Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number

Candidate Number

Monday 19 October 2020

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 5 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.

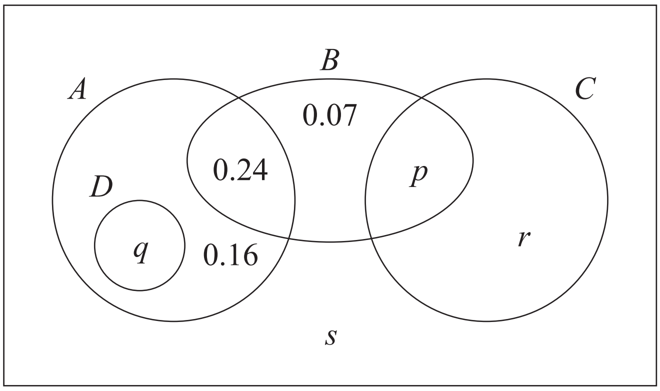
Turn over ▶

1/1/1/1/1/

P 6 6 7 8 8 A 0 1 2 0

Pearson

1. The Venn diagram shows the probabilities associated with four events, A, B, C and D



(a) Write down any pair of mutually exclusive events from A, B, C and D **(1)**

Event 1 and event 2 are mutually exclusive if they can't happen at the same time

Given that $P(B) = 0.4$

(b) find the value of p **(1)**

Given also that A and B are independent

(c) find the value of q **(2)**

Given further that $P(B' \mid C) = 0.64$

(d) find

(i) the value of r

(ii) the value of s **(4)**

a) A and C or D and C or D and B ① For any of those pairs

b) $0.24 + 0.07 + p = 0.4$

$p = 0.4 - 0.24 - 0.07$

$= 0.09$ ①

c) If A and B are independant P(A and B) = P(A) × P(B)

$P(A \text{ and } B) = 0.24$ ∴ $0.24 = P(A) \times 0.4$ ①

$P(B) = 0.4$ $P(A) = 0.6$

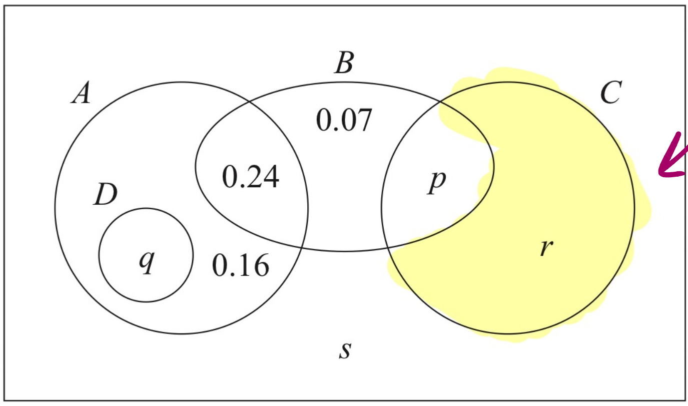
$0.24 + 0.16 + q = 0.6$

$q = 0.6 - 0.24 - 0.16$

$= 0.20$ ①

P 6 6 7 8 8 A 0 2 2 0

The Venn diagram shows the probabilities associated with four events, A, B, C and D



shaded area $P(C \cap B')$

d) $P(B' | C) = \dfrac{P(C \cap B')}{P(C)} = \dfrac{r}{p+r}$ ①

i) $\therefore 0.64 = \dfrac{r}{p+r}$ since $p = 0.09$ $0.64 = \dfrac{r}{0.09+r}$

$0.64(0.09+r) = r \Rightarrow 0.0576 + 0.64r = r$

$\Rightarrow 0.36r = 0.0576 \Rightarrow r = 0.16$ ①

ii) $0.4 + 0.16 + 0.36 + s = 1$ ①

$s = 1 - 0.4 - 0.16 - 0.36$

$= 0.08$ ①

(Total for Question 1 is 8 marks)

2. A random sample of 15 days is taken from the large data set for Perth in June and July 1987.

The scatter diagram in Figure 1 displays the values of two of the variables for these 15 days.

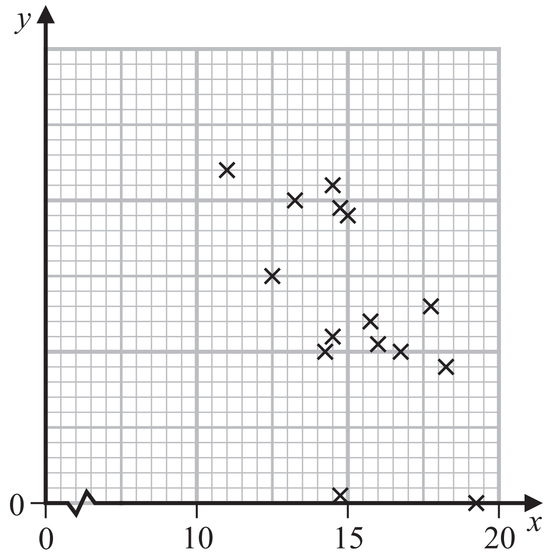


Figure 1

(a) Describe the correlation. Negative ①

(1)

The variable on the x-axis is Daily Mean Temperature measured in °C. ① rainfall or pressure

(b) Using your knowledge of the large data set, Rainfall → mm ① corresponding unit

(i) suggest which variable is on the y-axis, → Pressure → hPa, Pascals, hectopascals, mb, or millibars

(ii) state the units that are used in the large data set for this variable.

