Please check the examination details below before entering your candidate information

Candidate surname	Other names

Centre Number | Candidate Number

Pearson Edexcel Level 3 GCE

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 ▶

P72130A

Q:1/1/1/1/

P 7 2 1 3 0 A 0 1 2 0

Pearson

1. George throws a ball at a target 15 times.
Each time George throws the ball, the probability of the ball hitting the target is 0.48

The random variable X represents the number of times George hits the target in 15 throws.

(a) Find

(i) $P(X = 3)$

(ii) $P(X \geqslant 5)$

(3)

George now throws the ball at the target 250 times.

(b) Use a normal approximation to calculate the probability that he will hit the target more than 110 times.

(3)

$X \sim B(n, p)$ ↗ probability p

$X \sim B(15, 0.48)$

↘ number of trials

a)

i) $P(X = 3)$

Use binompd on calculator with x value 3

0.0197

ii) $P(X \geq 5)$

Way 1: use classwiz calculator

This needs to be turned into a $\leq$

$P(X \geq 5)$ means $X = 5, 6, 7, 8, 9, \ldots 15$

All probabilities add to 1

so $P(X \geq 5)$ is the same as saying

$= 1 - P(X \leq 4)$

All options

We subtract what we don't want from 1 in order to get what we want

Use binomcd on calculator to get $P(X \leq 4)$. x value is 4

$= 1 - 0.079869$

$= 0.920$

Way 2: Using Casio FX-CG50 calc.

$P(X \geq 5)$

We can use the calculator straight away

Use binomcd
Lower: 5
Upper: 15

$= 0.920$

b) n is large and p is close to 0.5, so can use a normal approximation

$\mu = np = 250(0.48) = 120$

$\sigma = npq = np(1-p) = 250(0.48)(0.52)$
$= 62.4$

$Y \sim N(120, 62.4)$

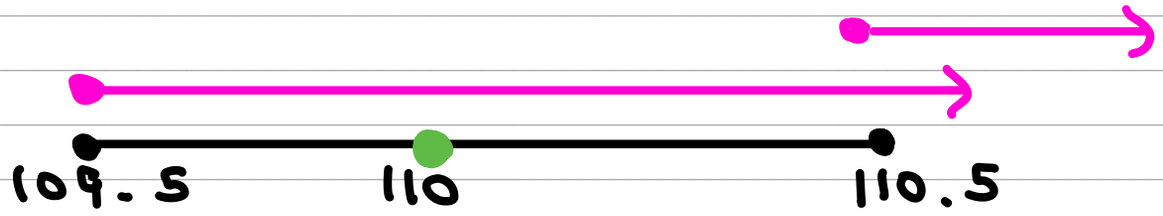
$P(Y > 110)$
Need to use a continuity correction since going from a discrete distribution (binomial where can only have whole numbers to normal where can have any number).

(Total for Question 1 is 6 marks)

Think back to when you first learnt bounds in GCSE's

Lower bound of 110 is 109.5
Upper bound of 110 is 110.5

our options are to be greater than either of these 2 bounds → since question



Which one do we choose? P(X ≥ 109.5) or P(X > 110.5)?

We want P(Y ≥ 110)

↑ no inequality sign here. so must choose the one that doesn't touch 110 hence P(Y > 110.5)

P(Y ≥ 110.5)

Use normcd on calculator
Lower: 110.5
Upper: 10000

0.886

(Total for Question 1 is 6 marks)

2. A manufacturer uses a machine to make metal rods.

The length of a metal rod, L cm, is normally distributed with

- a mean of 8 cm
- a standard deviation of x cm

Given that the proportion of metal rods less than 7.902 cm in length is 2.5%

(a) show that $x = 0.05$ to 2 decimal places.

(2)

(b) Calculate the proportion of metal rods that are between 7.94 cm and 8.09 cm in length.

(1)

The **cost** of producing a single metal rod is 20p

A metal rod

- where $L < 7.94$ is **sold** for scrap for 5p
- where $7.94 \leqslant L \leqslant 8.09$ is **sold** for 50p
- where $L > 8.09$ is shortened for an extra **cost** of 10p and then **sold** for 50p

(c) Calculate the expected profit per 500 of the metal rods.
Give your answer to the nearest pound.

(5)

The same manufacturer makes metal hinges in large batches.

