adamjee maths

A-LEVEL MATHS / EDEXCEL / FREE RESOURCES

Sampling

Question paper

14 questions

Exam questions and marking instructions remain credited to their board.

Contents

1 June 2025 - Paper 3 Statistics - Q3
2 June 2025 - Paper 3 Statistics - Q4
3 June 2024 - Paper 3 Statistics - Q3
4 June 2024 - Paper 3 Statistics - Q4
5 June 2023 - Paper 3 Statistics - Q3
6 June 2023 - Paper 3 Statistics - Q4
7 June 2022 - Paper 3 Statistics - Q3
8 June 2022 - Paper 3 Statistics - Q6
9 October 2021 - Paper 3 Statistics - Q1
10 October 2021 - Paper 3 Statistics - Q2
11 October 2021 - Paper 3 Statistics - Q3
12 October 2021 - Paper 3 Statistics - Q5
13 October 2020 - Paper 3 Statistics - Q2
14 June 2018 - Paper 3 - Q1

3. A manufacturer makes components for the car industry.

On average 1.5% of the components made are defective.

(a) Using a suitable model, find the probability that in a random sample of 36 components there are

(i) exactly 2 defective components,

(ii) more than 3 defective components.

(3)

Once every six months, the manufacturer tests whether the proportion of defective components has changed using hypotheses $H_0 : p = 0.015$ and $H_1 : p \neq 0.015$

A random sample of 260 components is taken and 8 are found to be defective.

(b) Using a 5% level of significance, complete the test.

(3)

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

Question 3 continued

(Total for Question 3 is 6 marks)

DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

4. Kay is studying the variables Daily Total Sunshine (x) and Daily Total Rainfall (y) from the large data set for Leeming in 2015

Kay starts with 5th May and then selects every 10th day thereafter.

(a) State the name of the sampling technique Kay uses.

(1)

Kay wants to find the regression line of y on x for these data.

(b) Using your knowledge of the large data set, explain how Kay might need to clean these data before finding the equation of the regression line.

(1)

The equation of the regression line Kay finds is $y = 0.741 + 0.199x$

(c) Using your knowledge of the large data set,

(i) state the units of the gradient of the regression line,

(ii) give an interpretation of the y-intercept of the regression line.

(2)

Kay's teacher claimed that the greater the amount of sunshine in a day the lower the amount of rain there should be.

(d) State, giving a reason, whether the teacher's claim is true for Kay's data.

(1)

The teacher used all the data for these variables from the large data set for Leeming in 2015, as a sample.
The teacher calculated the product moment correlation coefficient for x and y to be -0.160

In a suitable test to determine whether there is evidence to support the teacher's claim, the p-value was 0.015

(e) Using a 5% level of significance, state the hypotheses and conclusion for this test.

(2)

(f) For the test in part (e) describe

(i) the sample,

(ii) a possible population.

(2)

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

Question 4 continued

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

Question 4 continued

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

Question 4 continued

(Total for Question 4 is 9 marks)

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

3. Ming is studying the large data set for Perth in 2015

He intended to use all the data available to find summary statistics for the Daily Mean Air Temperature, x °C.
Unfortunately, Ming selected an incorrect variable on the spreadsheet.
This incorrect variable gave a mean of 5.3 and a standard deviation of 12.4

(a) Using your knowledge of the large data set, suggest which variable Ming selected.

(1)

The correct values for the Daily Mean Air Temperature are summarised as

$n = 184$ $\quad \sum x = 2801.2$ $\quad \sum x^2 = 44\,695.4$

(b) Calculate the mean and standard deviation for these data.

(3)

One of the months from the large data set for Perth in 2015 has

- mean $\bar{x} = 19.4$
- standard deviation $\sigma_x = 2.83$

for Daily Mean Air Temperature.

(c) Suggest, giving a reason, a month these data may have come from.

(2)

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

Question 3 continued

(Total for Question 3 is 6 marks)

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

4. The proportion of left-handed adults in a country is 10%
Freya believes that the proportion of left-handed adults under the age of 25 in this country is different from 10%
She takes a random sample of 40 adults under the age of 25 from this country to investigate her belief.

(a) Find the critical region for a suitable test to assess Freya's belief.
You should
- state your hypotheses clearly
- use a 5% level of significance
- state the probability of rejection in each tail

(4)

(b) Write down the actual significance level of your test in part (a)

(1)

In Freya's sample 7 adults were left-handed.

(c) With reference to your answer in part (a) comment on Freya's belief.

