Pearson Education accepts no responsibility whatsoever for the accuracy or method of working in the answers given.

Write your name here

Surname	Other names

Pearson Edexcel GCE

Centre Number	Candidate Number

Statistics S1

Advanced/Advanced Subsidiary

Wednesday 15 June 2016 – Morning
Time: 1 hour 30 minutes

Paper Reference
6683/01

You must have:
Mathematical Formulae and Statistical Tables (Pink)

Total Marks

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. 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). Coloured pencils and highlighter pens must not be used.
- **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 the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information

- The total mark for this paper is 75.
- 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

P46673A

1/1/1/1/

P 4 6 6 7 3 A 0 1 2 4

PEARSON

Leave blank

1. A biologist is studying the behaviour of bees in a hive. Once a bee has located a source of food, it returns to the hive and performs a dance to indicate to the other bees how far away the source of the food is. The dance consists of a series of wiggles. The biologist records the distance, d metres, of the food source from the hive and the average number of wiggles, w, in the dance.

Distance, d m	30	50	80	100	150	400	500	650
Average number of wiggles, w	0.725	1.210	1.775	2.250	3.518	6.382	8.185	9.555

[You may use $\sum w = 33.6$ $\sum dw = 13\,833$ $S_{dd} = 394\,600$ $S_{ww} = 80.481$ (to 3 decimal places)]

(a) Show that $S_{dw} = 5601$

(2)

(b) State, giving a reason, which is the response variable.

(1)

(c) Calculate the product moment correlation coefficient for these data.

(2)

(d) Calculate the equation of the regression line of w on d, giving your answer in the form $w = a + bd$

(4)

A new source of food is located 350 m from the hive.

(e) (i) Use your regression equation to estimate the average number of wiggles in the corresponding dance.

(ii) Comment, giving a reason, on the reliability of your estimate.

(2)

a) $S_{dw} = \sum dw - \frac{(\sum d)(\sum w)}{n}$

$= 13833 - \frac{(1960)(33.6)}{8}$

$= 5601$

b) w (the number of wiggles is a response to the distance)

Leave blank

Question 1 continued

c/ $r = \dfrac{S_{dw}}{\sqrt{S_{dd} S_{ww}}}$

$= \dfrac{5601}{\sqrt{(394600)(80.481)}}$

$= 0.994 \quad (3sf)$

d) $b = \dfrac{S_{dw}}{S_{dd}}$

$= \dfrac{5601}{394600}$

$= 0.0142 \ (3sf)$

$a = \bar{w} - b\bar{d}$

$= \dfrac{33.6}{8} - 0.0142\left(\dfrac{1960}{8}\right)$

$= 0.722 \ (3sf)$

$w = 0.722 + 0.0142d$

e)

i/ $w = 0.722 + 0.0142(350)$

$= 5.692$

$= 5.69 \ (3sf)$

ii/ It is reliable because 350 is within the range of data.

Leave blank

2. The discrete random variable X has the following probability distribution, where p and q are constants.

$\frac{1}{x}$	$-\frac{1}{2}$	-1	2	$\frac{2}{3}$	$\frac{1}{2}$
x	-2	-1	$\frac{1}{2}$	$\frac{3}{2}$	2
$\mathrm{P}(X = x)$	p	q	0.2	0.3	p

(a) Write down an equation in p and q

(1)

Given that $\mathrm{E}(X) = 0.4$

(b) find the value of q

(3)

(c) hence find the value of p

(2)

Given also that $\mathrm{E}(X^2) = 2.275$

(d) find $\mathrm{Var}(X)$

(2)

Sarah and Rebecca play a game.

A computer selects a single value of X using the probability distribution above.

Sarah's score is given by the random variable $S = X$ and Rebecca's score is given by the random variable $R = \frac{1}{X}$

(e) Find $\mathrm{E}(R)$

(3)

Sarah and Rebecca work out their scores and the person with the higher score is the winner. If the scores are the same, the game is a draw.

(f) Find the probability that

(i) Sarah is the winner,

(ii) Rebecca is the winner.

