$0 < x \leq 1$	$0 < x \leq 2$	$0 < x \leq 5$	$0 < x \leq 10$	$0 < x \leq 25$	$0 < x \leq 50$
Cumulative frequency	24					

(1)

- (b) Draw a cumulative frequency diagram for this information.



(3)



In the same survey, the median distance travelled by train for a sample of people was 11.3 miles.

Hamish says,

“The information collected in the survey shows that people in these samples travel on average further by train than by car.”

(c) Assess whether or not Hamish’s conclusion is appropriate.

(2)

Hamish recorded the number of passengers travelling on 250 buses.

Information about his results is shown in the cumulative frequency table below.

Number of passengers	35	36	37	38	39	40	41	42	43	44	45
Cumulative frequency	5	12	24	39	65	98	138	176	207	234	250

Hamish plans to draw a cumulative frequency step polygon for his results rather than the type of cumulative frequency diagram drawn in part (b).

(d) Explain why Hamish’s plan is appropriate.

(1)

Hamish wants to take a sample of the bus passengers.

Hamish plans to take a sample of 50 men and 50 women.

(e) Name this sampling method and explain why this might be an appropriate method to use.

(2)

(Total for Question 2 is 9 marks)



- 3 The back-to-back stem and leaf diagram gives information about the ages of a random sample of members of parliament in Canada and in the UK.

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

Key:

2|3|5 represents an age of 32 for a member of parliament in Canada and an age of 35 for a member of parliament in the UK

(Source: *en.wikipedia.org*)

- (a) Give a reason to support the use of a back-to-back stem and leaf diagram to represent this information.

(1)

Some information about the quartiles of these two distributions is given in the table below.

	Canada	UK
lower quartile	44	b
median	a	54
upper quartile	61	c

- (b) Find the value of a , the value of b and the value of c

$a =$

$b =$

$c =$

(3)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(c) Write down the proportion of members of parliament in the UK that are likely to be older than 54 years old.

Give a reason for your answer.

(1)

(d) Compare the spread of ages for members of parliament in Canada with the spread of ages for members of parliament in the UK.

State clearly the values of the statistic you use to make your comparison.

Interpret your comparison.

(3)

One member of parliament in the UK wants to investigate the ages of the people living in her constituency.

She suggests using the electoral register as a sample frame for her investigation.

(e) State one use of a sample frame in an investigation.

(1)

(f) Assess the suitability of using the electoral register as a sample frame for this investigation.

(2)

(Total for Question 3 is 11 marks)



- 4 The World Happiness Report 2017 is based on results obtained by surveying a sample of 1000 people in each of 155 countries.

In the report each country is given a happiness score based on the answers the sampled people give to a series of questions.

The table shows the happiness scores for the top 8 countries and the Gross Domestic Product (GDP) per capita of these countries.

For the sample data, the highest happiness score (Happiness rank 1) represents the country with the happiest people and the highest GDP represents the wealthiest country.

Country	GDP per capita (US Dollars)	Happiness score	Happiness rank
Norway	70 392	7.537	1
Denmark	53 744	7.522	2
Iceland	59 629	7.504	3
Switzerland	79 242	7.494	4
Finland	43 169	7.469	5
Netherlands	45 283	7.377	6
Canada	42 210	7.316	7
New Zealand	38 345	7.314	8

(Source: *www.worldhappiness.report*)

- (a) Calculate Spearman's rank correlation coefficient for these data.
Give your answer correct to 2 decimal places.

.....
(4)



(b) Describe and interpret the correlation found in part (a).

.....

.....

.....

(2)

George wants to compare the happiness scores for these top 8 countries with average life expectancies for those countries.

He works out Spearman's rank correlation coefficient for these data and gets a result of 1.3

(c) Explain how you know that this result is not correct.

.....

.....

(1)

George also compared the happiness scores for these countries with two other variables.

George calculated Spearman's rank correlation coefficient in each case.

His values are -0.8 and -0.5

(d) Compare these two correlation coefficients.

Give **two** comparisons.

.....

.....

.....

.....

(2)

(Total for Question 4 is 9 marks)



- 5 Laura owns a company that makes and sells packets of crisps.

