PhysicsAndMathsTutor.com

Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel Level 3 GCE

Tuesday 20 June 2023

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.

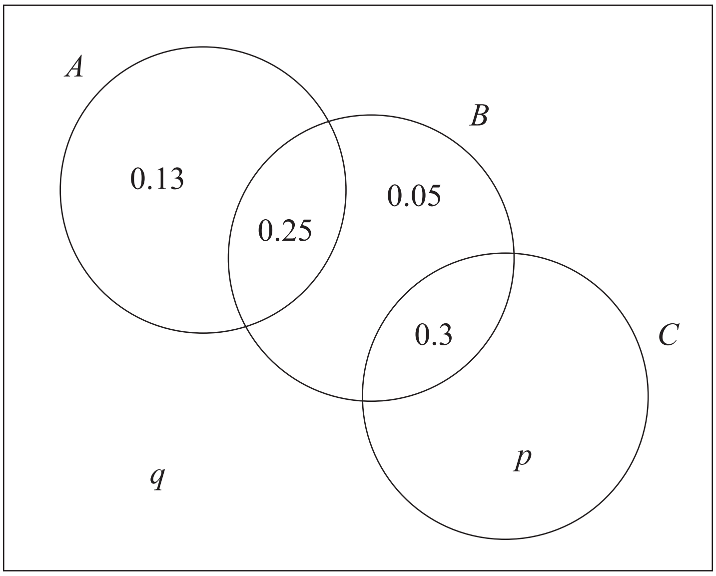
Turn over

©2023 Pearson Education Ltd.
N:1/1/1/

Pearson

PhysicsAndMathsTutor.com

1. The Venn diagram, where p and q are probabilities, shows the three events A, B and C and their associated probabilities.



(a) Find P(A) **(1)**

The events B and C are independent.

(b) Find the value of p and the value of q **(3)**

(c) Find $P(A|B')$ $\rightarrow \dfrac{P(A \cap B')}{P(B')}$ **(2)**

a) $P(A) = 0.13 + 0.25$

$= 0.38$ 1

b) $P(B \cap C) = P(B) \times P(C)$ — independent event

$0.3 = (0.3 + 0.25 + 0.05) \times (0.3 \times p)$ 1

$0.3 = 0.6 \times (0.3 + p)$

$0.3 = 0.18 + 0.6p$

$0.12 = 0.6p$

$\therefore p = 0.2$ 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

Σ probabilities $= 1$:

$0.13 + 0.25 + 0.05 + 0.3 + p + q = 1$

$0.13 + 0.25 + 0.05 + 0.3 + 0.2 + q = 1$

$0.93 + q = 1$

$\therefore q = 0.07$ (1)

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

$= \dfrac{0.13}{0.13 + 0.2 + 0.07}$ (1)

$= \dfrac{0.13}{0.4}$

$= 0.325$ (1)

(Total for Question 1 is 6 marks)

2. A machine fills packets with sweets and $\frac{1}{7}$ of the packets also contain a prize.

The packets of sweets are placed in boxes before being delivered to shops.
There are 40 packets of sweets in each box.

The random variable T represents the number of packets of sweets that contain a prize in each box.

(a) State a condition needed for T to be modelled by $\mathrm{B}(40, \frac{1}{7})$

(1)

A box is selected at random.

(b) Using $T \sim \mathrm{B}(40, \frac{1}{7})$ find

(i) the probability that the box has exactly 6 packets containing a prize,

(ii) the probability that the box has fewer than 3 packets containing a prize.

(2)

Kamil's sweet shop buys 5 boxes of these sweets.

(c) Find the probability that exactly 2 of these 5 boxes have fewer than 3 packets containing a prize.

(2)

Kamil claims that the proportion of packets containing a prize is less than $\frac{1}{7}$

A random sample of 110 packets is taken and 9 packets contain a prize.

(d) Use a suitable test to assess Kamil's claim.
You should

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

(4)

a) The probability of a packet containing a prize is constant. 1

b) $T \sim B(40, \frac{1}{7})$

(i) $P(T=6) = 0.1727\ldots = 0.173$ (3 s.f.) 1

(ii) $P(T<3) = P(T \leq 2)$

$= 0.061587\ldots = 0.0616$ (3 s.f.) 1

P 7 2 8 1 9 A 0 4 2 0

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

(c) Let r.v. K = number of boxes with fewer than 3 packets containing a prize.

