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.*

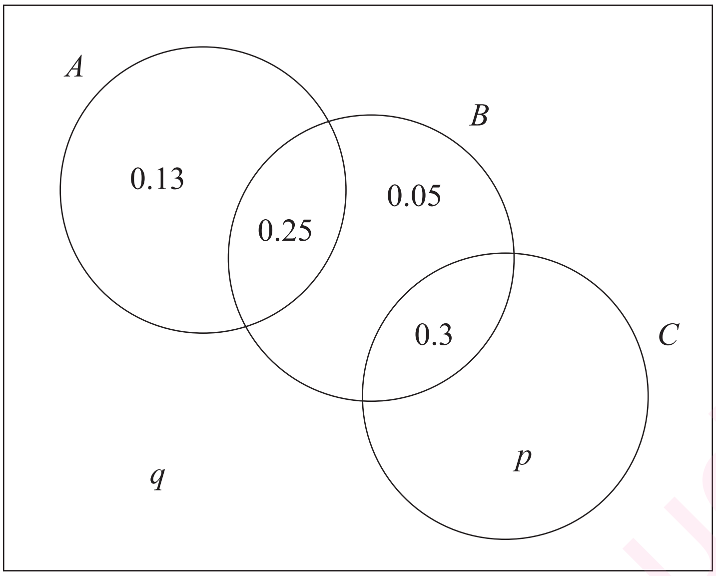
Advice

- 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

Pearson

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

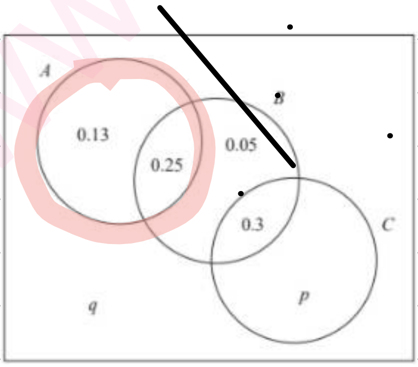
(3)

(c) Find $P(A|B')$

(2)

First thing worth noticing is that the Venn diagram shows us the probabilities of each event rather than the actual outcomes or number of outcomes

$\therefore$ (a) P(A) can just be read off from the Venn diagram (pink highlight).



$P(A) = 0.13 + 0.25$

$= \underline{\underline{0.38}}$

P 7 2 8 1 9 A 0 2 2 0

Question 1 continued

(b) remembering the formula for independent events

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

"A intersection B"
(read off from Venn diagram)

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

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

expanding the brackets

$0.3 = 0.18 + 0.6p$

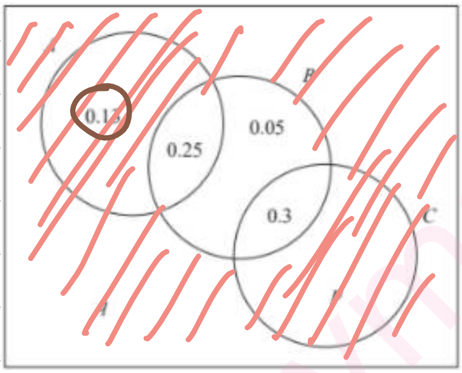
$0.6p = 0.12$

$\Rightarrow p = \underline{\underline{0.2}}$

$\therefore$ using fact that $\Sigma p = 1$: $q = 1 - (0.13 + 0.25 + 0.05 + 0.3 + 0.2)$

$= 1 - 0.93$

$\therefore q = \underline{\underline{0.07}}$



(c) we are dealing with conditional probability : $P(A|B')$

WAY 1: restricted sample space

consider shading just the event you are conditioning on so $P(B')$ - this will become the denominator of $P(A|B')$

$$P(A|B') = \frac{}{0.13 + 0.07 + 0.2}$$

and now GIVEN the shaded, see what's also in A: 0.13 (numerator)

$$\text{so } P(A|B') = \frac{0.13}{0.13 + 0.07 + 0.2} = \frac{0.13}{0.4} \text{ or } \frac{13}{40} \text{ //}$$

WAY 2: USING THE CONDITIONAL PROBABILITY FORMULA

subbing into the general formula $P(A|B') = \dfrac{P(A \cap B')}{P(B')}$

$$= \frac{0.13}{0.13 + 0.07 + 0.2} = \frac{0.13}{0.4} = \frac{13}{40} \text{ //}$$

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

(1)

A box is selected at random.

