Centre No.	1	0	6	5	8	Paper Reference							Surname BALOGUN	Initial(s) AB
Candidate No.	5	0	2	2		6	6	8	3	/	0	1	Signature	

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 17 May 2013 – Morning

Time: 1 hour 30 minutes

ND02770193 4

Examiner's use only		

Team Leader's use only		

Question Number	Leave Blank
1	12
2	10
3	10
4	12
5	15
6	5
Total	64

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

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 or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
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 for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 6 questions in this question paper. The total mark for this paper is 75.
There are 24 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

Printer's Log. No.
P42831A
W850/R6683/57570 5/5/5/

P 4 2 8 3 1 A 0 1 2 4

Turn over

PEARSON

843694/1

Leave blank

1. A meteorologist believes that there is a relationship between the height above sea level, h m, and the air temperature, t °C. Data is collected at the same time from 9 different places on the same mountain. The data is summarised in the table below.

h	1400	1100	260	840	900	550	1230	100	770
t	3	10	20	9	10	13	5	24	16

[You may assume that $\sum h = 7150$, $\sum t = 110$, $\sum h^2 = 7\,171\,500$, $\sum t^2 = 1716$, $\sum th = 64\,980$ and $S_{tt} = 371.56$]

(a) Calculate S_{th} and S_{hh}. Give your answers to 3 significant figures. **(3)**

(b) Calculate the product moment correlation coefficient for this data. **(2)**

(c) State whether or not your value supports the use of a regression equation to predict the air temperature at different heights on this mountain. Give a reason for your answer. **(1)**

(d) Find the equation of the regression line of t on h giving your answer in the form $t = a + bh$. **(4)**

(e) Interpret the value of b. **(1)**

(f) Estimate the difference in air temperature between a height of 500 m and a height of 1000 m. **(2)**

1 Q01aM
1 Q01aA1
1 Q01aA2
1 Q01bM
1 Q01bA
1 Q01cB
1 Q01dM1
1 Q01dA1
1 Q01dM2
1 Q01dA2
0 Q01eB
1 Q01fM
1 Q01fA

a)

$S_{hh} = \Sigma h^2 - \frac{(\Sigma h)^2}{n}$

$S_{hh} = 7171500 - \frac{(7150)^2}{9}$

$S_{hh} = 1491222.222$ $\therefore S_{hh} = 1490000$ (3s.f.)

$S_{th} = \Sigma th - \frac{\Sigma t \Sigma h}{n}$

$S_{th} = 64980 - \frac{7150 \times 110}{9} = -22408.\dot{8}$

$\therefore S_{th} = -22400$ (3sf)

P 4 2 8 3 1 A 0 2 2 4

843694/2

Leave blank

Question 1 continued

b) $r = \dfrac{S_{th}}{\sqrt{S_{tt} \times S_{hh}}}$ using answers to 3 significant figures

$\therefore\ r = \dfrac{-22400}{\sqrt{1490000 \times 371.56}}$ $r = -0.9520075816$

$r \approx -0.952$ (3d.p.)

c) A strong negative correlation therefore the use of a regression line would be appropriate

~~d)~~ $\cancel{b = \dfrac{S_{th}}{S_{tt}} = \dfrac{-22400}{371.56} = -60.28636021 \approx -60.3}$

$\cancel{a = \bar{y} - b\bar{x}}$

~~in terms of h and t~~ $\cancel{a = \bar{h} - b\bar{t}}$

$\cancel{a = \dfrac{\sum h}{n} - b\dfrac{\sum t}{n} = \dfrac{750}{9} + \left(60.. \times \dfrac{110}{9}\right)}$

$\cancel{a = 1531.277726}$

$\cancel{a = 1531\ (4s.f.)}$

$\cancel{h = 1531 - 60.3t}$

~~e) b is the gradient of the regression line~~

d) $t = a + bh$

$b = \dfrac{S_{th}}{S_{hh}} = \dfrac{-22400}{1490000} = -0.01503355705 \approx -0.015$

Leave blank

Question 1 continued

~~$\bar{h}$~~ $\bar{h} = \cancel{6250}\ \frac{7150}{9}$ $\quad \bar{t} = \frac{110}{9}$

$a = \bar{t} - b\bar{h}$

$a = \frac{110}{9} - \cancel{b}(\cancel{0.015}{-0.015\ldots} \times \frac{7150}{9})$

$a = 24.1655481$

$a \approx 24.2$ (3s.f.)

