

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	Centre Number	Candidate Number
Edexcel GCE	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Statistics S1		
Advanced/Advanced Subsidiary		
Wednesday 7 June 2017 – 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 ►

P48947A

©2017 Pearson Education Ltd.

1/1/1/1




Pearson

1. A clothes shop manager records the weekly sales figures, £ s , and the average weekly temperature, t °C, for 6 weeks during the summer. The sales figures were coded so that

$$w = \frac{s}{1000}$$

The data are summarised as follows

$$S_{ww} = 50 \quad \sum wt = 784 \quad \sum t^2 = 2435 \quad \sum t = 119 \quad \sum w = 42$$

- (a) Find S_{wt} and S_{tt} (3)
- (b) Write down the value of S_{ss} and the value of S_{st} (2)
- (c) Find the product moment correlation coefficient between s and t . (2)

The manager of the clothes shop believes that a linear regression model may be appropriate to describe these data.

- (d) State, giving a reason, whether or not your value of the correlation coefficient supports the manager's belief. (1)
- (e) Find the equation of the regression line of w on t , giving your answer in the form $w = a + bt$ (3)
- (f) Hence find the equation of the regression line of s on t , giving your answer in the form $s = c + dt$, where c and d are correct to 3 significant figures. (2)
- (g) Using your equation in part (f), interpret the effect of a 1°C increase in average weekly temperature on weekly sales during the summer. (1)

$$a) \quad S_{wt} = \sum wt - \frac{\sum w \sum t}{n} \quad S_{tt} = \sum t^2 - \frac{(\sum t)^2}{n}$$

$$S_{wt} = 784 - \frac{(42)(119)}{6} \quad S_{tt} = 2435 - \frac{(119)^2}{6}$$

$$= \underline{\underline{-49}} \quad = \frac{449}{6}$$

$$b/ \quad S_{ww} = 50$$

$$S_{ss} = 50 \times 1000^2$$

$$= \underline{\underline{50000000}}$$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 1 continued

$$\begin{aligned}
 S_{SE} &= S_{wt} \times 1000 \\
 &= -49 \times 1000 \\
 &= \underline{\underline{-49000}}
 \end{aligned}$$

$$c) \quad r = \frac{S_{wt}}{\sqrt{S_{ww} S_{tt}}}$$

$$= \frac{-49}{\sqrt{(50) \left(\frac{449}{6}\right)}}$$

$$= -0.801 \text{ (3sf)}$$

d) Yes this value supports the manager's belief because it is a strong correlation. Close to -1.

$$e) \quad \begin{aligned} y &= a + b0c \\ w &= a + bt \end{aligned}$$

$$b = \frac{S_{wt}}{S_{tt}} \quad a = \bar{w} - b\bar{t}$$

$$b = \frac{-49}{449/6} = \frac{-294}{449} = -0.655 \text{ (3sf)}$$

$$a = \frac{42}{6} - \text{Ans} \left(\frac{119}{6} \right)$$

$$= 20.0 \text{ (3sf)}$$

$$w = 20.0 - 0.655t \quad w = \frac{s}{1000}$$

$$f) \quad \frac{s}{1000} = 20.0 - 0.655t$$

$$\underline{\underline{s = 20000 - 655t}}$$

g) A 1°C increase in average weekly temp. would decrease sales by £655



2. An estate agent is studying the cost of office space in London. He takes a random sample of 90 offices and calculates the cost, £x per square foot. His results are given in the table below.

Cost (£x)	Frequency (f)	Midpoint (£y)	<i>f y</i>
$20 \leq x < 40$	12	30	360
$40 \leq x < 45$	13 <i>25</i>	42.5	552.5
$45 \leq x < 50$	25 <i>50</i>	47.5	1187.5
$50 \leq x < 60$	32 <i>82</i>	55	1760
$60 \leq x < 80$	8 <i>90</i>	70	560
			<u>4420</u>

(You may use $\sum f y^2 = 226687.5$)

A histogram is drawn for these data and the bar representing $50 \leq x < 60$ is 2 cm wide and 8 cm high.