(2)

Stav believes that there is a correlation between Daily Total Sunshine and Daily Maximum Relative Humidity at Heathrow.

He calculates the product moment correlation coefficient between these two variables for a random sample of 30 days and obtains $r = -0.377$

(c) Carry out a suitable test to investigate Stav's belief at a 5% level of significance. State clearly

- your hypotheses
- your critical value

(3)

On a random day at Heathrow the Daily Maximum Relative Humidity was 97%

(d) Comment on the number of hours of sunshine you would expect on that day, giving a reason for your answer.

(1)

Humidity is high and there is evidence of correlation ①
and $r < 0$ so would expect lower than average amount of sunshine

P 6 6 7 8 8 A 0 4 2 0

DO NOT WRITE IN THIS AREA

Question 2 continued

c) let ρ be the population correlation coefficient

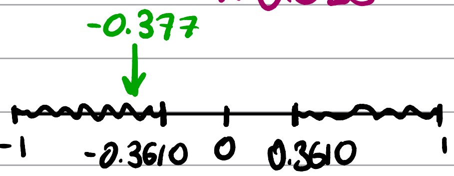
$H_0: \rho = 0$ ①

$H_1: \rho \neq 0$

$n = 30$

$SL = 5\%$

two-tailed test so $\div 2$ $\therefore 0.025$



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
0.2914	0.3687	0.4329	0.5034	0.5487	21
0.2841	0.3598	0.4227	0.4921	0.5368	22
0.2774	0.3515	0.4133	0.4815	0.5256	23
0.2711	0.3438	0.4044	0.4716	0.5151	24
0.2653	0.3365	0.3961	0.4622	0.5052	25
0.2598	0.3297	0.3882	0.4534	0.4958	26
0.2546	0.3233	0.3809	0.4451	0.4869	27
0.2497	0.3172	0.3739	0.4372	0.4785	28
0.2451	0.3115	0.3673	0.4297	0.4705	29
0.2407	0.3061	0.3610	0.4226	0.4629	30
0.2070	0.2638	0.3120	0.3665	0.4026	40
0.1843	0.2353	0.2787	0.3281	0.3610	50

Critical value $= -0.3610$ ①

$r = -0.377 < -0.3610$ ①

$\therefore$ Significant result and there is evidence of a correlation between Daily total Sunshine and Daily Max. Relative humidity

d)

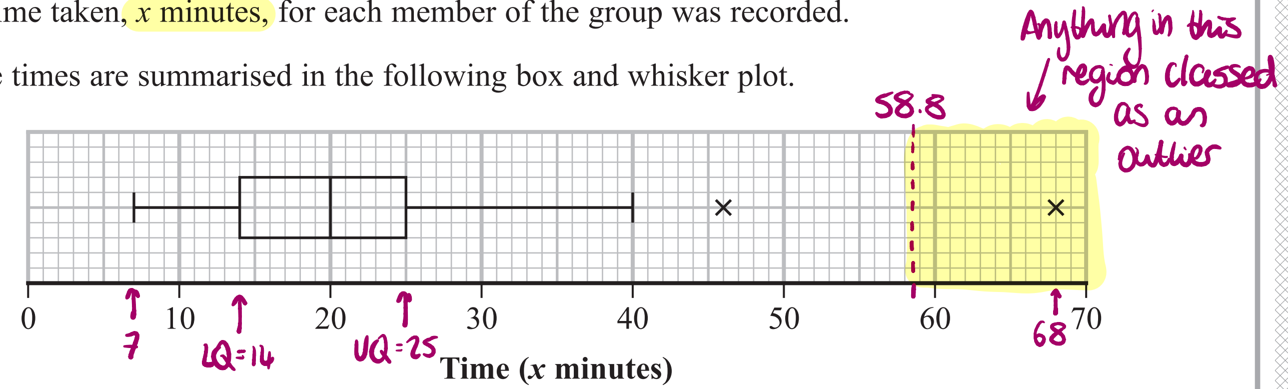
Humidity is high and there is evidence of correlation ① and $r < 0$ so would expect lower than average amount of sunshine

DO NOT WRITE IN THIS AREA

3. Each member of a group of 27 people was timed when completing a puzzle.

The time taken, x minutes, for each member of the group was recorded.