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

← hypothesis   population ✓parameter

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

↘ observed value (test statistic)

- state your hypotheses clearly
- use a 5% level of significance ← s.l

(4)

(a) we're asked to consider the assumptions necessary to model random variable T using a binomial distribution

·from the 4 assumptions that we know:

Binomial Distribution

Assumptions of Binomial Distribution

(1) Trials are repeated.
(2) All trials should be independent.
(3) The number of trial (n) should be fixed.
(4) There are two mutually exclusive outcomes in each trial, success and failure.

the most fitting for the context of packets of sweets are:
·'independent trials' → prizes must be placed in packets independently of each other

OR ·probability that packet contains a prize is constant for all packets of sweets

Question 2 continued

no. of trials

(b) $T \sim B(40, 1/7)$

↳ probability

(i) $P(T=6)$

WAY 1: USING BINOMIAL FORMULA

$P(T=r) = \binom{n}{r} p^r (1-p)^{n-r}$

$P(T=6) = \binom{40}{6}(1/7)^6(6/7)^{34}$ (in calc)

$= 0.172730\ldots$

$= 0.1727$ (4 s.f)

WAY 2: CALC (CLASSWIZ)

using CALC – 7. DISTRIBUTION – 4. BINOMIAL P.D – 2. VARIABLE

$x = 6$

$N = 40$

$p = 1/7$

$= 0.172730\ldots$

$= 0.1727$ (4 s.f)

(ii) $P(T<3)$ – here, key to remember that for binomial distribution, the cumulative probability function tells us the sum of all the individual probabilities up to and including t so $P(T \le t)$

→ hence, $P(T<3)$ is 0, 1, 2

∴ same as saying $P(T \le 2)$

USING CLASSWIZ CALC

7. DISTRIBUTIONS – 1. BINOMIAL C.D – 2. VARIABLE

$x = 2$

$N = 40$

$p = 1/7$

$\therefore P(T<3) = P(T \le 2) = 0.061587\ldots$

$= 0.06159$ (4 s.f)

(c) Now we're given a NEW RANDOM VARIABLE – no longer no. of packets of sweets that contain a prize but no. of boxes that have fewer than 3 packets containing a prize

let Y = no. of boxes that have fewer than 3 packets containing a prize

→ fixed probability (part (b)(ii))

$Y \sim B(5, 0.06159)$

no. of trials

$P(Y=2)$

WAY 1: USING BINOMIAL FORMULA

$P(Y=2) = \binom{5}{2}(0.06159)^2(0.93841)^3$

$= 0.031347\ldots$

$= 0.03135$ (4 s.f)

WAY 2: USING CALC CLASSWIZ

USING CALC – 7. DISTRIBUTION – 4. BINOMIAL P.D – 2. VARIABLE

$x = 2$

$N = 5$

$p = 0.06159$

$\therefore 0.03135$ (4 s.f)

Question 2 continued

(d) BINOMIAL HYPOTHESIS TEST

let X = no. of packets that contain a prize
p = probability that packet contains a prize

$X \sim B(110, p)$

$H_0 : p = 1/7$
$H_1 : p < 1/7$

assuming H_0 is true,

$X \sim B(110, 1/7)$

WAY 1: p-value way

key idea is to test whether sufficient evidence to reject null hypothesis by carrying out an experiment on a sample (test statistic) and comparing its p-value with the given significance level

i.e testing whether the observed value OR WORSE would've occured by chance (reject null hypothesis) or whether indeed something is wrong with the experiment (accept null hypothesis)