- (a) Calculate the width and height of the bar representing $20 \leq x < 40$ (3)
- (b) Use linear interpolation to estimate the median cost. (2)
- (c) Estimate the mean cost of office space for these data. (2)
- (d) Estimate the standard deviation for these data. (2)
- (e) Describe, giving a reason, the skewness. (1)

Rika suggests that the cost of office space in London can be modelled by a normal distribution with mean £50 and standard deviation £10

- (f) With reference to your answer to part (e), comment on Rika's suggestion. (1)
- (g) Use Rika's model to estimate the 80th percentile of the cost of office space in London. (3)

a/ 50 → 60 Freq. 32 F.d = $\frac{32}{10} = 3.2$
C. width = 10

width scale factor $\times \frac{0.2}{(\frac{2}{10})}$ height scale factor $\times \frac{2.5}{(\frac{8}{3.2})}$

DO NOT WRITE IN THIS AREA



Question 2 continued

$$20 - 40 \quad \text{class width} = \underline{\underline{20}} \quad F.d = \frac{12}{20} = \underline{\underline{0.6}}$$

$$\text{width } 20 \times 0.2 = \underline{\underline{4}} \text{ cm} \quad \text{Height } 0.6 \times 2.5 = \underline{\underline{1.5}} \text{ cm}$$

b/ 45th office.

$$45 + \frac{20}{25}(5) = \underline{\underline{49}}$$

$$\begin{aligned} \text{c/ } \frac{4420}{90} &= 49.1 \\ &= 49.1 \text{ (3sf)} \quad (\pounds 49.11) \end{aligned}$$

$$\text{d/ } \sigma^2 = \frac{\sum fy^2}{n} - \left(\frac{\sum fy}{n} \right)^2$$

$$\sigma^2 = \frac{226687.5}{90} - \left(\frac{4420}{90} \right)^2$$

$$\sigma^2 = 106.8487654$$

$$\begin{aligned} \sigma &= \sqrt{\text{Ans}} \\ &= \underline{\underline{10.3}} \text{ (3sf)} \quad (\pounds 10.34) \end{aligned}$$

e/ The mean is almost equal to the median.
There is no skew, symmetrical distribution

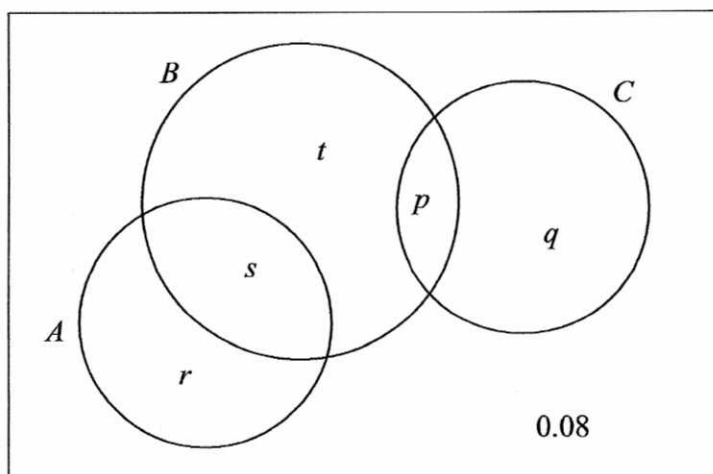
f/ Yes this could be an appropriate model
because it is a symmetrical distribution.

$$\begin{aligned} \text{g/ } p &= 0.2 \quad z = 0.8416 \quad z = \frac{x - \mu}{\sigma} \\ 0.8416 &= \frac{x - 50}{10} \end{aligned}$$

$$\begin{aligned} 8.416 &= x - 50 \\ x &= \underline{\underline{58.416}} \end{aligned}$$



3. The Venn diagram shows three events A , B and C , where p , q , r , s and t are probabilities.



$P(A) = 0.5$, $P(B) = 0.6$ and $P(C) = 0.25$ and the events B and C are independent.

