

Paper Reference(s)

**6683****Edexcel GCE****Statistics S1****(New Syllabus)****Advanced/Advanced Subsidiary****Friday 19 January 2001 – Afternoon****Time: 1 hour 30 minutes****Materials required for examination**Answer Book (AB16)  
Graph Paper (GP02)  
Mathematical Formulae**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has 6 questions. Pages 6, 7 and 8 are blank.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. The students in a class were each asked to write down how many CDs they owned. The student with the least number of CDs had 14 and all but one of the others owned 60 or fewer. The remaining student owned 65. The quartiles for the class were 30, 34 and 42 respectively.

Outliers are defined to be any values outside the limits of  $1.5(Q_3 - Q_1)$  below the lower quartile or above the upper quartile.

On graph paper draw a box plot to represent these data, indicating clearly any outliers. **(7 marks)**

---

2. The random variable  $X$  is normally distributed with mean 177.0 and standard deviation 6.4.

(a) Find  $P(166 < X < 185)$ . **(4 marks)**

It is suggested that  $X$  might be a suitable random variable to model the height, in cm, of adult males.

(b) Give two reasons why this is a sensible suggestion. **(2 marks)**

(c) Explain briefly why mathematical models can help to improve our understanding of real-world problems. **(2 marks)**

---

3. A fair six-sided die is rolled. The random variable  $Y$  represents the score on the uppermost, face.

(a) Write down the probability function of  $Y$ . **(2 marks)**

(b) State the name of the distribution of  $Y$ . **(1 mark)**

Find the value of

(c)  $E(6Y + 2)$ , **(4 marks)**

(d)  $\text{Var}(4Y - 2)$ . **(5 marks)**

---

4. The employees of a company are classified as management, administration or production. The following table shows the number employed in each category and whether or not they live close to the company or some distance away.

	Live close	Live some distance away
Management	6	14
Administration	25	10
Production	45	25

An employee is chosen at random.

Find the probability that this employee

- (a) is an administrator, **(2 marks)**  
 (b) lives close to the company, given that the employee is a manager. **(2 marks)**

Of the managers, 90% are married, as are 60% of the administrators and 80% of the production employees.

- (c) Construct a tree diagram containing all the probabilities. **(3 marks)**  
 (d) Find the probability that an employee chosen at random is married. **(3 marks)**

An employee is selected at random and found to be married.

- (e) Find the probability that this employee is in production. **(3 marks)**
- 

5. The following grouped frequency distribution summarises the number of minutes, to the nearest minute, that a random sample of 200 motorists were delayed by roadworks on a stretch of motorway.

Delay (mins)	Number of motorists
4—6	15
7—8	28
9	49
10	53
11—12	30
13—15	15
16—20	10

- (a) Using graph paper represent these data by a histogram. **(4 marks)**  
 (b) Give a reason to justify the use of a histogram to represent these data. **(1 mark)**  
 (c) Use interpolation to estimate the median of this distribution. **(2 marks)**  
 (d) Calculate an estimate of the mean and an estimate of the standard deviation of these data. **(6 marks)**

One coefficient of skewness is given by

$$\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

- (e) Evaluate this coefficient for the above data. **(2 marks)**  
 (f) Explain why the normal distribution may not be suitable to model the number of minutes that motorists are delayed by these roadworks. **(2 marks)**
-



1. Each of the 25 students on a computer course recorded the number of minutes  $x$ , to the nearest minute, spent surfing the internet during a given day. The results are summarised below.

$$\Sigma x = 1075, \Sigma x^2 = 46625.$$

(a) Find  $\mu$  and  $\sigma$  for these data. (3)

Two other students surfed the internet on the same day for 35 and 51 minutes respectively.

(b) Without further calculation, explain the effect on the mean of including these two students. (2)

---

2. On a particular day in summer 1993 at 0800 hours the height above sea level,  $x$  metres, and the temperature,  $y$  °C, were recorded in 10 Mediterranean towns. The following summary statistics were calculated from the results.

$$\Sigma x = 7300, \Sigma x^2 = 6599600, S_{xy} = -13060, S_{yy} = 140.9.$$

(a) Find  $S_{xx}$ . (2)

(b) Calculate, to 3 significant figures, the product moment correlation coefficient between  $x$  and  $y$ . (2)

(c) Give an interpretation of your coefficient. (1)

---

3. The continuous random variable  $Y$  is normally distributed with mean 100 and variance 256.

(a) Find  $P(Y < 80)$ . (3)

(b) Find  $k$  such that  $P(100 - k \leq Y \leq 100 + k) = 0.516$ . (5)

---

4. The discrete random variable  $X$  has the probability function shown in the table below.

$x$	-2	-1	0	1	2	3
$P(X = x)$	0.1	$\alpha$	0.3	0.2	0.1	0.1

Find

(a)  $\alpha$ . (1)

(b)  $P(-1 < X \leq 2)$ , (2)

(c)  $F(-0.4)$ , (1)

(d)  $E(3X + 4)$ , (4)

(e)  $\text{Var}(2X + 3)$ . (4)

---

5. A market researcher asked 100 adults which of the three newspapers  $A, B, C$  they read. The results showed that 30 read  $A$ , 26 read  $B$ , 21 read  $C$ , 5 read both  $A$  and  $B$ , 7 read both  $B$  and  $C$ , 6 read both  $C$  and  $A$  and 2 read all three.

(a) Draw a Venn diagram to represent these data. (6)

One of the adults is then selected at random.

Find the probability that she reads

(b) at least one of the newspapers, (2)

(c) only  $A$ , (1)

(d) only one of the newspapers, (2)

(e)  $A$  given that she reads only one newspaper. (2)

---

6. Three swimmers Alan, Diane and Gopal record the number of lengths of the swimming pool they swim during each practice session over several weeks. The stem and leaf diagram below shows the results for Alan.

Lengths	2   0 means 20	
2	0 1 2 2	(4)
2	5 5 6 7 7 8 9	(7)
3	0 1 2 2 4	(5)
3	5 6 6 7 9	(5)
4	0 1 3 3 3 3 4 4 4	(10)
4	5 5 6 6 6 7 7 8 8 9 9 9	(12)
5	0 0 0	(3)

(a) Find the three quartiles for Alan's results.

(4)

The table below summarises the results for Diane and Gopal.

	Diane	Gopal
Smallest value	35	25
Lower quartile	37	34
Median	42	42
Upper quartile	53	50
Largest value	65	57

(b) Using the same scale and on the same sheet of graph paper draw box plots to represent the data for Alan, Diane and Gopal.

(8)

(c) Compare and contrast the three box plots.

(4)

7. A music teacher monitored the sight-reading ability of one of her pupils over a 10 week period. At the end of each week, the pupil was given a new piece to sight-read and the teacher noted the number of errors  $y$ . She also recorded the number of hours  $x$  that the pupil had practised each week. The data are shown in the table below.

$x$	12	15	7	11	1	8	4	6	9	3
$y$	8	4	13	8	18	12	15	14	12	16

(a) Plot these data on a scatter diagram.

(3)

(b) Find the equation of the regression line of  $y$  on  $x$  in the form  $y = a + bx$ .

(You may use  $\Sigma x^2 = 746$ ,  $\Sigma xy = 749$ .)

(9)

(c) Give an interpretation of the slope and the intercept of your regression line.

(2)

(d) State whether or not you think the regression model is reasonable

(i) for the range of  $x$ -values given in the table,

(ii) for all possible  $x$ -values.

In each case justify your answer either by giving a reason for accepting the model or by suggesting an alternative model.

(2)

**END**

Paper Reference(s)

**6683****Edexcel GCE****Statistics S1****(New Syllabus)****Advanced/Advanced Subsidiary****Wednesday 16 January 2002 – Afternoon****Time: 1 hour 30 minutes****Materials required for examination**

Answer Book (AB16)

Graph Paper (ASG2)

Mathematical Formulae (Lilac)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has seven questions. Pages 6, 7 and 8 are blank.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. (a) Explain briefly what you understand by
- (i) a statistical experiment, (1)
- (ii) an event. (1)
- (b) State one advantage and one disadvantage of a statistical model. (2)
- 

2. A meteorologist measured the number of hours of sunshine, to the nearest hour, each day for 100 days. The results are summarised in the table below.

Hours of sunshine	Days
1	16
2–4	32
5–6	28
7	12
8	9
9–11	2
12	1

- (a) On graph paper, draw a histogram to represent these data. (5)
- (b) Calculate an estimate of the number of days that had between 6 and 9 hours of sunshine. (2)
- 

3. A discrete random variable  $X$  has the probability function shown in the table below.

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

- (a) Given that  $E(X) = \frac{5}{6}$ , find  $a$ . (3)
- (b) Find the exact value of  $\text{Var}(X)$ . (3)
- (c) Find the exact value of  $P(X \leq 15)$ . (1)
-

4. A contractor bids for two building projects. He estimates that the probability of winning the first project is 0.5, the probability of winning the second is 0.3 and the probability of winning both projects is 0.2.

- (a) Find the probability that he does not win either project. (3)
- (b) Find the probability that he wins exactly one project. (2)
- (c) Given that he does not win the first project, find the probability that he wins the second. (2)
- (d) By calculation, determine whether or not winning the first contract and winning the second contract are independent events. (3)
- 

5. The duration of the pregnancy of a certain breed of cow is normally distributed with mean  $\mu$  days and standard deviation  $\sigma$  days. Only 2.5% of all pregnancies are shorter than 235 days and 15% are longer than 286 days.

- (a) Show that  $\mu - 235 = 1.96\sigma$ . (2)
- (b) Obtain a second equation in  $\mu$  and  $\sigma$ . (3)
- (c) Find the value of  $\mu$  and the value of  $\sigma$ . (4)
- (d) Find the values between which the middle 68.3% of pregnancies lie. (2)
- 

6. Hospital records show the number of babies born in a year. The number of babies delivered by 15 male doctors is summarised by the stem and leaf diagram below.

Babies	(4   5 means 45)	Totals
0		(0)
1	9	(1)
2	1 6 7 7	(4)
3	2 2 3 4 8	(5)
4	5	(1)
5	1	(1)
6	0	(1)
7		(0)
8	6 7	(2)

- (a) Find the median and inter-quartile range of these data. (3)
- (b) Given that there are no outliers, draw a box plot on graph paper to represent these data. Start your scale at the origin. (4)
- (c) Calculate the mean and standard deviation of these data. (5)

The records also contain the number of babies delivered by 10 female doctors.

34	30	20	15	6
32	26	19	11	4

The quartiles are 11, 19.5 and 30.

- (d) Using the same scale as in part (b) and on the same graph paper draw a box plot for the data for the 10 female doctors. (3)
- (e) Compare and contrast the box plots for the data for male and female doctors. (2)
-

7. A number of people were asked to guess the calorific content of 10 foods. The mean  $s$  of the guesses for each food and the true calorific content  $t$  are given in the table below.

END

Food	$t$	$s$
Packet of biscuits	170	420
1 potato	90	160
1 apple	80	110
Crisp breads	10	70
Chocolate bar	260	360
1 slice white bread	75	135
1 slice brown bread	60	115
Portion of beef curry	270	350
Portion of rice pudding	165	390
Half a pint of milk	160	200

[You may assume that  $\Sigma t = 1340$ ,  $\Sigma s = 2310$ ,  $\Sigma ts = 396\,775$ ,  $\Sigma t^2 = 246\,050$ ,  $\Sigma s^2 = 694\,650$ .]

(a) Draw a scatter diagram, indicating clearly which is the explanatory (independent) and which is the response (dependent) variable.

(3)

(b) Calculate, to 3 significant figures, the product moment correlation coefficient for the above data.

(7)

(c) State, with a reason, whether or not the value of the product moment correlation coefficient changes if all the guesses are 50 calories higher than the values in the table.

(2)

The mean of the guesses for the portion of rice pudding and for the packet of biscuits are outside the linear relation of the other eight foods.

(d) Find the equation of the regression line of  $s$  on  $t$  excluding the values for rice pudding and biscuits.

(3)

[You may now assume that  $S_{ts} = 72\,587$ ,  $S_{tt} = 63\,671.875$ ,  $\bar{t} = 125.625$ ,  $\bar{s} = 187.5$ .]

(e) Draw the regression line on your scatter diagram.

(2)

(f) State, with a reason, what the effect would be on the regression line of including the values for a portion of rice pudding and a packet of biscuits.

