Paper Reference 1ST0/1H
Pearson Edexcel
Level 1/Level 2 GCSE (9–1)

Total Marks

Statistics

PAPER 1
Higher Tier
(Calculator)

Time: 1 hour 30 minutes

In the boxes below, write your name, centre number and candidate number.

Surname					
Other names					
Centre Number					
Candidate Number					

Q66367A

Pearson

YOU MUST HAVE

Ruler, protractor, compasses, writing and drawing equipment, scientific calculator.

YOU WILL BE GIVEN

Data Booklet
Formulae Pages

INSTRUCTIONS

Answer ALL questions.

Answer the questions in the Question Paper or on the separate data sheets – 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.

You must NOT write anything on the Formulae Pages. Anything you write on the Formulae Pages will gain NO credit.

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.

There may be spare copies of some data sheets in case you need them

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.

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1. Look at the diagram for Question 1 in the Data Booklet.

   It shows two stem and leaf diagrams.

   Bill is investigating how being grown in the shade and being grown in sunlight affects the heights of tree seedlings.

   The stem and leaf diagrams in the Data Booklet give information about the heights, in centimetres, of 17 tree seedlings grown in the shade and 17 tree seedlings grown in sunlight.

   The seedlings were all planted at the same time.

   Compare the average height of the tree seedlings grown in the shade with the average height of the tree seedlings grown in sunlight.

   State clearly the value of each average you use in order to make your comparison.

   (3 marks)

   Answer lines are on the next page.

1. continued.

(Total for Question 1 is 3 marks)

2. Look at the diagram for Question 2 in the Data Booklet.

It shows a population pyramid that gives information about the percentages of the population of the United Kingdom who are male and who are female in each age group for **2017**

Each percentage is given correct to one decimal place.

(a) Write down the percentage of the population who are female in the age group **50–54**

(1 mark)

_____ %

(continued on the next page)

2. continued.

(b) Work out the percentage of the population who are male in the age group **10–19**

(2 marks)

_____ %

(continued on the next page)

2. continued.

In 2017, the number of people age 100 and older (100+) in the United Kingdom was 13 310

Using the information above and information from the population pyramid,

(c) explain why the percentage of the population in the age group 100+ is given as 0·0% on the population pyramid.

You must show your working.

(3 marks)

(continued on the next page)

2. continued.

Jamie is carrying out research into the ages of people in the United Kingdom.

He uses the information in the population pyramid in the Data Booklet to claim,

“In the United Kingdom in 2017 the number of males who were older than 40 was greater than the number of females who were older than 40”

(d) Explain whether or not Jamie’s claim is correct.

(2 marks)

Answer lines continue on the next page.

2. (d) continued.

(Total for Question 2 is 8 marks)

3. Look at the table for Questions 3(a) and 3(b) in the Data Booklet.

Weronika works for a road traffic organisation.

One day she is investigating the speeds of cars and the speeds of motorcycles along a motorway.

The table shows part of the spreadsheet that Weronika used to record her results.

(a) Give a reason why Weronika will need to clean the data.

(1 mark)

Weronika concludes that the value of 124 in the spreadsheet must be wrong.

(b) Explain why.

(1 mark)

(continued on the next page)

3. continued.

Look at the table for Question 3(c) in the Data Booklet.

It shows the information about motorcycles from the spreadsheet with the data cleaned.

(c) Use linear interpolation to work out an estimate of the median speed of the motorcycles.

(3 marks)

_____ miles per hour

(continued on the next page)

3. continued.

Look at the diagram for Question 3(d) in the Data Booklet.

It shows the frequency polygon for the speeds of cars drawn on a grid.

(d) On the same grid in the Data Booklet, draw the frequency polygon for the speeds of the motorcycles using the values in the table for Question 3(c) in the Data Booklet.

(2 marks)

(continued on the next page)

3. continued.

(e) Using the two frequency polygons on the grid in the Data Booklet, compare the skew of the distribution of the speeds of the cars with the skew of the distribution of the speeds of the motorcycles.

Explain what your comparison means in context.

(2 marks)

(Total for Question 3 is 9 marks)

4. Look at the information for Question 4 in the Data Booklet.

It shows three sampling methods that might be used by a supermarket.