(4)

a/ $p + q + 0.2 + 0.3 + p = 1$

$2p + q + 0.5 = 1$

$2p + q = 0.5$

Leave blank

Question 2 continued

b) $-2p - q + \frac{1}{2}(0.2) + \frac{3}{2}(0.3) + 2p = 0.4$

$-q + 0.1 + 0.45 = 0.4$

$-q + 0.55 = 0.4$

$\underline{\underline{q = 0.15}}$

c) $2p + q = 0.5$

$2p + 0.15 = 0.5$

$2p = 0.35$

$\underline{\underline{p = 0.175}}$

d) $\mathrm{Var}(X) = E(X^2) - (E(X))^2$

$= 2.275 - 0.4^2$

$= \underline{\underline{2.115}}$

e) $E(R) = \frac{1}{-2}(0.175) + \frac{1}{-1}(0.15) + \frac{1}{0.5}(0.2) + \frac{1}{\cancel{3/2}\ 1.5}(0.3)$

$+ \frac{1}{2}(0.175)$

$= \underline{\underline{0.45}}$

f) Sarah wins when $X > \frac{1}{X}$

$x = \frac{3}{2}$ or 2

$0.3 + 0.175 = \underline{\underline{0.475}}$

P 4 6 6 7 3 A 0 7 2 4

Turn over

Question 2 continued

Leave blank

ii/ Rebecca wins when $\frac{1}{x} > x$

-2 or $\frac{1}{2}$

$0.175 + 0.2 = \underline{\underline{0.375}}$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

3. Before going on holiday to *Seapron*, Tania records the weekly rainfall (x mm) at *Seapron* for 8 weeks during the summer. Her results are summarised as

$$\sum x = 86.8 \qquad \sum x^2 = 985.88$$

(a) Find the standard deviation, σ_x, for these data.

(3)

Tania also records the number of hours of sunshine (y hours) per week at *Seapron* for these 8 weeks and obtains the following

$$\bar{y} = 58 \qquad \sigma_y = 9.461 \text{ (correct to 4 significant figures)} \qquad \sum xy = 4900.5$$

(b) Show that $S_{yy} = 716$ (correct to 3 significant figures)

(1)

(c) Find S_{xy}

(2)

(d) Calculate the product moment correlation coefficient, r, for these data.

(2)

During Tania's week-long holiday at *Seapron* there are 14 mm of rain and 70 hours of sunshine.

(e) State, giving a reason, what the effect of adding this information to the above data would be on the value of the product moment correlation coefficient.

(2)

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

$= \sqrt{\frac{(985.88)}{8} - \left(\frac{86.8}{8}\right)^2}$

$= 2.347871376$

$= \underline{2.35 \text{ (3sf)}}$

b/ $\sigma^2 = \frac{Syy}{n} \quad \therefore \quad Syy = \sigma^2 \times n$

$= (9.461)^2 \times 8$

$= \underline{716 \text{ (3sf)}}$

Leave blank

Question 3 continued

c)
$$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$$

$$= 4900.5 - \frac{(86.8)(58 \times 8)}{8}$$

$$= \cancel{133.9}$$

$$= -133.9$$

d)
$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

$$\cancel{S_{xx} = \sigma^2 \times 8}$$

$$= \frac{-133.9}{\sqrt{(44.1)(716)}}$$

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$= 985.88 - \frac{(86.8)^2}{8}$$

$$= -0.754 \text{ (3sf)}$$

$$= 44.1$$

e)
In this week there is above average sunshine and above average rain. Given the negative correlation we would expect one to be above average and one below average

This would move the pmcc closer to zero.

P46673A01124

Turn over

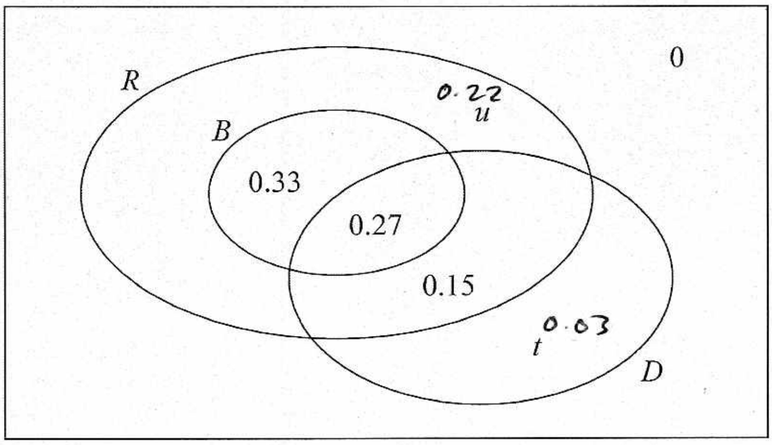
4. The Venn diagram shows the probabilities of customer bookings at Harry's hotel.

