Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number

Candidate Number

Pearson Edexcel Level 3 GCE

Monday 18 October 2021 – Afternoon

Paper reference **9MA0/31**

Mathematics

Advanced
PAPER 31: Statistics

You must have:
Mathematical Formulae and Statistical Tables (Green), calculator

Total Marks

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.

Turn over

P68828A

A:1/1/1/1/1/

P 6 8 8 2 8 A 0 1 2 0

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

(1)

p = 0.08

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

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

n = 36

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.

"and" ↳ multiply

(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. ∴ use p = 0.032

(2)

(a) It's not random and it might be biased B1

(b) $X \sim B(36, 0.08)$ M1

i. $P(X=4) = 0.167387 \longrightarrow 0.167$ to 3sf. A1

ii. $P(X \geq 7) = 1 - P(X \leq 6) = 0.022233 \longrightarrow 0.0222$ to 3sf A1

include 7 ∴ 7 - 1 = 6

(c) P(dance club and tango) = $0.08 \times 0.4 = 0.032$ B1

"and"

(d) T → people who dance tango define variable

$T \sim B(50, 0.032)$ from (c) M1

$P(T < 3) = P(T \leq 2) = 0.78508 \longrightarrow 0.785$ to 3sf A1

Question 1 continued.

(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.

(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)

(a) Weak Negative B1

(b) The suggestion is valid since there's negative correlation B1

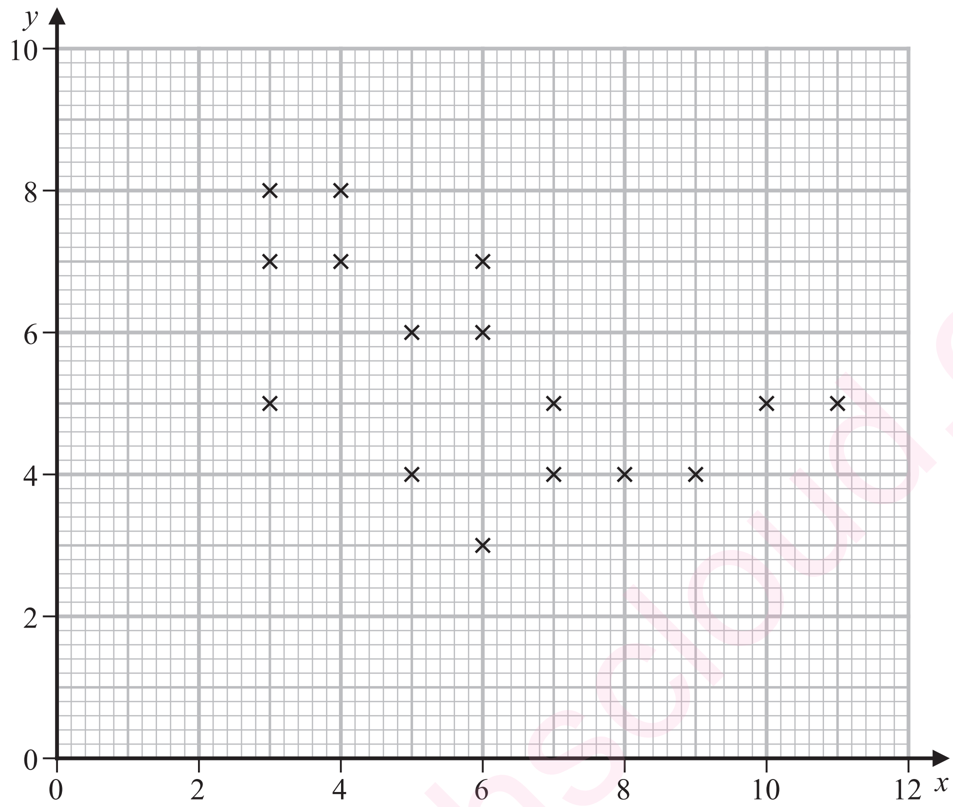
(c) $r = -0.54458 \longrightarrow -0.545$ to 3sf (use 'regression on your calculator) B1

(d) Hypotheses
$H_0: \rho = 0$
$H_1: \rho < 0$ B1

Critical value from tables: -0.4259 M1
as $-0.545 < -0.459$, there is sufficient evidence to reject H_0 ∴ there is evidence of negative correlation between the length of students' first and last names. A1

P 6 8 8 2 8 A 0 4 2 0

Question 2 continued.