$\therefore \quad t = 24.2 - 0.015h$

e) ~~The~~ b is the gradient of the regression line

f) $t = 24.16 - 0.015(500)$

$t = 31.68232663$

$t \approx 31.7^{\circ}$C (3s.f.) at 500m

$t = 24.16\ldots - 0.015\ldots(1000)$

$t = 39.19910515$

$t \approx 39.2^{\circ}$C (3s.f.)

Difference in air temp in 500m above sea and 1000m above sea $= 7.5^{\circ}$C (1d.p.)

Leave blank

Question 1 continued

(Total 13 marks)

Q1

12

Turn over

P42831A0524

2. The marks of a group of female students in a statistics test are summarised in Figure 1

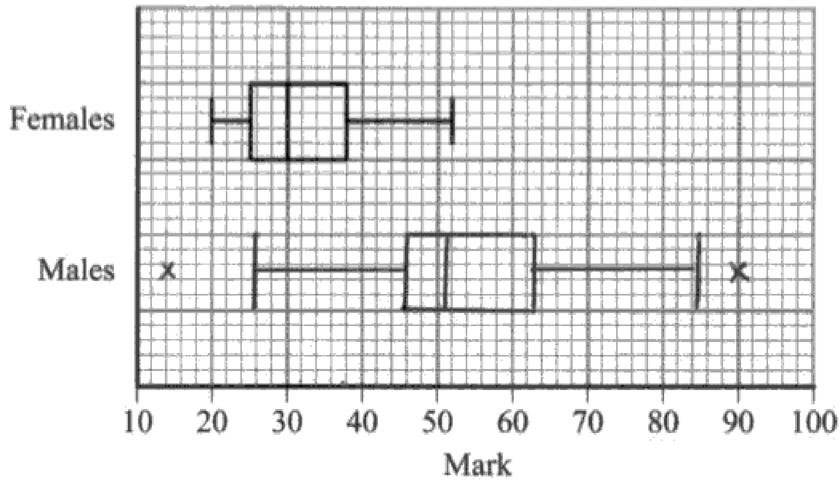


Figure 1

(a) Write down the mark which is exceeded by 75% of the female students.

(1)

The marks of a group of male students in the same statistics test are summarised by the stem and leaf diagram below.

Mark	(2|6 means 26)	Totals
1	4	(1)
2	6	(1)
3	4 4 7	(3)
4	0 6 6 7 7 8	(6)
5	0 0 1 1 1 3 6 7 7	(9)
6	2 2 3 3 3 8	(6)
7	0 0 8	(3)
8	5	(1)
9	0	(1)

(b) Find the median and interquartile range of the marks of the male students.

(3)

An outlier is a mark that is

either more than 1.5 × interquartile range above the upper quartile

or more than 1.5 × interquartile range below the lower quartile.

Leave blank

0 Q02aB

1 Q02bB
1 Q02bM
1 Q02bA
1 Q02cM1
1 Q02cA1
1 Q02cM2
1 Q02cA2
1 Q02cB
1 Q02dB1
1 Q02dB2

P 4 2 8 3 1 A 0 6 2 4

843694/6

Leave blank

(c) In the space provided on Figure 1 draw a box plot to represent the marks of the male students, indicating clearly any outliers.

(5)

(d) Compare and contrast the marks of the male and the female students.