R is the event that a customer books a room

B is the event that a customer books breakfast

D is the event that a customer books dinner

u and t are probabilities.



(a) Write down the probability that a customer books breakfast but does not book a room. **(1)**

Given that the events B and D are independent

(b) find the value of t **(4)**

(c) hence find the value of u **(2)**

(d) Find

(i) $\mathrm{P}(D \mid R \cap B)$

(ii) $\mathrm{P}(D \mid R \cap B')$ **(4)**

A coach load of 77 customers arrive at Harry's hotel.

Of these 77 customers

40 have booked a room and breakfast

37 have booked a room without breakfast

(e) Estimate how many of these 77 customers will book dinner. **(2)**

Question 4 continued

a/ 0

b/ $P(B) \times P(D) = P(B \cap D)$

$0.6 \times P(D) = 0.27$

$P(D) = \frac{0.27}{0.6}$

$= \underline{\underline{0.45}}$

$t = 0.45 - 0.15 - 0.27$

$= \underline{\underline{0.03}}$

c/ $u = 1 - 0.33 - 0.27 - 0.15 - 0.03$

$= \underline{\underline{0.22}}$

d/

i/ $\frac{0.27}{0.6} = \underline{\underline{0.45}}$

ii/ $\frac{0.15}{0.37} = \underline{\underline{\frac{15}{37}}}$

e/ $40 \times 0.45 + 37 \times \frac{15}{37} = \underline{\underline{33}}$

Leave blank

Leave blank

5. A midwife records the weights, in kg, of a sample of 50 babies born at a hospital. Her results are given in the table below.

Weight (w kg)	**Frequency (f)**	**Weight midpoint (x)**
$0 \leqslant w < 2$	1	1
$2 \leqslant w < 3$	8	2.5
$3 \leqslant w < 3.5$	17	3.25
$3.5 \leqslant w < 4$	17	3.75
$4 \leqslant w < 5$	7	4.5

[You may use $\sum fx^2 = 611.375$]

A histogram has been drawn to represent these data.

The bar representing the weight $2 \leqslant w < 3$ has a width of 1 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing a weight of $3 \leqslant w < 3.5$
(3)

(b) Use linear interpolation to estimate the median weight of these babies.
(2)

(c) (i) Show that an estimate of the mean weight of these babies is 3.43 kg.

(ii) Find an estimate of the standard deviation of the weights of these babies.
(3)

Shyam decides to model the weights of babies born at the hospital, by the random variable W, where $W \sim \mathrm{N}(3.43, 0.65^2)$

(d) Find $\mathrm{P}(W < 3)$
(3)

(e) With reference to your answers to (b), (c)(i) and (d) comment on Shyam's decision.
(3)

A newborn baby weighing 3.43 kg is born at the hospital.

(f) Without carrying out any further calculations, state, giving a reason, what effect the addition of this newborn baby to the sample would have on your estimate of the

(i) mean,

(ii) standard deviation.
(3)

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

Question 5 continued

a) $2 \leqslant w \leqslant 3$ frequency $= 8$

width $\times$ height $= 1 \times 4 = 4\text{cm}^2$

Area $= \frac{1}{2}$ Frequency

class width $\times 1 =$ width

$\frac{\text{freq}}{\text{class width}} \times 0.5 =$ height.