- (a) Find the value of p and the value of q . (2)
- (b) Find the value of r . (2)
- (c) Hence write down the value of s and the value of t . (2)
- (d) State, giving a reason, whether or not the events A and B are independent. (2)
- (e) Find $P(B | A \cup C)$. (3)

a/ B and C are independent, so $P(B) \times P(C) = P(B \cap C)$

$$0.6 \times 0.25 = 0.15$$

$$p = \underline{\underline{0.15}}$$

$$q = 0.25 - 0.15 = \underline{\underline{0.1}}$$

b/ $r = 1 - 0.6 - 0.1 - 0.08$
 $= \underline{\underline{0.22}}$

c/ $s = 0.5 - 0.22$ $t = 0.6 - 0.15 - 0.28$
 $= \underline{\underline{0.28}}$ $= \underline{\underline{0.17}}$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 3 continued

d/ If A and B are independent $P(A) \times P(B) = P(A \cap B)$

$$0.5 \times 0.6 = 0.3 \quad s = 0.28$$

$s \neq 0.3$ They are not independent.

$$\begin{aligned} e/ P(A \cup C) &= r + s + p + q \\ &= 0.22 + 0.28 + 0.15 + 0.1 \\ &= 0.75 \end{aligned}$$

$$\begin{aligned} P(B|A \cup C) &= \frac{s + p}{0.75} = \frac{0.28 + 0.15}{0.75} \\ &= \frac{43}{75} \end{aligned}$$



4. The discrete random variable X has probability distribution

x	-1	0	1	2
$P(X=x)$	$a \frac{1}{3}$	$b \frac{1}{4}$	$b \frac{1}{4}$	$c \frac{1}{6}$

The cumulative distribution function of X is given by

x	-1	0	1	2
$F(x)$	$\frac{1}{3}$	$d \frac{7}{12}$	$\frac{5}{6}$	$e \mid$

- (a) Find the values of a, b, c, d and e . (5)
- (b) Write down the value of $P(X^2 = 1)$. (1)

a) $a = \frac{1}{3}$

$e = 1$

$c = 1 - \frac{5}{6} = \frac{1}{6}$

$\frac{1}{3} + 2b + \frac{1}{6} = 1$
 $2b = \frac{1}{2}$
 $b = \frac{1}{4}$

$d = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$

b/ $X^2 = 1$ when $x = -1$ and $x = 1$

$\frac{1}{3} + \frac{1}{4} = \frac{7}{12}$



5. Yuto works in the quality control department of a large company. The time, T minutes, it takes Yuto to analyse a sample is normally distributed with mean 18 minutes and standard deviation 5 minutes.

(a) Find the probability that Yuto takes longer than 20 minutes to analyse the next sample. (3)

The company has a large store of samples analysed by Yuto with the time taken for each analysis recorded. Serena is investigating the samples that took Yuto longer than 15 minutes to analyse.

She selects, at random, one of the samples that took Yuto longer than 15 minutes to analyse.

(b) Find the probability that this sample took Yuto more than 20 minutes to analyse. (4)

Serena can identify, in advance, the samples that Yuto can analyse in under 15 minutes and in future she will assign these to someone else.

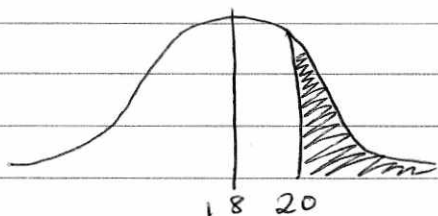
(c) Estimate the median time taken by Yuto to analyse samples in future. (5)

$$\mu = 18 \quad \sigma = 5$$

a) $P(X > 20)$

$$z = \frac{X - 18}{5}$$

$$= \frac{20 - 18}{5} = \underline{0.4}$$

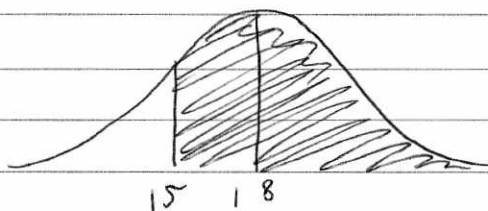


$$P(X > 20) = 1 - 0.6554 = \underline{\underline{0.3446}}$$

b) $P(X > 15)$

$$z = \frac{15 - 18}{5} = -0.6$$

$$P(X > 15) = 0.7257$$



$$\frac{0.3446}{0.7257} = 0.475 \quad (3sf)$$



Question 5 continued

$$c) \frac{0.7257}{2} = 0.36285$$

$$1 - 0.36285 = \underline{0.63715}$$

$$z = 0.35$$

$$0.35 = \frac{x - 18}{5}$$

$$1.75 = x - 18$$

$$\underline{\underline{x = 19.75}}$$



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

6. The score, X , for a biased spinner is given by the probability distribution

x	0	3	6
$P(X=x)$	$\frac{1}{12}$	$\frac{2}{3}$	$\frac{1}{4}$

Find

(a) $E(X)$ (2)