(2)

a) 38 marks

b) Q_2 = median

31 male students

$\therefore Q_2 = \frac{31}{2} = 15.5 \Rightarrow 16^{th}$ observation

$Q_2 = 51$

$Q_1 = \frac{31}{4} = 7.75 \Rightarrow 8^{th}$ observation

$Q_1 = 46$

$Q_3 = 3 \times \frac{31}{4} = 23.25 \Rightarrow 24^{th}$ observation

$Q_3 = 63$

$\therefore IQR = (Q_3 - Q_1) = (63 - 46) = 17$

c) $Q_3 + 1.5(IQR)$

$63 + 1.5(17) = 88.5$

$\therefore$ 90 is an outlier

$Q_1 - 1.5(IQR)$

$46 - 1.5(17) = 20.5$

$\therefore$ 14 is an outlier

Leave blank

Question 2 continued

d) - Males' and females' box plots are positively skewed
- Males' ~~box~~ boxplot has a greater median
- Males' box plot has a wider range
- IQR in males boxplot > IQR in females boxplot

Leave blank

Question 2 continued

(Total 11 marks)

Q2

10

P 4 2 8 3 1 A 0 9 2 4

Turn over

843694/9

Leave blank

3. In a company the 200 employees are classified as full-time workers, part-time workers or contractors.
The table below shows the number of employees in each category and whether they walk to work or use some form of transport.

	Walk	Transport
Full-time worker	2	8
Part-time worker	35	75
Contractor	30	50

The events F, H and C are that an employee is a full-time worker, part-time worker or contractor respectively. Let W be the event that an employee walks to work.

An employee is selected at random.

Find

(a) $P(H)$ **(2)** 1 Q03aM 1 Q03aA

(b) $P([F \cap W]')$ **(2)** 0 Q03bM 0 Q03bA

(c) $P(W \mid C)$ **(2)** 1 Q03cM 1 Q03cA

Let B be the event that an employee uses the bus.
Given that 10% of full-time workers use the bus, 30% of part-time workers use the bus and 20% of contractors use the bus,

(d) draw a Venn diagram to represent the events F, H, C and B, **(4)** 1 Q03dM 1 Q03dB1 1 Q03dB2 1 Q03dB3

(e) find the probability that a randomly selected employee uses the bus to travel to work. **(2)** 1 Q03eM 1 Q03eA