(1)

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

Question 4 continued

(Total for Question 4 is 6 marks)

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

3. Ben is studying the Daily Total Rainfall, x mm, in Leeming for 1987

He used all the data from the large data set and summarised the information in the following table.

x	0	0.1–0.5	0.6–1.0	1.1–1.9	2.0–4.0	4.1–6.9	7.0–12.0	12.1–20.9	21.0–32.0	tr
Frequency	55	18	18	21	17	9	9	6	2	29

(a) Explain how the data will need to be cleaned before Ben can start to calculate statistics such as the mean and standard deviation.

(2)

Using all 184 of these values, Ben estimates $\sum x = 390$ and $\sum x^2 = 4336$

(b) Calculate estimates for

(i) the mean Daily Total Rainfall,

(ii) the standard deviation of the Daily Total Rainfall.

(3)

Ben suggests using the statistic calculated in part (b)(i) to estimate the annual mean Daily Total Rainfall in Leeming for 1987

(c) Using your knowledge of the large data set,

(i) give a reason why these data would not be suitable,

(ii) state, giving a reason, how you would expect the estimate in part (b)(i) to differ from the actual annual mean Daily Total Rainfall in Leeming for 1987

(2)

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

Question 3 continued

(Total for Question 3 is 7 marks)

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

4. A study was made of adult men from region *A* of a country.
It was found that their heights were normally distributed with a mean of 175.4 cm and standard deviation 6.8 cm.

(a) Find the proportion of these men that are taller than 180 cm.

(1)

A student claimed that the mean height of adult men from region *B* of this country was different from the mean height of adult men from region *A*.

A random sample of 52 adult men from region *B* had a mean height of 177.2 cm

The student assumed that the standard deviation of heights of adult men was 6.8 cm both for region *A* and region *B*.

(b) Use a suitable test to assess the student's claim.
You should

- state your hypotheses clearly
- use a 5% level of significance

(4)

(c) Find the *p*-value for the test in part (b)

(1)

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

Question 4 continued

(Total for Question 4 is 6 marks)

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

3. Dian uses the large data set to investigate the Daily Total Rainfall, r mm, for Camborne.

(a) Write down how a value of $0 < r \leqslant 0.05$ is recorded in the large data set.

(1)

Dian uses the data for the 31 days of August 2015 for Camborne and calculates the following statistics

$$n = 31 \qquad \sum r = 174.9 \qquad \sum r^2 = 3523.283$$

(b) Use these statistics to calculate

(i) the mean of the Daily Total Rainfall in Camborne for August 2015,

(ii) the standard deviation of the Daily Total Rainfall in Camborne for August 2015.

(3)

Dian believes that the mean Daily Total Rainfall in August is less in the South of the UK than in the North of the UK.
The mean Daily Total Rainfall in Leuchars for August 2015 is 1.72 mm to 2 decimal places.

(c) State, giving a reason, whether this provides evidence to support Dian's belief.

(2)

Dian uses the large data set to estimate the proportion of days with no rain in Camborne for 1987 to be 0.27 to 2 decimal places.

(d) Explain why the distribution B(14, 0.27) might **not** be a reasonable model for the number of days without rain for a 14-day summer event.

(1)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Question 3 continued

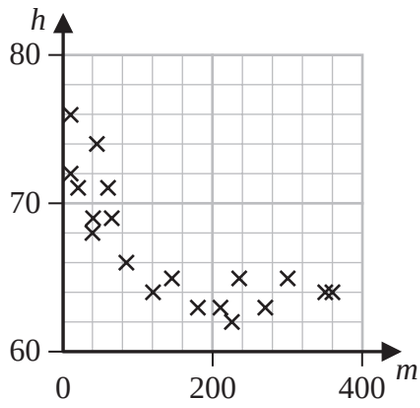
(Total for Question 3 is 7 marks)

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

6. Anna is investigating the relationship between exercise and resting heart rate. She takes a random sample of 19 people in her year at school and records for each person

- their resting heart rate, h beats per minute
- the number of minutes, m, spent exercising each week

Her results are shown on the scatter diagram.



(a) Interpret the nature of the relationship between h and m

(1)

Anna codes the data using the formulae

$$x = \log_{10} m$$
$$y = \log_{10} h$$

The product moment correlation coefficient between x and y is -0.897

(b) Test whether or not there is significant evidence of a negative correlation between x and y
You should

- state your hypotheses clearly
- use a 5% level of significance
- state the critical value used

(3)

The equation of the line of best fit of y on x is

$$y = -0.05x + 1.92$$

(c) Use the equation of the line of best fit of y on x to find a model for h on m in the form

$$h = am^k$$

where a and k are constants to be found.

(5)

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

Question 6 continued

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

Question 6 continued

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

Question 6 continued

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

Question 6 continued

(Total for Question 6 is 9 marks)

TOTAL FOR STATISTICS IS 50 MARKS

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

1. (a) State one disadvantage of using quota sampling compared with simple random sampling.

