Please check the examination details be	low before ente	ering your candidate information
Candidate surname		Other names
Centre Number Candidate N		
Pearson Edexcel Leve	I 3 GCE	
Monday 18 October 2021 – After	rnoon	
	Paper reference	9MA0/31
Mathematics		
Advanced PAPER 31: Statistics		
You must have: Mathematical Formulae and Statisti	cal Tables (G	reen), calculator

### 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

- Use black ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- Fill in the boxes at the top of this page with 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 – there may be more space than you need.
- 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.

### 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.

### 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.

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)  $P(X \ge 7)$

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.

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.

cannot be quota sampling isn't random it If used to make inferences. The probability of a student being in the club - 0.08. The number of students sampled = 36.  $X \sim R(36, 0.08)$  $_{50}$  P(X=4) = 0.167387... ) = 1 - P(X < 6) = 0.22233P(In club and dance trango)= 0.4×0.08=0.032. = Members of the club who can dance tango T~B(50,0.032)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(1)

(3)

(1)

(2)

Question 1 continued.

P(T < 3) = 0.7850815

(Total for Question 1 is 7 marks)



- 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.

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)

(1)

(3)

(1)

DO NOT WRITE IN THIS AREA

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.
- (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

) Negative Correlation. b) As the scatter diagram has a negative correlation, it is compatible with Marc's suggestion. c) Zx=97, Zy=88, Zx2=681, Zy2=520 Zixy= 502, n=16  $r = \frac{502 - (97 \cdot 88)/16}{\sqrt{681 - 97^2}, (520 - \frac{38^2}{16})}$ 54458266 ... 4

DO NOT WRITE IN

DO NOT WRITE IN THIS AREA



Question 2 continued	
Question 2 continued.	0
	NO
	OT
	WRI
	TE
	TN
	1 Alex
	ARE
	A
	X
	132
	Q
	NO
	WTC
	RIT
	E S
	E
	IS A
	REA
	12
	18
	125
	00
	ZO
	WIN
	RIT
	1.5
	H
	2
	RE
	1.1
	13
	12
	e du c

# Question 2 continued.

	Question 2 continued.
2	
Ē	
2	
L.	
NN	
5	
Ž	
ă	
V	
A BE	
2	
H	
Z	
1	
N	
5	
0	
0	
5	
19	
2	
1	
Z	
111	
M	
10	
NO	
0	
	(Total for Question 2 is 6 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

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

(a) State the units of the variable x

(1)

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE

AREA

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

(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	NE	E	W

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)

a) Hectopascal (hPa)  
b) 
$$\overline{x} = \overline{y} + 1010$$
,  $\overline{y} = 214/30$   
 $\overline{x} = 1017 \cdot 13$   
c) As  $y = x - 1010$ ,  $O_{xc} = O_{y}$ 



Question 3 continued.

As all of these pressures are nearly a standard deviation from the mean, we com Consider them high pressures. So the wird is circulating clockwise. The locations from north to south are Leuchars, Heathrow, Hurn. So the directions of their winds is as written on the table. (Total for Question 3 is 8 marks) 9 Turn over 🕨

 $\frac{5912}{30} - \left(\frac{214}{30}\right)^2 = 12.0905... = O_{\infty}.$ 

DO NOT WHITE IN THIS AREA

4. 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, where p, q, r and t are probabilities, gives the probability for each subset.



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

No students read all three magazines and P(G) = 0.25

(b) Find

- (i) the value of p
- (ii) the value of q

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

- (c) find
  - (i) the value of r

(ii) the value of t

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

a) 0.08+ 0.09+0.36=0.53



(1)

(3)

(4)

(3)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

NO NOT WRITE IN THIS AREA

Question 4 continued.

bi) As no student reads all 3 magazines, P(GnEnS) = p = 0.  $i:) P(G \cap S) = q$ P(G): P(GnS)+P(GnE)+P(JustG) 0.25=9+0.05+0.08=> q = 0.12ci)  $P(S|E) = 5/12 = P(S \cap E)$ P(E) = r + p + 0.05 + 0.09 = 0.14 + r $P(S \cap E) = r + p = r$  $\frac{5}{12} = \frac{1}{0.14 + r}$ =>  $12r = 5r + 5 \times 0.14$ => 4r = 0.7=> r = 0.1i) 0.08+0.05+0.09+9+p+r+0.36+t=1 q=0.12, p=0, r=0.1 => 0.8+t=1=> t=0.2.

