

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

Total Marks

Statistics
PAPER 2
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					

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 this 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.

Turn over

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 table for Question 1 in the Data Booklet.
It shows an incomplete table.

Tomasz found information about the weather on the
Isle of Skye for the **31** days of May **2019** and the
31 days of July **2019**

The table gives, for the Isle of Skye, some
information about the number of days with rainfall
during May.

- (a) Using the information in the table, find the
number of days with rainfall in May **2019**
You must show your working.
(2 marks)

_____ days

(continued on the next page)

Turn over

1. continued.

The relative risk of rainfall for a day in July compared with a day in May is 1.29

(b) Find the number of days with rainfall in July 2019

You must show your working.

(2 marks)

_____ days

(Total for Question 1 is 4 marks)

Turn over

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

It shows a box plot.

Bill investigated the heights of females competing in different sports.

Using data from the internet, he recorded the height, in cm, of each female in a sample of 15 female rugby players and the height, in cm, of each female in a sample of 15 female basketball players.

He drew the box plot in the Data Booklet for the recorded heights of the 15 female basketball players.

(continued on the next page)

2. continued.

The table gives information about the recorded heights of the **15** female rugby players.

Greatest height	180 cm
Median height	170 cm
Lower quartile	165 cm
Range	25 cm
Interquartile range	10 cm

(a) Using the information in the table, draw on the grid in the Data Booklet a box plot for the recorded heights of the **15 female rugby players.**

(3 marks)

(continued on the next page)

Turn over

2. continued.

(b) Compare the two distributions of heights.

Give THREE comparisons and interpret one of your comparisons.

(4 marks)

(continued on the next page)

Turn over

2. continued.

(c) Considering the method that Bill used to collect his data, comment on the reliability of your comparisons in part (b).

(2 marks)

(Total for Question 2 is 9 marks)

3. Look at the table for Question 3 in the Data Booklet. Irina is investigating whether the percentage of the population of a country living in urban areas has an effect on the life expectancy in that country.

(a) Suggest a hypothesis Irina could use for her investigation.

(1 mark)

(continued on the next page)

3. continued.

Irina collected the information about 10 countries shown in the table in the Data Booklet.

The Urban population (%) refers to the percentage of the population of the country who live in urban areas.

Irina used statistical software to draw a scatter diagram for the information in the table.

(b) Give a reason why a scatter diagram is an appropriate diagram to use.

(1 mark)

(continued on the next page)

Turn over

3. continued.

(c) For this investigation, which variable is the explanatory variable?

Give a reason for your answer.

(2 marks)

(continued on the next page)

3. continued.

Look at the diagram for Questions 3(d), (e) and (f) in the Data Booklet.

It shows the scatter diagram from the statistical software.

(d) Explain, giving a statistical reason, whether or not this scatter diagram supports your hypothesis in part (a).

(2 marks)

For these 10 countries, the double mean point of the data is $(63.9, 77.8)$.

(e) Using this information, draw a line of best fit on the scatter diagram in the Data Booklet.

(2 marks)

(continued on the next page)

Turn over

3. continued.

Using statistical software, Irina finds that the gradient of the line of best fit should be 0.19

(f) Interpret the gradient of the line of best fit.
(1 mark)

(continued on the next page)

3. continued.

Irina now finds that South Africa has Urban population **65% and Life expectancy **63** years.**

(g) Determine how this information for South Africa fits with the relationship shown in the scatter diagram for the other countries.

(2 marks)

(Total for Question 3 is 11 marks)

4. Look at Table 1 and Table 2 for Question 4 (a) in the Data Booklet.

A hospital department has three specialist consultants, **A**, **B** and **C**

The department manager uses **2**–digit random numbers to carry out a simulation of the number of consultations per week for each specialist.

According to hospital records, the average number of consultations per week for each specialist is shown in Table 1 in the Data Booklet.

(continued on the next page)

4. continued.

**The department manager allocates the
2–digit numbers to each specialist shown in Table 2
in the Data Booklet.**

**(a) Explain why this is an appropriate way for
the department manager to allocate the
2–digit numbers.**

(1 mark)

(continued on the next page)

4. continued.