$K \sim B(5, 0.0615\ldots)$ ①

$\therefore P(K=2) = 0.031344\ldots \approx 0.0313$ (3 s.f.) ①

d) Let r.v. X = number of packets containing a prize.

$X \sim B = (110, \frac{1}{7})$ ①

$H_0 : p = \frac{1}{7}$, $H_1 = p < \frac{1}{7}$ ①

$P(x \leq 9) = 0.038292\ldots$ (which is < 0.05) ①

$\therefore$ reject H_0 since there is evidence to support Kamil's claim. ①

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 9 marks)

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

n = 184

(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)

a) Replace 'tr' with a numerical value between 0 and 0.05. ①

For example, 0.025. ①

if the total amount of rainfall recorded is less than 0.05 mm, then it is recorded as 'tr'.

b) (i) mean, $\bar{x} = \dfrac{\sum x}{n}$

$= \dfrac{390}{184}$

$= 2.119\ldots = 2.12$ (3 s.f.) 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

(ii) Standard deviation $= \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$

$\bar{x}^2$

$= \sqrt{\frac{4336}{184} - 2.119..^2}$ 1

$= 4.367... = 4.37$ (3 s.f.) 1

c) (i) The data only covers May to October. So, it is not a representative of the whole year. 1

(ii) Winter months are missing when we'd expect more rain during this season. So, estimation in b(i) is expected to be an underestimate. 1

(Total for Question 3 is 7 marks)

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)

a) Let r.v. X = height from region A

$X \sim N(175.4, 6.8^2)$

$P(X > 180) = 0.2493\ldots = 0.249$ (3 s.f.) 1

b) Let r.v. Y = height from region B

$\bar{Y} \sim N\left(175.4, \frac{6.8^2}{52}\right)$ 1

$H_0 : \mu = 175.4$, $H_1 : \mu \neq 175.4$ 1

$\therefore P(\bar{Y} > 177.2) = 0.0281\ldots$ 1 > 0.025. (two-tailed, so $\alpha = 0.025$)

$\therefore$ do not reject H_0 as there is insufficient evidence to support student's claim. 1

P 7 2 8 1 9 A 0 1 0 2 0

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

Question 4 continued

c) $p = 2 \times 0.0281\ldots$

$= 0.05628,\ldots$ (1)

(Total for Question 4 is 6 marks)

5. Tisam is playing a game.
She uses a ball, a cup and a spinner.

The random variable X represents the number the spinner lands on when it is spun.
The probability distribution of X is given in the following table

x	20	50	80	100
$P(X = x)$	a	b	c	d

where a, b, c and d are probabilities.

To play the game

- the spinner is spun to obtain a value of x
- Tisam then stands x cm from the cup and tries to throw the ball into the cup

The event S represents the event that Tisam successfully throws the ball into the cup.

To model this game Tisam assumes that

- $P(S \mid \{X = x\}) = \dfrac{k}{x}$ where k is a constant
- $P(S \cap \{X = x\})$ should be the same whatever value of x is obtained from the spinner

Using Tisam's model,

(a) show that $c = \dfrac{8}{5}b$

(2)

(b) find the probability distribution of X

(5)

Nav tries, a large number of times, to throw the ball into the cup from a distance of 100 cm.
He successfully gets the ball in the cup 30% of the time.

(c) State, giving a reason, why Tisam's model of this game is not suitable to describe Nav playing the game for all values of X

(1)

a) To find equation with c and b, we use data when $x = 50$ and $x = 80$.

$P(S \cap \{X = 50\}) = P(S \cap \{X = 80\}) = \text{Constant}$

P 7 2 8 1 9 A 0 1 2 2 0

DO NOT WRITE IN THIS AREA

Question 5 continued

$P(S \mid \{X = x\}) = P(S)$

$P(S \cap \{X = x\}) = P(S) \times P(X = x)$

$= P(S \mid \{X = x\}) \times P(X = x)$

$= \frac{k}{x} \times P(X = x)$

when $x = 50$ and $x = 80$,

$\frac{k}{50} \times b = \frac{k}{80} \times c$ 1

$c = \frac{80}{50} b$

$\therefore \quad c = \frac{8}{5} b$ 1 (shown)

b) $b = \frac{5}{2} a$ 1, $c = 4a$, $d = 5a$ 1 find the probabilities in term of a.

