

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

Total Marks

Statistics
Paper 1
Higher Tier
(Calculator)

Monday 12 June 2023 – Afternoon

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

YOU WILL BE GIVEN

**Data Booklet
Formulae Pages**

Turn over

INSTRUCTIONS

Answer ALL questions.

Answer the questions in the spaces provided 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.

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.

Turn over

5

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

Turn over

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

It shows a back-to-back stem and leaf diagram.

The scores for Vitality Roses netball team for the **2017** season are listed below.

55	61	72	62	52
74	56	49	75	66
64	85	55	52	55
94	61	46	74	54

(continued on the next page)

1. continued.

(a) Use the data to complete the back-to-back stem and leaf diagram in the Data Booklet.

(3 marks)

(b) Work out the interquartile range of the scores for the 2018 season.

(2 marks)

(continued on the next page)

Turn over

1. continued.

The table gives information about the scores for the 2017 and 2018 seasons.

	Median	Interquartile range
2017 season	61	18·5
2018 season	48·5	

(continued on the next page)

Turn over

1. continued.

Naomi thinks that the results show that Vitality Roses had a higher and more consistent set of scores in the 2018 season than in the 2017 season.

(c) (i) Do the statistics support this conclusion?

You must give reasons for your answer.

(3 marks)

Answer lines continue on the next page.

1. (c) (i) continued.

(continued on the next page)

1. (c) continued.

(ii) Give a limitation of using the data provided to comment on whether or not Vitality Roses has improved as a team between the 2017 and 2018 seasons.

(1 mark)

(continued on the next page)

1. continued.

Naomi plans to process the data further.

She suggests checking for outliers, by performing calculations, before doing any more processing.

(d) Comment on the appropriateness of this suggestion.

(1 mark)

Answer lines continue on the next page.

Turn over

1. (d) continued.

(Total for Question 1 is 10 marks)

- 2. Kyle is investigating the heights and the weights of professional basketball players.**

He found the weight, in kilograms, of some professional basketball players from 1950 to 1959

(continued on the next page)

2. continued.

(a) Circle the word in the list below that describes weight, in kilograms, as a type of data.

(1 mark)

discrete

continuous

ordinal

categorical

(continued on the next page)

Turn over

2. continued.

**Look at the diagram for
Questions 2(b) and 2(c) in the
Data Booklet.**

It shows an incomplete histogram.

**The incomplete histogram in the
Data Booklet and incomplete grouped
frequency table on the following page
give information about the weights,
in kilograms, of the professional
basketball players from 1950 to
1959**

(continued on the next page)

Turn over

2. continued.

Weight (w kilograms)	Frequency
$65 < w \leq 70$	5
$70 < w \leq 75$	15
$75 < w \leq 80$	61
$80 < w \leq 85$	81
$85 < w \leq 90$	
$90 < w \leq 95$	
$95 < w \leq 100$	35
$100 < w \leq 105$	14
$105 < w \leq 110$	9
$110 < w \leq 115$	1

(continued on the next page)

Turn over

2. continued.

(b) Use the information in the histogram in the Data Booklet to complete the table on the previous page.

There are two spaces to fill.

(2 marks)

(c) Use the information in the table on the previous page to complete the histogram in the Data Booklet.

(2 marks)

(continued on the next page)

Turn over

2. continued.

Kyle also drew a histogram for the weights of professional basketball players from 2000 to 2009

This histogram was negatively skewed.

(d) Interpret the negative skew of the weights of professional basketball players from 2000 to 2009
(1 mark)

(continued on the next page)

Turn over

2. continued.

Look at the table for Question 2(e) in the Data Booklet.

Kyle also collected data about the heights of professional basketball players from 1950 to 1959 and the heights of professional basketball players from 2000 to 2009

The grouped frequency table in the Data Booklet gives information about the heights of professional basketball players from 2000 to 2009

(continued on the next page)

Turn over

2. continued.

The estimate of the mean height for professional basketball players from 1950 to 1959 is calculated to be 190.9 cm to one decimal place.

**(e) (i) Calculate an estimate of the mean height of basketball players from 2000 to 2009
(3 marks)**

_____ cm

(continued on the next page)

Turn over

2. (e) continued.

(ii) Comment on how the mean height of professional basketball players has changed between the two sets of data.

(1 mark)

(Total for Question 2 is 10 marks)

Turn over

- 3. Claire is investigating sales of different types of vehicle over time. She plans to collect data on the numbers of motorcycles first registered in the UK over time.**

(a) Write down a suitable hypothesis for this investigation.

(1 mark)

(continued on the next page)

Turn over

3. continued.

Look at Diagram 1 for Question 3 in the Data Booklet.

It shows a time series graph with some information about the numbers of motorcycles first registered in the UK from 2017 to 2019

(b) Identify and interpret one example of seasonal trend shown by the time series graph in

Diagram 1

(2 marks)

Answer lines are on the next page.

Turn over

3. (b) continued.

(continued on the next page)

3. continued.

Claire calculated 4–point moving averages for the information shown in the time series graph.

**(c) Explain why this is appropriate.
(1 mark)**

(continued on the next page)

Turn over

3. continued.