Look at the table for Question 4(b) in the Data Booklet.

Also, according to hospital records, the expected total number of consultations in a week for all three specialists is 50

The table below shows the first 45 results from the simulation.

Specialist	Simulated number of consultations
A	18
B	19
C	8

(continued on the next page)

Turn over

4. continued.

The last five random numbers used by the department manager are

06 74 59 14 29

(b) Using these last five numbers, complete the table for Question 4(b) in the Data Booklet to show the final result of the simulation.

There are three spaces to fill.

(1 mark)

(continued on the next page)

4. continued.

Each consultation is planned to last no longer than **20** minutes.

(c) According to the simulation, how much more time a week should be allocated to specialist **A** than to specialist **C** for all of their consultations?

(1 mark)

_____ minutes

(continued on the next page)

4. continued.

The department manager decides that the results of her simulation are sufficient for her to plan how much time is needed per week for consultations with each specialist.

**(d) Explain why the manager is NOT correct.
(2 marks)**

(Total for Question 4 is 5 marks)

Turn over

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

(a) Explain briefly why a control group should be used.

(1 mark)

(b) Explain how the manager could use matched pairs in order to select the people in the two groups.

(1 mark)

(continued on the next page)

Turn over

5. continued.

(c) Assess the reliability of any conclusions that the manager may reach as a result of using his plan.

Give TWO reasons for your answer.

(3 marks)

(Total for Question 5 is 5 marks)

6. Paul uses the Petersen capture recapture method in order to find an estimate for the number of trout living in a lake.

To do this, Paul catches an initial sample of **50** trout from the lake and he tags each trout.

He then releases the trout back into the lake.

For his second sample, Paul catches a sample of **80** trout from the lake.

He finds that **16** of these trout have tags.

- (a) Explain what needs to be true about his two samples for them to be valid for the capture recapture method.

(1 mark)

(continued on the next page)

Turn over

6. continued.

Using his results, Paul is able to work out an estimate for the number of trout living in the lake.

(b) Find Paul's estimate for the number of trout and discuss the validity and the reliability of his estimate.

(5 marks)

Answer lines continue on the next page.

Turn over

6. (b) continued.

(Total for Question 6 is 6 marks)

7. Look at the diagram for Question 7 in the Data Booklet.

It shows a histogram.

Andy is investigating the lateness of the trains that arrive at Manchester Piccadilly.

- (a) Explain why it is sensible for Andy to collect secondary data for his investigation.

(1 mark)

Andy found out data on the lateness of each of **238** trains arriving at Manchester Piccadilly from London on Monday to Friday for one week.

The histogram in the Data Booklet was drawn using Andy's data.

(continued on the next page)

7. continued.

A negative time for a train means that the train arrived early at Manchester Piccadilly.

In the week, 17 trains arrived early.

- (b) Find how many of the 238 trains did NOT arrive more than 10 minutes late at Manchester Piccadilly.**
(3 marks)
-

(continued on the next page)

7. continued.

For a long distance journey, such as London to Manchester Piccadilly, a train is only considered to be late when it arrives more than 10 minutes late.

Andy says that more than 1 in 3 trains from London arrive late at Manchester Piccadilly.

(c) Determine whether or not Andy is correct.

You should comment on the reliability of your conclusion.

(4 marks)

Answer space and lines continue on the next page.

7. (c) continued.

(Total for Question 7 is 8 marks)

8. Look at Diagram 1 and Diagram 2 for Question 8 in the Data Booklet.

An engineering production line is set up to produce machine parts with a diameter of **24.50 mm**

For quality control, random samples of size **4** are taken and the mean diameter of each sample is calculated.

The production line is set so that the sample means should be normally distributed with a mean of **24.50 mm** and a standard deviation of **0.005 mm**

- (a) Using this information, draw warning lines on the control chart for the sample means shown in Diagram 1 in the Data Booklet.

Action lines have already been drawn on the chart.

(2 marks)

(continued on the next page)

Turn over

8. continued.

A control chart for the sample ranges is also used and is shown on Diagram 2 in the Data Booklet.

(b) Explain why it is not appropriate to have lower warning lines and lower action lines on the control chart for sample ranges.

