

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

**Total Marks**

# **Statistics**

**Paper 2**  
**(Calculator)**  
**Higher Tier**

**Tuesday 18 June 2019 – Morning**

**Time: 1 hour 30 minutes plus your additional time allowance.**

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

<b>Surname</b>					
<b>Other names</b>					
<b>Centre Number</b>					
<b>Candidate Number</b>					

**YOU MUST HAVE**

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

**YOU WILL BE GIVEN**

**Data Book  
Formulae Pages**

**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.**

**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.**

## **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. At a university, 70% of students are undergraduates and 30% of students are postgraduates.**

**Amy and Robert want to do a survey.**

**Amy decides to use simple random sampling to collect a sample of 100 students.**

**She uses the university database as a sample frame and she numbers each student on the database. She then generates exactly 100 random numbers and uses these random numbers to select her sample.**

**(continued on the next page)**

1. continued.

(a) Give TWO reasons why Amy's method may NOT produce a sample of 100 students.

(2 marks)

---

---

---

---

---

---

(continued on the next page)

1. continued.

**Remember:**

**At a university, 70% of students are undergraduates and 30% of students are postgraduates.**

**Robert decides to use quota sampling to collect a sample of 100 students.**

**He plans to stand outside the main building until he has interviewed 70 undergraduates and 30 postgraduates.**

**(b) Give TWO advantages of using quota sampling.  
(2 marks)**

---

---

---

---

---

---

---

**(continued on the next page)**

**Turn over**

1. continued.

(c) Explain why this quota sample is NOT a random sample.

(1 mark)

---

---

---

(Total for Question 1 is 5 marks)

---



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

The population pyramid shows information about the numbers (in thousands) of drivers of each gender who made car insurance claims in the UK in 2015

- (a) How many female drivers aged 50–59 in the UK in 2015 made car insurance claims?  
(1 mark)

\_\_\_\_\_ thousand

(continued on the next page)

**2. continued.**

**The population pyramid shows that the age group which has the fewest number of drivers who made car insurance claims is the 17–19 age group.**

**(b) Suggest a reason why this should be so.**

**(1 mark)**

---

---

---

**(continued on the next page)**

**2. continued.**

**In 2014, the number of male drivers aged 20–49 in the UK who made car insurance claims was 66 700**

**(c) Compare the number of male drivers aged 20–49 in the UK who made car insurance claims in 2014 with the number of male drivers aged 20–49 in the UK who made car insurance claims in 2015**

**You must show your working.**

**(3 marks)**

---

---

---

**(continued on the next page)**

**Turn over**

**2. continued.**

**The SafeDrive insurance company charges young male drivers more for car insurance than it charges all other drivers.**

**(d) Explain TWO features of the population pyramid which SafeDrive might use as its justification for doing this.**

**(2 marks)**

---

---

---

---

---

---

**(continued on the next page)**

**2. continued.**

**Jeremy says,**

**“The population pyramid shows that the total number of male drivers in the UK in 2015 is greater than the total number of female drivers in the UK in 2015”**

**(e) Explain whether or not Jeremy’s conclusion is appropriate.**

**(1 mark)**

---

---

---

**(continued on the next page)**

**2. continued.**

**Vicki says,**

**“In the UK in 2019, there will be more male drivers who make car insurance claims than female drivers who make car insurance claims”**

**(f) Explain whether or not the information in the population pyramid can be used to support Vicki’s statement.**

**(1 mark)**

---

---

---

**(Total for Question 2 is 9 marks)**

---

3. Look at the diagram and at the table for Question 3 in the Data Book.

The frequency polygon shows information about the times taken by **48** athletes to run **400** metres at the **2017** World Championships.

- (a) Use the information in the frequency polygon to complete the table by writing in the missing times.

(1 mark)

None of the athletes ran **400** metres in exactly **53** seconds.

- (b) Find the number of athletes who ran **400** metres in less than **53** seconds.

(2 marks)

---

(continued on the next page)

Turn over

3. continued.

(c) Calculate an estimate for the mean time of the 48 athletes.

(3 marks)

\_\_\_\_\_ seconds

(Total for Question 3 is 6 marks)

---



4. Look at the diagram for Question 4 in the Data Book.

It shows a stem and leaf diagram.

Randall used the internet to find the number of points scored by each of the winning teams in the last 30 years of the American Super Bowl.

He drew the stem and leaf diagram in the Data Book for his data.