USING BINOMIAL C.D (CLASSWIZ CALC

$P(X \leq 9) = 0.03829\ldots < 0.05$

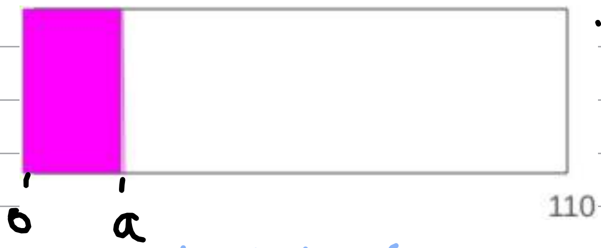
$\therefore$ sufficient evidence to reject null hypothesis, Kamil's claim is correct

NOTE: the conclusion talks about whether to accept or reject H_0 and then we specify what it means in the context after

Question 2 continued

WAY 2: CRITICAL REGIONS (CALC)

we reject when observed value is in the critical region



.. to find 'a'

$P(X \leq a) \leq 0.05$

using calculator (Binomial C.D list)

Casio Classwiz	Casio FX-CG50
Menu 7:Distribution BinomialCD List Write all X's under first column (write all the way to n or if n is big guess high enough) find n value corresponding to being closest to $\leq \alpha$	2:Statistics F5: Dist F5:Binomial F3 InvB Data: variable Area:α Numtrial: n P: probability

NOTE: could've used tables in formula book (5.statistical tables) but these only go up to n = 50 (we need n = 110)

calc gives us $a = 9$

$\therefore$ critical region is

$(X \leq 9)$ (9)

Now we need to check if our observed value Y is in the critical region

0 9 (right on the boundary) 110

so sufficient evidence to reject null hypothesis, Kamil's claim is correct //

(Total for Question 2 is 9 marks)

one of the variables in the LDS — 1 of 5 UK locations — LDS 2 time periods: May-Oct 1987, May-Oct 2015

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

(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) as soon as we see the variable 'daily total rainfall' we should pay close attention to the fact that not all data values will be numerical

- notice the 29 'tr' values (circled) i.e rainfall < 0.05 mm

∴ before start to calculate statistics such as mean or s.d need to convert tr values into zeros, 0.025 or anything < 0.05 //

(b)(i) formula for mean

$\bar{x} = \frac{\sum x}{n} = \frac{390}{184} = 2.11956... = 2.12$ (3 s.f)

Question 3 continued

(ii) formula for standard deviation

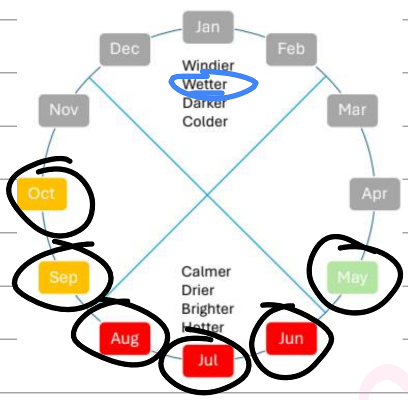
$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$= \sqrt{\frac{4336}{184} - (2.11956\ldots)^2}$$

$$= 4.36722\ldots$$

$$= \underline{4.37 \text{ (3 s.f)}}$$

(c) we know that the time range for the LDS only covers May - Oct ∴ unsuitable sample to then extrapolate to cover the whole year



(ii) the estimate for part (b)(i) would be an underestimate as the winter months are not included (these are usually wetter than rest of months

(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)

hypothesis — test statistic

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.

observed value

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

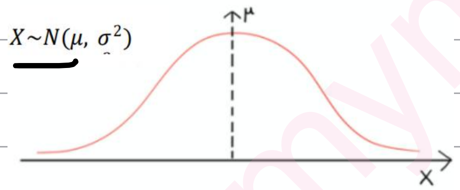
- state your hypotheses clearly
- use a 5% level of significance → s.l

(4)

(c) Find the p-value for the test in part (b)

(1)