(1 mark)

(continued on the next page)

Turn over

8. continued.

The first two samples taken have the following summary statistics.

	Sample 1	Sample 2
Sample mean (mm)	24·505	24·495
Sample range (mm)	0·006	0·008

**(c) Plot the summary statistics for these two samples on the control charts shown in Diagram 1 and Diagram 2 in the Data Booklet.
(2 marks)**

(continued on the next page)

8. continued.

The machine parts in Sample 3 had the following diameters:

24.498 mm 24.492 mm

24.501 mm 24.509 mm

- (d) Use these results to complete BOTH control charts shown in Diagram 1 and Diagram 2 in the Data Booklet for Sample 3
(4 marks)

(continued on the next page)

Turn over

8. continued.

(e) Describe what action needs to be taken after
Sample 3
(1 mark)

(Total for Question 8 is 10 marks)

9. Look at the diagram for Question 9 in the Data Booklet.

It shows two pie charts.

Three types of sausage are sold in a farm shop, pork, beef and vegan.

The comparative pie charts show information about the number of packs of these sausages sold in the farm shop in May **2015** and in May **2019**

The pie charts are drawn accurately and the angles for the vegan sectors are shown.

The radius of the pie chart for May **2019** is greater than the radius of the pie chart for May **2015**

- (a) Explain what can be deduced from this information.

(1 mark)

(continued on the next page)

Turn over

9. continued.

30 packs of vegan sausages were sold in the farm shop in May **2015**

(b) Work out the number of packs of vegan sausages that were sold in the farm shop in May **2019**
(3 marks)

(Total for Question 9 is 4 marks)

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

Changes in the cost of living in the United Kingdom are measured by the consumer price index (CPI).

The table in the Data Booklet shows the index numbers for June 2019, using 2015 as base year, of the items that contribute to the CPI.

The table also shows the weightings for how the spending by the Jones family is divided between these items.

Using 2015 as base year, the CPI for June 2019 was 107.9

(continued on the next page)

10. continued.

Compare the overall change in the cost of living for the Jones family between 2015 and June 2019 with the change in the CPI between 2015 and June 2019

Explain whether or not it is possible to conclude from your calculations that the Jones family is financially better off or worse off in June 2019 as compared with 2015

(5 marks)

Answer lines are on the next page.

10. continued.

(Total for Question 10 is 5 marks)

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

The mean of the daily winter temperatures and the mean of the daily summer temperatures in the UK have been recorded for each year from 1910 to 2018

This information has been used to model the two distributions of the mean temperatures.

The graphs in the Data Booklet give information about these models.

- (a) Write down the name of the distribution that is suggested by each of these graphs.**
(1 mark)
-

(continued on the next page)

Turn over

11. continued.

**(b) Comment on the difference between the means
of these two distributions.**

(2 marks)

(continued on the next page)

11. continued.

The standard deviation for the distribution of the summer temperatures is 0.8°C , correct to one decimal place.

- (c) Using the graph for the winter temperatures, calculate an estimate for the standard deviation of the distribution of the winter temperatures, correct to one decimal place.**
- (2 marks)**

_____ $^{\circ}\text{C}$

(continued on the next page)

11. continued.

The maximum height of the graph above the temperature axis for the winter temperatures is less than the maximum height of the graph above the temperature axis for the summer temperatures.

(d) Explain why.

(1 mark)

(continued on the next page)

Turn over

11. continued.

**Carol says that the standard deviation of
DAILY temperatures in summer for the years from
1910 to 2018 will also be 0.8°C**

**(e) Explain whether or not Carol is correct.
(2 marks)**

(continued on the next page)

11. continued.

Greta assumes that the means of the daily temperatures each year are independent.

She works out the probability that the means of the daily summer temperatures are greater than 15.6°C for two consecutive years.

She concludes that this probability is less than 0.001

**(f) Using the model for the daily summer temperatures, assess Greta's conclusion.
(5 marks)**

Answer space continues on the next page.

11. (f) continued.

(Total for Question 11 is 13 marks)

TOTAL FOR PAPER IS 80 MARKS

END OF PAPER