a) $P(H) = \frac{35+75}{200} = \frac{11}{20}$

$P(H) = 0.55$

b) $P([F \cap W]') =$ not full time worker and does not walk to work

$P([F \cap W]') = \frac{35+50}{200} = \frac{125}{200} = 0.625$

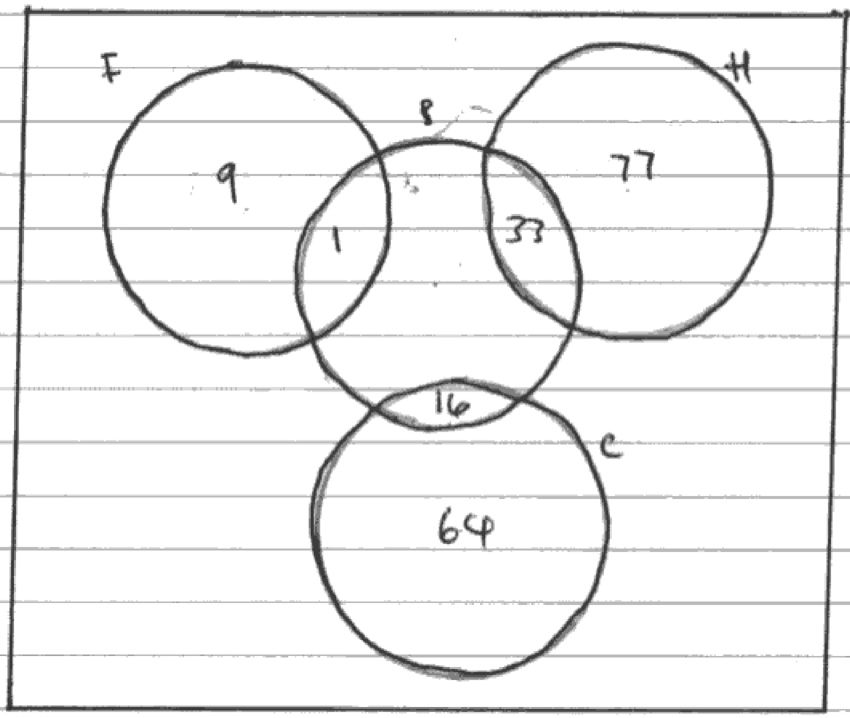
Question 3 continued

c) $P(W|C) = \dfrac{P(W \cap C)}{P(C)}$

$$= \frac{\left(\frac{30}{200}\right)}{\left(\frac{80}{200}\right)}$$

$P(W|C) = \frac{3}{8} = 0.375$

d) 10% full time = 1 person takes bus
30% part-time = 33 people take bus
20% contractors = 16 people take bus



e) $\dfrac{33+16+1}{200} = \dfrac{50}{200}$

probability that an employee takes the bus is $= \frac{1}{4}$.

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Q3

10

(Total 12 marks)

P 4 2 8 3 1 A 0 1 3 2 4

Turn over

843694/13

Leave blank

4. The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

[You may use $\sum ft^2 = 134281.25$]

(a) Estimate the mean and standard deviation of these data. **(5)**

(b) Use linear interpolation to estimate the value of the median. **(2)**

(c) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures. **(1)**

(d) Estimate the interquartile range of this distribution. **(2)**

(e) Give a reason why the mean and standard deviation are not the most appropriate summary statistics to use with these data. **(1)**

The person timing the exam made an error and each student actually took 5 minutes less than the times recorded above. The table below summarises the actual times.

Time (minutes) t	6 – 15	16 – 20	21 – 25	26 – 30	31 – 40	41 – 55
Number of students f	62	88	16	13	11	10

(f) Without further calculations, explain the effect this would have on each of the estimates found in parts (a), (b), (c) and (d). **(3)**

1 Q04aB
1 Q04aM1
1 Q04aA1
1 Q04aM2
1 Q04aA2
1 Q04bM
1 Q04bA
1 Q04cB
1 Q04dB1
1 Q04dB2
1 Q04eB
1 Q04fB1
0 Q04fB2
0 Q04fB3

a)

Time (minutes) t	f	midpoint ~~x~~ t	ft
11 – 20	62	15.5	961
21 – 25	88	23	2024
26 – 30	16	28	448
31 – 35	13	33	429
36 – 45	11	40.5	445.5
46 – 60	10	53	530

$\sum ft = 4837.5$

P 4 2 8 3 1 A 0 1 4 2 4

843694/14

Leave blank

Question 4 continued

$$\mu = \frac{\sum ft}{\sum f} = \frac{4837.5}{200} = 24.1875$$

$$\sigma = \sqrt{\frac{\sum ft^2}{\sum f} - \mu^2}$$

$\sigma = 9.293604992$

$\sigma = 9.29$ (3s.f.)

b) $Q_2 = \frac{1}{2}n$

$\therefore$ Q_2 is in the modal class 16–20

$$Q_2 = \cancel{15.5}\ 20.5 + \left(\frac{100-62}{88}\right)\times 5$$

$$Q_2 = \cancel{15.5}\ 20.5 + \frac{95}{44}$$

$$Q_2 = \cancel{\tfrac{997}{44}}\ 22.65909091$$

$Q_2 \approx$ ~~17.7 (3s.f.)~~ 22.7 (3s.f.)