For each of the three methods in the Data Booklet, identify the sampling method and discuss whether the sampling method is an appropriate way to select the employees to be in the survey.

As part of your discussion you should also state, with reasons, which of the three sampling methods is the most appropriate method for the directors to use.

(6 marks)

Answer lines continue on the next two pages.

4. continued.

4. continued.

(Total for Question 4 is 6 marks)

5. Look at the diagram for Question 5 in the Data Booklet.

It shows a blank grid.

David is investigating the prices of houses in Streetly, which is in the West Midlands.

David took a sample of 40 houses in Streetly from a property website and recorded the price of each of these houses.

The cumulative frequency table on the next page shows information about the prices of these houses.

(continued on the next page)

5. continued.

Price (£P thousand)	Cumulative frequency
$240 < P \leq 280$	8
$240 < P \leq 320$	16
$240 < P \leq 360$	24
$240 < P \leq 400$	32
$240 < P \leq 440$	36
$240 < P \leq 480$	40

(a) Draw a cumulative frequency diagram for this information on the grid in the Data Booklet.

(2 marks)

(continued on the next page)

5. continued.

In David's sample there were N houses with a price in the interval $300 < P \leq 340$

(b) Find an estimate for the value of N

(2 marks)

(continued on the next page)

5. continued.

David wants to use his investigation to predict the number of houses in Central London that have a price between £300 000 and £340 000

(c) Assess whether or not it is sensible for David to make this prediction using the results of his investigation.

(2 marks)

(Total for Question 5 is 6 marks)

6. Look at the table for Question 6 in the Data Booklet. It shows a table Gaby is using to research the final position of a football team in the English Premier League and the mean value of all the players in the team.

(a) Suggest a diagram that Gaby could draw to see if there is an association between the final position and the mean value.
(1 mark)

(continued on the next page)

6. continued.

The table gives information about the data for the **2018–2019** season that Gaby used.

(b) Calculate Spearman's rank correlation coefficient for the information in the table. You may complete the d column and d^2 column of the table in the Data Booklet to help you.

(3 marks)

(continued on the next page)

6. continued.

(c) Interpret your answer to part (b) in the context of Gaby's research.

You should refer to the effects of any anomalous data.

(2 marks)

(continued on the next page)

6. continued.

Amelia suggests that Pearson's product moment correlation coefficient should be used instead of Spearman's rank correlation coefficient to measure the correlation between the data Gaby is researching.

(d) Discuss whether or not Amelia's suggestion is appropriate.

(3 marks)

(Total for Question 6 is 9 marks)

7. Look at the table for Question 7 in the Data Booklet. It shows a table that gives information about the numbers, in thousands, of overseas visitors to the United Kingdom from North America.

The table also gives the 4–point moving averages for the data in the table.

(a) Explain why it is appropriate to calculate 4–point moving averages for the data in the table.

(1 mark)

(continued on the next page)

7. continued.

Look at the diagram for Questions 7(b) and 7(c) in the Data Booklet.

It shows a time series graph that has been plotted for the data in the table in the Data Booklet.

Twelve of the 4–point moving averages for the data in the table have also been plotted on the grid so that the final 4–point moving average is missing from the grid.

(b) Plot this 4–point moving average on the grid.

(1 mark)

(c) Draw a trend line for the time series graph.

(1 mark)

(continued on the next page)

7. continued.

Jo wants to predict the number of visitors to the United Kingdom from North America for the first quarter of 2019

She is going to use the time series graph, the trend line and an average seasonal effect.

(d) Describe how Jo should use these to obtain her prediction.

(3 marks)

Answer lines continue on the next page.

7. (d) continued.

(Total for Question 7 is 6 marks)

8. Look at the table for Question 8 in the Data Booklet.

The table gives the chain base index number for the average price of houses in the United Kingdom in January for each of the years 2016 to 2019

(a) Describe what the chain base index numbers show about the average price of houses in January for the years 2016 to 2019

(2 marks)

(continued on the next page)

8. continued.

In January 2017, the average price of houses in the United Kingdom was £215 243

(b) Calculate the average price of houses in the United Kingdom in January 2016

Give your answer correct to the nearest £

(2 marks)

£ _____

(Total for Question 8 is 4 marks)

9. There are 8 coloured discs in bag X and 9 coloured discs in bag Y

In bag X, there are 3 red discs and 5 yellow discs.
In bag Y, there are 5 red discs and 4 yellow discs.