Laura wants to find out how sales of different flavours of crisps have changed over the past five years.

Here is part of a spreadsheet showing information about the percentage of total sales for each flavour sold by the company for each year from 2013 to 2017

Flavour	Percentage of total sales for each year				
	2013	2014	2015	2016	2017
Salt and vinegar	14	15.5	15	15.8	16
Prawn cocktail	6.9	7.7	8.1	8	8.2
Beef	6	6.8	6.5	6	6.1
Chicken	9.4	10.2	10	10.5	10.6
Cheese and onion	176	17.5	17.1	17	17.3
Ready salted	16	15.1	15.3	14.9	14.2
Pickled onion	5	4.3	4.1	3.9	3.7
Cheese	13.1	12.2	12.9	12.5	13
Smoky bacon	12	10.7	11	11.4	10.9

- (a) Explain one way that Laura can use the spreadsheet to check whether any of the data needs to be cleaned.

(1)

- (b) Circle the cell in the spreadsheet that needs to be cleaned.
Write down what you think the correct value should be.

.....%

(1)

Laura thinks that the company should stop making one of their flavours of crisps.

Laura notes that there is a downward trend in the percentage of total sales of ready salted crisps. She also notes that there is a decrease of 1.8% between 2013 and 2017 and that this is greater than the decrease for any other flavour.

One of Laura's employees, Ben, concludes that this means that it might be a good idea to stop making ready salted crisps.

- (c) Assess the validity of Ben's conclusion.

(2)



- 6 The two tables show information about the race times of the 57 wheelchair competitors in the 2017 London Marathon.

The tables show information about the same data but the tables use different classes.

Table 1

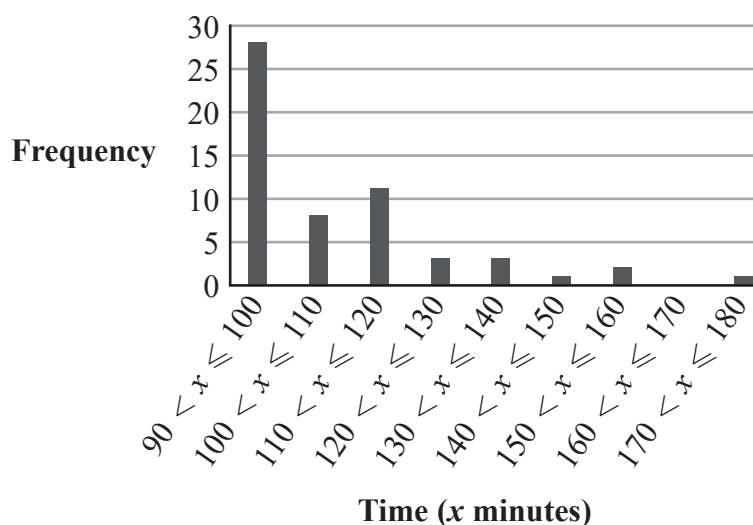
Time (x minutes)	Frequency
$90 < x \leq 100$	28
$100 < x \leq 110$	8
$110 < x \leq 120$	11
$120 < x \leq 130$	3
$130 < x \leq 140$	3
$140 < x \leq 150$	1
$150 < x \leq 160$	2
$160 < x \leq 170$	0
$170 < x \leq 180$	1

Table 2

Time (x minutes)	Frequency
$90 < x \leq 95$	22
$95 < x \leq 100$	6
$100 < x \leq 110$	8
$110 < x \leq 120$	11
$120 < x \leq 140$	6
$140 < x \leq 180$	4

(Source: www.virginmoneylondonmarathon.com)

Susan uses statistical software to produce this bar chart to represent the data in **Table 1**



Susan decides to draw a histogram for the data in **Table 2**

- (a) Explain whether the bar chart or a histogram is the most appropriate way to represent the data. Give reasons for your answer.

.....

.....