c) $Q_1 = \frac{200}{4}$, $\therefore$ in modal class 6–15

$$Q_1 = \cancel{5.5}\ 10.5 + \left(\frac{50-}{62}\right)\times 10$$

$$Q_1 = 10.5 + \frac{500}{62}$$

$Q_1 = 18.56451613$

$\therefore$ $Q_1 \approx 18.6$

Leave blank

Question 4 continued

d) $Q_3 = \frac{3}{4}(200) = 150$

$Q_3 =$ in modal class 21–25

$Q_3 = 20.5 + \left(\frac{150-62}{88}\right) \times 5$

$Q_3 = 25.5$

$IQR = \cancel{25.5}\ (Q_3 - Q_1) = 6.9$

e) the data is skewed

~~negative~~ skew

$Q_3 - Q_2 \cancel{<} Q_2 - Q_1$

$2.8 < 4.1$

f)

a- mean would change because it is affected by extreme values

b- ~~mode~~ would be the same

c- would remain the same.

d- would be the same

Leave blank

Question 4 continued

Q4

(Total 14 marks)

12

Leave blank

5. A biased die with six faces is rolled. The discrete random variable X represents the score on the uppermost face. The probability distribution of X is shown in the table below.

x	1	2	3	4	5	6
$P(X = x)$	a	a	a	b	b	0.3

(a) Given that $E(X) = 4.2$ find the value of a and the value of b. **(5)**

(b) Show that $E(X^2) = 20.4$ **(1)**

(c) Find $Var(5 - 3X)$ **(3)**

A biased die with five faces is rolled. The discrete random variable Y represents the score which is uppermost. The cumulative distribution function of Y is shown in the table below.

y	1	2	3	4	5
$F(y)$	$\frac{1}{10}$	$\frac{2}{10}$	$3k$	$4k$	$5k$

(d) Find the value of k. **(1)**

(e) Find the probability distribution of Y. **(3)**

Each die is rolled once. The scores on the two dice are independent.

(f) Find the probability that the sum of the two scores equals 2 **(2)**

1 Q05aM1
1 Q05aM2
1 Q05aM3
1 Q05aB1
1 Q05aB2
1 Q05bB
1 Q05cM1
1 Q05cM2
1 Q05cA
1 Q05dB
1 Q05eB
1 Q05eM
1 Q05eA
1 Q05fM
1 Q05fA

a) $E(X) = 4.2$

$4.2 = a + 2a + 3a + 4b + 5b + 1.8$
$2.4 = 6a + 9b$

$3a + 2b + 0.3 = 1$
$3a + 2b = 0.7$

P 4 2 8 3 1 A 0 1 8 2 4

843694/18

Question 5 continued

$6a + 9b = 2.4$

$6a + 4b = 1.4$

$5b = 1$

$b = \frac{1}{5} = 0.2$

$\therefore a = 0.1$

b)

x	1	2	3	4	5	6
$P(X=x)$	0.1	0.1	0.1	0.2	0.2	0.3

$(1^2 \times 0.1) + (2^2 \times 0.1) + (3^2 \times 0.1) + (4^2 \times 0.2) + (5^2 \times 0.2) + (6^2 \times 0.3)$

$0.1 + 0.4 + 0.9 + \cancel{3}3.2 + 5 + 10.8$

$= 20.4$

c) $20.4 - (4.2)^2 = Var(X)$

$Var(X) = \cancel{2.76}\ 2.76$

$Var(5 - 3X) = 3^2 \times 2.76 = 24.84$

d) At 5 $F(y) = 1$

$\cancel{5}$

$\therefore 5k = 1$

$k = \frac{1}{5}$

e)

y	1	2	3	4	5
$P(Y=y)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{4}{10}$	$\frac{2}{10}$	$\frac{2}{10}$

f) $\frac{1}{10} \times \frac{1}{10} = 0.01$

Leave blank

Question 5 continued

Leave blank

Question 5 continued

(Total 15 marks)