(1)

In a university 8% of students are members of the university dance club.

A random sample of 36 students is taken from the university.

The random variable X represents the number of these students who are members of the dance club.

(b) Using a suitable model for X, find

(i) $P(X = 4)$

(ii) $P(X \geqslant 7)$

(3)

Only 40% of the university dance club members can dance the tango.

(c) Find the probability that a student is a member of the university dance club and can dance the tango.

(1)

A random sample of 50 students is taken from the university.

(d) Find the probability that fewer than 3 of these students are members of the university dance club and can dance the tango.

(2)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Question 1 continued.

(Total for Question 1 is 7 marks)

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

2. Marc took a random sample of 16 students from a school and for each student recorded

- the number of letters, x, in their last name
- the number of letters, y, in their first name

His results are shown in the scatter diagram on the next page.

(a) Describe the correlation between x and y.

(1)

Marc suggests that parents with long last names tend to give their children shorter first names.

(b) Using the scatter diagram comment on Marc's suggestion, giving a reason for your answer.

(1)

The results from Marc's random sample of 16 observations are given in the table below.

x	3	6	8	7	5	3	11	3	4	5	4	9	7	10	6	6
y	7	7	4	4	6	8	5	5	8	4	7	4	5	5	6	3

(c) Use your calculator to find the product moment correlation coefficient between x and y for these data.

(1)

(d) Test whether or not there is evidence of a negative correlation between the number of letters in the last name and the number of letters in the first name.

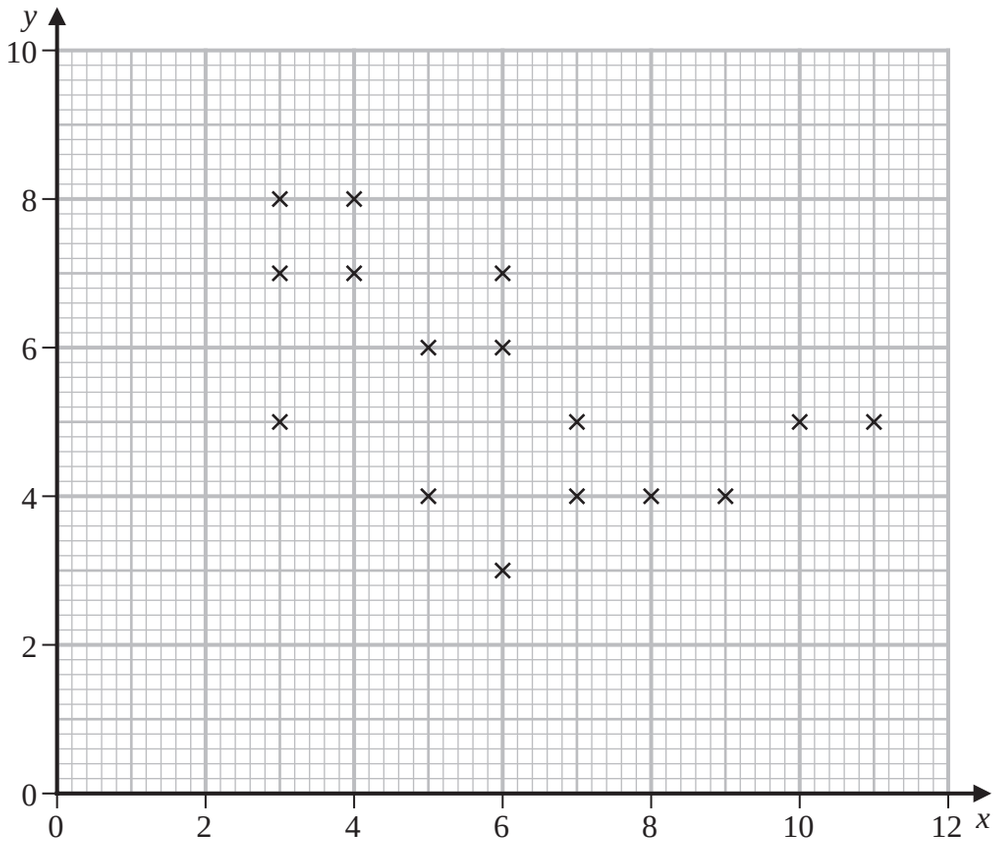
You should

- state your hypotheses clearly
- use a 5% level of significance

(3)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Question 2 continued.



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

Question 2 continued.

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

Question 2 continued.

(Total for Question 2 is 6 marks)

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

3. Stav is studying the large data set for September 2015

He codes the variable Daily Mean Pressure, x, using the formula $y = x - 1010$

The data for all 30 days from Hurn are summarised by

$$\sum y = 214 \qquad \sum y^2 = 5912$$

(a) State the units of the variable x

(1)

(b) Find the mean Daily Mean Pressure for these 30 days.

