

**Paper Reference 9MA0/31**  
**Pearson Edexcel Level 3 GCE**

**Mathematics**  
**Advanced**  
**PAPER 31: Statistics**

**Monday 18 October 2021 – Afternoon**

**YOU MUST HAVE**  
**Mathematical Formulae and Statistical Tables (Green),**  
**calculator**

**YOU WILL BE GIVEN**  
**Data Booklet**  
**Answer Booklet**

**X68828A**

**Candidates may use any calculator allowed by Pearson regulations. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.**

## **INSTRUCTIONS**

**In the boxes on the Answer Booklet and on the Data Booklet, write your name, centre number and candidate number.**

**Answer ALL questions and ensure that your answers to parts of questions are clearly labelled.**

**Answer the questions in the spaces provided in the Answer Booklet or on the separate data sheets – there may be more space than you need.**

**Do NOT write on the Question Paper.**

**You should show sufficient working to make your methods clear. Answers without working may not gain full credit.**

**Values from statistical tables should be quoted in full. If a calculator is used instead of tables the value should be given to an equivalent degree of accuracy.**

**Inexact answers should be given to three significant figures unless otherwise stated.**

**Turn over**

**INFORMATION**

**A booklet ‘Mathematical Formulae and Statistical Tables’ is provided.**

**The total mark for this part of the examination is 50  
There are 6 questions.**

**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.**

---

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

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)  $P(X \geq 7)$

(3 marks)

(continued on the next page)

Turn over

1. continued.

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 mark)

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 marks)

(Total for Question 1 is 7 marks)

---

2. Refer to the diagram for Question 2 in the Data Booklet.

It shows a scatter diagram.

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 in the Data Booklet.

- (a) Describe the correlation between  $x$  and  $y$   
(1 mark)

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

- (b) Using the scatter diagram in the Data Booklet, comment on Marc's suggestion, giving a reason for your answer.  
(1 mark)

(continued on the next page)

Turn over

**2. continued.**

**Refer to the table for Question 2(c) in the Data Booklet.**

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

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

**(1 mark)**

**(continued on the next page)**

**2. continued.**

- (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 marks)**

**(Total for Question 2 is 6 marks)**

---



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

$$\sum y = 214$$

$$\sum y^2 = 5912$$

- (a) State the units of the variable  $x$   
(1 mark)
- (b) Find the mean Daily Mean Pressure for these 30 days.  
(2 marks)

(continued on the next page)

Turn over

**3. continued.**

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

**(3 marks)**

**Refer to the table for Question 3(d) in the Data Booklet.**

**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 in the Data Booklet gives the Daily Mean Pressure for 3 locations from the large data set on 26/09/2015**

**(continued on the next page)**

**3. continued.**

**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.**

**There are three spaces to fill.**

**Give a reason for your answer.**

**(2 marks)**

**(Total for Question 3 is 8 marks)**

---

4. Refer to the diagram for Question 4 in the Data Booklet.

It shows a Venn diagram.

A large college produces three magazines.

One magazine is about green issues, one is about equality and one is about sports.

A student at the college is selected at random and the events **G**, **E** and **S** are defined as follows

**G** is the event that the student reads the magazine about green issues

**E** is the event that the student reads the magazine about equality

**S** is the event that the student reads the magazine about sports

The Venn diagram in the Data Booklet, where **p**, **q**, **r** and **t** are probabilities, gives the probability for each subset.

(continued on the next page)

4. continued.

(a) Find the proportion of students in the college who read exactly one of these magazines.

(1 mark)

No students read all three magazines and

$$P(G) = 0.25$$

(b) Find

(i) the value of  $p$

(ii) the value of  $q$

(3 marks)

(continued on the next page)

4. continued.

Given that  $P(S | E) = \frac{5}{12}$

(c) find

(i) the value of  $r$

(ii) the value of  $t$

(4 marks)

(d) Determine whether or not the events  $(S \cap E')$  and  $G$  are independent.

Show your working clearly.

(3 marks)

(Total for Question 4 is 11 marks)

---

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 marks)

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

(continued on the next page)

**5. continued.**

**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 marks)**

**(continued on the next page)**



**5. continued.**

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 marks)**

**(Total for Question 5 is 11 marks)**

---

6. Refer to the table for Question 6 in the Data Booklet.  
The discrete random variable  $X$  has the probability distribution shown in the Data Booklet.

(a) Find

(i) the value of  $a$

(ii) the value of  $b$

(iii) the value of  $c$

Show your working clearly.

(5 marks)

(continued on the next page)

6. continued.

The independent random variables  $X_1$  and  $X_2$  each have the same distribution as  $X$

(b) Find  $P(X_1 = X_2)$   
(2 marks)

(Total for Question 6 is 7 marks)

---

**TOTAL FOR STATISTICS IS 50 MARKS**

**END OF PAPER**

---