- (a) Write down two improvements Randall should make to the stem and leaf diagram.

(2 marks)

1 \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

2 \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

(continued on the next page)

**4. continued.**

**(b) Assess the reliability of using the internet as the source of Randall's data.**

**(1 mark)**

---

---

---

**Randall believes that the number of points scored by the winning team is increasing over time.**

**(c) Comment on whether or not Randall has represented his data in an appropriate diagram in order to support his belief.**

**(1 mark)**

---

---

---

---

---

**(Total for Question 4 is 4 marks)**

---

**Turn over**

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

Compare the amount of UK aid given to Ethiopia and the amount of UK aid given to Pakistan for the years 2009 to 2015

You may use

$$219^2 + 263^2 + 344^2 + 265^2 + 329^2 + 321^2 + 338^2 = 631\,217$$

(5 marks)

Answer space continues on the next page.

5. continued.

(Total for Question 5 is 5 marks)

---

Turn over

6. Look at the table for Question 6(a) in the Data Book. Using July 2015 as the base month, the table shows some of the index numbers for the consumer price index (CPI) from July 2015 to July 2017

- (a) Give an interpretation of the index number 103·2 in the table.  
(2 marks)

---

---

---

---

---

---

(continued on the next page)

**6. continued.**

**Look at the information for Question 6(b) in the Data Book.**

**Using July 2015 as the base month, the index number for goods in July 2017 was 100.7**

**(b) Calculate the index number for services in July 2017**

**Give your answer to the nearest whole number.**

**(2 marks)**

---

**(continued on the next page)**

6. continued.

Remember:

Using July 2015 as the base month, the index number for goods in July 2017 was 100·7

One of the items included in the goods index is food.

- (c) Assess whether or not it is appropriate to use 100·7 in order to find an estimate of the change in the price of food from July 2015 to July 2017 (2 marks)

---

---

---

---

---

(Total for Question 6 is 6 marks)

---

7. Susie is investigating the amount of time students at her school spend watching television.

She selects a random sample of **50** students from her school and asks them the following question.

How much time, in hours to the nearest hour, did you spend watching television last Saturday?

☐

2

☐

3

☐

4 to 5

☐

6 to 7

☐

8 to 11

(continued on the next page)



7. continued.

Susie received answers from only **43** of the students.

The table below shows a summary of her results.

Time (t hours)	Frequency
2	14
3	11
4 to 5	8
6 to 7	6
8 to 11	4

(continued on the next page)

**7. continued.**

**Look at the diagram for Question 7(a) and Question 7(b) in the Data Book.**

**The incomplete histogram shows information about some of Susie's results.**

**(a) Give a reason to support using a histogram to represent the amounts of time students at her school spend watching television.**

**(1 mark)**

---

---

---

**(b) Complete the histogram.**

**(3 marks)**

**(continued on the next page)**

**7. continued.**

**A student is selected at random from those students in the sample who watched less than 5.5 hours of television last Saturday.**

**(c) Estimate the probability that this student watched less than 2 hours of television last Saturday.**

**(2 marks)**

---

**(continued on the next page)**

**7. continued.**

**Using class midpoints, Susie calculated the mean time spent watching television to be 4 hours.**

**She concludes that the mean time spent watching television each day is 4 hours for ALL students.**

**(d) Assess how Susie's question affects the accuracy of the mean time.**

**(2 marks)**

---

---

---

---

---

---

---

---

**(continued on the next page)**

**7. continued.**

**(e) Assess how Susie's method for collecting her data affects the validity of her conclusion.**

**(3 marks)**

---

---

---

---

---

---

---

---

---

---

---

**(Total for Question 7 is 11 marks)**

---

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

It shows a time series graph.

The table below and time series graph give information about the number, in thousands, of visitors to the British Museum each quarter from Quarter 1, 2016 to Quarter 3, 2017

	Quarter			
Year	1	2	3	4
2016	1490	1670	1800	1460
2017	1300	1650	1690	

(continued on the next page)

8. continued.

Four of the five 4–point moving averages for the data are also plotted on the grid.

- (a) Explain why it is appropriate to calculate 4–point moving averages for this set of data.  
(1 mark)

---

---

---

The eighth 4–point moving average for the data is 1525

- (b) Plot this moving average on the grid.  
(1 mark)

(continued on the next page)

8. continued.

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

The average seasonal effect for Quarter 4 for the years 2015 and 2016 is found to be **−140 thousand**.