(2)

Paper Reference(s)

**6683****Edexcel GCE****Statistics S1****Advanced/Advanced Subsidiary****Tuesday 28 May 2002 – Afternoon****Time: 1 hour 30 minutes****Materials required for examination**

Answer Book (AB16)

Graph Paper (ASG2)

Mathematical Formulae (Lilac)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has seven questions. Pages 6, 7 and 8 are blank.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. An unbiased die has faces numbered 1 to 6 inclusive. The die is rolled and the number that appears on the uppermost face is recorded.

(a) State the probability of not recording a 6 in one roll of the die.

(1)

The die is thrown until a 6 is recorded.

(b) Find the probability that a 6 occurs for the first time on the third roll of the die.

(3)

2. Statistical models can be used to describe real world problems. Explain the process involved in the formulation of a statistical model.

(4)

3. For the events  $A$  and  $B$ ,

(a) explain in words the meaning of the term  $P(B|A)$ ,

(2)

(b) sketch a Venn diagram to illustrate the relationship  $P(B|A) = 0$ .

(2)

Three companies operate a bus service along a busy main road. Amber buses run 50% of the service and 2% of their buses are more than 5 minutes late. Blunder buses run 30% of the service and 10% of their buses are more than 5 minutes late. Clipper buses run the remainder of the service and only 1% of their buses run more than 5 minutes late.

Jean is waiting for a bus on the main road.

(c) Find the probability that the first bus to arrive is an Amber bus that is more than 5 minutes late.

(2)

Let  $A$ ,  $B$  and  $C$  denote the events that Jean catches an Amber bus, a Blunder bus and a Clipper bus respectively. Let  $L$  denote the event that Jean catches a bus that is more than 5 minutes late.

(d) Draw a Venn diagram to represent the events  $A$ ,  $B$ ,  $C$  and  $L$ . Calculate the probabilities associated with each region and write them in the appropriate places on the Venn diagram.

(4)

(e) Find the probability that Jean catches a bus that is more than 5 minutes late.

(2)

4. A discrete random variable  $X$  takes only positive integer values. It has a cumulative distribution function  $F(x) = P(X \leq x)$  defined in the table below.

$X$	1	2	3	4	5	6	7	8
$F(x)$	0.1	0.2	0.25	0.4	0.5	0.6	0.75	1

- (a) Determine the probability function,  $P(X = x)$ , of  $X$ . (3)
- (b) Calculate  $E(X)$  and show that  $\text{Var}(X) = 5.76$ . (6)
- (c) Given that  $Y = 2X + 3$ , find the mean and variance of  $Y$ . (3)
- 

5. A random variable  $X$  has a normal distribution.

- (a) Describe two features of the distribution of  $X$ . (2)

A company produces electronic components which have life spans that are normally distributed. Only 1% of the components have a life span less than 3500 hours and 2.5% have a life span greater than 5500 hours.

- (b) Determine the mean and standard deviation of the life spans of the components. (6)

The company gives warranty of 4000 hours on the components.

- (c) Find the proportion of components that the company can expect to replace under the warranty. (4)
- 

6. The labelling on bags of garden compost indicates that the bags weigh 20 kg. The weights of a random sample of 50 bags are summarised in the table below.

Weight in kg	Frequency
14.6 – 14.8	1
14.8 – 18.0	0
18.0 – 18.5	5
18.5 – 20.0	6
20.0 – 20.2	22
20.2 – 20.4	15
20.4 – 21.0	1

- (a) On graph paper, draw a histogram of these data. (4)

- (b) Using the coding  $y = 10(\text{weight in kg} - 14)$ , find an estimate for the mean and standard deviation of the weight of a bag of compost. (6)

$$[\text{Use } \Sigma fy^2 = 171\,503.75]$$

- (c) Using linear interpolation, estimate the median. (2)

The company that produces the bags of compost wants to improve the accuracy of the labelling. The company decides to put the average weight in kg on each bag.

- (d) Write down which of these averages you would recommend the company to use. Give a reason for your answer. (2)
-

7. An ice cream seller believes that there is a relationship between the temperature on a summer day and the number of ice creams sold. Over a period of 10 days he records the temperature at 1 p.m.,  $t$  °C, and the number of ice creams sold,  $c$ , in the next hour. The data he collects is summarised in the table below.

$t$	$c$
13	24
22	55
17	35
20	45
10	20
15	30
19	39
12	19
18	36
23	54

[Use  $\Sigma t^2 = 3\,025$ ,  $\Sigma c^2 = 14\,245$ ,  $\Sigma ct = 6\,526$ .]

- (a) Calculate the value of the product moment correlation coefficient between  $t$  and  $c$ .  
(7)
- (b) State whether or not your value supports the use of a regression equation to predict the number of ice creams sold. Give a reason for your answer.  
(2)
- (c) Find the equation of the least squares regression line of  $c$  on  $t$  in the form  $c = a + bt$ .  
(2)
- (d) Interpret the value of  $b$ .  
(1)
- (e) Estimate the number of ice creams sold between 1 p.m. and 2 p.m. when the temperature at 1 p.m. is 16 °C.  
(3)
- (f) At 1 p.m. on a particular day, the highest temperature for 50 years was recorded. Give a reason why you should not use the regression equation to predict ice cream sales on that day.  
(1)

**END**

Paper Reference(s)

**6683**

## Edexcel GCE

### Statistics S1

#### Advanced/Advanced Subsidiary

Tuesday 5 November 2002 – Morning

Time: 1 hour 30 minutes

**Materials required for examination**

Answer Book (AB16)  
Graph Paper (ASG2)  
Mathematical Formulae (Lilac)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has seven questions. Pages 6, 7 and 8 are blank.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. (a) Explain briefly why statistical models are used when attempting to solve real-world problems. (2)
- (b) Write down the name of the distribution you would recommend as a suitable model for each of the following situations.
- (i) The weight of marmalade in a jar.
- (ii) The number on the uppermost face of a fair die after it has been rolled. (2)
- 
2. There are 125 sixth-form students in a college, of whom 60 are studying only arts subjects, 40 only science subjects and the rest a mixture of both.
- Three students are selected at random, *without replacement*.
- Find the probability that
- (a) all three students are studying only arts subjects, (4)
- (b) exactly one of the three students is studying only science subjects. (3)
- 
3. The events  $A$  and  $B$  are independent such that  $P(A) = 0.25$  and  $P(B) = 0.30$ .
- Find
- (a)  $P(A \cap B)$ , (2)
- (b)  $P(A \cup B)$ , (2)
- (c)  $P(A | B')$ . (4)
- 

4. Strips of metal are cut to length  $L$  cm, where  $L \sim N(\mu, 0.5^2)$ .
- (a) Given that 2.5% of the cut lengths exceed 50.98 cm, show that  $\mu = 50$ . (5)
- (b) Find  $P(49.25 < L < 50.75)$ . (4)
- Those strips with length either less than 49.25 cm or greater than 50.75 cm cannot be used.
- Two strips of metal are selected at random.
- (c) Find the probability that both strips cannot be used. (2)
- 
5. An agricultural researcher collected data, in appropriate units, on the annual rainfall  $x$  and the annual yield of wheat  $y$  at 8 randomly selected places.
- The data were coded using  $s = x - 6$  and  $t = y - 20$  and the following summations were obtained.
- $$\Sigma s = 48.5, \quad \Sigma t = 65.0, \quad \Sigma s^2 = 402.11, \quad \Sigma t^2 = 701.80, \quad \Sigma st = 523.23$$
- (a) Find the equation of the regression line of  $t$  on  $s$  in the form  $t = p + qs$ . (7)
- (b) Find the equation of the regression line of  $y$  on  $x$  in the form  $y = a + bx$ , giving  $a$  and  $b$  to 3 decimal places. (3)
- The value of the product moment correlation coefficient between  $s$  and  $t$  is 0.943, to 3 decimal places.
- (c) Write down the value of the product moment correlation coefficient between  $x$  and  $y$ . Give a justification for your answer. (2)
-

6. The discrete random variable  $X$  has the following probability distribution.

$x$	-2	-1	0	1	2
$P(X = x)$	$\alpha$	0.2	0.1	0.2	$\beta$

- (a) Given that  $E(X) = -0.2$ , find the value of  $\alpha$  and the value of  $\beta$ . (6)
  - (b) Write down  $F(0.8)$ . (1)
  - (a) Evaluate  $\text{Var}(X)$ . (4)
- Find the value of
- (d)  $E(3X - 2)$ , (2)
  - (e)  $\text{Var}(2X + 6)$ . (2)
- 

7. The following stem and leaf diagram shows the aptitude scores  $x$  obtained by all the applicants for a particular job.

Aptitude score	3   1 means 31
3   1 2 9	(3)
4   2 4 6 8 9	(5)
5   1 3 3 5 6 7 9	(7)
6   0 1 3 3 3 5 6 8 8 9	(10)
7   1 2 2 2 4 5 5 5 6 8 8 8 9	(14)
8   0 1 2 3 5 8 8 9	(8)
9   0 1 2	(3)

- (a) Write down the modal aptitude score. (1)
  - (b) Find the three quartiles for these data. (3)
- Outliers can be defined to be outside the limits  $Q_1 - 1.0(Q_3 - Q_1)$  and  $Q_3 + 1.0(Q_3 - Q_1)$ .
- (c) On a graph paper, draw a box plot to represent these data. (7)
- For these data,  $\sum x = 3363$  and  $\sum x^2 = 238305$ .
- (d) Calculate, to 2 decimal places, the mean and the standard deviation for these data. (3)
  - (e) Use two different methods to show that these data are negatively skewed. (4)

---

END

Paper Reference(s)

**6683****Edexcel GCE****Statistics S1****Advanced/Advanced Subsidiary****Wednesday 15 January 2003 – Morning****Time: 1 hour 30 minutes****Materials required for examination**

Answer Book (AB16)

Graph Paper (ASG2)

Mathematical Formulae (Lilac)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has seven questions. Pages 6, 7 and 8 are blank.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. The total amount of time a secretary spent on the telephone in a working day was recorded to the nearest minute. The data collected over 40 days are summarised in the table below.

Time (mins)	90–139	140–149	150–159	160–169	170–179	180–229
No. of days	8	10	10	4	4	4

Draw a histogram to illustrate these data

(4)

2. A car dealer offers purchasers a three year warranty on a new car. He sells two models, the Zippy and the Nifty. For the first 50 cars sold of each model the number of claims under the warranty is shown in the table below.

	Claim	No claim
Zippy	35	15
Nifty	40	10

One of the purchasers is chosen at random. Let  $A$  be the event that no claim is made by the purchaser under the warranty and  $B$  the event that the car purchased is a Nifty.

- (a) Find  $P(A \cap B)$ .

(2)

- (b) Find  $P(A')$ .

(2)

Given that the purchaser chosen does not make a claim under the warranty,

- (c) find the probability that the car purchased is a Zippy.

(2)

- (d) Show that making a claim is not independent of the make of the car purchased.

Comment on this result.

(3)

3. A drinks machine dispenses coffee into cups. A sign on the machine indicates that each cup contains 50 ml of coffee. The machine actually dispenses a mean amount of 55 ml per cup and 10% of the cups contain less than the amount stated on the sign. Assuming that the amount of coffee dispensed into each cup is normally distributed find

(a) the standard deviation of the amount of coffee dispensed per cup in ml, (4)

(b) the percentage of cups that contain more than 61 ml. (3)

Following complaints, the owners of the machine make adjustments. Only 2.5% of cups now contain less than 50 ml. The standard deviation of the amount dispensed is reduced to 3 ml.

Assuming that the amount of coffee dispensed is still normally distributed,

(c) find the new mean amount of coffee per cup. (4)

4. A restaurant owner is concerned about the amount of time customers have to wait before being served. He collects data on the waiting times, to the nearest minute, of 20 customers. These data are listed below.

15, 14, 16, 15, 17, 16, 15, 14, 15, 16,  
17, 16, 15, 14, 16, 17, 15, 25, 18, 16

(a) Find the median and inter-quartile range of the waiting times. (5)

An outlier is an observation that falls either  $1.5 \times$  (inter-quartile range) above the upper quartile or  $1.5 \times$  (inter-quartile range) below the lower quartile.

(b) Draw a boxplot to represent these data, clearly indicating any outliers. (7)

(c) Find the mean of these data. (2)

(d) Comment on the skewness of these data. Justify your answer. (2)

5. The discrete random variable  $X$  has probability function

$$P(X=x) = \begin{cases} k(2-x), & x=0, 1, 2, \\ k(x-2), & x=3, \\ 0, & \text{otherwise,} \end{cases}$$

where  $k$  is a positive constant.

(a) Show that  $k = 0.25$ . (2)

(b) Find  $E(X)$  and show that  $E(X^2) = 2.5$ . (4)

(c) Find  $\text{Var}(3X - 2)$ . (3)

Two independent observations  $X_1$  and  $X_2$  are made of  $X$ .