11

Turn over 🕨

Question 4 continued. d)  $P(s \cap E') = 0.36 \cdot q = 0.36 + 0.12$ = 0.48. P((SnE')nG) = q = 0.12P(G) = 0.25 $P(S \cap E') \times P(G) = 0.48 \times 0.25 = 0.12$  $=P((S \cap E') \cap G)$ Hence they are independent.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 4 continued		
e e andre e a a a a a a a a a a a a a a a a a a		
	(Total for Question 4 is	s 11 marks)
and the second second second second second second second		

10	13C.CO.uk
5.	<ul> <li>The heights of females from a country are normally distributed with</li> <li>a mean of 166.5 cm</li> <li>a standard deviation of 6.1 cm</li> </ul>
	Given that 1% of females from this country are shorter than k cm,
	(a) find the value of $k$
	(b) Find the proportion of females from this country with heights between 150 cm and 175 cm
	A female, from this country, is chosen at random from those with heights between $150 \text{ cm}$ and $175 \text{ cm}$
	(c) Find the probability that her height is more than 160 cm
	The heights of females from a different country are normally distributed with a stand 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 .

a) Let 
$$F \sim N(166.5, 6.1^2)$$
  
 $P(F < k) = 0.01$   
 $= > \frac{K - 166.5}{6.1} = -2.3263$   
 $= > K = 152.3 \text{ cm}.$   
 $b) P(150 < F < 175) = 0.914840...$ 



(4)

(2)

(1)

(4)

standard

DO NOT WRITE IN THIS AREA

Question 5 continued. c) P(F>160 150<F<175)  $= \frac{P(160 < F < 175)}{P(150 < F < 175)}$ P(160 < F < 175) = 0.7749487...=>P(F& 60/150< F<175) = 0.7749487... = 0.84708 ... Null hypothesis: Ho: M=166.5. M: as hypothesis: H1: M<166.5. het X=height from this country  $\overline{X} \sim N\left(166.5, \left(\frac{7.4}{\sqrt{51}}\right)^2\right)$ P(X<164.6)=0.03472. <0.05, so we can reject Ho. => There is evidence to support Mia's belief.

OO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Mastion 5 and in 1	
acsuon 5 continued.	

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

6. The discrete random variable X has the following probability distribution

x	а	b	С
P(X = x)	$\log_{36} a$	$\log_{36} b$	$\log_{36} c$

where

- a, b and c are distinct integers (a < b < c)
- all the probabilities are greater than zero

(a) Find

- (i) the value of a
- (ii) the value of b
- (iii) the value of c

Show your working clearly.

(5)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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

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

a) 
$$log_{36}(a) + log_{36}(b) + log_{36}(c) = 1$$
.  
=>  $log_{36}(abc) = 1$ .  
=>  $abc = 36$ .  
As all the probabilities are greater than 0,  
 $1 \le a, b, c$ .  
Prime factors of  $36 = 2 \times 2 \times 3 \times 3$ .  
As we need 3 distinct integers that multiply  
to 36 we get  
 $36 = 2 \times 3 \times 6$   
=>  $a = 2$ ,  $b = 3$ ,  $c = 6$  (as  $acb = c$ )



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 6 continued.

b) 3 ways they can be equal:  $P(X_1 = X_2 = a) = (log_{36} a)^2 [P(X_1 = a) \times P(X_2 = a)]$  $P(X_1 = X_2 = b) = (log_{36} b)^2, [P(X_1 = b) \times P(X_1 = b)]$  $P(X_1 = X_2 = C) = (\log_{36} C)^2 [P(X_1 = C) \times P(X_2 = C)]$ So P(X,=X2)=(10936a)2+(10936b)2+(10936c)2 =0.381401 ... 19 

oro		$\sim$	111/2
EE	VISE.	LU.	un
 0.0			

(Total for Question 6 is 7 marks)	T	OTAL FOR STATISTICS IS 50 MARKS
		(Total for Question 6 is 7 marks)