Sonia takes at random one disc from bag X and she takes at random one disc from bag Y

Calculate the probability that the two discs taken by Sonia are the same colour.

(3 marks)

Answer space continues on the next two pages.

9. continued.

9. continued.

(Total for Question 9 is 3 marks)

10. Look at the table for Question 10 in the Data Booklet.

Huan has applied for a job at a college.

As part of the interview process all candidates must take a numeracy test and a literacy test.

The table in the Data Booklet gives Huan's score for the numeracy test and his score for the literacy test. The table also gives the mean and the standard deviation of the scores for each of these two tests for all candidates who were interviewed for the job.

Huan concludes that because he scored a higher mark in literacy than in numeracy he performed better in literacy than in numeracy as compared to the other candidates.

Use statistical calculations to assess Huan's conclusion.

(5 marks)

Answer space and answer lines are on the next two pages.

10. continued.

10. continued.

(Total for Question 10 is 5 marks)

11. Look at the table for Question 11 in the Data Booklet.

It shows a grouped frequency table.

Viktoria collects information about the times taken, in seconds, by **50** boys to run **400** metres.

She records her results in the grouped frequency table, and gives the grouped frequency table to her friend Emeka.

Viktoria asks Emeka to use only the information in the grouped frequency table to work out an estimate of the standard deviation of the times taken.

(a) Explain why Emeka is only able to work out an ESTIMATE of the standard deviation.

(1 mark)

(continued on the next page)

11. continued.

Emeka uses the information in the table in the Data Booklet to work out the following estimates of the summary statistics

$\sum ft = 3542 \cdot 5$ and $\sum ft^2 = 252\,331 \cdot 25$

Mean $= 70 \cdot 85$ and Median $= 71 \cdot 18$

(b) Using Emeka's estimates of the summary statistics, calculate an estimate of the skew of the distribution of the times taken by the **50** boys.

Give your answer to **2** decimal places.

(3 marks)

Answer space continues on the next page.

11. (b) continued.

(c) **Interpret your answer to part (b) in the context of the data in the table in the Data Booklet.**

(1 mark)

(Total for Question 11 is 5 marks)

12. A game is played with 4 ordinary 6–faced dice. Each dice is to be rolled once and the number of dice that land on a six is recorded.

(a) Write down TWO conditions needed so that a binomial distribution is a suitable model for the number of sixes recorded.

(2 marks)

(continued on the next page)

12. continued.

(b) Calculate the probability that all of the 4 dice land on a six.

Give your answer as a fraction.

(2 marks)

(continued on the next page)

12. continued.

(c) Calculate the probability that at least 2 of the 4 dice land on a six.

Give your answer as a fraction.

(3 marks)

(Total for Question 12 is 7 marks)

13. Look at the table for Question 13 in the Data Booklet.

Suzie is investigating the profits made by two different shops, shop **A** and shop **B**

Suzie has obtained the annual percentage profits made by shop **A** for the years **2015** to **2019** and the annual percentage profits made by shop **B** for the years **2016** to **2019**

The table in the Data Booklet gives this information.

Suzie concludes that the average annual percentage profit made by shop **B** over the **4** years is greater than the average annual percentage profit made by shop **A** over the **5** years.

By using appropriate geometric means, assess Suzie's conclusion.

You must show your working.

(5 marks)

Answer lines are on the next page.

13. continued.

(Total for Question 13 is 5 marks)

14. Look at the diagram for Question 14 in the Data Booklet.

It shows an incomplete Venn diagram.

Olga is investigating the types of exercise taken by a group of 120 people.

Olga asked each of these 120 people whether or not, for exercise, they go to a gym (G) or walk (W) or cycle (C).

The incomplete Venn diagram gives some information about Olga's results.

One of the people that Olga asked is to be chosen at random.

(continued on the next page)

14. continued.

Given that, for this person,

$$P(W \mid G) = \frac{9}{52}$$

$$P(W \text{ or } G) = \frac{41}{60}$$

complete the Venn diagram.

(4 marks)

Space for working continues on the next page.

14. continued.

(Total for Question 14 is 4 marks)

TOTAL FOR PAPER IS 80 MARKS

END OF PAPER