(a) considering height is a continuous random variable, it's no surprise that it's modelled using a NORMAL DISTRIBUTION (knowing the general formula and characteristics of the Normal distribution)

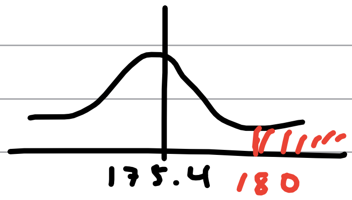


- symmetrical about mean = median = mode
- asymptotic at ends
- area under gives probability and sums to 1

NOW APPLYING THE CONTEXT OF THE QUESTION

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

mean — variance $(s.d^2)$



and using normal values from CALC CLASSWIZ

$P(H > 180)$

NOTE: for normal values doesn't matter '<' or '>' thanks to the 'lower' and 'upper' values on calc

LOWER = 180
UPPER = large no. (eg 100,000)
$\sigma = 6.8$
$\mu = 175.4$
$\therefore P(H > 180) = 0.24937 09\ldots$
$= 0.249$ (3 s.f)

(b) first thing worth noticing is that this is a two-tailed NORMAL hypothesis test ("is different to")

WAY 1: p-value method

this involves testing a hypothesis made about value of population's mean by carrying out an experiment on the sample mean - calculating its p-value and comparing with s.l to see whether sufficient evidence to reject null hypothesis

$n = 52$ $\sigma = 6.8$
$s.l = 0.025$ (halved due to two-tailed)

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

assuming H_0 is true,
using $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

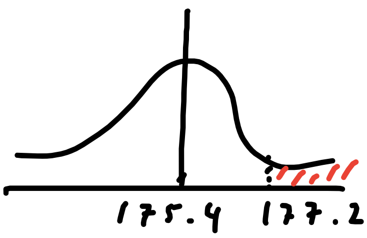
↳ calc input $\frac{\sigma}{\sqrt{n}}$

$\bar{X} \sim N(175.4, \frac{6.8^2}{52})$
↳ $\frac{6.8}{\sqrt{52}}$

Now you can see whether the observed value OR WORSE would occur by chance (reject null hypothesis) or if there is something wrong with the initial hypothesis (accept null hypothesis)

↳ this determines our '<'/'>' direction, but we want 177.2 or worse (as far away from mean as possible

$P(\bar{H} > 177.2)$



Using CLASSWIZ - NORMAL C.D

$P(\bar{H} > 177.2) = 0.02814\ldots > 0.025$

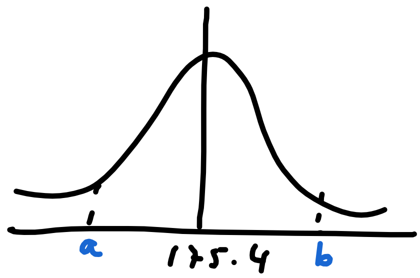
$\therefore$ insufficient evidence to reject null hypothesis

NO EVIDENCE to support student's claim

NOTE: conclusion always talks about whether accept or reject H_0 and then specify what that means in the context after

WAY 2: CRITICAL REGIONS

we reject when the observed value (177.2) is in either of the critical regions



for lower critical, need 'a' such that $P(X < a) \leq 0.025$

$\therefore$ for boundary, need $P(X < a) = 0.025$

Using INVERSE NORMAL (calc)

area $= 0.025$

$\sigma = 6.8$

$\mu = 175.4$

$\Rightarrow$ calc gives $a = 162.07$

now for upper critical, need 'b' such that $P(X > b) \leq 0.025$

but inverse normal deals with cumulative probability

so $P(X < a) = p$

$\therefore$ if have $P(X > a)$, need to input into calc as

$P(X < a) = 1 - p$

$\therefore P(X < b) = 0.975$

using INVERSE NORMAL ON CALC CLASSWIZ

area $= 0.975$

$\sigma = 6.8$

$\mu = 175.4$

$\Rightarrow$ calc gives $b = 188.72$

$\therefore$ critical region is

$\bar{m} < 162.07$ and $\bar{m} > 188.72$

177.2 not in critical region

$\therefore$ insufficient evidence to reject null hypothesis, student's claim is INCORRECT