P 6 8 8 2 8 A 0 5 2 0

Question 2 continued.

Question 2 continued.

(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)**

(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	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) The units for daily mean pressure are Hectopascal B1

(b) Formula for mean: $\bar{x} = \frac{\Sigma x}{n}$

We are looking for $\bar{x}$, however the data is coded using $y=x-1010$.

★ Coding

Mean: both +/- and ×/÷ applied

SD: only ×/÷ applied

$\bar{y} = \bar{x} - 1010$

$\bar{x} = \bar{y} + 1010$

$\bar{y} = \frac{\Sigma y}{n} = \frac{214}{30}$

$\bar{x} = \frac{214}{30} + 1010 = 1017.133 \rightarrow 1017$ M1A1

Question 3 continued.

(c) This coding does not affect the standard deviation, σ_x M1

Formula for SD: $\sigma_x = \sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2}$

$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

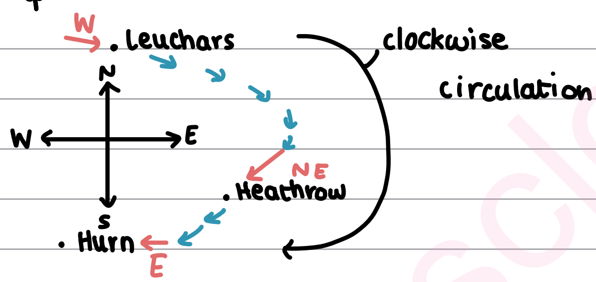
$= \sqrt{\frac{5912}{30} - \left(\frac{214}{30}\right)^2}$

$= \sqrt{146.18...}$ M1

$= 12.0905 \longrightarrow$ 12.1 to 3sf A1

(d) Pressure is high ∴ clockwise direction B1

On the map:



∴ Blows FROM :

Leuchars : W

Heathrow : NE

Hurn : E B1

(Total for Question 3 is 8 marks)

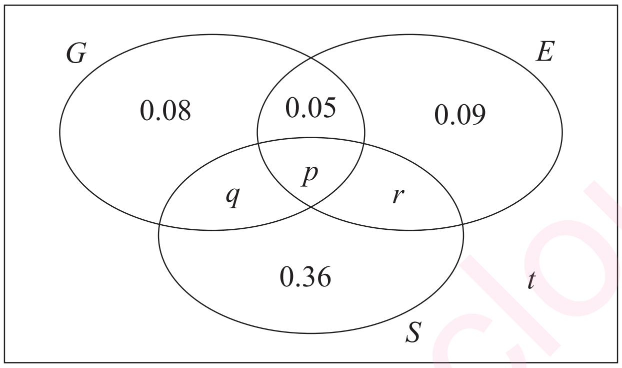
P 6 8 8 2 8 A 0 9 2 0

Turn over

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.
(1)

No students read all three magazines and $\mathrm{P}(G) = 0.25$

(b) Find

(i) the value of p

(ii) the value of q
(3)

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

(c) find

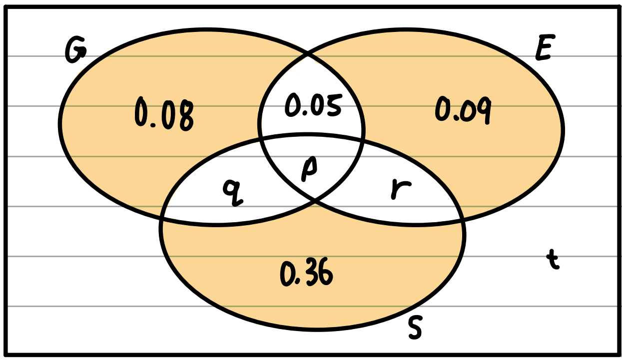
(i) the value of r

(ii) the value of t
(4)

(d) Determine whether or not the events $(S \cap E')$ and G are independent.
Show your working clearly.
(3)

Question 4 continued.