(d) Show that  $P(X_1 + X_2 = 5) = 0$ . (1)

(e) Find the complete probability function for  $X_1 + X_2$ . (3)

(f) Find  $P(1.3 \leq X_1 + X_2 \leq 3.2)$ . (3)

6. The chief executive of Rex cars wants to investigate the relationship between the number of new car sales and the amount of money spent on advertising. She collects data from company records on the number of new car sales,  $c$ , and the cost of advertising each year,  $p$  (£000). The data are shown in the table below.

Year	Number of new car sale, $c$	Cost of advertising (£000), $p$
1990	4240	120
1991	4380	126
1992	4420	132
1993	4440	134
1994	4430	137
1995	4520	144
1996	4590	148
1997	4660	150
1998	4700	153
1999	4790	158

- (a) Using the coding  $x = (p - 100)$  and  $y = \frac{1}{10}(c - 4000)$ , draw a scatter diagram to represent these data. Explain why  $x$  is the explanatory variable. (5)
- (b) Find the equation of the least squares regression line of  $y$  on  $x$ .  
[Use  $\Sigma x = 402$ ,  $\Sigma y = 517$ ,  $\Sigma x^2 = 17\,538$  and  $\Sigma xy = 22\,611$ .] (7)
- (c) Deduce the equation of the least squares regression line of  $c$  on  $p$  in the form  $c = a + bp$ . (3)
- (d) Interpret the value of  $a$ . (2)
- (e) Predict the number of extra new cars sales for an increase of £2000 in advertising budget. Comment on the validity of your answer. (2)

**END**

Paper Reference(s)

**6683**

**Edexcel GCE**

**Statistics S1**

**Advanced/Advanced Subsidiary**

**Thursday 5 June 2003 – Morning**

**Time: 1 hour 30 minutes**

**Materials required for examination**

Answer Book (AB16)  
Graph Paper (ASG2)  
Mathematical Formulae (Lilac)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has seven questions.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. In a particular week, a dentist treats 100 patients. The length of time, to the nearest minute, for each patient's treatment is summarised in the table below.

Time (minutes)	4 – 7	8	9 – 10	11	12 – 16	17 – 20
Number of patients	12	20	18	22	15	13

Draw a histogram to illustrate these data.

(5)

2. The lifetimes of batteries used for a computer game have a mean of 12 hours and a standard deviation of 3 hours. Battery lifetimes may be assumed to be normally distributed.

Find the lifetime,  $t$  hours, of a battery such that 1 battery in 5 will have a lifetime longer than  $t$ .

(6)

3. A company owns two petrol stations  $P$  and  $Q$  along a main road. Total daily sales in the same week for  $P$  (£ $p$ ) and for  $Q$  (£ $q$ ) are summarised in the table below.

	$p$	$q$
Monday	4760	5380
Tuesday	5395	4460
Wednesday	5840	4640
Thursday	4650	5450
Friday	5365	4340
Saturday	4990	5550
Sunday	4365	5840

When these data are coded using  $x = \frac{p - 4365}{100}$  and  $y = \frac{q - 4340}{100}$ ,

$$\Sigma x = 48.1, \Sigma y = 52.8, \Sigma x^2 = 486.44, \Sigma y^2 = 613.22 \text{ and } \Sigma xy = 204.95.$$

- (a) Calculate  $S_{xy}$ ,  $S_{xx}$  and  $S_{yy}$ . (4)
- (b) Calculate, to 3 significant figures, the value of the product moment correlation coefficient between  $x$  and  $y$ . (3)
- (c) (i) Write down the value of the product moment correlation coefficient between  $p$  and  $q$ . (3)
- (ii) Give an interpretation of this value. (3)

4. The discrete random variable  $X$  has probability function

$$P(X = x) = \begin{cases} k(x^2 - 9), & x = 4, 5, 6 \\ 0, & \text{otherwise,} \end{cases}$$

where  $k$  is a positive constant.

- (a) Show that  $k = \frac{1}{50}$ . (3)
- (b) Find  $E(X)$  and  $\text{Var}(X)$ . (6)
- (c) Find  $\text{Var}(2X - 3)$ . (2)

5. The random variable  $X$  represents the number on the uppermost face when a fair die is thrown.

- (a) Write down the name of the probability distribution of  $X$ . (1)
- (b) Calculate the mean and the variance of  $X$ . (3)

Three fair dice are thrown and the numbers on the uppermost faces are recorded.

- (c) Find the probability that all three numbers are 6. (2)
- (d) Write down all the different ways of scoring a total of 16 when the three numbers are added together. (4)
- (e) Find the probability of scoring a total of 16. (2)

6. The number of bags of potato crisps sold per day in a bar was recorded over a two-week period. The results are shown below.

20, 15, 10, 30, 33, 40, 5, 11, 13, 20, 25, 42, 31, 17

- (a) Calculate the mean of these data. (2)
- (b) Draw a stem and leaf diagram to represent these data. (3)
- (c) Find the median and the quartiles of these data. (3)

An outlier is an observation that falls either  $1.5 \times$  (interquartile range) above the upper quartile or  $1.5 \times$  (interquartile range) below the lower quartile.

- (d) Determine whether or not any items of data are outliers. (3)
- (e) On graph paper draw a box plot to represent these data. Show your scale clearly. (3)
- (f) Comment on the skewness of the distribution of bags of crisps sold per day. Justify your answer. (2)

**TURN OVER FOR QUESTION 7**

7. Eight students took tests in mathematics and physics. The marks for each student are given in the table below where  $m$  represents the mathematics mark and  $p$  the physics mark.

		Student							
		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>
Mark	<i>m</i>	9	14	13	10	7	8	20	17
	<i>p</i>	11	23	21	15	19	10	31	26

A science teacher believes that students' marks in physics depend upon their mathematical ability. The teacher decides to investigate this relationship using the test marks.

- (a) Write down which is the explanatory variable in this investigation. (1)
- (b) Draw a scatter diagram to illustrate these data. (3)
- (c) Showing your working, find the equation of the regression line of  $p$  on  $m$ . (8)
- (d) Draw the regression line on your scatter diagram. (2)

A ninth student was absent for the physics test, but she sat the mathematics test and scored 15.

- (e) Using this model, estimate the mark she would have scored in the physics test. (2)

**END**

Paper Reference(s)

**6683**

# Edexcel GCE

## Statistics S1

### Advanced/Advanced Subsidiary

**Tuesday 4 November 2003 – Morning**

**Time: 1 hour 30 minutes**

**Materials required for examination**

Answer Book (AB16)  
Graph Paper (ASG2)  
Mathematical Formulae (Lilac)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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. This paper has six questions.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled. You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. A company wants to pay its employees according to their performance at work. The performance score  $x$  and the annual salary,  $y$  in £100s, for a random sample of 10 of its employees for last year were recorded. The results are shown in the table below.

$x$	15	40	27	39	27	15	20	30	19	24
$y$	216	384	234	399	226	132	175	316	187	196

[You may assume  $\Sigma xy = 69\,798$ ,  $\Sigma x^2 = 7\,266$ ]

- (a) Draw a scatter diagram to represent these data. (4)
- (b) Calculate exact values of  $S_{xy}$  and  $S_{xx}$ . (4)
- (c) (i) Calculate the equation of the regression line of  $y$  on  $x$ , in the form  $y = a + bx$ .  
Give the values of  $a$  and  $b$  to 3 significant figures.  
(ii) Draw this line on your scatter diagram. (5)
- (d) Interpret the gradient of the regression line. (1)

The company decides to use this regression model to determine future salaries.

- (e) Find the proposed annual salary for an employee who has a performance score of 35. (2)
- 

2. A fairground game involves trying to hit a moving target with a gunshot. A round consists of up to 3 shots. Ten points are scored if a player hits the target, but the round is over if the player misses. Linda has a constant probability of 0.6 of hitting the target and shots are independent of one another.

- (a) Find the probability that Linda scores 30 points in a round. (2)

The random variable  $X$  is the number of points Linda scores in a round.

- (b) Find the probability distribution of  $X$ . (5)
- (c) Find the mean and the standard deviation of  $X$ . (5)

A game consists of 2 rounds.

- (d) Find the probability that Linda scores more points in round 2 than in round 1. (6)
- 

3. Cooking sauces are sold in jars containing a stated weight of 500 g of sauce. The jars are filled by a machine. The actual weight of sauce in each jar is normally distributed with mean 505 g and standard deviation 10 g.

- (a) (i) Find the probability of a jar containing less than the stated weight.  
(ii) In a box of 30 jars, find the expected number of jars containing less than the stated weight. (5)

The mean weight of sauce is changed so that 1% of the jars contain less than the stated weight. The standard deviation stays the same.

- (b) Find the new mean weight of sauce. (4)
-

4. Explain what you understand by

(a) a sample space,

(1)

(b) an event.

(1)

Two events  $A$  and  $B$  are independent, such that  $P(A) = \frac{1}{3}$  and  $P(B) = \frac{1}{4}$ .

Find

(c)  $P(A \cap B)$ ,

(1)

(d)  $P(A | B)$ ,

(2)

(e)  $P(A \cup B)$ .

(2)

5. The random variable  $X$  has the discrete uniform distribution

$$P(X = x) = \frac{1}{n}, \quad x = 1, 2, \dots, n.$$

Given that  $E(X) = 5$ ,

(a) show that  $n = 9$ .

(3)

Find

(b)  $P(X < 7)$ ,

(2)

(c)  $\text{Var}(X)$ .

(4)

6. A travel agent sells holidays from his shop. The price, in £, of 15 holidays sold on a particular day are shown below.

299	1050	2315	999	485
350	169	1015	650	830
99	2100	689	550	475

For these data, find

(a) the mean and the standard deviation,

(3)

(b) the median and the inter-quartile range.

(4)

An outlier is an observation that falls either more than  $1.5 \times$  (inter-quartile range) above the upper quartile or more than  $1.5 \times$  (inter-quartile range) below the lower quartile.

(c) Determine if any of the prices are outliers.

(3)

The travel agent also sells holidays from a website on the Internet. On the same day, he recorded the price, £ $x$ , of each of 20 holidays sold on the website. The cheapest holiday sold was £98, the most expensive was £2400 and the quartiles of these data were £305, £1379 and £1805. There were no outliers.

(d) On graph paper, and using the same scale, draw box plots for the holidays sold in the shop and the holidays sold on the website.

(4)

(e) Compare and contrast sales from the shop and sales from the website.

(2)

END

Paper Reference(s)

**6683****Edexcel GCE****Statistics S1****Advanced/Advanced Subsidiary****Wednesday 14 January 2004 – Morning****Time: 1 hour 30 minutes****Materials required for examination**

Answer Book (AB16)

Graph Paper (ASG2)

Mathematical Formulae (Lilac)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has six questions.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. An office has the heating switched on at 7.00 a.m. each morning. On a particular day, the temperature of the office,  $t$  °C, was recorded  $m$  minutes after 7.00 a.m. The results are shown in the table below.

$m$	0	10	20	30	40	50
$t$	6.0	8.9	11.8	13.5	15.3	16.1

- (a) Calculate the exact values of  $S_m$  and  $S_{mm}$ . (4)
- (b) Calculate the equation of the regression line of  $t$  on  $m$  in the form  $t = a + bm$ . (3)
- (c) Use your equation to estimate the value of  $t$  at 7.35 a.m. (2)
- (d) State, giving a reason, whether or not you would use the regression equation in (b) to estimate the temperature
- (i) at 9.00 a.m. that day,
- (ii) at 7.15 a.m. one month later. (4)

2. The random variable  $X$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$ .

- (a) Write down 3 properties of the distribution of  $X$ . (3)
- Given that  $\mu = 27$  and  $\sigma = 10$
- (b) find  $P(26 < X < 28)$ . (4)

3. A discrete random variable  $X$  has the probability function shown in the table below.

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

Find

- (a)  $P(1 < X \leq 3)$ , (2)
- (b)  $F(2.6)$ , (1)
- (c)  $E(X)$ , (2)
- (d)  $E(2X - 3)$ , (2)
- (e)  $\text{Var}(X)$  (3)

4. The events  $A$  and  $B$  are such that  $P(A) = \frac{2}{5}$ ,  $P(B) = \frac{1}{2}$  and  $P(A | B') = \frac{4}{5}$ .