Question 4 continued

(c) the p-value of the two-tailed test would be 2 × p-value that compared with the s.l during the p-value method used in (b)

$2 \times 0.02814...$

$= 0.05628...$

$= 0.0563$ (3 s.f)

(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 (TABLE)
- 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

- $P(S \mid \{X = x\}) = \frac{k}{x}$ 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 $c = \frac{8}{5}b$

(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)

Let's look at the question in parts

First consider the table

x = no. spinner lands on (random variable)

probability that the random variable takes a specific value

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

so
P(spinner lands on 20) = a
P(spinner lands on 50) = b
P(spinner lands on 80) = c
P(spinner lands on 100) = d

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

Question 5 continued

now we have the event S which is the probability ball thrown successfully

given a conditional probability : $P(S \mid \{X=x\}) = \frac{k}{x}$ — ①

AND an intersection : $P(S \cap \{X=x\})$ equal for all x in table — ②

(HINTS at use of conditional probability formula)

Now looking at the 'show that' we see 'c' and 'b', so know we're interested in the $\begin{array}{c|c} 50 & 80 \\ \hline b & c \end{array}$ part of given table

Starting with the condition ① - let's use conditional probability formula

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

subbing in $X=50$ (table)

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

$$P(S \mid \{X=50\}) = \frac{P(S \cap \{X=50\})}{P(X=50)} = \frac{k}{50}$$

$$= \frac{P(S \cap \{X=50\})}{b} = \frac{k}{50}$$

$\therefore$ rearranging to get $P(S \cap \{X=50\})$ [condition ②]

$$= P(S \cap \{X=50\}) = \frac{bk}{50}$$

we know from condition ② that this is the same for all x values in the table.

$\therefore$ if we repeated the process with $X=80$ we can equate the final expressions subbing in X = 80 (table)

$$P(S \mid \{X=80\}) = \frac{P(S \cap \{X=80\})}{P(X=80)} = \frac{k}{80}$$

$$= \frac{P(S \cap \{X=80\})}{c} = \frac{k}{80}$$

$$= P(S \cap \{X=80\}) = \frac{ck}{80}$$

now can equate the two expressions

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

cancelling the 'k's'

and rearranging for 'c' ($\times 80$)

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

$\div 10$

$$\Rightarrow c = \frac{8}{5}b \quad \text{as required} \;//$$

(b) finding the probability distribution of X requires you to find the values 'a' 'b' 'c' 'd' currently UNKNOWN in the probability distribution table

using (a), let's replace c with $\frac{8}{5}b$ in the table

x	20	50	80	100
$P(X=x)$	a	b	$\frac{8}{5}b$	d

And using the same process as in (a), let's find similar expressions for d in terms of a

NOTICE pattern that for 'b' final expression was $\frac{bk}{50}$ and for 'c' it's $\frac{ck}{80}$ $\therefore$ something in form of $\frac{P(X=x)k}{x}$

so for 'd' $\Rightarrow \frac{dk}{100}$

for 'a' $\Rightarrow \frac{ak}{20}$

equating

$$\frac{a\cancel{k}}{20} = \frac{d\cancel{k}}{100}$$

$$d = \frac{100}{20}a$$

$$\Rightarrow d = 5a$$

Question 5 continued

So now the table becomes

x	20	50	80	100
$P(X=x)$	a	b	$\frac{8}{5}b$	$5a$

now we're quite helpfully down to 2 variables - but still possible to get down to 1 variable
∴ let's repeat process for 'a' and 'b'

$$\frac{ka}{20} = \frac{kb}{50}$$

$$b = \frac{50a}{20}$$

$$= \frac{5}{2}a$$

∴ now we finally have a table only in terms of 1 variable - here in terms of 'a'

x	20	50	80	100
$P(X=x)$	a	$\frac{5}{2}a$	$\frac{8}{5}\times\left(\frac{5}{2}a\right)$ $= 4a$	$5a$

