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

YORKSHIRE MATHS TUTOR

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. 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)

a) ${}^{15}C_3 \times (0.48)^3 \times (0.52)^{12}$

$= 0.019668 \ldots\ldots$

b) $P(P \geq 5) = 1 - P(x \leq 4)$

$1 - 0.07986789$ on calculator

$= 0.92013 \ldots$ Binomial CD

$x = 4$

$N = 15$

$p = 0.48$

b) $n = 250 \quad p = 0.48$

Normal approximation

$\mu = np = 250 \times 0.48 = 120$

$\sigma = \sqrt{np(1-p)} = \sqrt{120(1-0.48)}$

$\sigma = 7.8994 \quad (4dp)$

$Y \sim N(\mu, \sigma^2)$

$Y \sim N(120, 62.4)$

We need continuity correction

$P(Y > 110.5)$

Question 1 continued

On Calculator

Normal CD

Lower 110.5

Upper 2000

σ = 7.8994 (4dp)

p = 0.885438875

(Total for Question 1 is 6 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

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 $\mu = 8$
- a standard deviation of x cm $\sigma = x$

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   40p

(c) Calculate the expected profit per 500 of the metal rods.
Give your answer to the nearest pound.
(5)

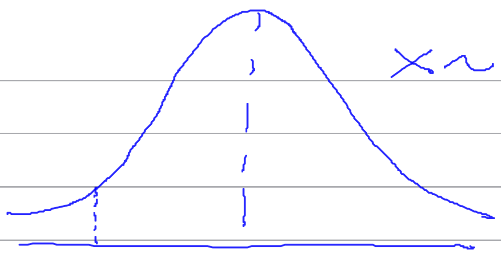
The same manufacturer makes metal hinges in large batches.

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.

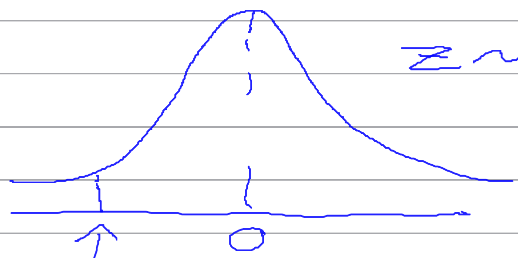
The manufacturer's aim is for 95% of batches to be accepted.

(d) Explain whether the manufacturer is likely to achieve its aim.
(4)



$X \sim N(8, \sigma^2)$

7.902 $\mu = 8$



$Z \sim N(0, 1^2)$

0

$Z = -1.959964$

Calculator
Inverse normal
Area = 0.025
$\sigma = 1$
$\mu = 0$

$Z = -1.959964$

Question 2 continued

$$Z = \frac{X - \mu}{\sigma}$$

$$\sigma = \frac{X - \mu}{Z} = \frac{7.902 - 8}{-1.959964028}$$

$$\sigma = 0.050000911$$

$\therefore x = 0.05$ (2dp) as required

b) Normal CD
lower 7.94
upper 8.09
$\sigma = 0.0500$
$\mu = 8$

$p = 0.849$ (3sf)

c) $500 \times 0.2 = £100$ to produce (1)

$P(L < 7.94) = 0.115069702$
lower -1000
upper 7.94

$P(7.94 \leq L \leq 8.09) = 0.849000106$

$P(L > 8.09) = 0.035930314$

Sells for
$(0.115069702 \times 500 \times 0.05)$
$+ (0.849000106 \times 500 \times 0.5)$
$+ (0.0359303194 \times 500 \times 0.4)$

$= 222.3128083$ (2)

Profit $=$ (2) $-$ (1) $= 122.3128$
$= £122$ (nearest pound)

Question 2 continued

d) $p = 0.015$ of fault

$X \sim B(200, 0.015)$

Binomial CD $P(X \leq 5) = 0.917...$

$x = 5$

$N = 200$

$p = 0.015$

To be accepted 5 or less have to be faulty

as $0.917 < 0.95$ manufacturer is unlikely to achieve their aim

Question 2 continued

(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) $0 < r \leq 0.05$ recorded as tr

b i) mean $= \bar{r} = \frac{\sum r}{n} = \frac{174.9}{31} = 5.6419354$

(ii) $\sigma = \sqrt{\frac{\sum r^2}{n} - \bar{r}^2} = \sqrt{\frac{3523.283}{31} - (\text{Ans})^2}$

$\sigma = 9.045598616$

c) Leuchars is in the North and Camborne is in the South

The mean is smaller for Leuchars (1.72mm) than Camborne (5.64mm) so there is no evidence to support Dian's belief

Question 3 continued

d) $\rho = 0.27$ is unlikely to be constant for all 1987

(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.