The hinges each have a probability of 0.015 of having a fault.

A random sample of 200 hinges is taken from each batch and the batch is accepted if fewer than 6 hinges are faulty.

The manufacturer's aim is for 95% of batches to be accepted.

(d) Explain whether the manufacturer is likely to achieve its aim.

(4)

$L \sim N(\mu, \sigma^2)$
$L \sim N(8, x^2)$

a) $P(L < 7.902) = 0.025$

We must use standardised values since σ is unknown

use invnorm on calc
area: 0.025 (left)
$\mu = 0$
$\sigma = 1$
This gives $z = -1.960$
(note: this doesn't give us L since using standardised values)

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

We know $Z = \frac{L-\mu}{\sigma}$

$\Rightarrow \frac{L-\mu}{\sigma} = -1.960$

Fill in everything we know

$\frac{7.902 - 8}{x} = -1.960$

$\frac{-0.098}{x} = -1.960$

$-1.960x = -0.098$

$x = 0.05$ as required

b) $P(7.94 < L < 8.09)$

Use normcd on calculator
Lower: 7.94
Upper: 8.09
$\mu = 8$
$\sigma = 0.05$

0.849

c)

- $L < 7.94$: Profit $= -20p + 5p = -15p = -£0.15$
- $7.94 \leq L \leq 8.09$: Profit $= -20p + 50p = 30p = £0.30$
- $L > 8.09$: Profit $= -20p - 10p + 50p = 20p = £0.20$

- $P(L < 7.94)$

Use Normcd on calculator
Lower: -1000
Upper: 7.94
$\mu = 8$
$\sigma = 0.05$
$= 0.115$ (this is for -£0.15 profit)

- $P(7.94 \leq L \leq 8.09) = 0.849$ (from b) (this is for £0.30 profit)

- $P(L > 8.09)$

Use Normcd on calculator
Lower: 8.09
Upper: 1000
$\mu = 8$
$\sigma = 0.05$

$= 0.0359$ (this is for £0.30 profit)

- Expected profit = Sum of all the profits x their probabilities

profit	-£0.15	£0.30	£0.20
probability	0.115	0.849	0.0359

$= -0.15(0.115) + 0.30(0.849) + 0.20(0.0359)$

$= 0.24463$

We have 500 roads

$500(0.24+63)$

$= 122.315$

£122

d) $X \sim B(200 \quad 0.015)$

$P(X < 6)$

Turn into ≤ in order to use the calculator

$P(X \leq 5)$

Use binomcd on calculator with x value 5

$= 0.9176$

$\approx 92\%$

92% < 95% so the manufacturer is not likely to achieve the aim

(Total for Question 2 is 12 marks)

3. Dian uses the large data set to investigate the Daily Total Rainfall, r mm, for Camborne.

(a) Write down how a value of $0 < r \leqslant 0.05$ is recorded in the large data set.

(1)

Dian uses the data for the 31 days of August 2015 for Camborne and calculates the following statistics

$$n = 31 \qquad \sum r = 174.9 \qquad \sum r^2 = 3523.283$$

(b) Use these statistics to calculate

(i) the mean of the Daily Total Rainfall in Camborne for August 2015,

(ii) the standard deviation of the Daily Total Rainfall in Camborne for August 2015.

(3)

Dian believes that the mean Daily Total Rainfall in August is less in the South of the UK than in the North of the UK.
The mean Daily Total Rainfall in Leuchars for August 2015 is 1.72 mm to 2 decimal places.

(c) State, giving a reason, whether this provides evidence to support Dian's belief.

(2)

Dian uses the large data set to estimate the proportion of days with no rain in Camborne for 1987 to be 0.27 to 2 decimal places.

(d) Explain why the distribution B(14, 0.27) might **not** be a reasonable model for the number of days without rain for a 14-day summer event.

(1)

a) recorded as tr (trace)

b)

i) $\text{Mean} = \frac{\sum r}{n} = \frac{174.9}{31} = 5.6419 = 5.64\text{ mm}$

ii) $\text{S.D} = \sqrt{\text{variance}}$

$$\text{S.D} = \sqrt{\frac{\sum r^2}{n} - \text{Mean}^2} = \sqrt{\frac{3523.283}{31} - 5.6419^2}$$

$$= 9.0456$$

$$= 9.05\text{ mm}$$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 3 continued

c) No, this does not provide enough evidence to support the claim.