NOW using following fact that probabilities in a statistical distributions table sum to 1

$$a + \frac{5}{2}a + 4a + 5a$$

$$= 12.5a = 1$$

$$\Rightarrow a = \frac{2}{25}$$

... from this:

x	20	50	80	100
$P(X=x)$	$\frac{2}{25}$	$\frac{5}{2}\times\frac{2}{25}$ $= \frac{1}{5}$	$4\times\frac{2}{25}$ $= \frac{8}{25}$	$5\times\frac{2}{25}$ $= \frac{2}{5}$

Question 5 continued

$\therefore a = 2/25,\ b = 1/5,\ c = 8/25,\ d = 2/5$ //

(c) subbing in $X = 100$ into the process used in (a)

$$P(S \mid \{x = 100\}) = \frac{P(S \cap \{X = 100\})}{P(X = 100)} = \frac{k}{100} = 0.3$$

$\Rightarrow k = 30$

but then if were to compare 'd' with 'a'

e.g. $\frac{ka}{20}$

$= \frac{30}{20}a$ but $\frac{30}{20} = 3/2 > 1$

$\therefore$ Tisam's model is not suitable for all x values as all probabilities should sum to 1 (not possible if one already exceeds 1.

(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.

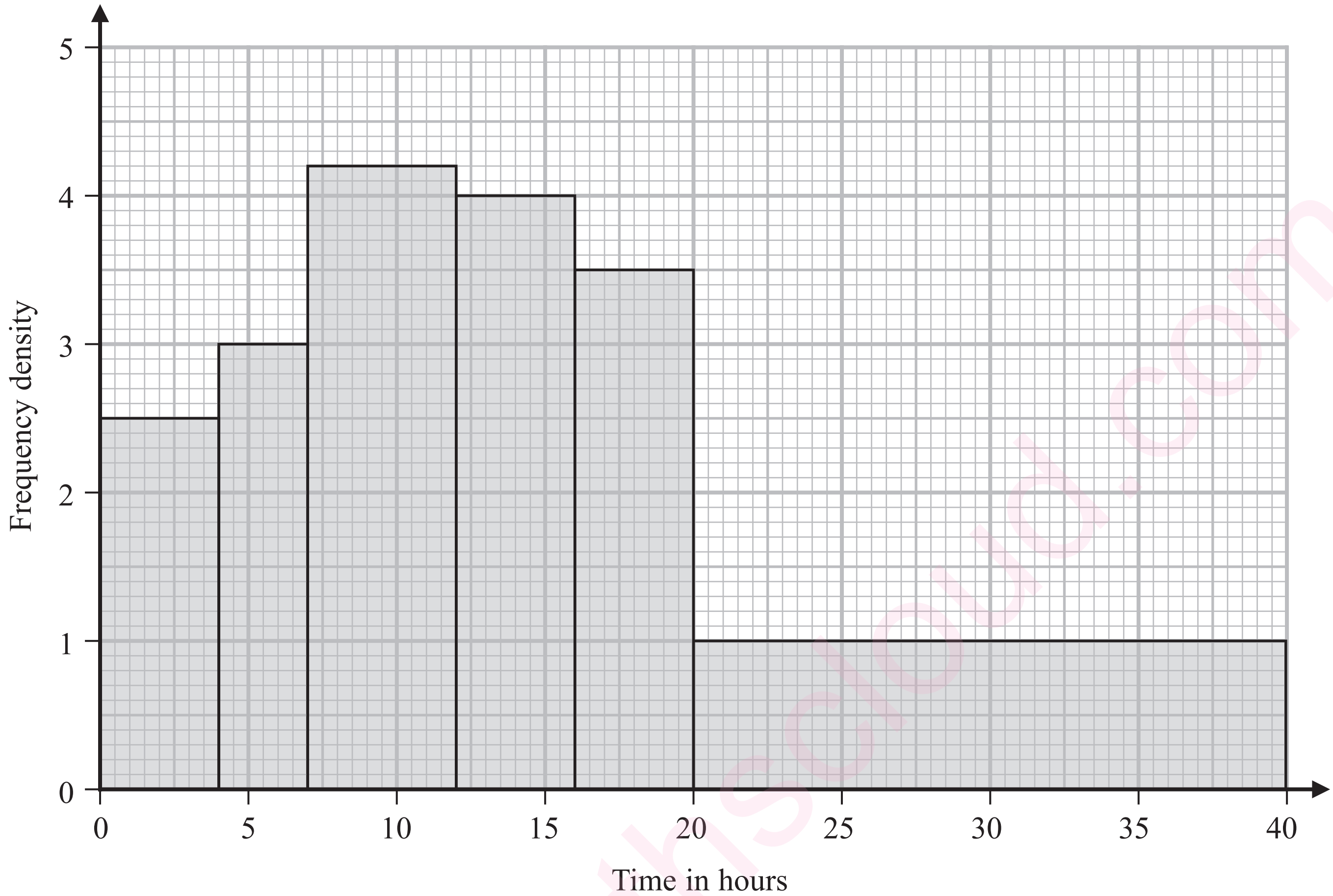
(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 6 continued