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) $H_0 : p = 0.1$
$H_1 : p \neq 0.1$

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

Binomial CD
$P(X \leq 0) = 0.0051537$
$P(X \leq 1) = 0.0337 > 0.025$
testing lower tail

$P(X \leq 8) = 0.94213$ upper tail
$P(X \leq 9) = 0.97546 > 0.975$

Critical region $X = 0$ with probability of rejection 0.00515 (3sf)

1

DO NOT WRITE IN THIS AREA

Question 4 continued

also $X \geqslant 10$
with probability of rejection
$1 - 0.97546 = 0.0245$ (2) (3 sf)

c) actual significance level
$=$ (1) $+$ (2) $= 0.00515 + 0.0245$
$= 0.02965$
$= 0.0297$ (3 sf)

d) as $X \geqslant 10$ critical region

15 is in critical region, so
reject H_0, accept H_1

There is evidence to support
the manager's belief that
there has been a change.

(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***
Professional	740	380
Skilled	275	90
Elementary	260	80

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
- 40% of skilled employees in both area A and area B work from home
- 5% of elementary employees in both area A and area B work from home
- 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

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

(4)

(d) Find $\mathrm{P}(R' \cap F)$

(1)

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

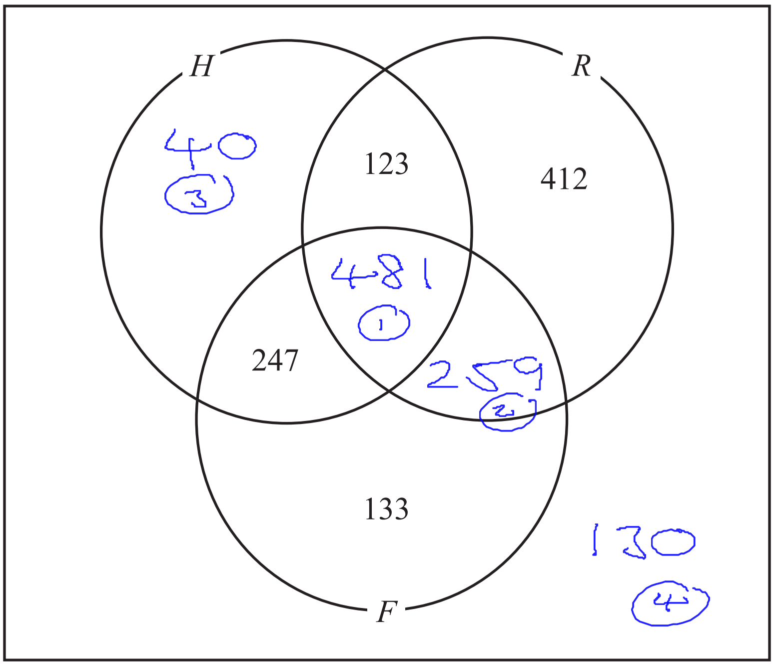
(1)

(f) Find $\mathrm{P}(F \mid H)$

(2)

H – Home F – Professional
R – area A

Question 5 continued



Work out from centre

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

a) $\frac{275 + 90}{1825} = \frac{365}{1825}$

b) $\frac{90 + 80}{1825} = \frac{170}{1825}$

c) Work on Venn digram out from centre
Centre is Home, Area A and professional
$0.65 \times 740 = 481$ 1

R is total of area A
$= 740 + 275 + 260 = 1275$
(from table)

Missing in R is
$1275 - (123 + 412 + 481)$
$= 259$ 2

Home employee total for H
$0.65 \times (740 + 380) + 0.4 \times (275 + 90)$
$+ 0.05 \times (260 + 80) = 891$

Question 5 continued

Missing space in H

$891 - (123 + 481 + 247) = 40$ (3)

Add all numbers on diagram

$40 + 123 + 481 + 247 + 412 + 259$
$+ 133 = 1695$

Number outside $= 1825 - 1695 = 130$ (4)