Q5 15

Leave blank

6. The weight, in grams, of beans in a tin is normally distributed with mean μ and standard deviation 7.8

Given that 10% of tins contain less than 200 g, find

(a) the value of μ **(3)**

(b) the percentage of tins that contain more than 225 g of beans. **(3)**

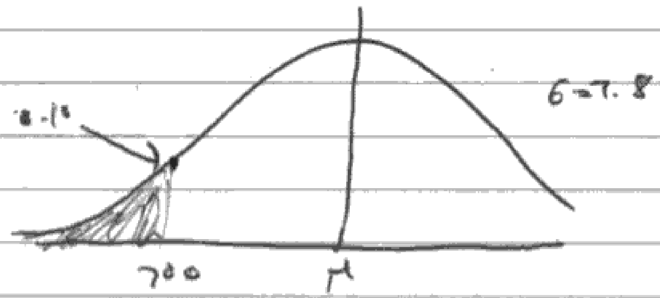
The machine settings are adjusted so that the weight, in grams, of beans in a tin is normally distributed with mean 205 and standard deviation σ.

(c) Given that 98% of tins contain between 200 g and 210 g find the value of σ. **(4)**

1 Q06aM
1 Q06aB
1 Q06aA
0 Q06bM
0 Q06bA1
0 Q06bA2
1 Q06cM1
0 Q06cB
1 Q06cM2
0 Q06cA

6 $W \sim (\mu, 7.8^2)$

a)



$Z = \dfrac{200-\mu}{7.8}$

$P(W<200) = P\left(Z < \dfrac{200-\mu}{7.8}\right)$

$= 0.10$

$\dfrac{200-\mu}{7.8} = -1.2816$

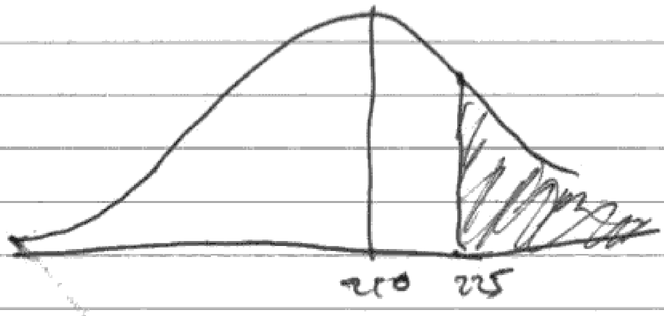
$200-\mu = 9.99648$

$\mu = 209.99648$

$\mu = 210$ (3s.f.)

Question 6 continued

b)

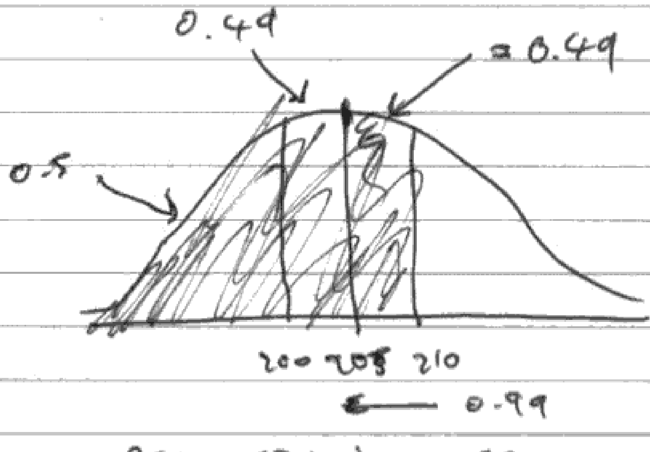


$z = \frac{225-210}{7.68} = 1.953125$

$z = 1.95$

$P(Z > 1.95) = 1 - \Phi(1.95)$
$= 1 - 0.9744$
$= 0.0256$

c) ~~W6~~ $W \sim N(205, 6)$



$P(W < 210) = 0.99$

$z = \frac{210-205}{6}$

$z = \frac{5}{6}$

$\frac{5}{6} = 2.32$

$5 = 2.326$

$6 = 0.464$

Turn over

P42831A02324

Leave blank

Question 6 continued

Q6

5

(Total 10 marks)

TOTAL FOR PAPER: 75 MARKS

END