(2)

(c) Find the standard deviation of Daily Mean Pressure for these 30 days.

(3)

Stav knows that, in the UK, winds circulate

- in a **clockwise** direction around a region of **high** pressure
- in an **anticlockwise** direction around a region of **low** pressure

The table gives the Daily Mean Pressure for 3 locations from the large data set on 26/09/2015

Location	Heathrow	Hurn	Leuchars
Daily Mean Pressure	1029	1028	1028
Cardinal Wind Direction			

The Cardinal Wind Directions for these 3 locations on 26/09/2015 were, in random order,

W NE E

You may assume that these 3 locations were under a single region of pressure.

(d) Using your knowledge of the large data set, place each of these Cardinal Wind Directions in the correct location in the table.
Give a reason for your answer.

(2)

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Question 3 continued.

(Total for Question 3 is 8 marks)

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

5. The heights of females from a country are normally distributed with

- a mean of 166.5 cm
- a standard deviation of 6.1 cm

Given that 1% of females from this country are shorter than k cm,

(a) find the value of k

(2)

(b) Find the proportion of females from this country with heights between 150 cm and 175 cm

(1)

A female, from this country, is chosen at random from those with heights between 150 cm and 175 cm

(c) Find the probability that her height is more than 160 cm

(4)

The heights of females from a different country are normally distributed with a standard deviation of 7.4 cm

Mia believes that the mean height of females from this country is less than 166.5 cm

Mia takes a random sample of 50 females from this country and finds the mean of her sample is 164.6 cm

(d) Carry out a suitable test to assess Mia's belief.
You should

- state your hypotheses clearly
- use a 5% level of significance

(4)

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

Question 5 continued.

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

Question 5 continued.

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

Question 5 continued.

(Total for Question 5 is 11 marks)

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

2. A random sample of 15 days is taken from the large data set for Perth in June and July 1987.

The scatter diagram in Figure 1 displays the values of two of the variables for these 15 days.

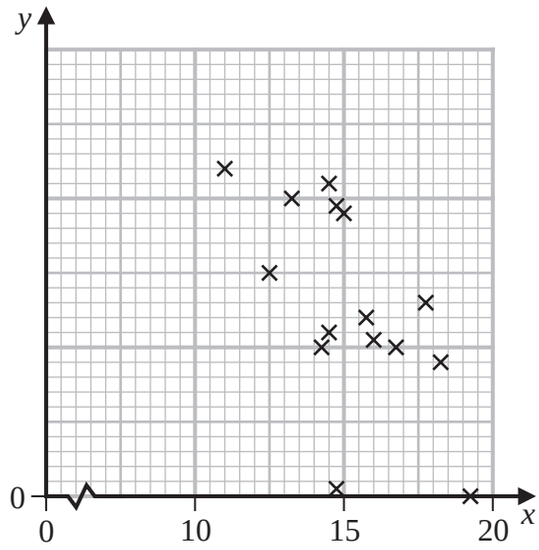


Figure 1

(a) Describe the correlation.

(1)

The variable on the x-axis is Daily Mean Temperature measured in °C.

(b) Using your knowledge of the large data set,

(i) suggest which variable is on the y-axis,

(ii) state the units that are used in the large data set for this variable.

(2)

Stav believes that there is a correlation between Daily Total Sunshine and Daily Maximum Relative Humidity at Heathrow.

He calculates the product moment correlation coefficient between these two variables for a random sample of 30 days and obtains $r = -0.377$

(c) Carry out a suitable test to investigate Stav's belief at a 5% level of significance. State clearly

- your hypotheses
- your critical value

(3)

On a random day at Heathrow the Daily Maximum Relative Humidity was 97%

(d) Comment on the number of hours of sunshine you would expect on that day, giving a reason for your answer.

(1)

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

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

Question 2 continued

Question 2 continued

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

Question 2 continued

(Total for Question 2 is 7 marks)

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

SECTION A: STATISTICS

Answer ALL questions. Write your answers in the spaces provided.

1. Helen believes that the random variable C, representing cloud cover from the large data set, can be modelled by a discrete uniform distribution.

(a) Write down the probability distribution for C.

(2)

(b) Using this model, find the probability that cloud cover is less than 50%

(1)

Helen used all the data from the large data set for Hurn in 2015 and found that the proportion of days with cloud cover of less than 50% was 0.315

(c) Comment on the suitability of Helen's model in the light of this information.

(1)

(d) Suggest an appropriate refinement to Helen's model.

(1)

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

Question 1 continued

(Total for Question 1 is 5 marks)

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