(a) We know that the histogram illustrates the FREQUENCY of the continuous variable i.e the time in his patient stays in hospital

↳ hence we would need to use the frequency given to us by the histogram to then find the probability

$$P(10<T<30) = \frac{\text{frequency}}{\text{total of outcomes}}$$

· usually in A-level:

frequency $\propto$ area of bars

i.e $f = kA$

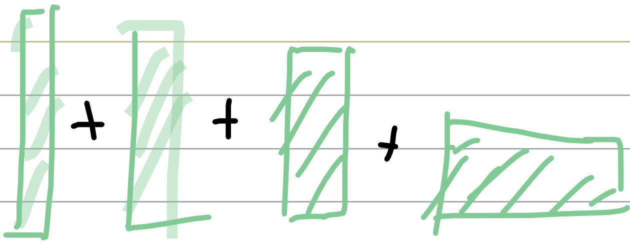
but because no further information given to us about frequency of other bars given their area, then can assume that $k = 1$

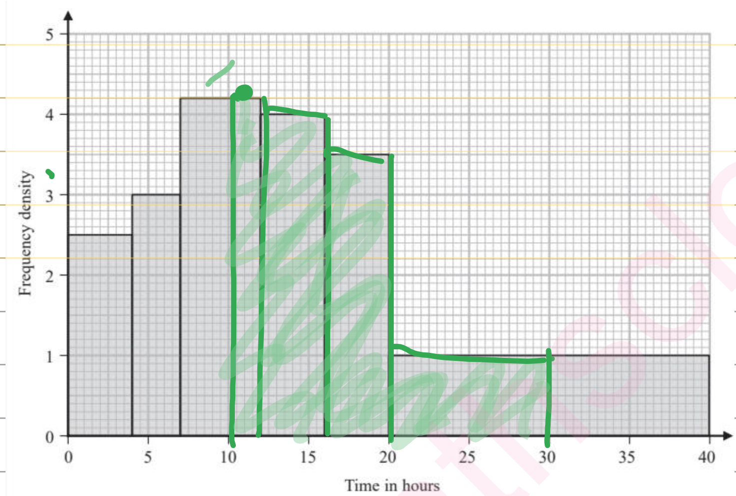
so $f = A$

Question 6 continued

$\therefore$ identifying the areas of bars where $10 < T < 30$ (shaded below)

frequency =





$= \frac{2}{5}(5 \times 4.2) + (4 \times 4) + (4 \times 3.5) + (10 \times 1)$

$= 8.4 + 16 + 14 + 10$

$= 48.4$

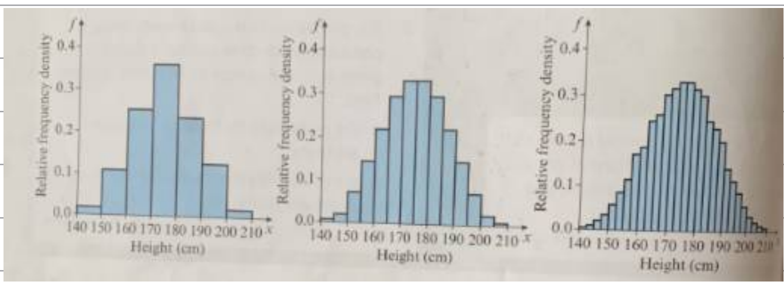
probability $= \frac{\text{frequency}}{\text{total}}$