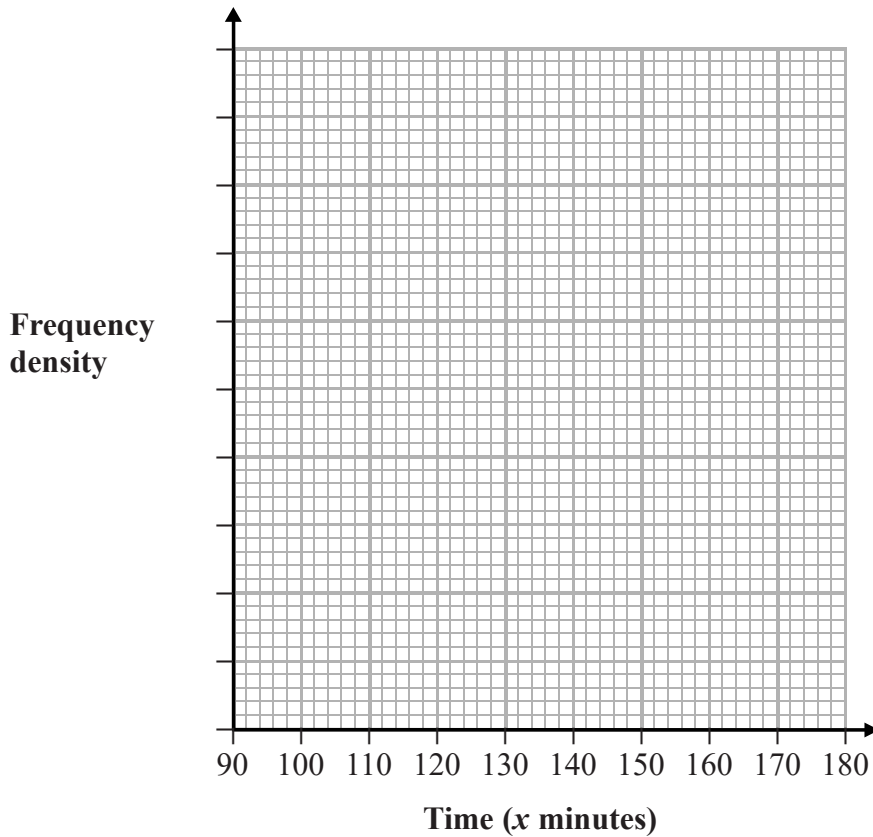
.....

(2)



(b) On the grid below, draw a histogram for the data in **Table 2**

Time (x minutes)	Frequency
$90 < x \leq 95$	22
$95 < x \leq 100$	6
$100 < x \leq 110$	8
$110 < x \leq 120$	11
$120 < x \leq 140$	6
$140 < x \leq 180$	4



(4)

Susan says that an estimate for the number of wheelchair competitors with a race time of between 105 minutes and 150 minutes is 22

(c) Show that Susan is correct.

(2)

(Total for Question 6 is 8 marks)



7 Sanjit writes this hypothesis.

At my school, younger boys are faster at running 400 metres than older girls.

He decides to collect some data from the students at his school to see if his hypothesis is correct.

He plans to ask each student their name, their gender, their best time to run 400 m and their age.

One of these variables is an extraneous variable.

(a) Which one?

.....
(1)

Sanjit decides to take a stratified sample of the students at his school.

He decides to stratify by gender.

(b) By considering Sanjit's hypothesis, explain whether or not this is the best way to stratify.

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



The table gives the mean, the standard deviation and the median of the best times taken by all the children in Sanjit's school to run 400 m.

Mean (seconds)	Standard deviation (seconds)	Median (seconds)
120	25	124

Sanjit carries out this calculation using the information in the table.

$$\frac{3 \times (120 - 124)}{25} = -0.48$$

- (c) Describe what conclusion you can make about the spread of values less than the median compared with the spread of the values greater than the median.
Give a reason for your answer.

.....

.....

.....

.....

(2)

(Total for Question 7 is 5 marks)



8 Nina suggests that each of the following problems can be answered by using a binomial distribution $B(n, p)$.

Problem 1

A manufacturer claims that, in his boxes of balloons, 30% are red, 35% are blue, 20% are green and 10% are yellow.

Find the probability that when 3 boxes are chosen at random exactly 1 box contains no yellow balloons.