Look at Diagram 2 for Question 3 in the Data Booklet.

Claire also collected data on the numbers of cars first registered in the UK from 2017 to 2019

The time series graph in the Data Booklet shows some information about the numbers of cars first registered in the UK from 2017 to 2019 together with the first seven 4–point moving averages.

(continued on the next page)

Turn over

3. continued.

(d) Compare the seasonal trend shown for the numbers of motorcycles first registered in the UK with the seasonal trend for the numbers of cars first registered in the UK.

(1 mark)

(continued on the next page)

Turn over

3. continued.

The last three 4–point moving averages (thousands) for the number of cars registered in the UK from 2017 to 2019 are

576·0

575·3

573·9

(continued on the next page)

Turn over

3. continued.

**(e) Plot these three moving averages
on the time series graph in
Diagram 2 and draw a trend line.
(3 marks)**

(continued on the next page)

3. continued.

- (f) Describe and interpret the trend in the numbers of cars first registered in the UK from 2017 to 2019**
(2 marks)

(Total for Question 3 is 10 marks)

Turn over

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

Maciek is investigating the performance of athletes in the long jump.

He wants to find out how the height of an athlete affects the distance jumped.

It shows in the Data Booklet how Maciek plans to collect, process and present his data.

(continued on the next page)

4. continued.

Discuss whether Maciek's plans for collecting, processing and presenting data are appropriate.

(5 marks)

Answer lines continue on the next page.

4. continued.

(Total for Question 4 is 5 marks)

5. A local arts festival has a needlecraft competition and a flower arranging competition.

Mr Smith judges both competitions, giving each entry a rank.

Mr Smith thinks there is an association between the ranks given to competitors in the needlecraft competition and the ranks given to competitors in the flower arranging competition.

(continued on the next page)

5. continued.

He plans to calculate Spearman's rank correlation coefficient for the ranks in the two competitions.

(a) Explain why this may NOT be appropriate.

(1 mark)

(continued on the next page)

5. continued.

Look at the table for Question 5(b) in the Data Booklet.

The arts festival also has a photography competition.

In the photography competition, the 8 entries were given the ranks shown in the table in the Data Booklet by two judges, Mrs John and Mr Nowak.

(continued on the next page)

Turn over

5. continued.

(b) Using suitable calculations, investigate the level of agreement between the two judges.

You may use the blank columns in the table in the Data Booklet for your working.

(5 marks)

Answer space and lines continue on the next two pages.

Turn over

5. (b) continued.

Turn over

5. (b) continued.

(Total for Question 5 is 6 marks)

6. Look at the table for Question 6 in the Data Booklet.

It gives information about the monthly average price per troy ounce, in pounds, of gold over a period of six months in 2019

The table also gives some of the chain base index numbers, correct to one decimal place, for this information.

(continued on the next page)

6. continued.

- (a) Calculate the chain base index number for June and write it in the table in the Data Booklet.**
- Give your answer correct to one decimal place.**

(2 marks)

(continued on the next page)

Turn over

6. continued.

(b) (i) Calculate the geometric mean of the five chain base index numbers.

You must show your working.

Give your answer correct to one decimal place.

(2 marks)

Answer space continues on the next page.

Turn over

6. (b) (i) continued.

(continued on the next page)

Turn over

6. (b) continued.

**(ii) Interpret your answer to
part (i) in context.
(2 marks)**

(Total for Question 6 is 6 marks)

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

A class is learning about the Petersen capture–recapture method.

Their teacher provides them with the data shown in the table in the Data Booklet for burbot fish in Tolsona Lake, Alaska.

(continued on the next page)

7. continued.

The data was collected by tagging all of the fish in the initial sample before releasing them, then counting the number of fish with tags in the second sample.

(a) Calculate an estimate of the total number of burbot fish in Tolsona Lake.

(2 marks)

7. continued.

Sophie is planning her own investigation using the Petersen capture–recapture method. Here is her plan for data collection.

Data collection to estimate the population of frogs in the science garden.

Monday 9 am

Capture 5 frogs and tag them using stick–on tags.

Release them.

(continued on the next page)

Turn over

7. continued.

Friday 9 am

Capture 5 frogs and record how many have tags on.

**(b) Discuss the appropriateness of Sophie's plan for data collection.
(3 marks)**

Answer lines continue on the next page.

Turn over

7. (b) continued.

(Total for Question 7 is 5 marks)

Turn over

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

It shows a probability tree diagram.

Ian collects information about

whether drivers have had advanced

driver training and whether they

have exceeded the speed limit in the

last month.

He finds that 18% of the drivers he

surveyed have had advanced driver

training and that 16% of these

drivers have exceeded the speed limit

in the last month.

(continued on the next page)

Turn over

8. continued.

For the drivers who have not had advanced driver training 26% of them have exceeded the speed limit in the last month.

(a) Complete the probability tree diagram in the Data Booklet for this information.

(1 mark)

(continued on the next page)

8. continued.

(b) Calculate

$P(\text{Advanced driver} \mid \text{Not exceeded speed limit})$.