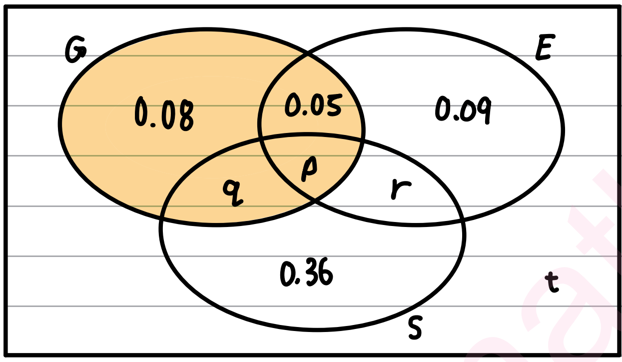
(a) Only 1 magazine → no intersections



P(only 1) = 0.08 + 0.09 + 0.36
= 0.53 B1

(b) i. "No students read all 3 magazines" ∴ $P(G \cap E \cap S) = 0$, p = 0 B1

ii.



$P(G) = 0.25$

$\therefore 0.08 + 0.05 + \cancel{p} + q = 0.25$ (p = 0) M1

$0.13 + q = 0.25$

$q = 0.12$ A1

(c) i. $P(S \mid E) = \frac{P(S \cap E)}{P(E)} = \frac{\cancel{p} + r}{0.05 + 0.09 + \cancel{p} + r} = \frac{5}{12}$ (p = 0) M1 A1

$\frac{r}{0.14 + r} = \frac{5}{12}$ ↓ solve for r

$12r = 5(0.14 + r)$

$12r = 0.7 + 5r$

$7r = 0.7 \rightarrow r = 0.1$ A1

ii. $\sum$ all probabilities = 1

$\therefore 0.08 + 0.05 + 0.09 + q + p + r + 0.36 + t = 1$

$0.58 + 0.12 + 0 + t + 0.1 = 1$ ← substitute from above

$0.8 + t = 1 \rightarrow t = 0.2$ B1

Question 4 continued.

(d) When two events are independent:

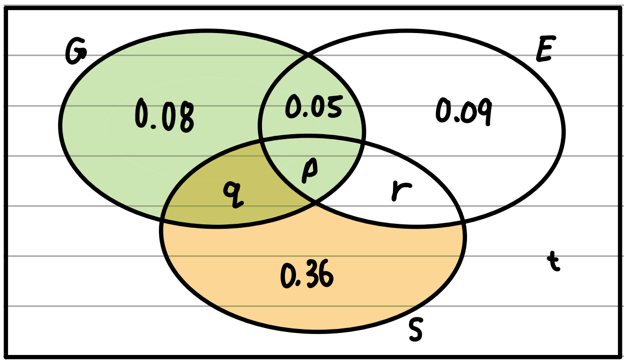
$P(A \cap B) = P(A) \times P(B)$

$\therefore$ $P((S \cap E') \cap G) = P(S \cap E') \times P(G)$ must be true:

$P(G) = 0.25$

$P(S \cap E') = 0.12 + 0.36 = 0.48$ B1

$P(G) \times P(S \cap E') = 0.25 \times 0.48 = 0.12$ M1



$P((S \cap E') \cap G)$:
$P(S \cap E')$ in orange
$P(G)$ in green
Their intersection $\rightarrow q = 0.12$

$\therefore$

Hence, as $P(G) \times P(S \cap E') = 0.12$ and $P((S \cap E') \cap G) = 0.12$,
$0.12 = P((S \cap E') \cap G) = P(S \cap E') \times P(G)$ is true

$P(S \cap E')$ and $P(G)$ are independent A1

Question 4 continued.