- (a) Find
- (i)  $P(A \cap B')$ ,
- (ii)  $P(A \cap B)$ ,
- (iii)  $P(A \cup B)$ ,
- (iv)  $P(A | B)$ . (7)
- (b) State, with a reason, whether or not  $A$  and  $B$  are
- (i) mutually exclusive, (2)
- (ii) independent. (2)

5. The values of daily sales, to the nearest £, taken at a newsagents last year are summarised in the table below.

Sales	Number of days
1 – 200	166
201 – 400	100
401 – 700	59
701 – 1000	30
1001 – 1500	5

- (a) Draw a histogram to represent these data. (5)
- (b) Use interpolation to estimate the median and inter-quartile range of daily sales. (5)
- (c) Estimate the mean and the standard deviation of these data. (6)
- The newsagent wants to compare last year's sales with other years.
- (d) State whether the newsagent should use the median and the inter-quartile range or the mean and the standard deviation to compare daily sales. Give a reason for your answer. (2)

6. One of the objectives of a computer game is to collect keys. There are three stages to the game. The probability of collecting a key at the first stage is  $\frac{2}{3}$ , at the second stage is  $\frac{1}{2}$ , and at the third stage is  $\frac{1}{4}$ .
- (a) Draw a tree diagram to represent the 3 stages of the game. (4)
- (b) Find the probability of collecting all 3 keys. (2)
- (c) Find the probability of collecting exactly one key in a game. (5)
- (d) Calculate the probability that keys are not collected on at least 2 successive stages in a game. (5)

**END**

Paper Reference(s)

**6683****Edexcel GCE****Statistics S1****Advanced/Advanced Subsidiary****Friday 11 June 2004 – Morning****Time: 1 hour 30 minutes****Materials required for examination**

Answer Book (AB16)

Graph Paper (ASG2)

Mathematical Formulae (Lilac)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has six questions.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

N17022A

This publication may only be reproduced in accordance with London Qualification Limited  
©2004 London Qualification Limited

1. A fair die has six faces numbered 1, 2, 2, 3, 3 and 3. The die is rolled twice and the number showing on the uppermost face is recorded each time.

Find the probability that the sum of the two numbers recorded is at least 5.

(5)

2. A researcher thinks there is a link between a person's height and level of confidence. She measured the height  $h$ , to the nearest cm, of a random sample of 9 people. She also devised a test to measure the level of confidence  $c$  of each person. The data are shown in the table below.

$h$	179	169	187	166	162	193	161	177	168
$c$	569	561	579	561	540	598	542	565	573

[You may use  $\Sigma h^2 = 272\,094$ ,  $\Sigma c^2 = 2\,878\,966$ ,  $\Sigma hc = 884\,484$ ]

- (a) Draw a scatter diagram to illustrate these data. (4)
- (b) Find exact values of  $S_{hc}$ ,  $S_{hh}$  and  $S_{cc}$ . (4)
- (c) Calculate the value of the product moment correlation coefficient for these data. (3)
- (d) Give an interpretation of your correlation coefficient. (1)
- (e) Calculate the equation of the regression line of  $c$  on  $h$  in the form  $c = a + bh$ . (3)
- (f) Estimate the level of confidence of a person of height 180 cm. (2)
- (g) State the range of values of  $h$  for which estimates of  $c$  are reliable. (1)

N17022A

2

3. A discrete random variable  $X$  has a probability function as shown in the table below, where  $a$  and  $b$  are constants.

$x$	0	1	2	3
$P(X = x)$	0.2	0.3	$b$	$a$

Given that  $E(X) = 1.7$ ,

- (a) find the value of  $a$  and the value of  $b$ . (5)

Find

- (b)  $P(0 < X < 1.5)$ , (1)

- (c)  $E(2X - 3)$ . (2)

- (d) Show that  $\text{Var}(X) = 1.41$ . (3)

- (e) Evaluate  $\text{Var}(2X - 3)$ . (2)
- 

4. The attendance at college of a group of 18 students was recorded for a 4-week period.

The number of students actually attending each of 16 classes are shown below.

18	18	17	17
16	17	16	18
18	14	17	18
15	17	18	16

- (a) (i) Calculate the mean and the standard deviation of the number of students attending these classes. (5)

- (ii) Express the mean as a percentage of the 18 students in the group.

In the same 4-week period, the attendance of a different group of 20 students is shown below.

20	16	18	19
15	14	14	15
18	15	16	17
16	18	15	14

- (b) Construct a back-to-back stem and leaf diagram to represent the attendance in both groups. (5)

- (c) Find the mode, median and inter-quartile range for each group of students. (6)

The mean percentage attendance and standard deviation for the second group of students are 81.25 and 1.82 respectively.

- (d) Compare and contrast the attendance of these 2 groups of students. (3)
-

5. A health club lets members use, on each visit, its facilities for as long as they wish. The club's records suggest that the length of a visit can be modelled by a normal distribution with mean 90 minutes. Only 20% of members stay for more than 125 minutes.

- (a) Find the standard deviation of the normal distribution. (4)
- (b) Find the probability that a visit lasts less than 25 minutes. (3)

The club introduce a closing time of 10:00 pm. Tara arrives at the club at 8:00 pm.

- (c) Explain whether or not this normal distribution is still a suitable model for the length of her visit. (2)

6. Three events  $A$ ,  $B$  and  $C$  are defined in the sample space  $S$ . The events  $A$  and  $B$  are mutually exclusive and  $A$  and  $C$  are independent.

- (a) Draw a Venn diagram to illustrate the relationships between the 3 events and the sample space. (3)

Given that  $P(A) = 0.2$ ,  $P(B) = 0.4$  and  $P(A \cup C) = 0.7$ , find

- (b)  $P(A|C)$ , (2)
- (c)  $P(A \cup B)$ , (2)
- (d)  $P(C)$ . (4)

END

Paper Reference(s)

6683

## Edexcel GCE

### Statistics S1

### Advanced/Advanced Subsidiary

Tuesday 2 November 2004 – Morning

Time: 1 hour 30 minutes

**Materials required for examination**

Answer Book (AB16)  
Graph Paper (ASG2)  
Mathematical Formulae (Lilac)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has seven questions.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. As part of their job, taxi drivers record the number of miles they travel each day. A random sample of the mileages recorded by taxi drivers Keith and Asif are summarised in the back-to-back stem and leaf diagram below.

Totals	Keith		Asif	Totals
(9)	8 7 7 4 3 2 1 1 0	18	4 4 5 7	(4)
(11)	9 9 8 7 6 5 4 3 3 1 1	19	5 7 8 9 9	(5)
(6)	8 7 4 2 2 0	20	0 2 2 4 4 8	(6)
(6)	9 4 3 1 0 0	21	2 3 5 6 6 7 9	(7)
(4)	6 4 1 1	22	1 1 2 4 5 5 8	(7)
(2)		23	1 1 3 4 6 6 7 8	(8)
(2)		24	2 4 8 9	(4)
(1)		25	4	(1)
(2)		26		(0)

Key: 0 | 18 | 4 means 180 for Keith and 184 for Asif

The quartiles for these two distributions are summarised in the table below.

	Keith	Asif
Lower quartile	191	$a$
Median	$b$	218
Upper quartile	221	$c$

- (a) Find the values of  $a$ ,  $b$  and  $c$ . (3)

Outliers are values that lie outside the limits

$$Q_1 - 1.5(Q_3 - Q_1) \text{ and } Q_3 + 1.5(Q_3 - Q_1).$$

- (b) On graph paper, and showing your scale clearly, draw a box plot to represent Keith's data. (8)
- (c) Comment on the skewness of the two distributions. (3)
- 

2. An experiment carried out by a student yielded pairs of  $(x, y)$  observations such that

$$\bar{x} = 36, \bar{y} = 28.6, S_{xx} = 4402, S_{xy} = 3477.6$$

- (a) Calculate the equation of the regression line of  $y$  on  $x$  in the form  $y = a + bx$ . Give your values of  $a$  and  $b$  to 2 decimal places. (3)

- (b) Find the value of  $y$  when  $x = 45$ . (1)
- 

3. The random variable  $X \sim N(\mu, \sigma^2)$ .

It is known that

$$P(X \leq 66) = 0.0359 \text{ and } P(X \geq 81) = 0.1151.$$

- (a) In the space below, give a clearly labelled sketch to represent these probabilities on a Normal curve. (1)

- (b) (i) Show that the value of  $\sigma$  is 5.
- (ii) Find the value of  $\mu$ . (8)

- (c) Find  $P(69 \leq X \leq 83)$ . (3)
-

4. The discrete random variable  $X$  has probability function

$$P(X = x) = \begin{cases} 0.2, & x = -3, -2, \\ \alpha, & x = -1, 0 \\ 0.1, & x = 1, 2. \end{cases}$$

Find

- (a)  $\alpha$ , (2)
- (b)  $P(-1 \leq X < 2)$ , (1)
- (c)  $F(0.6)$ , (1)
- (d) the value of  $a$  such that  $E(aX + 3) = 1.2$ , (4)
- (e)  $\text{Var}(X)$ , (4)
- (f)  $\text{Var}(3X - 2)$ . (2)

5. The events  $A$  and  $B$  are such that  $P(A) = \frac{1}{2}$ ,  $P(B) = \frac{1}{3}$  and  $P(A \cap B) = \frac{1}{4}$ .

- (a) Using the space below, represent these probabilities in a Venn diagram. (4)

Hence, or otherwise, find

- (b)  $P(A \cup B)$ , (1)
- (c)  $P(A \mid B')$  (2)

6. Students in Mr Brawn's exercise class have to do press-ups and sit-ups. The number of press-ups  $x$  and the number of sit-ups  $y$  done by a random sample of 8 students are summarised below.

$$\Sigma x = 272, \quad \Sigma x^2 = 10\,164, \quad \Sigma xy = 11\,222,$$

$$\Sigma y = 320, \quad \Sigma y^2 = 13\,464.$$

- (a) Evaluate  $S_{xx}$ ,  $S_{yy}$  and  $S_{xy}$ . (4)
- (b) Calculate, to 3 decimal places, the product moment correlation coefficient between  $x$  and  $y$ . (3)
- (c) Give an interpretation of your coefficient. (2)
- (d) Calculate the mean and the standard deviation of the number of press-ups done by these students. (4)

Mr Brawn assumes that the number of press-ups that can be done by any student can be modelled by a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ . Assuming that  $\mu$  and  $\sigma$  take the same values as those calculated in part (d),

- (e) find the value of  $a$  such that  $P(\mu - a < X < \mu + a) = 0.95$ . (3)
- (f) Comment on Mr Brawn's assumption of normality. (2)

7. A college organised a 'fun run'. The times, to the nearest minute, of a random sample of 100 students who took part are summarised in the table below.

Time	Number of students
40–44	10
45–47	15
48	23
49–51	21
52–55	16
56–60	15

- (a) Give a reason to support the use of a histogram to represent these data. (1)
- (b) Write down the upper class boundary and the lower class boundary of the class 40–44. (1)
- (c) On graph paper, draw a histogram to represent these data. (4)

---

END

Paper Reference(s)

**6683**

# Edexcel GCE

## Statistics S1

### Advanced Subsidiary

**Friday 14 January 2005 – Morning**

**Time: 1 hour 30 minutes**

**Materials required for examination**

Mathematical Formulae (Lilac)  
Graph Paper (ASG2)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

#### Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has seven questions.

The total mark for this paper is 75.

#### Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

N16741A

This publication may only be reproduced in accordance with London Qualifications Limited copyright policy.  
©2005 London Qualifications Limited

1. A company assembles drills using components from two sources. Goodbuy supplies 85% of the components and Amart supplies the rest. It is known that 3% of the components supplied by Goodbuy are faulty and 6% of those supplied by Amart are faulty.

(a) Represent this information on a tree diagram. (3)

An assembled drill is selected at random.

(b) Find the probability that it is not faulty. (3)

---

2. The number of caravans on Seaview caravan site on each night in August last year is summarised in the following stem and leaf diagram.

Caravans	1   0 means 10	Totals
1	0 5	(2)
2	1 2 4 8	(4)
3	0 3 3 3 4 7 8 8	(8)
4	1 1 3 5 8 8 8 9 9	(9)
5	2 3 6 6 7	(5)
6	2 3 4	(3)

(a) Find the three quartiles of these data. (3)

During the same month, the least number of caravans on Northcliffe caravan site was 31. The maximum number of caravans on this site on any night that month was 72. The three quartiles for this site were 38, 45 and 52 respectively.

(b) On graph paper and using the same scale, draw box plots to represent the data for both caravan sites. You may assume that there are no outliers. (6)

(c) Compare and contrast these two box plots. (3)

(d) Give an interpretation to the upper quartiles of these two distributions. (2)

---

3. The following table shows the height  $x$ , to the nearest cm, and the weight  $y$ , to the nearest kg, of a random sample of 12 students.

