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

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 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.*

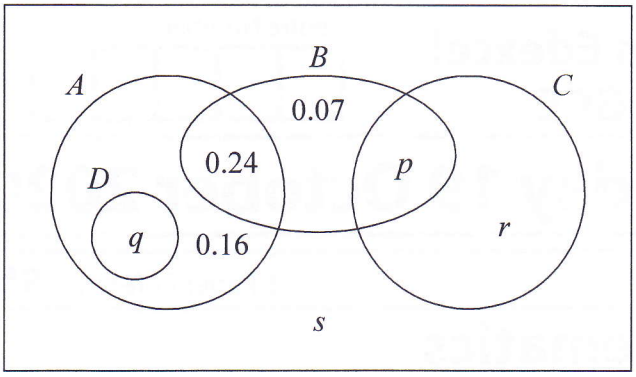
Advice

- 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

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)

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 $P(B' | C) = 0.64$

(d) find

(i) the value of r

(ii) the value of s

(4)

a) Mutually exclusive A and C
or D and B
or D and C

b) $P(B) = 0.4$ $p = 0.4 - 0.24 - 0.07 = 0.09$

c) If A and B independent
$P(A) \times P(B) = P(A \cap B)$
$P(A) \times 0.4 = 0.24$
$P(A) = 0.6$
$q = 0.6 - 0.24 - 0.16 = 0.2$

Question 1 continued

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

$0.64 = \frac{r}{p+r}$ $\quad p = 0.09$

$0.64(0.09 + r) = r$

$0.0576 + 0.64r = r$

$0.0576 = r - 0.64r$

$0.0576 = 0.36r$

$r = 0.16$

$s = 1 - (0.6 + 0.07 + 0.09 + 0.16)$

$s = 1 - 0.92 = 0.08$

(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.

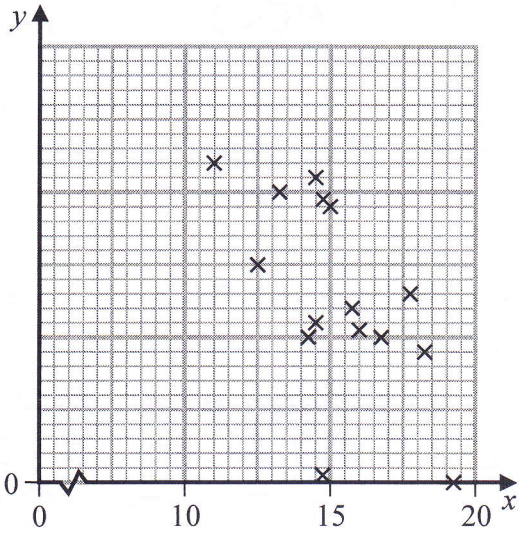


Figure 1

(a) Describe the correlation.

(1)

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

(b) Using your knowledge of the large data set,

(i) suggest which variable is on the y-axis,

(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)

Question 2 continued

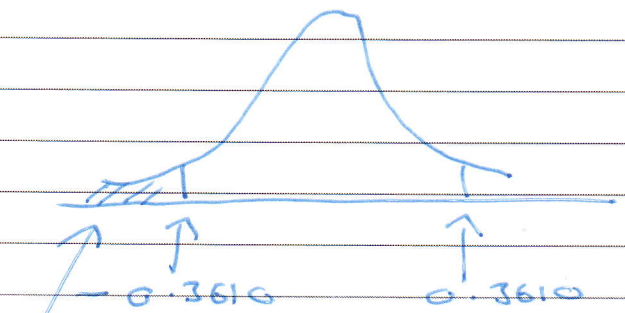
2a) Negative correlation

b) (i) Rainfall
(ii) mm

c) $r = -0.377$

PMCC tables $n = 30$ 2.5% (2 tail)
PMCC $= 0.3610$

$H_0: \rho = 0$ $H_1: \rho \neq 0$
two tail test



Critical value
$= -0.3610$ in lower tail

$r = -0.377$
(test statistic)

As $r = -0.377$ is in the lower tail
reject H_0, accept H_1

There is a correlation between the Daily Total Sunshine and Daily Maximum Relative humidity at Heathrow.

d) As humidity is high (97%) and there is negative correlation ($r < 0$), we would expect the number of hours of sunshine at Heathrow to be lower than the average.