(b) $\text{Var}(X)$ (3)

A biased coin has one face labelled 2 and the other face labelled 5
The score, Y , when the coin is spun has

$$P(Y=5) = p \quad \text{and} \quad E(Y) = 3$$

(c) Form a linear equation in p and show that $p = \frac{1}{3}$ (3)

(d) Write down the probability distribution of Y . (1)

Sam plays a game with the spinner and the coin.
Each is spun once and Sam calculates his score, S , as follows

$$\begin{aligned} \text{if } X=0 \text{ then } S &= Y^2 \\ \text{if } X \neq 0 \text{ then } S &= XY \end{aligned}$$

(e) Show that $P(S=30) = \frac{1}{12}$ (2)

(f) Find the probability distribution of S . (3)

(g) Find $E(S)$. (2)

Charlotte also plays the game with the spinner and the coin.
Each is spun once and Charlotte ignores the score on the coin and just uses X^2 as her score.
Sam and Charlotte each play the game a large number of times.

(h) State, giving a reason, which of Sam and Charlotte should achieve the higher total score. (2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 6 continued

a/ $E(X) = 0\left(\frac{1}{12}\right) + 3\left(\frac{2}{3}\right) + 6\left(\frac{1}{4}\right)$
 $= \underline{\underline{3.5}}$

b/ $E(X^2) = 0^2\left(\frac{1}{12}\right) + 3^2\left(\frac{2}{3}\right) + 6^2\left(\frac{1}{4}\right)$
 $= \underline{\underline{15}}$

$Var(X) = E(X^2) - (E(X))^2$
 $= 15 - 3.5^2$
 $= \underline{\underline{\frac{11}{4}}}$

c/

y	2	5
$P(Y=y)$	q	p

$p + q = 1$ $2q + 5p = 3$
 $q = 1 - p$

$2(1 - p) + 5p = 3$
 $2 - 2p + 5p = 3$
 $2 + 3p = 3$
 $3p = 1$
 $p = \underline{\underline{\frac{1}{3}}}$

d/

y	2	5
$P(Y=y)$	$\frac{2}{3}$	$\frac{1}{3}$

e/ The only way to score 30 is 6 (on spinner) and 5 (with coin)

$\frac{1}{4} \times \frac{1}{3} = \underline{\underline{\frac{1}{12}}}$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 6 continued

f)	If	$X=0$	$Y=2$	$S=4$	$\frac{1}{12} \times \frac{2}{3}$	✓
		$X=0$	$Y=5$	$S=25$	$\frac{1}{12} \times \frac{1}{3}$	✓
		$X=3$	$Y=2$	$S=6$	$\frac{2}{3} \times \frac{2}{3}$	✓
		$X=3$	$Y=5$	$S=15$	$\frac{2}{3} \times \frac{1}{3}$	✓
		$X=6$	$Y=2$	$S=12$	$\frac{1}{4} \times \frac{2}{3}$	✓
		$X=6$	$Y=5$	$S=30$	$\frac{1}{4} \times \frac{1}{3}$	✓

S	4	6	12	15	25	30
$P(S=S)$	$\frac{1}{18}$	$\frac{4}{9}$	$\frac{2}{6}$	$\frac{2}{9}$	$\frac{1}{36}$	$\frac{1}{12}$

$$g) 4\left(\frac{1}{18}\right) + 6\left(\frac{4}{9}\right) + 12\left(\frac{2}{6}\right) + 15\left(\frac{2}{9}\right) + 25\left(\frac{1}{36}\right) + 30\left(\frac{1}{12}\right)$$

$$= \frac{137}{12} = \underline{\underline{11.4}} \text{ (3sf)}$$

w) Charlotte will have the higher score.
The expected value of X^2 is 15.

$$15 > 11.4$$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