$x$	148	164	156	172	147	184	162	155	182	165	175	152
$y$	39	59	56	77	44	77	65	49	80	72	70	52

(a) On graph paper, draw a scatter diagram to represent these data. (3)

(b) Write down, with a reason, whether the correlation coefficient between  $x$  and  $y$  is positive or negative. (2)

The data in the table can be summarised as follows.

$$\Sigma x = 1962, \quad \Sigma y = 740, \quad \Sigma y^2 = 47\,746, \quad \Sigma xy = 122\,783, \quad S_{xx} = 1745.$$

(c) Find  $S_{xy}$ . (2)

The equation of the regression line of  $y$  on  $x$  is  $y = -106.331 + bx$ .

(d) Find, to 3 decimal places, the value of  $b$ . (2)

(e) Find, to 3 significant figures, the mean  $\bar{y}$  and the standard deviation  $s$  of the weights of this sample of students. (3)

(f) Find the values of  $\bar{y} \pm 1.96s$ . (2)

(g) Comment on whether or not you think that the weights of these students could be modelled by a normal distribution. (1)

---

4. The random variable  $X$  has probability function

$$P(X = x) = kx, \quad x = 1, 2, \dots, 5.$$

(a) Show that  $k = \frac{1}{15}$ .

(2)

Find

(b)  $P(X < 4)$ ,

(2)

(c)  $E(X)$ ,

(2)

(d)  $E(3X - 4)$ .

(2)

5. Articles made on a lathe are subject to three kinds of defect,  $A$ ,  $B$  or  $C$ . A sample of 1000 articles was inspected and the following results were obtained.

- 31 had a type  $A$  defect
- 37 had a type  $B$  defect
- 42 had a type  $C$  defect
- 11 had both type  $A$  and type  $B$  defects
- 13 had both type  $B$  and type  $C$  defects
- 10 had both type  $A$  and type  $C$  defects
- 6 had all three types of defect.

(a) Draw a Venn diagram to represent these data.

(6)

Find the probability that a randomly selected article from this sample had

(b) no defects,

(1)

(c) no more than one of these defects.

(2)

An article selected at random from this sample had only one defect.

(d) Find the probability that it was a type  $B$  defect.

(2)

Two different articles were selected at random from this sample.

(e) Find the probability that both had type  $B$  defects.

(2)

6. A discrete random variable is such that each of its values is assumed to be equally likely.

(a) Write down the name of the distribution that could be used to model this random variable.

(1)

(b) Give an example of such a distribution.

(1)

(c) Comment on the assumption that each value is equally likely.

(2)

(d) Suggest how you might refine the model in part (a).

(2)

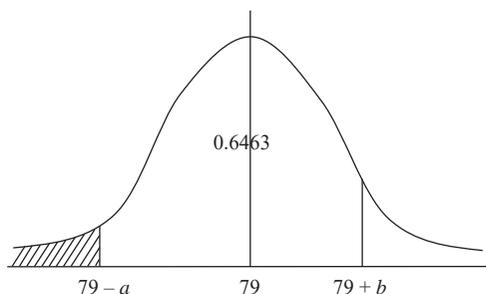
7. The random variable  $X$  is normally distributed with mean 79 and variance 144.

Find

(a)  $P(X < 70)$ , (3)

(b)  $P(64 < X < 96)$ . (3)

It is known that  $P(79 - a \leq X \leq 79 + b) = 0.6463$ . This information is shown in the figure below.



Given that  $P(X \geq 79 + b) = 2P(X \leq 79 - a)$ ,

(c) show that the area of the shaded region is 0.1179. (3)

(d) Find the value of  $b$ . (4)

---

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683**

**Edexcel GCE**

**Statistics S1**

**Advanced Subsidiary**

**Thursday 9 June 2005 – Morning**

**Time: 1 hour 30 minutes**

Materials required for examination

Mathematical Formulae (Lilac)

Graph Paper (ASG2)

Items included with question papers

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has seven questions.

The total mark for this paper is 75.

**Advice to Candidates**

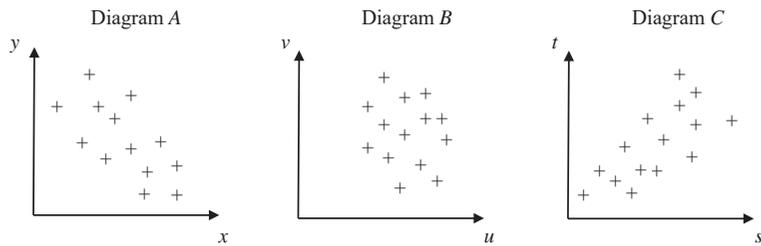
You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

N20910A

This publication may only be reproduced in accordance with London Qualifications Limited copyright policy.  
©2005 London Qualifications Limited

1. The scatter diagrams below were drawn by a student.



The student calculated the value of the product moment correlation coefficient for each of the sets of data.

The values were

0.68      -0.79      0.08

Write down, with a reason, which value corresponds to which scatter diagram.

(6)

2. The following table summarises the distances, to the nearest km, that 134 examiners travelled to attend a meeting in London.

Distance (km)	Number of examiners
41–45	4
46–50	19
51–60	53
61–70	37
71–90	15
91–150	6

(a) Give a reason to justify the use of a histogram to represent these data. (1)

(b) Calculate the frequency densities needed to draw a histogram for these data. (2)  
**(DO NOT DRAW THE HISTOGRAM)**

(c) Use interpolation to estimate the median  $Q_2$ , the lower quartile  $Q_1$ , and the upper quartile  $Q_3$  of these data.

The mid-point of each class is represented by  $x$  and the corresponding frequency by  $f$ . Calculations then give the following values

$$\sum fx = 8379.5 \quad \text{and} \quad \sum fx^2 = 557489.75$$

(d) Calculate an estimate of the mean and an estimate of the standard deviation for these data. (4)

One coefficient of skewness is given by

$$\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

(e) Evaluate this coefficient and comment on the skewness of these data. (4)

(f) Give another justification of your comment in part (e). (1)

3. A long distance lorry driver recorded the distance travelled,  $m$  miles, and the amount of fuel used,  $f$  litres, each day. Summarised below are data from the driver's records for a random sample of 8 days.

The data are coded such that  $x = m - 250$  and  $y = f - 100$ .

$$\sum x = 130 \quad \sum y = 48 \quad \sum xy = 8880 \quad S_{xx} = 20487.5$$

- (a) Find the equation of the regression line of  $y$  on  $x$  in the form  $y = a + bx$ . (6)
- (b) Hence find the equation of the regression line of  $f$  on  $m$ . (3)
- (c) Predict the amount of fuel used on a journey of 235 miles. (1)
- 

4. Aeroplanes fly from City A to City B. Over a long period of time the number of minutes delay in take-off from City A was recorded. The minimum delay was 5 minutes and the maximum delay was 63 minutes. A quarter of all delays were at most 12 minutes, half were at most 17 minutes and 75% were at most 28 minutes. Only one of the delays was longer than 45 minutes.

An outlier is an observation that falls either  $1.5 \times$  (interquartile range) above the upper quartile or  $1.5 \times$  (interquartile range) below the lower quartile.

- (a) On graph paper, draw a box plot to represent these data. (7)
- (b) Comment on the distribution of delays. Justify your answer. (2)
- (c) Suggest how the distribution might be interpreted by a passenger who frequently flies from City A to City B. (1)
- 

5. The random variable  $X$  has probability function

$$P(X = x) = \begin{cases} kx, & x = 1, 2, 3, \\ k(x+1), & x = 4, 5, \end{cases}$$

where  $k$  is a constant.

- (a) Find the value of  $k$ . (2)
- (b) Find the exact value of  $E(X)$ . (2)
- (c) Show that, to 3 significant figures,  $\text{Var}(X) = 1.47$ . (4)
- (d) Find, to 1 decimal place,  $\text{Var}(4 - 3X)$ . (2)
- 

6. A scientist found that the time taken,  $M$  minutes, to carry out an experiment can be modelled by a normal random variable with mean 155 minutes and standard deviation 3.5 minutes.

Find

- (a)  $P(M > 160)$ , (3)
- (b)  $P(150 \leq M \leq 157)$ , (4)
- (c) the value of  $m$ , to 1 decimal place, such that  $P(M \leq m) = 0.30$ . (4)
-

7. In a school there are 148 students in Years 12 and 13 studying Science, Humanities or Arts subjects. Of these students, 89 wear glasses and the others do not. There are 30 Science students of whom 18 wear glasses. The corresponding figures for the Humanities students are 68 and 44 respectively.

A student is chosen at random.

Find the probability that this student

- (a) is studying Arts subjects, (4)
- (b) does not wear glasses, given that the student is studying Arts subjects. (2)

Amongst the Science students, 80% are right-handed. Corresponding percentages for Humanities and Arts students are 75% and 70% respectively.

A student is again chosen at random.

- (c) Find the probability that this student is right-handed. (3)
- (d) Given that this student is right-handed, find the probability that the student is studying Science subjects. (3)

---

**TOTAL FOR PAPER:75 MARKS**

**END**

Paper Reference(s)

**6683/01**

# Edexcel GCE

## Statistics S1

### Advanced/Advanced Subsidiary

**Monday 16 January 2006 – Morning**

**Time: 1 hour 30 minutes**

Materials required for examination  
Mathematical Formulae (Green or Lilac)

Items included with question papers  
Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

#### Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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 7 questions on this paper. The total mark for this paper is 75.

#### Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. Over a period of time, the number of people  $x$  leaving a hotel each morning was recorded. These data are summarised in the stem and leaf diagram below.

Number leaving	3	2 means 32	Totals
2	7 9 9		(3)
3	2 2 3 5 6		(5)
4	0 1 4 8 9		(5)
5	2 3 3 6 6 6 8		(7)
6	0 1 4 5		(4)
7	2 3		(2)
8	1		(1)

For these data,

- (a) write down the mode, (1)
- (b) find the values of the three quartiles. (3)

Given that  $\Sigma x = 1335$  and  $\Sigma x^2 = 71\,801$ , find

- (c) the mean and the standard deviation of these data. (4)

One measure of skewness is found using

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

- (d) Evaluate this measure to show that these data are negatively skewed. (2)
- (e) Give two other reasons why these data are negatively skewed. (4)
- 

2. The random variable  $X$  has probability distribution

$x$	1	2	3	4	5
$P(X = x)$	0.10	$p$	0.20	$q$	0.30

- (a) Given that  $E(X) = 3.5$ , write down two equations involving  $p$  and  $q$ . (3)

Find

- (b) the value of  $p$  and the value of  $q$ , (3)
- (c)  $\text{Var}(X)$ , (4)
- (d)  $\text{Var}(3 - 2X)$ . (2)
- 

3. A manufacturer stores drums of chemicals. During storage, evaporation takes place. A random sample of 10 drums was taken and the time in storage,  $x$  weeks, and the evaporation loss,  $y$  ml, are shown in the table below.

$x$	3	5	6	8	10	12	13	15	16	18
$y$	36	50	53	61	69	79	82	90	88	96

- (a) On graph paper, draw a scatter diagram to represent these data. (3)
- (b) Give a reason to support fitting a regression model of the form  $y = a + bx$  to these data. (1)
- (c) Find, to 2 decimal places, the value of  $a$  and the value of  $b$ .

(You may use  $\Sigma x^2 = 1352$ ,  $\Sigma y^2 = 53\,112$  and  $\Sigma xy = 8354$ .) (7)

- (d) Give an interpretation of the value of  $b$ . (1)
- (e) Using your model, predict the amount of evaporation that would take place after
- (i) 19 weeks, (2)
- (ii) 35 weeks. (2)
- (f) Comment, with a reason, on the reliability of each of your predictions. (4)
-

4. A bag contains 9 blue balls and 3 red balls. A ball is selected at random from the bag and its colour is recorded. The ball is not replaced. A second ball is selected at random and its colour is recorded.

(a) Draw a tree diagram to represent the information. (3)

Find the probability that

(a) the second ball selected is red, (2)

(b) both balls selected are red, given that the second ball selected is red. (2)

---

5. (a) Write down two reasons for using statistical models. (2)

(b) Give an example of a random variable that could be modelled by

(i) a normal distribution,

(ii) a discrete uniform distribution. (2)

---

6. For the events  $A$  and  $B$ ,

$$P(A \cap B') = 0.32, P(A' \cap B) = 0.11 \text{ and } P(A \cup B) = 0.65.$$

(a) Draw a Venn diagram to illustrate the complete sample space for the events  $A$  and  $B$ . (3)

(b) Write down the value of  $P(A)$  and the value of  $P(B)$ . (3)

(c) Find  $P(A | B')$ . (2)

(d) Determine whether or not  $A$  and  $B$  are independent. (3)

---

7. The heights of a group of athletes are modelled by a normal distribution with mean 180 cm and a standard deviation 5.2 cm. The weights of this group of athletes are modelled by a normal distribution with mean 85 kg and standard deviation 7.1 kg.