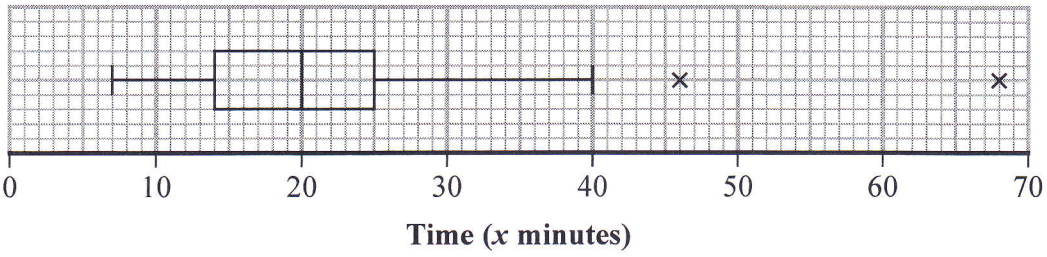
P 6 6 7 8 8 A 0 5 2 0

Turn over

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.

(1)

(b) Find the interquartile range of the times.

(1)

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

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

(1)

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

(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.

(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)

3 a) range = 68 − 7 = 61

b) IQR = UQ − LQ = 25 − 14 = 11

P 6 6 7 8 8 A 0 8 2 0

Question 3 continued

c) mean $= \frac{\Sigma x}{n} = \frac{607.5}{27} = 22.5$

d) $\sigma = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} = \sqrt{\frac{17623.25}{27} - 22.5^2}$

$\sigma = 12.1$

e) $3 \times 12.1 = 36.3 = 3$ sd

upper $22.5 + 36.3 = 58.8$
lower $22.5 - 36.3 = -13.8$

Only one outlier (68) above 58.8 minutes

f) 7 . . . . . 20 22 23 68

added

14th 15th (new median)

median

current total time $= 22.5 \times 27 = 607.5$
add 2 people
total time $= 22.5 \times 29 = 652.5$
(we've added 45)
say, 22 + 23
New median will be 15th (22)
- Median increased from 20 to 22
- Mean the same (22.5)

g) Both 22 and 23 are less than 1 s.d. (12.1 units) from the mean, so s.d. of all 29 values will be smaller.

P 6 6 7 8 8 A 0 9 2 0

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)

A single observation of D is made.

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)

4a) $k\left(\frac{1}{10} + \frac{1}{20} + \frac{1}{30} + \frac{1}{40} + \frac{1}{50}\right) = 1$

$\frac{137}{600}k = 1$

$k = \frac{600}{137}$ (as required)

b) For $D_1 + D_2 = 80$

$D_1 = 30 \quad D_2 = 50$

or $D_1 = 50 \quad D_2 = 30$

or $D_1 = 40 \quad D_2 = 40$

$= \frac{k}{30} \times \frac{k}{50} + \frac{k}{50} \times \frac{k}{30} + \frac{k}{40} \times \frac{k}{40}$

$= k^2\left(\frac{1}{1500} + \frac{1}{1500} + \frac{1}{1600}\right)$

$= \left(\frac{600}{137}\right)^2\left(\frac{1}{1500} + \frac{1}{1500} + \frac{1}{1600}\right)$

$= \frac{705}{18769} = 0.03756 = 0.0376$ (3 sf)

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

Question 4 continued

	1st	2nd	3rd	4th
c)	a	$a+d$	$a+2d$	$a+3d$

$S_n = \frac{n}{2}(2a + (n-1)d)$

$n = 4$, $S_n = 360$ (angles in quadrilateral)