Problem 2

Out of 64 matches that Roger Federer has played at Wimbledon, 6 of the matches were five-set matches.

(Source: www.wimbledon.com)

Find the probability that, out of the next 8 matches that Roger Federer plays at Wimbledon, exactly 3 of them are five-set matches.

(a) Assess the suitability of using a binomial distribution model to answer each of the problems.

If a binomial model is suitable, you should consider any assumptions that must be made and give possible values of n and p for the model.

If a binomial model is not suitable, you should explain why.

Problem 1

.....

.....

.....

.....

.....

.....

Problem 2

.....

.....

.....

.....

.....

.....

(6)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

The probability that a particular machine has a fault is 7%.

A random sample of 6 of these machines are taken and each machine is checked.

- (b) Work out the probability that at least 2 of these machines have the fault.
Give your answer correct to 3 decimal places.

.....
(3)

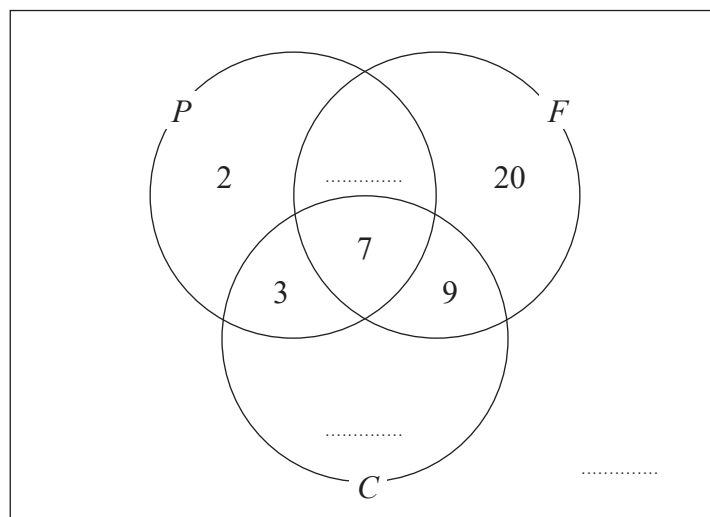
(Total for Question 8 is 9 marks)



- 9 Sam wants to investigate what musical instruments people play.

He asked a sample of 100 people whether they play any or none of the piano (P), the flute (F) or the clarinet (C).

The incomplete Venn diagram shows some information about his results.



- (a) Explain fully what the number 3 represents in the Venn diagram.

(1)

Of the 100 people Sam asked,

37 played the piano and the flute
31 played the clarinet.

- (b) Complete the Venn diagram.

(2)

One of the 100 people Sam asked is chosen at random.

Given that this person played at least two of the instruments,

- (c) find the probability that this person played all three of the instruments.

(2)

(Total for Question 9 is 5 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

10 The table gives information about the change in the cost of a second class stamp from 2011 to 2017

Year	2011	2013	2015	2017
Cost of second class stamp (pence)	36	50	54	56
Chain base index number		138.9	108	

(Source: www.2ndclassstamp.co.uk)

By working out the geometric mean of 3 appropriate chain base index numbers, what can be deduced about the average two-yearly change in the cost of a second class stamp from 2011 to 2017?

.....

.....

.....

.....

(Total for Question 10 is 6 marks)



11 Amelia collects data about the time it takes people to complete a puzzle.

Amelia calculates the lower quartile (LQ), the upper quartile (UQ) and the interquartile range (IQR) of the data.

The table shows information, in seconds, about the data collected by Amelia.

Three greatest times taken	LQ	Value of $LQ - 1.5 \times IQR$	Value of $UQ + 1.5 \times IQR$
42, 58, 63	20	- 4	60

Debbie says,

“The information in the table shows that there is only one outlier in the data collected by Amelia.”

- (a) Explain whether or not Debbie is correct.
You must give reasons for your answer.

.....

.....

.....

.....

(2)

- (b) Work out the value of the upper quartile for the data collected by Amelia.

.....seconds

(2)

(Total for Question 11 is 4 marks)

TOTAL FOR PAPER IS 80 MARKS