Leuchars is north of Camborne, however the mean daily total rainfall is less

d) The probability of 0.27 calculated for the <u>whole</u> year might not be a reliable indication of the probability of rain in the summer where there might be less rain

(Total for Question 3 is 7 marks)

4. A dentist knows from past records that 10% of customers arrive late for their appointment.

A new manager believes that there has been a change in the proportion of customers who arrive late for their appointment.

two tailed ≠ since doesn't mention if more than/less than so can't be one tailed

A random sample of 50 of the dentist's customers is taken.

(a) Write down

- a null hypothesis corresponding to no change in the proportion of customers who arrive late
- an alternative hypothesis corresponding to the manager's belief

(1)

(b) Using a 5% level of significance, find the critical region for a two-tailed test of the null hypothesis in (a)
You should state the probability of rejection in each tail, which should be less than 0.025

(3)

(c) Find the actual level of significance of the test based on your critical region from part (b)

(1)

The manager observes that 15 of the 50 customers arrived late for their appointment.

(d) With reference to part (b), comment on the manager's belief.

(1)

a) $X \sim B(50, 0.1)$

$H_0 : p = 0.1$

$H_1 : p \neq 0.1$

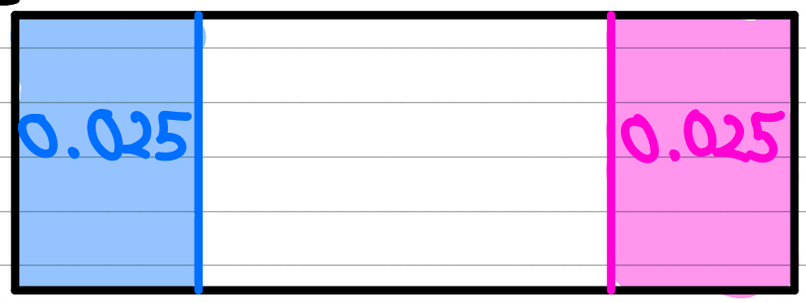
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 4 continued

b) We reject when in either of the shaded region aka critical region



To find a:
$P(X \leq a) \leq 0.025$
use tables or calc

$a = 0$

To find b:
$P(X \geq b) \leq 0.025$
$1 - P(X \leq b-1) \leq 0.025$
$P(X \leq b-1) \geq 0.975$
use tables or calc
$b - 1 = 9$
$b = 10$

So critical regions are

- $X \leq 0$ i.e $X = 0$
- $X \geq 10$ i.e $10 \leq X \leq 50$

We need to find their probabilities too

$P(X = 0)$
use binomcd
$= 0.0052$

$P(X \geq 10)$
$= 1 - P(X \leq 9)$
use binomcd
$= 1 -$
$= 0.0245$

(Total for Question 4 is 6 marks)

Question 4 continued

c) actual significance level

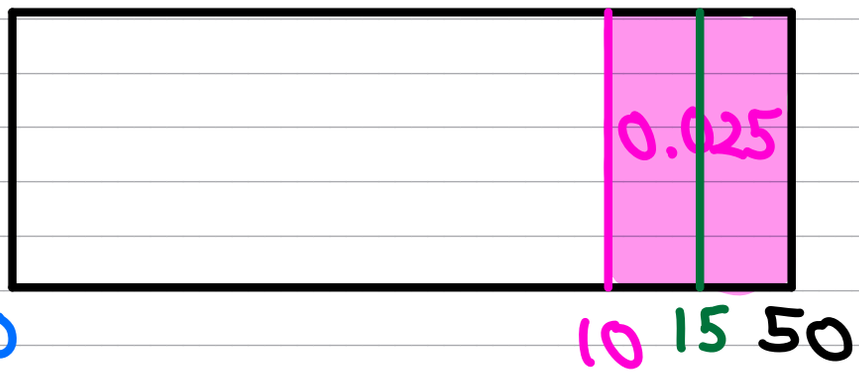
$= 0.0052 + 0.0245$

$= 0.0297$

$= 2.97\%$

(This is close to 2.5% like we expect)

d)



15 is in the critical region

∴ there is sufficient evidence to reject the null hypothesis and we can conclude that the managers claim that there has been a change in the proportion of customers who arrive late is justified

Note: The conclusion always talks about whether accept or reject H_0 and then we specify what that means in context after

(Total for Question 4 is 6 marks)

5. A company has 1825 employees.
The employees are classified as professional, skilled or elementary.

The following table shows

- the number of employees in each classification
- the two areas, A or B, where the employees live

	A	B	Total
Professional	740	380	1120
Skilled	275	90	365
Elementary	260	80	340
Total	1275	550	1825

An employee is chosen at random.