$\sum$ probabilities $= 1$:

$a + b + c + d = 1$

$2 \times \left(a + \frac{5}{2} a + 4a + 5a\right) = 1 \times 2$ 1

$\therefore \quad 2a + 5a + 8a + 10a = 2$

$\therefore \quad 25a = 2$

$\therefore \quad a = \frac{2}{25}$ 1

Question 5 continued

$\therefore b = \frac{\cancel{8}1}{\cancel{2}1}\left(\frac{\cancel{2}1}{\cancel{25}5}\right) = \frac{1}{5}$

$\therefore c = 4\left(\frac{2}{25}\right) = \frac{8}{25}$

$\therefore d = \cancel{5}1\left(\frac{2}{\cancel{25}5}\right) = \frac{2}{5}$ 1

x	20	50	80	100
$P(X=x)$	$\frac{2}{25}$	$\frac{1}{5}$	$\frac{8}{25}$	$\frac{2}{5}$

c) $P(S \mid \{X = 20\}) = \frac{k}{20}$

$\therefore \frac{k}{100} = 0.3$

$\therefore k = 30 \Rightarrow \frac{30}{20} = 1.5$ (which is > 1)

1

$\therefore$ For a distance of 20 cm, this would give a probability of greater than 1, which is impossible.

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 8 marks)

6. A medical researcher is studying the number of hours, T, a patient stays in hospital following a particular operation.

The histogram on the page opposite summarises the results for a random sample of 90 patients.

(a) Use the histogram to estimate $P(10 < T < 30)$

(2)

For these 90 patients the time spent in hospital following the operation had

- a mean of 14.9 hours
- a standard deviation of 9.3 hours

Tomas suggests that T can be modelled by $N(14.9, 9.3^2)$

(b) With reference to the histogram, state, giving a reason, whether or not Tomas' model could be suitable.

(1)

Xiang suggests that the frequency polygon based on this histogram could be modelled by a curve with equation

$$y = kxe^{-x} \quad 0 \leqslant x \leqslant 4$$

where

- x is measured in **tens of hours**
- k is a constant

(c) Use algebraic integration to show that

$$\int_0^n xe^{-x}\,dx = 1 - (n+1)e^{-n}$$

(4)

(d) Show that, for Xiang's model, $k = 99$ to the nearest integer.

(3)

(e) Estimate $P(10 < T < 30)$ using

(i) Tomas' model of $T \sim N(14.9, 9.3^2)$

(1)

(ii) Xiang's curve with equation $y = 99xe^{-x}$ and the answer to part (c)

(2)