These times are summarised in the following box and whisker plot.



(a) Find the range of the times.

Range = Highest value − Lowest Value $\quad 68-7=61$ ①

(1)

(b) Find the interquartile range of the times.

$IQR = UQ - LQ = 25-14 = 11$ ①

(1)

For these 27 people $\sum x = 607.5$ and $\sum x^2 = 17\,623.25$

$\bar{x}$ represents "the mean"

(c) calculate the mean time taken to complete the puzzle,

$\bar{x} = \dfrac{607.5}{27} = 22.5$ ①

(1)

(d) calculate the standard deviation of the times taken to complete the puzzle.

$\sigma = \sqrt{\dfrac{\Sigma x^2}{n} - (\bar{x}^2)} = \sqrt{\dfrac{17623.25}{27} - (22.5^2)}$ ① $= 12.10218... = 12.1$ (1dp) ①

(2)

Taruni defines an outlier as a value more than 3 standard deviations above the mean.

(e) State how many outliers Taruni would say there are in these data, giving a reason for your answer.

$\bar{x} + 3\sigma = 22.5 + 3(12.10218...) = 58.8$ (1dp) $\therefore$ only one Outlier ①

(1)

Adam and Beth also completed the puzzle in a minutes and b minutes respectively, where $a > b$.
When their times are included with the data of the other 27 people

- the median time increases
- the mean time does not change

(f) Suggest a possible value for a and a possible value for b, explaining how your values satisfy the above conditions.

(3)

(g) Without carrying out any further calculations, explain why the standard deviation of all 29 times will be lower than your answer to part (d).

(1)

f) Median increases implies that both values must be > 20 ①

Since current median is 20

$\dfrac{y}{29} = 22.5$ because we know mean doesn't change

let y = new total sum of times

$y = 652.5$

P 6 6 7 8 8 A 0 8 2 0

Question 3 continued

new total sum of times = 652.5

$652.5 - 607.5 = 45 \quad \therefore \quad a+b=45$ (1)

"old sum"

$a+b=45$
$a>20, b>20$
$a>b$

Possible values could be: $a=24$ $b=21$ (1)

g) $\bar{x}=22.5 \quad \sigma=12.1$

$22.5+12.1=34.6$
$22.5-12.1=10.4$

because of conditions on a, b impossible for either to be outside
$10.4<a<34.6$ and $10.4<b<34.6$
(so both values less than 1 standard deviation from mean) ← so values "less spread out"
so smaller standard deviation

Both values will be less than 1 standard deviation from the mean and so the standard deviation of all values will be smaller (1)

4. The discrete random variable D has the following probability distribution

d	10	20	30	40	50
$P(D = d)$	$\frac{k}{10}$	$\frac{k}{20}$	$\frac{k}{30}$	$\frac{k}{40}$	$\frac{k}{50}$

where k is a constant.

(a) Show that the value of k is $\frac{600}{137}$

(2)

The random variables D_1 and D_2 are independent and each have the same distribution as D.

(b) Find $P(D_1 + D_2 = 80)$
Give your answer to 3 significant figures.

(3)

The value obtained, d, is the common difference of an arithmetic sequence.

The first 4 terms of this arithmetic sequence are the angles, measured in degrees, of quadrilateral Q

(c) Find the exact probability that the smallest angle of Q is more than 50°

(5)

a) $\frac{k}{10} + \frac{k}{20} + \frac{k}{30} + \frac{k}{40} + \frac{k}{50} = 1$ ① $\quad \frac{137k}{600} = 1 \quad k = \frac{600}{137}$ as needed ①

b) $40 + 40 = 80$
or $50 + 30 = 80$

multiplication

$D_1 = 40, D_2 = 40$ For 'And' $(\times)$ ✓
$D_1 = 50, D_2 = 30$ For 'Or' $(+)$ ✓ addition
$D_1 = 30, D_2 = 50$ ①

$P(D_1 + D_2 = 80) = P(D_1 = 40 \text{ and } D_2 = 50) + P(D_1 = 30 \text{ and } D_2 = 50) + P(D_1 = 50 + D_2 = 30)$

$= \left(\frac{k}{40} \times \frac{k}{40}\right) + \left(\frac{k}{50} \times \frac{k}{30}\right) + \left(\frac{k}{30} \times \frac{k}{50}\right) = \frac{k^2}{1600} + \frac{k^2}{1500} + \frac{k^2}{1500}$

① $= \frac{47k^2}{24000}$ since $k = \frac{600}{137}$

$= \frac{47\left(\frac{600}{137}\right)^2}{24000} = 0.0375619\ldots = 0.0376$ (3sf) ①