(2 marks)

(continued on the next page)

Turn over

8. continued.

(c) (i) Calculate the relative risk of exceeding the speed limit for the not advanced drivers compared to the advanced drivers.

(2 marks)

Answer space continues on the next page.

Turn over

8. (c) (i) continued.

(continued on the next page)

8. (c) continued.

(ii) Interpret this relative risk in context.

(1 mark)

(continued on the next page)

8. continued.

Ian researches the relative risk of exceeding the speed limit for newly qualified drivers compared to drivers who are not newly qualified.

He finds that in 2019 the relative risk was 0·95 and in 2020 the relative risk was 0·87

Ian concludes that the proportion of newly qualified drivers exceeding the speed limit has decreased between 2019 and 2020

(continued on the next page)

Turn over

8. continued.

(d) Do the statistics support Ian's conclusion?

You must give a reason for your answer.

(2 marks)

Answer lines continue on the next page.

Turn over

8. (d) continued.

(Total for Question 8 is 8 marks)

9. Look at Diagram 1 and Diagram 2 for Question 9 in the Data Booklet.

Ligita collected data on the handspans, in cm, of men and the handspans, in cm, of women.

She processed her data using statistical software and her results are shown in the Data Booklet.

Diagram 1 shows male handspans.

Diagram 2 shows female handspans.

(continued on the next page)

9. continued.

**(a) Calculate the skew of the female
handspan data.**

(2 marks)

(continued on the next page)

Turn over

9. continued.

**(b) Compare in context the
distribution of male handspans
and the distribution of female
handspans.**

(5 marks)

**Answer lines continue on the
next page.**

Turn over

9. (b) continued.

(continued on the next page)

Turn over

9. continued.

Ligita plans to use the mean for the males and the mean for the females to find all of the handspans.

(c) Explain why she should use a weighted mean for this calculation.

(1 mark)

(Total for Question 9 is 8 marks)

Turn over

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

A study took place in the southeastern United States to find if there was a relationship between the snout–vent length, in cm, and the body weight, in grams, of adult snakes.

The researchers found the equations of the regression lines for the relationship between snout–vent length (x centimetres) and the body weight (y grams) for each species of snake.

10. continued.

For two species of snake, the timber rattlesnake and the eastern racer, there was a difference in the relationship between the two variables for female snakes and the relationship between the two variables for male snakes.

The table in the Data Booklet gives the equations of the regression lines for these snakes.

(continued on the next page)

Turn over

10. continued.

**(a) Interpret in context the figure
19.61 in the regression equation
for female timber rattlesnakes.**

(1 mark)

(continued on the next page)

Turn over

10. continued.

**(b) Compare the relationships
between snout–vent length and
body weight in the two species of
snake.**

**Include in your comparisons
reference to male snakes and
female snakes.**

(3 marks)

**Answer lines continue on the
next page.**

Turn over

10. (b) continued.

(continued on the next page)

10. continued.

The researchers would like to use a normal distribution as a model for the weights of male timber rattlesnakes.

They plan to draw a histogram, calculate averages and the standard deviation for these weights to check whether a normal distribution is a suitable model.

(continued on the next page)

Turn over

10. continued.

(c) (i) Explain how they could use a histogram to check whether a normal distribution is a suitable model.

(1 mark)

(continued on the next page)

Turn over

10. (c) continued.

(ii) Explain how they could use averages and the standard deviation to check whether the normal distribution is a suitable model.

(2 marks)

Answer lines continue on the next page.

Turn over

10. (c) (ii) continued.

(Total for Question 10 is 7 marks)

11. The manager of a large delivery company wants to investigate whether employees have dealt with personal errands whilst at work in the last 6 months.

He plans to ask employees whether they have dealt with personal errands during the work day.

(continued on the next page)

11. continued.

Libby suggests using the random response technique.

(a) Explain why it is appropriate to use the random response technique in this case.

(1 mark)

(continued on the next page)

Turn over

11. continued.

**Look at the information for
Question 11(b) in the Data Booklet.
It is a random response question
designed by the manager.**

**(b) Comment on the appropriateness
of this random response
question.**

Give reasons for your answer.

(2 marks)

**Answer lines continue on the
next page.**

Turn over

11. (b) continued.

(continued on the next page)

11. continued.

Look at the table for Question 11(c) in the Data Booklet.

The final questionnaire will be sent to a sample of the delivery company employees.

The delivery company has office staff, warehouse staff and delivery drivers.

The staff work either full time or part time.

The table in the Data Booklet gives the number of each type of employee.

(continued on the next page)

Turn over

11. continued.

The manager wants to take a sample stratified by employment role and contract type.

He wants to sample at least 15 people from each strata.

If his calculated sample for one of the strata is a decimal he will round this to the nearest integer.

(continued on the next page)

11. continued.

(c) Work out the smallest total sample size that gives at least 15 people from each strata.

(2 marks)

Answer space continues on the next page.

11. (c) continued.

(Total for Question 11 is 5 marks)

TOTAL FOR PAPER IS 80 MARKS

END OF PAPER

Sources

Question 1

(Source: www.englandnetball.co.uk)