$= \frac{48.4}{90} = \frac{121}{225} = 0.5377\ldots$

$= 0.54$ (2 d.p)

(b) Remember that the smooth, symmetrical bell curve of normal distribution is formed by many histograms with their c.w getting smaller

↓



· symmetrical
· asymptotes

However, in our histogram, data is NOT symmetrical (skewed), so Tomas' model is not suitable //

(c) reminding yourself what a frequency polygon looks like (connecting m.ps of each of c.ws - see below)

assumed equation

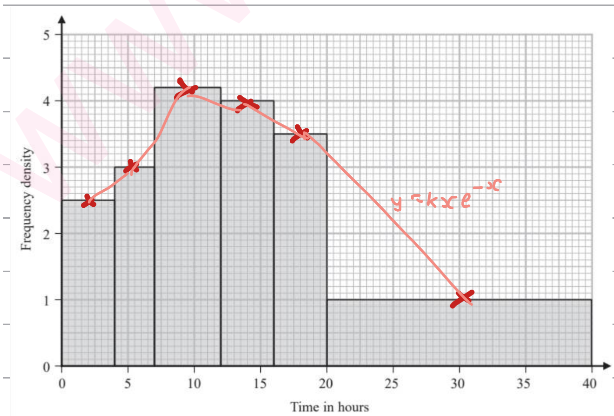
$y = kxe^{-x}$ for $0 < x < 4$

tens of hrs (so actually $0 < x < 40$ on graph)

finding the area under the frequency polygon

∴ integrating by PARTS!

$$\int xe^{-x}\,dx$$



...'u':

Logs

Inverse functions (FM)

Algebraic expressions ✓

Trig

Exponentials

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

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

remembering I.B.P

$\int uv\,dx = uv - \int vu'\,dx$

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

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

$= -xe^{-x} - e^{-x} + c$

definite integration with limits

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

$= \{(-ne^{-n} - e^{-n}) - (-e^{-0})\}$

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

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

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

(d) for 'show that' - need known facts

know that total frequency = 90

(when 40hrs

so $n = 4$)

subbing into (d)

$90 = k(1 - e^{-4}(4+1))$

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

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

$\Rightarrow k = 99.07291..$

$\therefore k = 99$ (nearest integer)

(e)(i) using NORMAL C.D on CALC - CLASSWIZ

7. DISTRIBUTIONS - 2. Normal CD

lower : 10

upper : 30

$\sigma = 9.3$

$\mu = 14.9$

$\therefore P(10 < T < 30) = 0.64863...$

$= 0.649$ (3 s.f)

(ii) given equation, so need to use integration from (c)
$x = 1$ and 3 (tens of hrs) become LIMITS
↳ finds frequency

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

$$= 99\left[\{1-4e^{-3}\}-\{1-2e^{-1}\}\right]$$

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

$$= 53.12445\ldots$$

using probability = $\frac{\text{frequency}}{\text{total outcomes}}$

$$= \frac{53.124\ldots}{90}$$

$$= 0.590271\ldots$$

$$= 0.590 \text{ (3 s.f)}$$

f) When using Xiang's model for $P(a<T<b)$ need to consider that model only valid for

$0 \le x \le 4$

(so $0 \le x \le 40$ hrs)

but a patient may stay longer than 40 hrs ∴ limitation //

(Total for Question 6 is 14 marks)

TOTAL FOR STATISTICS IS 50 MARKS