Find the probability that a randomly chosen athlete

(a) is taller than 188 cm, (3)

(b) weighs less than 97 kg. (2)

(c) Assuming that for these athletes height and weight are independent, find the probability that a randomly chosen athlete is taller than 188 cm and weighs more than 97 kg. (3)

(d) Comment on the assumption that height and weight are independent. (1)

---

**END** **TOTAL FOR PAPER: 75 MARKS**

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced/Advanced Subsidiary****Wednesday 24 May 2006 – Afternoon****Time: 1 hour 30 minutes****Materials required for examination**

Mathematical Formulae (Green)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, initials and signature. 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. There are 6 questions in this question paper. The total mark for this paper is 75.

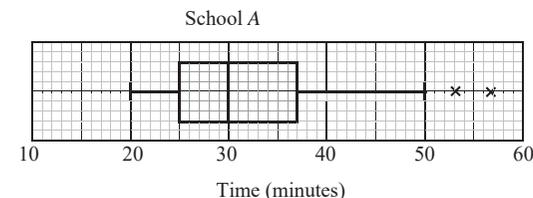
**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled. You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

**N22337A** This publication may only be reproduced in accordance with London Qualifications copyright policy. ©2006 London Qualifications Limited.

1. (a) Describe the main features and uses of a box plot. (3)

Children from schools *A* and *B* took part in a fun run for charity. The times, to the nearest minute, taken by the children from school *A* are summarised in Figure 1.

**Figure 1**

- (b) (i) Write down the time by which 75% of the children in school *A* had completed the run. (2)
- (ii) State the name given to this value. (2)
- (c) Explain what you understand by the two crosses (x) on Figure 1. (2)

For school *B* the least time taken by any of the children was 25 minutes and the longest time was 55 minutes. The three quartiles were 30, 37 and 50 respectively.

- (d) On graph paper, draw a box plot to represent the data from school *B*. (4)
- (e) Compare and contrast these two box plots. (4)

2. Sunita and Shelley talk to each other once a week on the telephone. Over many weeks they recorded, to the nearest minute, the number of minutes spent in conversation on each occasion. The following table summarises their results.

Time (to the nearest minute)	Number of conversations
5–9	2
10–14	9
15–19	20
20–24	13
25–29	8
30–34	3

Two of the conversations were chosen at random.

- (a) Find the probability that both of them were longer than 24.5 minutes. (2)

The mid-point of each class was represented by  $x$  and its corresponding frequency by  $f$ , giving  $\sum fx = 1060$ .

- (b) Calculate an estimate of the mean time spent on their conversations. (2)

During the following 25 weeks they monitored their weekly conversation and found that at the end of the 80 weeks their overall mean length of conversation was 21 minutes.

- (c) Find the mean time spent in conversation during these 25 weeks. (4)
- (d) Comment on these two mean values. (2)
- 

3. A metallurgist measured the length,  $l$  mm, of a copper rod at various temperatures,  $t$  °C, and recorded the following results.

$t$	$l$
20.4	2461.12
27.3	2461.41
32.1	2461.73
39.0	2461.88
42.9	2462.03
49.7	2462.37
58.3	2462.69
67.4	2463.05

The results were then coded such that  $x = t$  and  $y = l - 2460.00$ .

- (a) Calculate  $S_{xy}$  and  $S_{xx}$ .  
(You may use  $\sum x^2 = 15965.01$  and  $\sum xy = 757.467$ ) (5)
- (b) Find the equation of the regression line of  $y$  on  $x$  in the form  $y = a + bx$ . (5)
- (c) Estimate the length of the rod at 40 °C. (3)
- (d) Find the equation of the regression line of  $l$  on  $t$ . (2)
- (e) Estimate the length of the rod at 90 °C. (1)
- (f) Comment on the reliability of your estimate in part (e). (2)
-

4. The random variable  $X$  has the discrete uniform distribution

$$P(X = x) = \frac{1}{5}, \quad x = 1, 2, 3, 4, 5.$$

- (a) Write down the value of  $E(X)$  and show that  $\text{Var}(X) = 2$ .

(3)

Find

- (b)  $E(3X - 2)$ ,

(2)

- (c)  $\text{Var}(4 - 3X)$

(2)

5. From experience a high jumper knows that he can clear a height of at least 1.78 m once in 5 attempts. He also knows that he can clear a height of at least 1.65 m on 7 out of 10 attempts.

Assuming that the heights the high jumper can reach follow a Normal distribution,

- (a) draw a sketch to illustrate the above information,

(3)

- (b) find, to 3 decimal places, the mean and the standard deviation of the heights the high jumper can reach,

(6)

- (c) calculate the probability that he can jump at least 1.74 m.

(3)

6. A group of 100 people produced the following information relating to three attributes. The attributes were wearing glasses, being left-handed and having dark hair.

Glasses were worn by 36 people, 28 were left-handed and 36 had dark hair. There were 17 who wore glasses and were left-handed, 19 who wore glasses and had dark hair and 15 who were left-handed and had dark hair. Only 10 people wore glasses, were left-handed and had dark hair.

- (a) Represent these data on a Venn diagram.

(6)

A person was selected at random from this group.

Find the probability that this person

- (b) wore glasses but was not left-handed and did not have dark hair,

(1)

- (c) did not wear glasses, was not left-handed and did not have dark hair,

(1)

- (d) had only two of the attributes,

(2)

- (e) wore glasses, given they were left-handed and had dark hair.

(3)

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced/Advanced Subsidiary****Tuesday 16 January 2007 – Morning****Time: 1 hour 30 minutes****Materials required for examination**  
Mathematical Formulae (Green or Lilac)**Items included with question papers**  
Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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 7 questions on this paper. The total mark for this paper is 75.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

N23957A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.  
©2007 Edexcel Limited.

1. As part of a statistics project, Gill collected data relating to the length of time, to the nearest minute, spent by shoppers in a supermarket and the amount of money they spent. Her data for a random sample of 10 shoppers are summarised in the table below, where  $t$  represents time and  $£m$  the amount spent over £20.

$t$ (minutes)	$£m$
15	-3
23	17
5	-19
16	4
30	12
6	-9
32	27
23	6
35	20
27	6

- (a) Write down the actual amount spent by the shopper who was in the supermarket for 15 minutes. (1)

- (b) Calculate  $S_{tt}$ ,  $S_{mm}$  and  $S_{tm}$ .

(You may use  $\Sigma t^2 = 5478$ ,  $\Sigma m^2 = 2101$ , and  $\Sigma tm = 2485$ )

- (c) Calculate the value of the product moment correlation coefficient between  $t$  and  $m$ . (3)
- (d) Write down the value of the product moment correlation coefficient between  $t$  and the actual amount spent. Give a reason to justify your value. (2)

On another day Gill collected similar data. For these data the product moment correlation coefficient was 0.178.

- (e) Give an interpretation to both of these coefficients. (2)
- (f) Suggest a practical reason why these two values are so different. (1)

2. In a factory, machines  $A$ ,  $B$  and  $C$  are all producing metal rods of the same length. Machine  $A$  produces 35% of the rods, machine  $B$  produces 25% and the rest are produced by machine  $C$ . Of their production of rods, machines  $A$ ,  $B$  and  $C$  produce 3%, 6% and 5% defective rods respectively.

(a) Draw a tree diagram to represent this information. (3)

(b) Find the probability that a randomly selected rod is

(i) produced by machine  $A$  and is defective,

(ii) is defective. (5)

(c) Given that a randomly selected rod is defective, find the probability that it was produced by machine  $C$ . (3)

3. The random variable  $X$  has probability function

$$P(X = x) = \frac{(2x-1)}{36} \quad x = 1, 2, 3, 4, 5, 6.$$

(a) Construct a table giving the probability distribution of  $X$ . (3)

Find

(b)  $P(2 < X \leq 5)$  (2)

(c) the exact value of  $E(X)$ . (2)

(d) Show that  $\text{Var}(X) = 1.97$  to 3 significant figures. (4)

(e) Find  $\text{Var}(2 - 3X)$ . (2)

4. Summarised below are the distances, to the nearest mile, travelled to work by a random sample of 120 commuters.

Distance (to the nearest mile)	Number of commuters
0 – 9	10
10 – 19	19
20 – 29	43
30 – 39	25
40 – 49	8
50 – 59	6
60 – 69	5
70 – 79	3
80 – 89	1

For this distribution,

(a) describe its shape, (1)

(b) use linear interpolation to estimate its median. (2)

The mid-point of each class was represented by  $x$  and its corresponding frequency by  $f$  giving

$$\sum fx = 3550 \text{ and } \sum fx^2 = 138020$$

(c) Estimate the mean and standard deviation of this distribution. (3)

One coefficient of skewness is given by

$$\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

(d) Evaluate this coefficient for this distribution. (3)

(e) State whether or not the value of your coefficient is consistent with your description in part (a). Justify your answer. (2)

(f) State, with a reason, whether you should use the mean or the median to represent the data in this distribution. (2)

(g) State the circumstance under which it would not matter whether you used the mean or the median to represent a set of data. (1)

5. A teacher recorded, to the nearest hour, the time spent watching television during a particular week by each child in a random sample. The times were summarised in a grouped frequency table and represented by a histogram.

One of the classes in the grouped frequency distribution was 20–29 and its associated frequency was 9. On the histogram the height of the rectangle representing that class was 3.6 cm and the width was 2 cm.

- (a) Give a reason to support the use of a histogram to represent these data. (1)
- (b) Write down the underlying feature associated with each of the bars in a histogram. (1)
- (c) Show that on this histogram each child was represented by  $0.8 \text{ cm}^2$ . (3)

The total area under the histogram was  $24 \text{ cm}^2$ .

- (d) Find the total number of children in the group. (2)

6. (a) Give two reasons to justify the use of statistical models. (2)

It has been suggested that there are 7 stages involved in creating a statistical model. They are summarised below, with stages 3, 4 and 7 missing.

Stage 1. The recognition of a real-world problem.

Stage 2. A statistical model is devised.

Stage 3.

Stage 4.

Stage 5. Comparisons are made against the devised model.

Stage 6. Statistical concepts are used to test how well the model describes the real-world problem.

Stage 7.

- (b) Write down the missing stages. (3)

7. The measure of intelligence, IQ, of a group of students is assumed to be Normally distributed with mean 100 and standard deviation 15.

- (a) Find the probability that a student selected at random has an IQ less than 91. (4)

The probability that a randomly selected student has an IQ of at least  $100 + k$  is 0.2090.

- (b) Find, to the nearest integer, the value of  $k$ . (6)

**END**

**TOTAL FOR PAPER: 75 MARKS**

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced/Advanced Subsidiary****Tuesday 5 June 2007 – Afternoon****Time: 1 hour 30 minutes****Materials required for examination**

Mathematical Formulae (Green)

**Items included with question papers**

Nil

Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.

**Instructions to Candidates**

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, initials and signature.

**Information for Candidates**

A booklet 'Mathematical Formulae and Statistical Tables' is provided.  
Full marks may be obtained for answers to ALL questions.  
There are 7 questions in this question paper. The total mark for this paper is 75.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.  
You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. A young family were looking for a new 3 bedroom semi-detached house. A local survey recorded the price  $x$ , in £1000, and the distance  $y$ , in miles, from the station of such houses. The following summary statistics were provided

$$S_{xx} = 113\,573, S_{yy} = 8.657, S_{xy} = -808.917$$

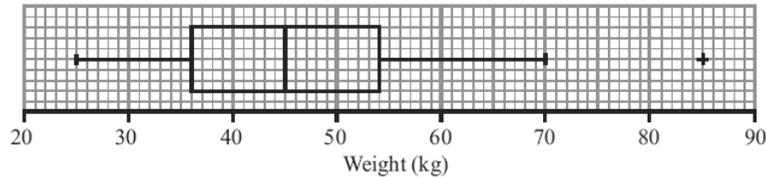
- (a) Use these values to calculate the product moment correlation coefficient. (2)
- (b) Give an interpretation of your answer to part (a). (1)

Another family asked for the distances to be measured in km rather than miles.

- (c) State the value of the product moment correlation coefficient in this case. (1)
-

2. The box plot in Figure 1 shows a summary of the weights of the luggage, in kg, for each musician in an orchestra on an overseas tour.