- (d) (i) Give an interpretation of this value.  
(1 mark)

---

---

---

(continued on the next page)



8. (d) continued.

- (ii) Calculate an estimate of the number of  
visitors to the British Museum in  
Quarter 4, 2017  
(2 marks)

\_\_\_\_\_ thousand

(continued on the next page)

**8. continued.**

**(e) Write down two assumptions that have to be made so that your estimate in part (d)(ii) is valid.**

**(2 marks)**

**1** \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**2** \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**(Total for Question 8 is 8 marks)**

---

9. Look at the table for Question 9 in the Data Book.  
To apply for a job at a bank, an applicant must first take a numeracy test.

The table gives information about the results of the test taken by all applicants.

To get an interview, an applicant must score a standardised score of at least  $1.5$  for the test.

Mithra scored **68** in the test.

- (a) Determine whether or not Mithra gets an interview.

(3 marks)

Answer space continues on the next page.

9. (a) continued.

(continued on the next page)

9. continued.

Alexi's standardised score for the test is  $-1.25$

Fiona's standardised score for the test is  $-1$

(b) Compare Alexi's performance in the test with  
Fiona's performance in the test.

Give a reason for your answer.

(2 marks)

---

---

---

---

---

---

(Total for Question 9 is 5 marks)

---

**10. Look at the information and table for Question 10 in the Data Book.**

**Noah is investigating the reaction times of some athletes.**

**Noah decides to clean the data before proceeding with his investigation.**

**He models the reaction times using a normal distribution.**

**Using the summary statistics, describe how Noah should clean the data and justify why he needs to clean the data.**

**(6 marks)**

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

---

---

---

---

---

---

---

**10. continued.**

---

---

---

---

---

---

---

---

---

---

**(Total for Question 10 is 6 marks)**

---

11. Gisele collected data about the age and the salary of each employee at a small company.

She used statistical software to draw a scatter diagram for her data.

- (a) Give one advantage of using statistical software when representing data.

(1 mark)

---

---

---

(continued on the next page)



11. continued.

Gisele calculated correlation coefficients for her data.

She obtained the following results.

<b>Spearman's rank correlation coefficient</b>	<b>0·95</b>
<b>Pearson's product moment correlation coefficient</b>	<b>0·77</b>

(b) (i) Describe and interpret the type of correlation represented by **0·95** in the table.

(2 marks)

---

---

---

---

---

---

(continued on the next page)

Turn over

**11. (b) continued.**

**(ii) Which of the two correlation coefficients in the table represents the stronger correlation?**

**You must give a reason for your answer.**

**(1 mark)**

---

---

---

**(continued on the next page)**

**11. continued.**

**Look at the diagrams for Question 11(c) in the Data Book.**

**Diagram 1 and Diagram 2 show two possible scatter diagrams for Gisele's data.**

**There are ten points plotted on each graph.**

**(c) Which one of these two diagrams most likely represents the data?**

**You must give a reason for your answer.**

**(2 marks)**

---

---

---

---

---

---

**(continued on the next page)**

**11. continued.**

**Brett wants to use a Pearson's product moment correlation coefficient (PMCC) to compare the salaries of male employees with the salaries of female employees.**

**(d) Explain whether or not it is appropriate to use the PMCC to make this comparison.**

**(1 mark)**

---

---

---

---

---

**(Total for Question 11 is 7 marks)**

---

**12. Jasper has 3 coins.**

**In an experiment, Jasper flips each of the 3 coins and records the total number of heads that he gets.**

**Jasper believes that each coin is biased so that the number of heads he gets can be modelled by the binomial distribution,  $B(3, 0.4)$**

**(a) Show that**

$$\mathbf{P(0 \text{ heads}) = 0.216}$$

**(1 mark)**

**(continued on the next page)**

**12. continued.**

**(b) Work out the probability that the outcome of the experiment is exactly 1 head.**

**(2 marks)**

---

**(continued on the next page)**

**12. continued.**

**Look at the table for Question 12(c) in the Data Book.**

**Jasper carries out the experiment 100 times.**

**The table shows information about the number of heads he gets for each experiment.**

**(c) Determine whether or not the model  $B(3, 0.4)$  is suitable for Jasper's experiment.**

**(5 marks)**

**Answer space continues on the next page.**

12. (c) continued.

(Total for Question 12 is 8 marks)

---

**TOTAL FOR PAPER IS 80 MARKS**

**END OF PAPER**

---