$360 = 2(2a + 3d)$

$\therefore 2a + 3d = 180$

From table

if $d = 10$, $a = \frac{180-30}{2} = 75$ ✓

$d = 20$, $a = \frac{180-60}{2} = 60$ ✓

$d = 30$, $a = \frac{180-90}{2} = 45$ ✗

a must be > 50

$\therefore d = 10$ or $d = 20$

$= \frac{k}{10} + \frac{k}{20}$

$= \frac{600}{137}\left(\frac{1}{10} + \frac{1}{20}\right)$

$= \frac{90}{137}$ exact probability

P 6 6 7 8 8 A 0 1 3 2 0

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) $X \sim N(10, 4^2)$
$P(x > 15)$

Normal CD
lower = 15
upper = 1000
σ = 4
μ = 10
$p = 0.1056$

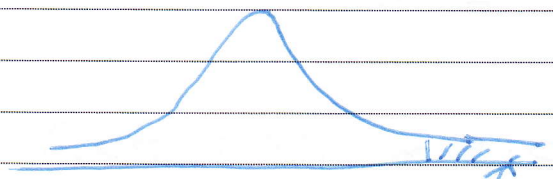
P 6 6 7 8 8 A 0 1 6 2 0

Question 5 continued

b) $n = 20$ $\quad \sigma = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{20}} = 0.8944$

$\bar{x} = 11.5$ $\quad H_0: \mu = 10 \quad H_1: \mu > 10$

5% 1 tail test



Inverse normal
Area = 0.95
$\sigma = 0.8944$
$\mu = 10$

0.95

test statistic
$\bar{x} = 11.5$ minutes
in tail

critical value

$x = 11.4711$

As 11.5 is in tail, reject H_0, accept H_1, there is evidence to support complaint that $\mu > 10$ minutes for appointments now.

c) (i) $T \sim N(5, 3.5^2)$
$P(T < 2)$

Normal CD
lower = −1000
upper = 2
$\sigma = 3.5$
$\mu = 5$

$p = 0.1957$ (4dp)

(ii) $P(0 < T < 2) = 0.1191$

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

$= \frac{0.1191}{0.9234}$

$P(T<2)$

lower 0
upper 1000

$= 0.1290$ (4dp)

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

Question 5 continued

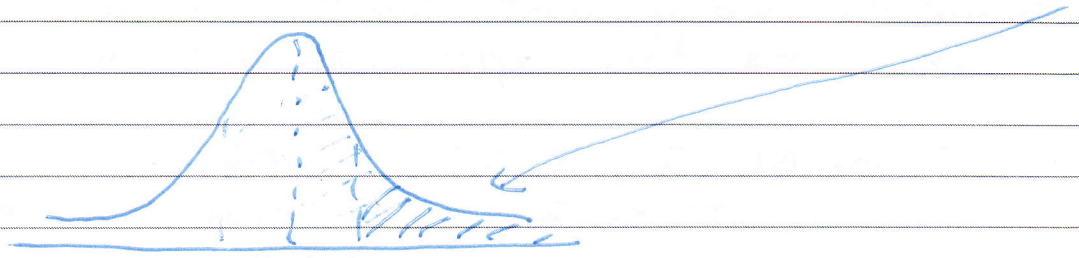
5a) median $p = 0.5$

$\therefore \dfrac{P(T > t)}{P(T > 2)} = 0.5$

$P(T > t) = 0.5 \times P(T > 2)$

Lower = 2
upper = 1000
$\sigma = 3.5$
$\mu = 5$
$p = 0.8043$

$P(T > t) = 0.5 \times 0.8043 = 0.4022$



t $p = 0.4022$

$1 - 0.4022 = 0.5978$

0.5978

Inverse normal
Area = 0.5978
$\sigma = 3.5$
$\mu = 5$

$t = 5.8667$
$t = 5.9$ minutes (1dp)

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