Figure 1



The airline's recommended weight limit for each musician's luggage was 45 kg.

Given that none of the musician's luggage weighed exactly 45 kg,

- (a) state the proportion of the musicians whose luggage was below the recommended weight limit. (1)

A quarter of the musicians had to pay a charge for taking heavy luggage.

- (b) State the smallest weight for which the charge was made. (1)

- (c) Explain what you understand by the + on the box plot in Figure 1, and suggest an instrument that the owner of this luggage might play. (2)

- (d) Describe the skewness of this distribution. Give a reason for your answer. (2)

One musician of the orchestra suggests that the weights of the luggage, in kg, can be modelled by a normal distribution with quartiles as given in Figure 1.

- (c) Find the standard deviation of this normal distribution. (4)
- 

3. A student is investigating the relationship between the price ( $y$  pence) of 100g of chocolate and the percentage ( $x\%$ ) of the cocoa solids in the chocolate. The following data is obtained

Chocolate brand	A	B	C	D	E	F	G	H
$x$ (% cocoa)	10	20	30	35	40	50	60	70
$y$ (pence)	35	55	40	100	60	90	110	130

(You may use:  $\sum x = 315$ ,  $\sum x^2 = 15\,225$ ,  $\sum y = 620$ ,  $\sum y^2 = 56\,550$ ,  $\sum xy = 28\,750$ )

- (a) Draw a scatter diagram to represent these data. (2)

- (b) Show that  $S_{xy} = 4337.5$  and find  $S_{xx}$ . (3)

The student believes that a linear relationship of the form  $y = a + bx$  could be used to describe these data.

- (c) Use linear regression to find the value of  $a$  and the value of  $b$ , giving your answers to 1 decimal place. (4)

- (d) Draw the regression line on your diagram. (2)

The student believes that one brand of chocolate is overpriced.

- (e) Use the scatter diagram to
- (i) state which brand is overpriced,
  - (ii) suggest a fair price for this brand.
- Give reasons for both your answers. (4)
-

4. A survey of the reading habits of some students revealed that, on a regular basis, 25% read quality newspapers, 45% read tabloid newspapers and 40% do not read newspapers at all.

(a) Find the proportion of students who read both quality and tabloid newspapers. (3)

(b) Draw a Venn diagram to represent this information. (3)

A student is selected at random. Given that this student reads newspapers on a regular basis,

(c) find the probability that this student only reads quality newspapers. (3)

---

5.

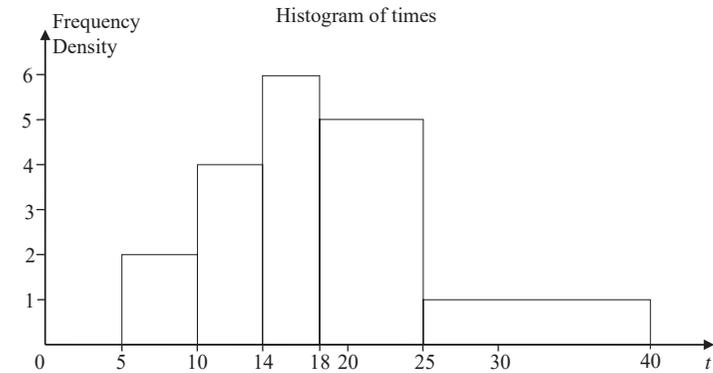


Figure 2

Figure 2 shows a histogram for the variable  $t$  which represents the time taken, in minutes, by a group of people to swim 500 m.

(a) Copy and complete the frequency table for  $t$ .

$t$	5 – 10	10 – 14	14 – 18	18 – 25	25 – 40
Frequency	10	16	24		

(2)

(b) Estimate the number of people who took longer than 20 minutes to swim 500 m.

(2)

(c) Find an estimate of the mean time taken.

(4)

(d) Find an estimate for the standard deviation of  $t$ .

(3)

(e) Find the median and quartiles for  $t$ .

(4)

One measure of skewness is found using  $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$ .

(f) Evaluate this measure and describe the skewness of these data.

(2)

---

6. The random variable  $X$  has a normal distribution with mean 20 and standard deviation 4.

(a) Find  $P(X > 25)$ . (3)

(b) Find the value of  $d$  such that  $P(20 < X < d) = 0.4641$ . (4)

7. The random variable  $X$  has probability distribution

$x$	1	3	5	7	9
$P(X = x)$	0.2	$p$	0.2	$q$	0.15

(a) Given that  $E(X) = 4.5$ , write down two equations involving  $p$  and  $q$ . (3)

Find

(b) the value of  $p$  and the value of  $q$ , (3)

(c)  $P(4 < X \leq 7)$ . (2)

Given that  $E(X^2) = 27.4$ , find

(d)  $\text{Var}(X)$ , (2)

(e)  $E(19 - 4X)$ , (1)

(f)  $\text{Var}(19 - 4X)$ . (2)

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683/01**

**Edexcel GCE**

**Statistics S1**

**Advanced Subsidiary**

**Tuesday 15 January 2008 – Morning**

**Time: 1 hour 30 minutes**

Materials required for examination

Mathematical Formulae (Green)

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

#### Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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 7 questions on this paper. The total mark for this paper is 75.

#### Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. A personnel manager wants to find out if a test carried out during an employee’s interview and a skills assessment at the end of basic training is a guide to performance after working for the company for one year.

The table below shows the results of the interview test of 10 employees and their performance after one year.

Employee	A	B	C	D	E	F	G	H	I	J
Interview test, $x$ %	65	71	79	77	85	78	85	90	81	62
Performance after one year, $y$ %	65	74	82	64	87	78	61	65	79	69

[You may use  $\sum x^2 = 60\,475$ ,  $\sum y^2 = 53\,122$ ,  $\sum xy = 56\,076$ ]

- (a) Showing your working clearly, calculate the product moment correlation coefficient between the interview test and the performance after one year. (5)

The product moment correlation coefficient between the skills assessment and the performance after one year is  $-0.156$  to 3 significant figures.

- (b) Use your answer to part (a) to comment on whether or not the interview test and skills assessment are a guide to the performance after one year. Give clear reasons for your answers. (2)
- 

2. Cotinine is a chemical that is made by the body from nicotine which is found in cigarette smoke. A doctor tested the blood of 12 patients, who claimed to smoke a packet of cigarettes a day, for cotinine. The results, in appropriate units, are shown below.

Patient	A	B	C	D	E	F	G	H	I	J	K	L
Cotinine level, $x$	160	390	169	175	125	420	171	250	210	258	186	243

[You may use  $\sum x^2 = 724\,961$ ]

- (a) Find the mean and standard deviation of the level of cotinine in a patient’s blood. (4)
- (b) Find the median, upper and lower quartiles of these data. (3)

A doctor suspects that some of his patients have been smoking more than a packet of cigarettes per day. He decides to use  $Q_3 + 1.5(Q_3 - Q_1)$  to determine if any of the cotinine results are far enough away from the upper quartile to be outliers.

- (c) Identify which patient(s) may have been smoking more than a packet of cigarettes a day. Show your working clearly. (4)

Research suggests that cotinine levels in the blood form a skewed distribution.

One measure of skewness is found using  $\frac{(Q_1 - 2Q_2 + Q_3)}{(Q_3 - Q_1)}$ .

- (d) Evaluate this measure and describe the skewness of these data. (3)
-

3. The histogram in Figure 1 shows the time taken, to the nearest minute, for 140 runners to complete a fun run.

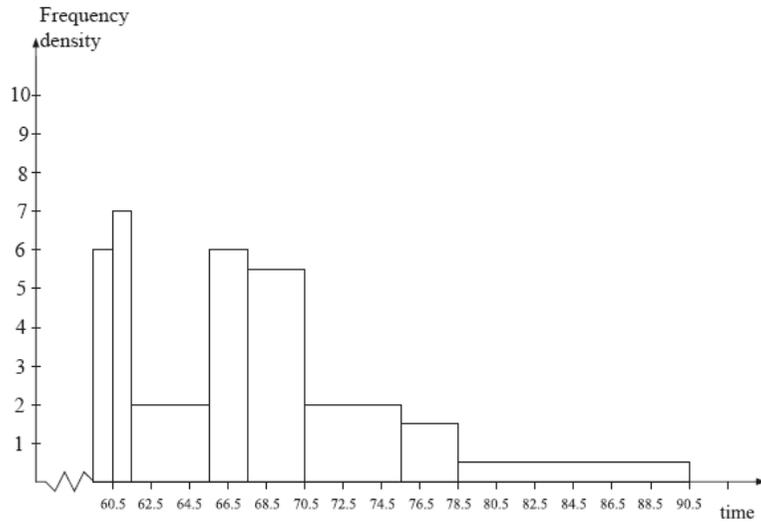


Figure 1

Use the histogram to calculate the number of runners who took between 78.5 and 90.5 minutes to complete the fun run.

(5)

4. A second hand car dealer has 10 cars for sale. She decides to investigate the link between the age of the cars,  $x$  years, and the mileage,  $y$  thousand miles. The data collected from the cars are shown in the table below.

Age, $x$ (years)	2	2.5	3	4	4.5	4.5	5	3	6	6.5
Mileage, $y$ (thousands)	22	34	33	37	40	45	49	30	58	58

[You may assume that  $\sum x = 41$ ,  $\sum y = 406$ ,  $\sum x^2 = 188$ ,  $\sum xy = 1818.5$ ]

- (a) Find  $S_{xx}$  and  $S_{xy}$ . (3)
- (b) Find the equation of the least squares regression line in the form  $y = a + bx$ . Give the values of  $a$  and  $b$  to 2 decimal places. (4)
- (c) Give a practical interpretation of the slope  $b$ . (1)
- (d) Using your answer to part (b), find the mileage predicted by the regression line for a 5 year old car. (2)

5. The following shows the results of a wine tasting survey of 100 people.

- 96 like wine A,
- 93 like wine B,
- 96 like wine C,
- 92 like A and B,
- 91 like B and C,
- 93 like A and C,
- 90 like all three wines.

(a) Draw a Venn Diagram to represent these data.

(6)

Find the probability that a randomly selected person from the survey likes

- (b) none of the three wines,
- (c) wine A but not wine B,
- (d) any wine in the survey except wine C,
- (e) exactly two of the three kinds of wine.

(1)

(2)

(2)

(2)

Given that a person from the survey likes wine A,

(f) find the probability that the person likes wine C.

(3)

6. The weights of bags of popcorn are normally distributed with mean of 200 g and 60% of all bags weighing between 190 g and 210 g.

(a) Write down the median weight of the bags of popcorn.

(1)

(b) Find the standard deviation of the weights of the bags of popcorn.

(5)

A shopkeeper finds that customers will complain if their bag of popcorn weighs less than 180 g.

(c) Find the probability that a customer will complain.

(3)

7. Tetrahedral dice have four faces. Two fair tetrahedral dice, one red and one blue, have faces numbered 0, 1, 2, and 3 respectively. The dice are rolled and the numbers face down on the two dice are recorded. The random variable  $R$  is the score on the red die and the random variable  $B$  is the score on the blue die.

(a) Find  $P(R = 3 \text{ and } B = 0)$ .

(2)

The random variable  $T$  is  $R$  multiplied by  $B$ .

(b) Complete the diagram below to represent the sample space that shows all the possible values of  $T$ .

<b>3</b>					
<b>2</b>		2			
<b>1</b>	0				
<b>0</b>					
<b>B</b>	<b>R</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>

Sample space diagram of  $T$

(3)

The table below represents the probability distribution of the random variable  $T$ .

$t$	0	1	2	3	4	6	9
$P(T = t)$	$a$	$b$	$\frac{1}{8}$	$\frac{1}{8}$	$c$	$\frac{1}{8}$	$d$

(c) Find the values of  $a, b, c$  and  $d$ .

(3)

Find the values of

(d)  $E(T)$ ,

(2)

(e)  $\text{Var}(T)$ .

(4)

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced/Advanced Subsidiary****Thursday 15 May 2008 – Morning****Time: 1 hour 30 minutes****Materials required for examination**

Mathematical Formulae (Green)

**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, differentiation and integration, or have retrievable mathematical formulae stored in them.

**Instructions to Candidates**

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, initials and signature. 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. There are 7 questions in this question paper. The total mark for this paper is 75.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled. You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

**H32582A** This publication may only be reproduced in accordance with Edexcel Limited copyright policy. ©2008 Edexcel Limited.

1. A disease is known to be present in 2% of a population. A test is developed to help determine whether or not someone has the disease.

Given that a person has the disease, the test is positive with probability 0.95.

Given that a person does not have the disease, the test is positive with probability 0.03.