(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 μ
- a standard deviation of 6.1 cm σ

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 "given that.."

(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

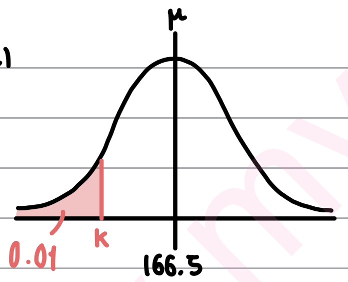
Mia takes a random sample of 50 females from this country and finds the mean of her sample is 164.6 cm

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

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

(4)

(a)



to get k:

InvN(0.01) with parameters N(166.5, 6.1²) M1

InvN(0.01) = 152.309

→ k = 152.3 to 1dp A1

(b) P(150 < F < 175) = 0.91480 → 0.915 to 3sf B1

(c) P(F > 160 | 150 < F < 175) = $\frac{P(160 < F < 175)}{P(150 < F < 175)} = \frac{0.7749487...}{0.91480...}$ = 0.84708 → 0.847 to 3sf A1

("given that" M1) M1 (from (b))

P 6 8 8 2 8 A 0 1 4 2 0

Question 5 continued.

(c) This part talks about "mean" ∴ we will use (sample mean) variable

Formula for Sample mean:

$X \sim N(\mu, \sigma^2) \longrightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$X \rightarrow$ height of a female from country 2 define variable

enter $\frac{7.4}{\sqrt{50}} = \sigma$ into your calculator!

Apply the formula.

$X \sim N(166.5, 7.4^2) \longrightarrow \bar{X} \sim N(166.5, \frac{7.4^2}{50})$ M1

found value — significance level

Hypotheses

$H_0: \mu = 166.5$ B1

$H_1: \mu < 166.5$

$P(\bar{X} < 164.6) = 0.03472... < 0.05$ A1 A1

∴ 164.6 does fall in the critical region and there is sufficient evidence to reject H_0.

Her claim is supported A1

Turn over

P 6 8 8 2 8 A 0 1 5 2 0

Question 5 continued.

Question 5 continued.

(Total for Question 5 is 11 marks)

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

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

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)

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) Σ all probabilities $= 1$

M1 $\log_{36}a + \log_{36}b + \log_{36}c = 1$ Use log law: $\log a + \log b = \log a \times b$

$\log_{36}abc = 1$

$abc = 36^1$

A1 $abc = 36$

"all probabilities are bigger than 0" $\therefore a > 1,\ b > 1,\ c > 1$ because if they were < 1, $\log_{36}a$ or $\log_{36}b$ or $\log_{36}c$ would be negative. B1

Which combinations of 3 distinct numbers give 36?

$36 = 2^2 \times 3^2$ dM1

$2 \times 2 \times 9 = 36$ ✗ not distinct!

$36 = 2 \times 3 \times 6$ $\therefore$ as $a < b < c$:

i. $a = 2$

ii. $b = 3$

iii. $c = 6$ A1

(b) For $X_1 = X_2$ both outcomes would have to be the same, e.g. $P(X=a) \times P(X=a)$.

$\therefore P(X_1 = X_2) = (\log_{36}a)^2 + (\log_{36}b)^2 + (\log_{36}c)^2$ M1

$= (\log_{36}2)^2 + (\log_{36}3)^2 + (\log_{36}6)^2 = 0.3814 \longrightarrow 0.381$ to 3sf A1

DO NOT WRITE IN THIS AREA

P 6 8 8 2 8 A 0 1 8 2 0

Question 6 continued.

Question 6 continued.

(Total for Question 6 is 7 marks)

TOTAL FOR STATISTICS IS 50 MARKS