$3 \leqslant w < 3.5$ class width $= 0.5$

$\therefore$ width $= \underline{\underline{0.5}}$ cm

$\frac{\text{freq}}{\text{c.w}} \times 0.5 =$ height

$\frac{17}{0.5} \times 0.5 =$ height

height $= \underline{\underline{17 \text{ cm}}}$

b) 25^{th} baby.

$$3 + \frac{16}{17}(0.5)$$

$= 3.47$ kg (3sf)

c i) $\frac{1(1) + 8(2.5) + 17(3.25) + 17(3.75) + 7(4.5)}{50}$

$= \underline{\underline{3.43 \text{ kg}}}$

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

Leave blank

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

Question 5 continued

$$\sigma = \sqrt{\frac{611.375}{50} - (3.43)^2}$$

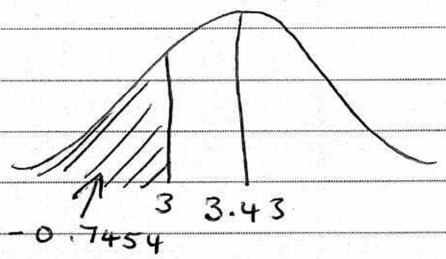
$$= \underline{0.680} \quad (3sf)$$

d/

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

$$z = \frac{3 - 3.43}{0.65}$$

$$= -0.66$$



$$P(W<3) = 1 - 0.7454$$
$$= \underline{0.2546}$$

e/ The mean is almost equal to the median so there is no skew indicating that a normal distribution is a good model.

$\frac{9}{50}$ babies were less than 3kg (18%)

this is lower than (25%) suggesting the model may not be suitable.

Q5

(Total 17 marks)

Leave blank

5
Question ~~4~~ continued

f i) The Mean would be unchanged as the new value is equal to the mean.

ii The standard deviation would decrease the difference from the mean would remain the same with more data (51 instead of 50)

Q4

17
(Total ~~13~~ marks)

P46673A01724

Turn over

Leave blank

6. The time, in minutes, taken by men to run a marathon is modelled by a normal distribution with mean 240 minutes and standard deviation 40 minutes.

(a) Find the proportion of men that take longer than 300 minutes to run a marathon.
(3)

Nathaniel is preparing to run a marathon. He aims to finish in the first 20% of male runners.

(b) Using the above model estimate the longest time that Nathaniel can take to run the marathon and achieve his aim.
(3)

The time, W minutes, taken by women to run a marathon is modelled by a normal distribution with mean μ minutes.

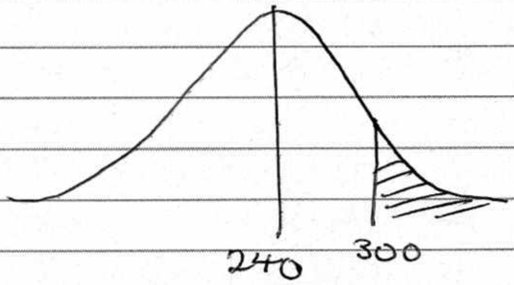
Given that $P(W < \mu + 30) = 0.82$

(c) find $P(W < \mu - 30 \mid W < \mu)$
(3)

$z = \dfrac{x - \mu}{\sigma}$

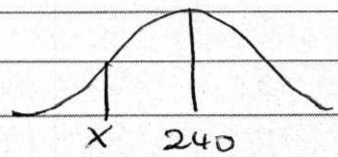
$= \dfrac{300 - 240}{40}$

$= 1.5$



$1 - 0.9332 = \underline{\underline{0.0668}}$

b/ $z = -0.8416$



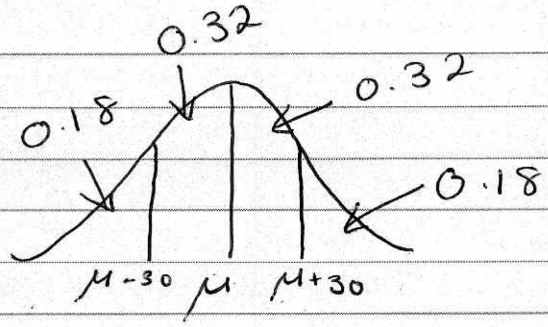
$-0.8416 = \dfrac{x - 240}{40}$

$40(-0.8416) = x - 240$

$x = 206$ minutes (3sf)

Question 6 continued

c/



$$\frac{0.18}{0.5} = \underline{\underline{0.36}}$$

P 4 6 6 7 3 A 0 2 3 2 4