The researcher decides to use Xiang's curve to model $P(a < T < b)$

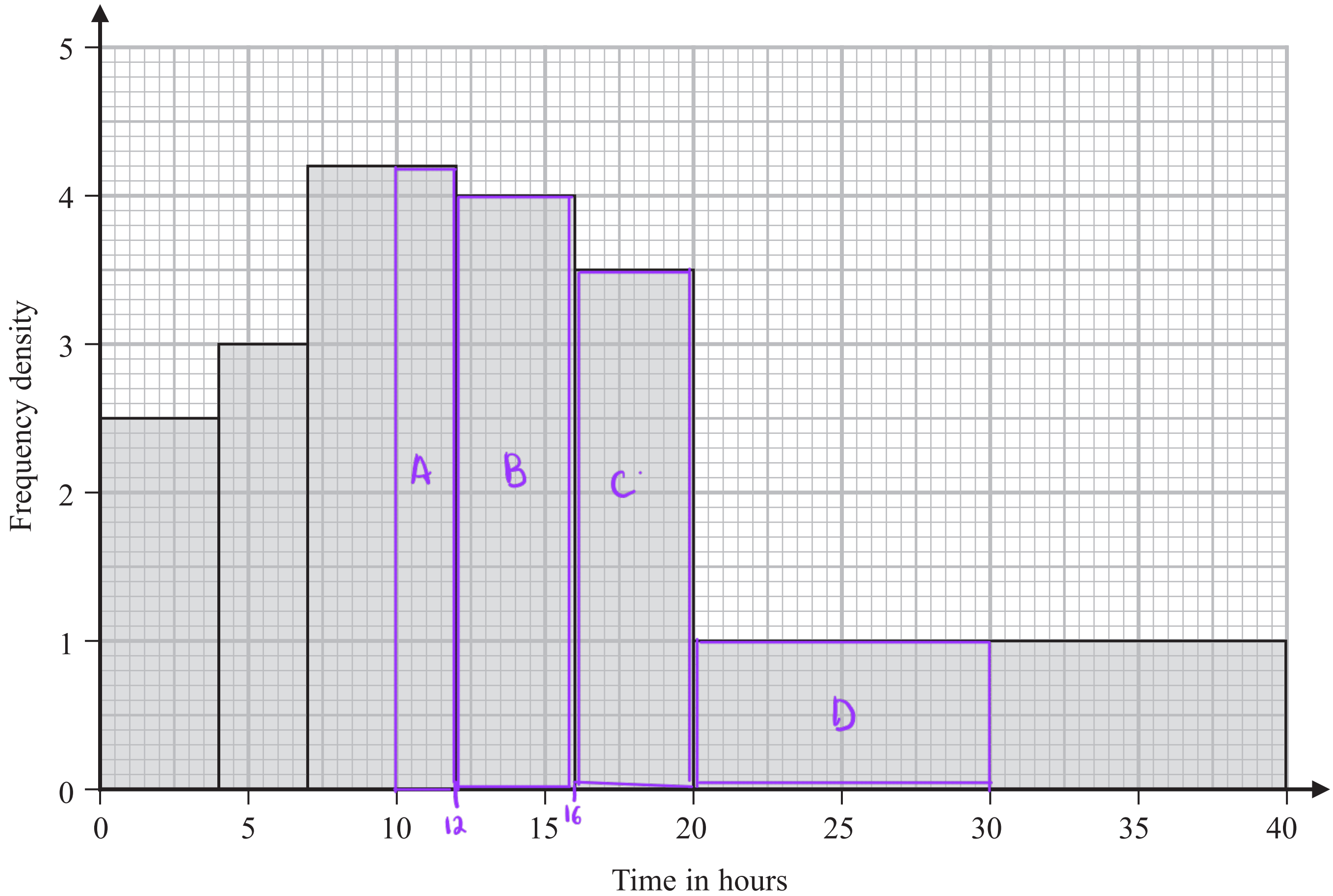
(f) State one limitation of Xiang's model.

(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 6 continued



a) $P(10 < T < 30) = P(A) + P(B) + P(C) + P(D)$

$= \dfrac{(2\times 4.2) + (4\times 4) + (4\times 3.5) + (16\times 1)}{90}$ 1

$= \dfrac{8.4 + 16 + 14 + 10}{90}$

$= \dfrac{48.4}{90}$

$= 0.5377\ldots = 0.54\ (2\text{s.f.})$ 1

(b) It does not look suitable because a normal distribution is symmetrical and the histogram is not. 1

Question 6 continued

c) $u = x \qquad v = -e^{-x}$

$u' = 1 \qquad v' = e^{-x}$ (1)

$u \times v - \int v u'$

$\therefore \int x e^{-x}\, dx = -x e^{-x} - \int (-e^{-x})\, dx$ (1)

$= -x e^{-x} + \int e^{-x}\, dx$

$= -x e^{-x} - e^{-x}$

$\therefore \int_0^n x e^{-x}\, dx = \left[-x e^{-x} - e^{-x}\right]_0^n$

$= \left(-n e^{-n} - e^{-n}\right) - \left(-e^{0}\right)$ (1)

$= -n e^{-n} - e^{-n} + 1$

$= 1 - (n+1) e^{-n}$ (1) (shown)

d) Area under frequency polygon = 90 when $n = 4$

$\therefore k \int x e^{-x}\, dx = 90$ (1)

$\therefore k \left\{1 - (n+1) e^{-n}\right\} = 90$

when $n = 4$: $k(1 - 5e^{-4}) = 90$ (1)

$k = \dfrac{90}{1 - 5e^{-4}}$

$k = 99.07\ldots$ (1)

$\therefore k \approx 99$

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

e) (i) $T \sim N(14.9, 9.3^2)$

$P(10 < T < 30) = 0.6486.... = 0.649$ (3s.f.) 1

(ii) No° of patients

$= \int_1^3 99xe^{-x}\,dx$

$= 99\{(1-4e^{-3}) - (1-2e^{-1})\}$

$= 53.1....$ 1

$\therefore$ Probability $= \dfrac{53.1....}{90}$

$= 0.590...$ 1

(f) Some patients might stay longer than 40 hours. 1

Question 6 continued

(Total for Question 6 is 14 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