Find the probability that this employee

(a) is skilled,

(1)

(b) lives in area B and is not a professional.

(1)

Some classifications of employees are more likely to work from home.

- 65% of professional employees in both area A and area B work from home . $65(1120) = 728$
- 40% of skilled employees in both area A and area B work from home . $40(365) = 146$
- 5% of elementary employees in both area A and area B work from home . $05(340) = 17$
- Event F is that the employee is a professional
- Event H is that the employee works from home
- Event R is that the employee is from area A

So home $= 728 + 146 + 17 = 891$

(c) Using this information, complete the Venn diagram on the opposite page.

(4)

(d) Find $P(R' \cap F)$ a) $\frac{365}{1825} = \frac{1}{5}$

(1)

(e) Find $P([H \cup R]')$

(1)

(f) Find $P(F \mid H)$ b) $\frac{90+80}{1825} = \frac{170}{1825} = \frac{34}{365}$

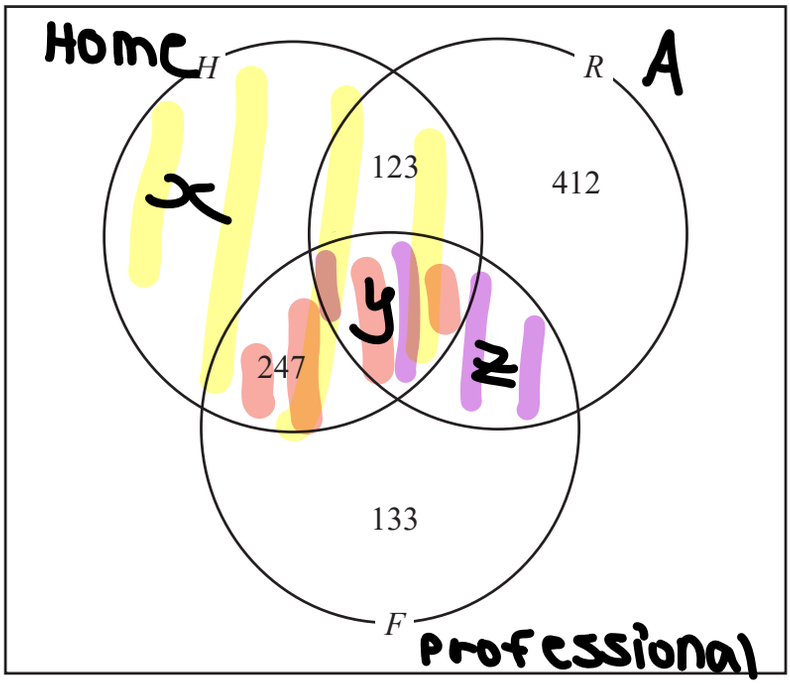
(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 5 continued



Turn over for a spare diagram if you need to redraw your Venn diagram.

Build equations

- 740 in A and professional:

  $y + z = 740$ ①

- $728 + 146 + 17 = 891$ Work from home:

  $x + 123 + 247 + y = 891$

  $x + y = 521$ ②

- 728 of professional Work from home:

