Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel Level 3 GCE

Thursday 20 June 2024

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 permitted by Pearson regulations. Calculators must not have the facility for symbolic algebraic 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 ▶

P74093A

F:1/1/1/1/

P 7 4 0 9 3 A 0 1 2 0

Pearson

1. Xian rolls a fair die 10 times.

The random variable X represents the number of times the die lands on a six.

(a) Using a suitable distribution for X, find

(i) $P(X = 3)$

(ii) $P(X < 3)$

(3)

Xian repeats this experiment each day for 60 days and records the number of days when $X = 3$

(b) Find the probability that there were at least 12 days when $X = 3$

(3)

(c) Find an estimate for the total number of sixes that Xian will roll during these 60 days.

(1)

(d) Use a normal approximation to estimate the probability that Xian rolls a total of more than 95 sixes during these 60 days.

(4)

a die has 6 sides, so P(lands on six) = 1/6

a) $X \sim B(10, 1/6)$ ①

(i) $P(X=3) = 0.155045...$
$= 0.155$ (3sf) ①

(ii) $P(X<3) = P(X \leq 2) = 0.775226...$
$= 0.775$ (3sf) ①

b) Let D = the number of days when $X = 3$

$D \sim B(60, 0.155...)$ ①

$P(D \geq 12) = 1 - P(D \leq 11)$ ①
$= 1 - 0.78819...$
$= 0.212$ (3sf) ①

c) Xian rolls 10 dice 60 times = 600 dice rolled

sixes 1/6 of the time = $600 \times \frac{1}{6} = 100$ sixes ①

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 1 continued

d) let S be the normal approximation for D.

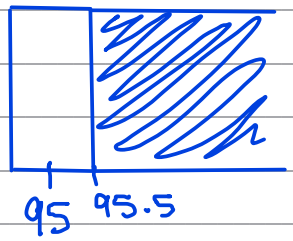
$\mu = np = 600 \times \frac{1}{6} = 100$

$\sigma = \sqrt{np(1-p)} = \sqrt{\frac{5}{6} \times 100} = \sqrt{\frac{250}{3}}$

$S \sim N\left(100, \sqrt{\frac{250}{3}}^2\right)$ (1) (1)

$P(D > 95) \approx P(S > 95.5) = 0.688976\ldots$ (1)

$= 0.689$ (3sf) (1)



Question 1 continued

Question 1 continued

(Total for Question 1 is 11 marks)

2. Amar is studying the flight of a bird from its nest.

He measures the bird's height above the ground, h metres, at time t seconds for 10 values of t
Amar finds the equation of the regression line for the data to be $h = 38.6 - 1.28t$

(a) Interpret the gradient of this line.

(1)

The product moment correlation coefficient between h and t is -0.510

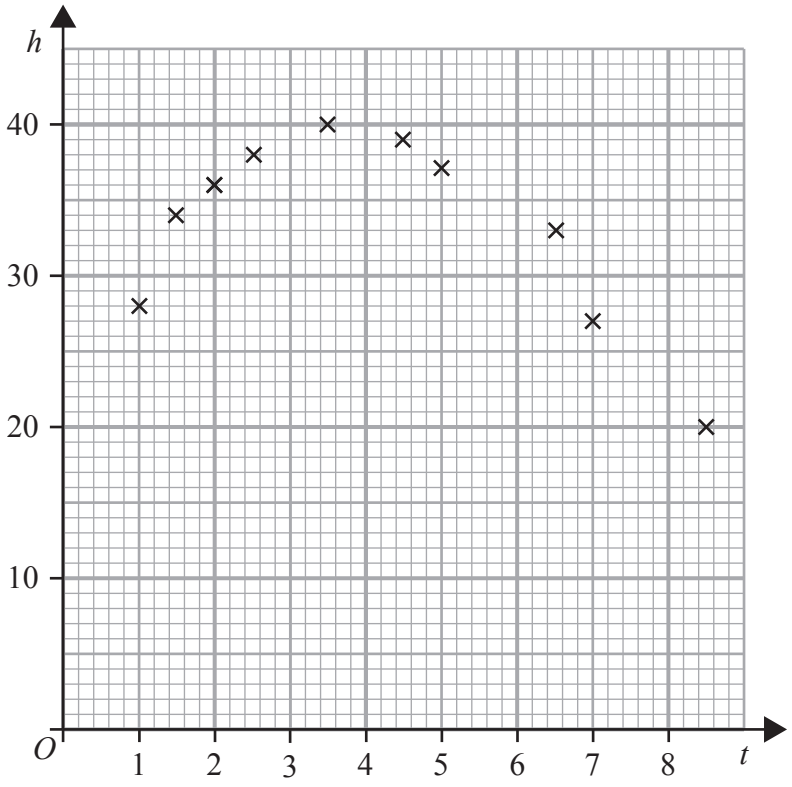
(b) Test whether or not there is evidence of a negative correlation between the height above the ground and the time during the flight.
You should

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

$H_1: \rho < 0$

(3)

Jane draws the following scatter diagram for Amar's data.