DO NOT WRITE IN THIS AREA

Question 4 continued

c) let a = first term in sequence

Angles: $a, a+d, a+2d, a+3d$ ①

Interior angles in a quadrilateral add to 360°

$a + a + d + a + 2d + a + 3d = 360$ ①

$4a + 6d = 360$

$\div 2$

$2a + 3d = 180$ ①

For $a > 50$ only possible cases (see working)

$d = 10, a = 75$ or $d = 20, a = 60$ ①

let $d = D$

Sub $k = \frac{600}{137}$

$P(D = 10 \text{ or } D = 20) = \left(\frac{k}{10} + \frac{k}{20}\right) = \left(\frac{3k}{20}\right) = \frac{90}{137}$ ①

Possibe cases:

Since $2a + 3d = 180$

when $d = 10$

$2a + 3(10) = 180$

$2a = 180 - 30$

$2a = 150$

$a = 75$

meets condition $a > 50$

when $d = 20$

$2a + 3(20) = 180$

$2a = 180 - 60$

$2a = 120$

$a = 60$

meets condition $a > 50$

when $d = 30$

$2a + 3(30) = 180$

$2a = 180 - 90$

$2a = 90$

$a = 45$

doesn't meet condition $a > 50$ and no point trying any more possibilities for d since as d gets bigger a gets smaller

5. A health centre claims that the time a doctor spends with a patient can be modelled by a normal distribution with a mean of 10 minutes and a standard deviation of 4 minutes.

(a) Using this model, find the probability that the time spent with a randomly selected patient is more than 15 minutes.

(1)

Some patients complain that the mean time the doctor spends with a patient is more than 10 minutes.

The receptionist takes a random sample of 20 patients and finds that the mean time the doctor spends with a patient is 11.5 minutes.

(b) Stating your hypotheses clearly and using a 5% significance level, test whether or not there is evidence to support the patients' complaint.

(4)

The health centre also claims that the time a dentist spends with a patient during a routine appointment, T minutes, can be modelled by the normal distribution where $T \sim N(5, 3.5^2)$

(c) Using this model,

(i) find the probability that a routine appointment with the dentist takes less than 2 minutes

(1)

(ii) find $P(T < 2 \mid T > 0)$

(3)

(iii) hence explain why this normal distribution may not be a good model for T.

(1)

The dentist believes that she cannot complete a routine appointment in less than 2 minutes.

She suggests that the health centre should use a refined model only including values of $T > 2$

(d) Find the median time for a routine appointment using this new model, giving your answer correct to one decimal place.

(5)

a) let X = time spent $X \sim N(10, 4^2)$
μ σ^2

$P(X > 15) = 0.105649... = 0.106$ (3dp) 1

b) $H_0: \mu = 10$ 1
$H_1: \mu > 10$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ where n = sample size

Significance level = 5%

$\bar{X} \sim N(10, \frac{4^2}{20})$ 2 $\frac{4^2}{20} = \sigma^2 \Rightarrow \frac{4}{\sqrt{20}} = \sigma$

1

$P(\bar{X} > 11.5) = 0.046766...$ 0.0468 (4dp) < 0.05 so there is evidence to support the patients complaint

P 6 6 7 8 8 A 0 1 6 2 0

DO NOT WRITE IN THIS AREA

Question 5 continued

c) $P(T<2) = 0.1956\ldots = 0.196$ (3dp) (1)

i)

$T \sim N(5, 3.5^2)$

ii) $P(T<2 \mid T>0)$ — 'given that'

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

(3)

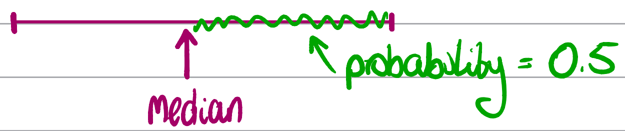
$$P(T<2 \mid T>0) = \frac{P(0<T<2)}{P(T>0)} = \frac{0.119119\ldots}{0.923436\ldots} = 0.1289955\ldots = 0.129 \text{ (3dp)}$$

iii) The current model suggests non-negligible probability of T values < 0 which is impossible (1)

$P(T<0) \approx 0.077$

(length of time can't be negative so suggests model unsuitable)

d)



$P(T>t \mid T>2) = 0.5$ (1)

median

(1)

$$P(T>t \mid T>2) = \frac{P(T>t)}{P(T>2)} = 0.5$$

using conditional probability formula

$T \sim N(5, 3.5^2)$

Note: $P(T>t \cap T>2) = P(T>t)$ since new model only includes $T>2$ so median definetly more than 2

$$\frac{P(T>t)}{0.8043\ldots} = 0.5$$

(1)

$P(T>t) = 0.40215\ldots$

(1)

$P(T<t) = 1 - 0.40215\ldots = 0.5978\ldots$

Need to do this step for calculators that can only test the 'lower tail'

$t = 5.867\ldots$

$= 5.9$ (1dp) (1)

use inverse normal on calculator