d) $P(R' \cap F) = \frac{247 + 133}{1825} = \frac{380}{1825}$

not R intersection F

e) $P([H \cup R]') = \frac{133 + 130}{1825}$

not $H \cup R$

$= \frac{263}{1825}$

f) $P(F \mid H) = \frac{247 + 481}{40 + 123 + 481 + 247}$

F given picking from H

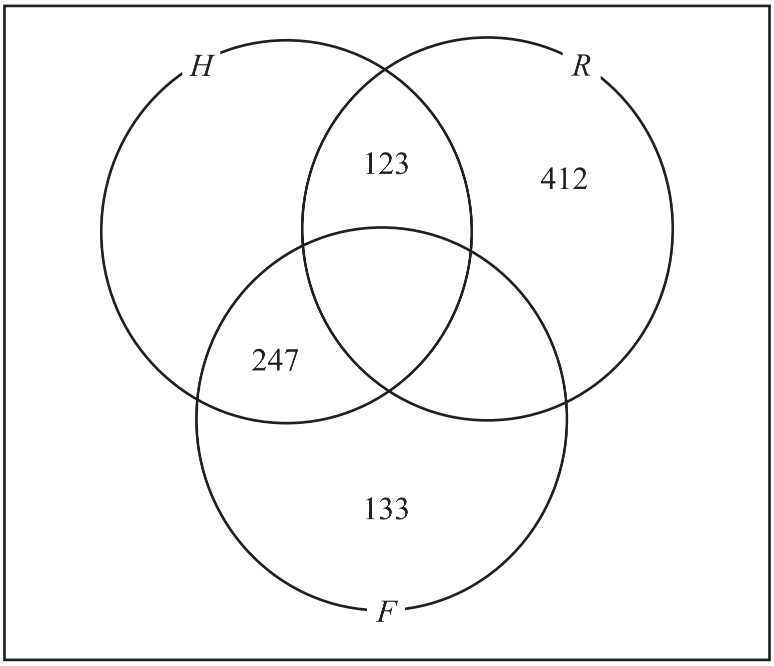
$= \frac{728}{891}$

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

P 7 2 1 3 0 A 0 1 4 2 0

Question 5 continued

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

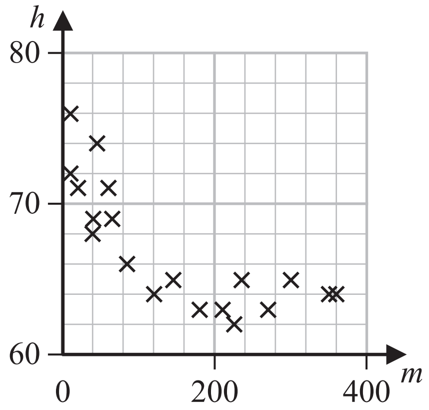


(Total for Question 5 is 10 marks)

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)

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

Question 6 continued

a) As the number of minutes exercising increases the resting heart rate decreases

b) $H_0 : \rho = 0$
$H_1 : \rho < 0$ testing for negative correlation

Product Moment Coefficient					
Level					Sample size, n
0.10	0.05	0.025	0.01	0.005	
0.8000	0.9000	0.9500	0.9800	0.9900	4
0.6870	0.8054	0.8783	0.9343	0.9587	5
0.6084	0.7293	0.8114	0.8822	0.9172	6
0.5509	0.6694	0.7545	0.8329	0.8745	7
0.5067	0.6215	0.7067	0.7887	0.8343	8
0.4716	0.5822	0.6664	0.7498	0.7977	9
0.4428	0.5494	0.6319	0.7155	0.7646	10
0.4187	0.5214	0.6021	0.6851	0.7348	11
0.3981	0.4973	0.5760	0.6581	0.7079	12
0.3802	0.4762	0.5529	0.6339	0.6835	13
0.3646	0.4575	0.5324	0.6120	0.6614	14
0.3507	0.4409	0.5140	0.5923	0.6411	15
0.3383	0.4259	0.4973	0.5742	0.6226	16
0.3271	0.4124	0.4821	0.5577	0.6055	17
0.3170	0.4000	0.4683	0.5425	0.5897	18
0.3077	0.3887	0.4555	0.5285	0.5751	19
0.2992	0.3783	0.4438	0.5155	0.5614	20

From pmcc table critical value for $n = 19$ at 5% level is -0.3887

As $-0.897 < -0.3887$ it is in the critical region

Reject H_0, accept H_1

There is a negative correlation

c) $y = -0.05x + 1.92$

$x = \log_{10} m$
$y = \log_{10} h$ make h the subject

$\log_{10} h = -0.05 \log_{10} m + 1.92$

$\log_{10} h + 0.05 \log_{10} m = 1.92$
$\log_{10} h + \log_{10} m^{0.05} = 1.92$

Question 6 continued

$\log_{10} h m^{0.05} = 1.92$

$h m^{0.05} = 10^{1.92}$

$h = 10^{1.92} m^{-0.05}$

This is in form

$h = a m^{k}$

where $a = 10^{1.92} = 83.176$

$k = -0.05$

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