(c) With reference to the scatter diagram, state, giving a reason, whether or not the regression line $h = 38.6 - 1.28t$ is an appropriate model for these data.

(1)

Jane suggests an improved model using the variable $u = (t - k)^2$ where k is a constant.

She obtains the equation $h = 38.1 - 0.78u$

(d) Choose a suitable value for k to write Jane's improved model for h in terms of t only.

(1)

Question 2 continued

a) for every second of the flight, the height of the bird decreases by 1.28m. (1)

b) $H_0: \rho = 0$ $H_1: \rho < 0$ (1)

1 tailed ∴ sig. level = 5% = 0.05
sample size = 10
from table: critical value = 0.5494 (1)

$-0.5494 < -0.510$

so there is insufficient evidence of a negative correlation between height and time. (1)

c) No, the points seem to follow a quadratic curve, not a line. (1)

d) $h = 38.1 - 0.78(t - k)^2$

turning point of model: $(k, 38.1)$

turning point on graph has $t \approx 3.5$
so set $k = 3.5$ (1)

Question 2 continued

Question 2 continued

(Total for Question 2 is 6 marks)

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)

a) Rainfall ①
OR daily mean windspeed

there are only 5 measured variables for Perth:
- D.M. air temperature
- Rainfall
- D.M. pressure (always around 1000 hPa)
- D.M. windspeed
- D.M. windspeed Beaufort conversion.

b) $\bar{x} = \frac{\Sigma x}{n} = \frac{2801.2}{184} = 15.2239\ldots$
$= 15.2$ (3sf) ①

$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - \left[\frac{\Sigma x}{n}\right]^2} = \sqrt{\frac{44695.4}{184} - (15.22\ldots)^2}$ ①
$= 3.34$ (3sf) ①

Question 3 continued

c) $\bar{x} = 19.4$ which indicates a summer / spring month (1)

Perth is in the southern hemisphere so summer months start in October. (1)

(Total for Question 3 is 6 marks)

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)

a) $H_0: p = 0.1$ $H_1: p \neq 0.1$ (1)

two tailed $\therefore$ significance level $= 2.5\% = 0.025$

Assume H_0 is correct. Let X be the number of left-handed adults under the age of 25.

$X \sim B(40, 0.1)$

$P(X = 0) = 0.0148 < 0.025$ (1)
$P(X \leq 1) = 0.0805 > 0.025$

$P(X \geq 8) = 1 - P(X \leq 7) = 0.0419 > 0.025$
$P(X \geq 9) = 1 - P(X \leq 8) = 0.0155 < 0.025$ (1)

$CR = \{X = 0\} \cup \{X \geq 9\}$ (1)

b) $0.0148 + 0.0155 = 0.0303$ (1)

c) 7 is not in the critical region, so there is insufficient evidence to support Freya's belief. (1)

Question 4 continued

(Total for Question 4 is 6 marks)

5. The records for a school athletics club show that the height, H metres, achieved by students in the high jump is normally distributed with mean 1.4 metres and standard deviation 0.15 metres.

(a) Find the proportion of these students achieving a height of more than 1.6 metres.

(1)

The records also show that the time, T seconds, to run 1500 metres is normally distributed with mean 330 seconds and standard deviation 26 seconds.

The school's Head would like to use these distributions to estimate the proportion of students from the school athletics club who can jump higher than 1.6 metres **and** can run 1500 metres in less than 5 minutes.

(b) State a necessary assumption about H and T for the Head to calculate an estimate of this proportion.

(1)

(c) Find the Head's estimate of this proportion.

(3)

Students in the school athletics club also throw the discus.

The random variable $D \sim N(\mu, \sigma^2)$ represents the distance, in metres, that a student can throw the discus.

Given that $P(D < 16.3) = 0.30$ and $P(D > 29.0) = 0.10$

(d) calculate the value of μ and the value of σ

(5)

a) $H \sim N(1.4, 0.15^2)$

$P(H > 1.6) = 0.091211...$

$= 0.0912$ (3sf) (1)

b) H and T must be independent (1)

c) $T \sim N(330, 26^2)$

$P(T < 300) = 0.12428...$ (1)

(1)

probability of both $= 0.0912 \times 0.124 = 0.011335...$

$= 0.0113$ (3sf) (1)

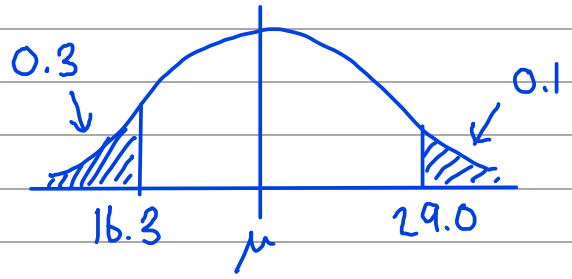
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