  $247 + y = 728$

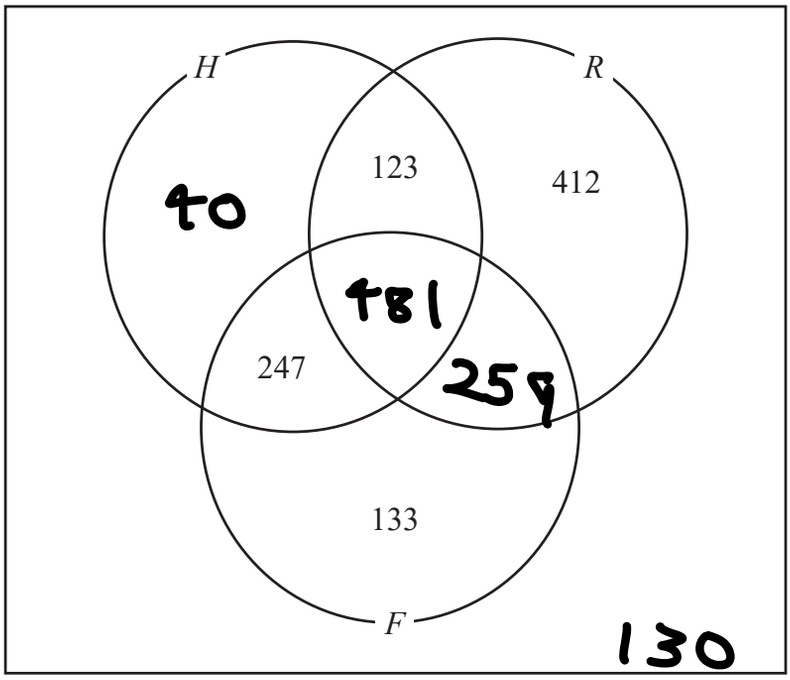
  $y = 481$

Sub y into ② $\Rightarrow x + 481 = 521 \Rightarrow x = 40$

Sub x, y into ① $\Rightarrow 481 + z = 740 \Rightarrow z = 259$

Question 5 continued

Only use this diagram if you need to redraw your Venn diagram.



The rest:
$1825 - 40 - 123 - 412 - 181 - 247 - 259 - 133 = 130$

d) $\frac{247+13}{1825} = \frac{76}{365}$

e) $\frac{133+130}{1825} = \frac{263}{1825}$

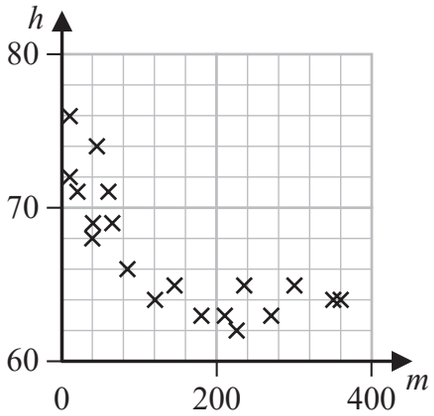
f) $\frac{247+81}{1825} = \frac{728}{891}$

(Total for Question 5 is 10 marks)

6. Anna is investigating the relationship between exercise and resting heart rate. She takes a random sample of 19 people in her year at school and records for each person

- their resting heart rate, h beats per minute
- the number of minutes, m, spent exercising each week

Her results are shown on the scatter diagram.



(a) Interpret the nature of the relationship between h and m

(1)

Anna codes the data using the formulae

$$x = \log_{10} m$$
$$y = \log_{10} h$$

The product moment correlation coefficient between x and y is -0.897

(b) Test whether or not there is significant evidence of a negative correlation between x and y
You should

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

(3)

The equation of the line of best fit of y on x is

$$y = -0.05x + 1.92$$

(c) Use the equation of the line of best fit of y on x to find a model for h on m in the form

$$h = am^k$$

where a and k are constants to be found.

(5)

a) There is a negative correlation between h and m. As one increases, the other decreases.
Note: it is not necessary to specify whether weak or strong

b) let p be the correlation for the population

$H_0: p = 0$

$H_1: p < 0$

- since < test

Critical value from tables $= -0.3887$

Test statistic/observed value $= -0.897$ (given)

We reject if in the shaded region

$-0.897 < -0.3887$ therefore there is sufficient evidence at the 5% level to reject H_0 and we can conclude that there is a negative correlation between x and y

c) $h = am^k$

log both sides to turn this into a linear model

$\log h = \log am^k$

Use log rule $\log ab = \log a + \log b$

Question 6 continued

$\log h = \log a + \log m^k$

use log rule $\log_a b^c = c\log_a b$

$\log h = \log a + k\log m$

$\log h = k\log m + \log a$

Compare this to $y = -0.05x + 1.92$

We can see that $y \rightarrow \log h$

$x \rightarrow \log m$

$k = -0.05$ and $\log a = 1.92$

$\log_{10} a = 1.92$

$10^{1.92} = a$

$a = 83.18$

$a = 83.2, k = -0.05$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 6 continued

Question 6 continued

(Total for Question 6 is 9 marks)

TOTAL FOR STATISTICS IS 50 MARKS