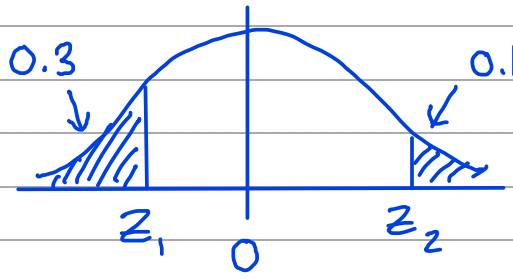
DO NOT WRITE IN THIS AREA

Question 5 continued

d) $D \sim N(\mu, \sigma^2)$



standardise:



$z_1 = \Phi^{-1}(0.3) = -0.5244...$

$z_2 = \Phi^{-1}(0.9) = 1.2815...$

$\frac{16.3-\mu}{\sigma} = -0.5244...$ (1) $\quad \frac{29-\mu}{\sigma} = 1.2815...$ (1)

$\Rightarrow -0.524\sigma + \mu = 16.3$ (1) $\quad \Rightarrow 1.282\sigma + \mu = 29$ (2)

(2) − (1): $29 - 16.3 = \sigma(1.282 + 0.524)$ (1)

$\sigma = 7.032115...$
$= 7.03$ (3sf) (1)

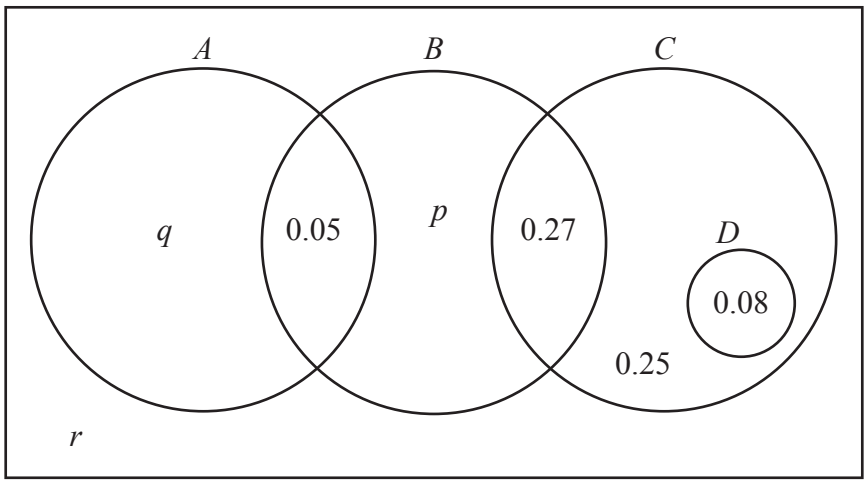
$\mu = 19.9876...$
$= 20.0$ (3sf) (1)

Question 5 continued

Question 5 continued

(Total for Question 5 is 10 marks)

6. The Venn diagram, where p, q and r are probabilities, shows the events A, B, C and D and associated probabilities.



(a) State any pair of mutually exclusive events from A, B, C and D

(1)

The events B and C are independent.

(b) Find the value of p

(2)

(c) Find the greatest possible value of $\mathrm{P}(A \mid B')$

(3)

Given that $\mathrm{P}(B \mid A') = 0.5$

(d) find the value of q and the value of r

(3)

(e) Find $\mathrm{P}\left(\left[A \cup B\right]' \cap C\right)$

(1)

(f) Use set notation to write an expression for the event with probability p

(1)

a) A and D (1)

b) $P(B)P(C) = P(B \cap C)$

$P(C) = 0.6 \quad P(B) = p + 0.32 \quad P(B \cap C) = 0.27$ (1)

$\Rightarrow 0.6(p + 0.32) = 0.27$

$p + 0.32 = 0.45$

$p = 0.13$ (1)

Question 6 continued

c) $P(A|B') = \dfrac{P(A \cap B')}{P(B')}$

$= \dfrac{q}{q+r+0.25+0.08}$ (1) $= \dfrac{q}{0.55}$

$q + r = 1 - (0.05 + 0.13 + 0.27 + 0.25 + 0.08)$
$= 0.22$ (1)

$\max P(A|B') \longrightarrow \max q$

$r \geq 0$ so max q happens when $r = 0 \Rightarrow q = 0.22$

$\max P(A|B') = \dfrac{0.22}{0.55} = \dfrac{2}{5}$ (1)

d) $P(B|A') = \dfrac{P(B \cap A')}{P(A')} = \dfrac{0.27 + 0.13}{0.6 + 0.13 + r}$

$\Rightarrow \dfrac{0.4}{r + 0.73} = 0.5$ (1)

$r + 0.73 = 0.8$

$r = 0.07$ (1)
$q = 0.22 - r$
$= 0.15$ (1)

e) $P([A \cup B]' \cap C) = 0.25 + 0.08 = 0.33$ (1)

"is not in A or B AND is in C"

Question 6 continued

f) $B \cap [A \cup C]'$ (1)

"is not in A or C AND is in B"

(Total for Question 6 is 11 marks)

TOTAL FOR STATISTICS IS 50 MARKS