- (a) Draw a tree diagram to represent this information. (3)

A person is selected at random from the population and tested for this disease.

- (b) Find the probability that the test is positive. (3)

A doctor randomly selects a person from the population and tests him for the disease. Given that the test is positive,

- (c) find the probability that he does not have the disease. (2)

- (d) Comment on the usefulness of this test. (1)
-

2. The age in years of the residents of two hotels are shown in the back to back stem and leaf diagram below.

Abbey Hotel	8   5   0 means 58 years in Abbey Hotel and 50 years in Balmoral Hotel	Balmoral Hotel
(1)	2   0	
(4)	9 7 5 1   1	
(4)	9 8 3 1   2 6	(1)
(11)	9 9 9 9 7 6 6 5 3 3 2   3 4 4 7	(3)
(6)	9 8 7 7 5 0   4 0 0 5 5 6 9	(6)
(1)	8   5 0 0 0 0 1 3 6 6 7	(9)
	6   2 3 3 4 5 7	(6)
	7   0 1 5	(3)

For the Balmoral Hotel,

- (a) write down the mode of the age of the residents, (1)
- (b) find the values of the lower quartile, the median and the upper quartile. (3)
- (c) (i) Find the mean,  $\bar{x}$ , of the age of the residents.
- (ii) Given that  $\sum x^2 = 81\,213$ , find the standard deviation of the age of the residents. (4)

One measure of skewness is found using

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

- (d) Evaluate this measure for the Balmoral Hotel. (2)

For the Abbey Hotel, the mode is 39, the mean is 33.2, the standard deviation is 12.7 and the measure of skewness is  $-0.454$ .

- (e) Compare the two age distributions of the residents of each hotel. (3)
- 

3. The random variable  $X$  has probability distribution given in the table below.

$x$	-1	0	1	2	3
$P(X = x)$	$p$	$q$	0.2	0.15	0.15

Given that  $E(X) = 0.55$ , find

- (a) the value of  $p$  and the value of  $q$ , (5)
- (b)  $\text{Var}(X)$ , (4)
- (c)  $E(2X - 4)$ . (2)
- 

4. Crickets make a noise. The pitch,  $v$  kHz, of the noise made by a cricket was recorded at 15 different temperatures,  $t$  °C. These data are summarised below.

$$\sum t^2 = 10\,922.81, \quad \sum v^2 = 42.3356, \quad \sum tv = 677.971, \quad \sum t = 401.3, \quad \sum v = 25.08$$

- (a) Find  $S_{tt}$ ,  $S_{vv}$  and  $S_{tv}$  for these data. (4)
- (b) Find the product moment correlation coefficient between  $t$  and  $v$ . (3)
- (c) State, with a reason, which variable is the explanatory variable. (2)
- (d) Give a reason to support fitting a regression model of the form  $v = a + bt$  to these data. (1)
- (e) Find the value of  $a$  and the value of  $b$ . Give your answers to 3 significant figures. (4)
- (f) Using this model, predict the pitch of the noise at 19 °C. (1)
-

5. A person's blood group is determined by whether or not it contains any of 3 substances  $A$ ,  $B$  and  $C$ .

A doctor surveyed 300 patients' blood and produced the table below.

Blood contains	No. of Patients
only $C$	100
$A$ and $C$ but not $B$	100
only $A$	30
$B$ and $C$ but not $A$	25
only $B$	12
$A$ , $B$ and $C$	10
$A$ and $B$ but not $C$	3

- (a) Draw a Venn diagram to represent this information. (4)
- (b) Find the probability that a randomly chosen patient's blood contains substance  $C$ . (2)

Harry is one of the patients. Given that his blood contains substance  $A$ ,

- (c) find the probability that his blood contains all 3 substances. (2)

Patients whose blood contains none of these substances are called universal blood donors.

- (d) Find the probability that a randomly chosen patient is a universal blood donor. (2)
- 

6. The discrete random variable  $X$  can take only the values 2, 3 or 4. For these values the cumulative distribution function is defined by

$$F(x) = \frac{(x+k)^2}{25} \text{ for } x = 2, 3, 4,$$

where  $k$  is a positive integer.

- (a) Find  $k$ . (2)
- (b) Find the probability distribution of  $X$ . (3)
- 

7. A packing plant fills bags with cement. The weight  $X$  kg of a bag of cement can be modelled by a normal distribution with mean 50 kg and standard deviation 2 kg.

- (a) Find  $P(X > 53)$ . (3)
- (b) Find the weight that is exceeded by 99% of the bags. (5)

Three bags are selected at random.

- (c) Find the probability that two weigh more than 53 kg and one weighs less than 53 kg. (4)
- 

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced Subsidiary****Monday 19 January 2009 – Afternoon****Time: 1 hour 30 minutes****Materials required for examination**

Mathematical Formulae (Green)

**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, differentiation and integration, or have retrievable mathematical formulae stored in them.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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 on this paper. The total mark for this paper is 75.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

**N32680A**

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.  
©2009 Edexcel Limited.

1. A teacher is monitoring the progress of students using a computer based revision course. The improvement in performance,  $y$  marks, is recorded for each student along with the time,  $x$  hours, that the student spent using the revision course. The results for a random sample of 10 students are recorded below.

$x$ hours	1.0	3.5	4.0	1.5	1.3	0.5	1.8	2.5	2.3	3.0
$y$ marks	5	30	27	10	-3	-5	7	15	-10	20

[You may use  $\sum x = 21.4$ ,  $\sum y = 96$ ,  $\sum x^2 = 57.22$ ,  $\sum xy = 313.7$ ]

- (a) Calculate  $S_{xx}$  and  $S_{xy}$ . (3)
- (b) Find the equation of the least squares regression line of  $y$  on  $x$  in the form  $y = a + bx$ . (4)
- (c) Give an interpretation of the gradient of your regression line. (1)

Rosemary spends 3.3 hours using the revision course.

- (d) Predict her improvement in marks. (2)
- Lee spends 8 hours using the revision course claiming that this should give him an improvement in performance of over 60 marks.
- (e) Comment on Lee's claim. (1)
-

2. A group of office workers were questioned for a health magazine and  $\frac{2}{5}$  were found to take regular exercise. When questioned about their eating habits  $\frac{2}{3}$  said they always eat breakfast and, of those who always eat breakfast  $\frac{9}{25}$  also took regular exercise.

Find the probability that a randomly selected member of the group

- (a) always eats breakfast and takes regular exercise, (2)
- (b) does not always eat breakfast and does not take regular exercise. (4)
- (c) Determine, giving your reason, whether or not always eating breakfast and taking regular exercise are statistically independent. (2)
- 

3. When Rohit plays a game, the number of points he receives is given by the discrete random variable  $X$  with the following probability distribution.

$x$	0	1	2	3
$P(X=x)$	0.4	0.3	0.2	0.1

- (a) Find  $E(X)$ . (2)
- (b) Find  $F(1.5)$ . (2)
- (c) Show that  $\text{Var}(X) = 1$ . (4)
- (d) Find  $\text{Var}(5 - 3X)$ . (2)

Rohit can win a prize if the total number of points he has scored after 5 games is at least 10. After 3 games he has a total of 6 points. You may assume that games are independent.

- (e) Find the probability that Rohit wins the prize. (6)
- 

4. In a study of how students use their mobile telephones, the phone usage of a random sample of 11 students was examined for a particular week.

The total length of calls,  $y$  minutes, for the 11 students were

17, 23, 35, 36, 51, 53, 54, 55, 60, 77, 110

- (a) Find the median and quartiles for these data. (3)

A value that is greater than  $Q_3 + 1.5 \times (Q_3 - Q_1)$  or smaller than  $Q_1 - 1.5 \times (Q_3 - Q_1)$  is defined as an outlier.

- (b) Show that 110 is the only outlier. (2)
- (c) Draw a box plot for these data indicating clearly the position of the outlier. (3)

The value of 110 is omitted.

- (d) Show that  $S_{yy}$  for the remaining 10 students is 2966.9 (3)

These 10 students were each asked how many text messages,  $x$ , they sent in the same week. The values of  $S_{xx}$  and  $S_{xy}$  for these 10 students are  $S_{xx} = 3463.6$  and  $S_{xy} = -18.3$ .

- (e) Calculate the product moment correlation coefficient between the number of text messages sent and the total length of calls for these 10 students. (2)

A parent believes that a student who sends a large number of text messages will spend fewer minutes on calls.

- (f) Comment on this belief in the light of your calculation in part (e). (1)
-

5. In a shopping survey a random sample of 104 teenagers were asked how many hours, to the nearest hour, they spent shopping in the last month. The results are summarised in the table below.

Number of hours	Mid-point	Frequency
0 – 5	2.75	20
6 – 7	6.5	16
8 – 10	9	18
11 – 15	13	25
16 – 25	20.5	15
26 – 50	38	10

A histogram was drawn and the group (8 – 10) hours was represented by a rectangle that was 1.5 cm wide and 3 cm high.

- (a) Calculate the width and height of the rectangle representing the group (16 – 25) hours. (3)
- (b) Use linear interpolation to estimate the median and interquartile range. (5)
- (c) Estimate the mean and standard deviation of the number of hours spent shopping. (4)
- (d) State, giving a reason, the skewness of these data. (2)
- (e) State, giving a reason, which average and measure of dispersion you would recommend to use to summarise these data. (2)
- 
6. The random variable  $X$  has a normal distribution with mean 30 and standard deviation 5.
- (a) Find  $P(X < 39)$ . (2)
- (b) Find the value of  $d$  such that  $P(X < d) = 0.1151$ . (4)
- (c) Find the value of  $e$  such that  $P(X > e) = 0.1151$ . (2)
- (d) Find  $P(d < X < e)$ . (2)

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683/01**

**Edexcel GCE**

**Statistics S1**

**Advanced/Advanced Subsidiary**

**Wednesday 20 May 2009 – Afternoon**

**Time: 1 hour 30 minutes**

**Materials required for examination**  
Mathematical Formulae (Orange Green)

**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, differentiation and integration, or have retrievable mathematical formulae stored in them.

#### Instructions to Candidates

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, initials and signature. Answer ALL questions. 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. There are 8 questions in this question paper. The total mark for this paper is 75.

#### Advice to Candidates

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

**H34279A** This publication may only be reproduced in accordance with Edexcel Limited copyright policy. ©2009 Edexcel Limited.

1. The volume of a sample of gas is kept constant. The gas is heated and the pressure,  $p$ , is measured at 10 different temperatures,  $t$ . The results are summarised below.

$$\Sigma p = 445 \quad \Sigma p^2 = 38\,125 \quad \Sigma t = 240 \quad \Sigma t^2 = 27\,520 \quad \Sigma pt = 26\,830$$

- (a) Find  $S_{pp}$  and  $S_{pt}$ . (3)

Given that  $S_{tt} = 21\,760$ ,

- (b) calculate the product moment correlation coefficient. (2)
- (c) Give an interpretation of your answer to part (b). (1)
- 

2. On a randomly chosen day the probability that Bill travels to school by car, by bicycle or on foot is  $\frac{1}{2}$ ,  $\frac{1}{6}$  and  $\frac{1}{3}$  respectively. The probability of being late when using these methods of travel is  $\frac{1}{5}$ ,  $\frac{2}{5}$  and  $\frac{1}{10}$  respectively.

- (a) Draw a tree diagram to represent this information. (3)
- (b) Find the probability that on a randomly chosen day
- (i) Bill travels by foot and is late,
- (ii) Bill is not late. (4)
- (c) Given that Bill is late, find the probability that he did not travel on foot. (4)
- 

3. The variable  $x$  was measured to the nearest whole number. Forty observations are given in the table below.

$x$	10 – 15	16 – 18	19 –
Frequency	15	9	16

A histogram was drawn and the bar representing the 10 – 15 class has a width of 2 cm and a height of 5 cm. For the 16 – 18 class find

- (a) the width, (1)
- (b) the height (2)

of the bar representing this class.

---

4. A researcher measured the foot lengths of a random sample of 120 ten-year-old children. The lengths are summarised in the table below.

Foot length, $l$ , (cm)	Number of children
$10 \leq l < 12$	5
$12 \leq l < 17$	53
$17 \leq l < 19$	29
$19 \leq l < 21$	15
$21 \leq l < 23$	11
$23 \leq l < 25$	7

- (a) Use interpolation to estimate the median of this distribution. (2)
- (b) Calculate estimates for the mean and the standard deviation of these data. (6)

One measure of skewness is given by

$$\text{Coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

- (c) Evaluate this coefficient and comment on the skewness of these data. (3)

Greg suggests that a normal distribution is a suitable model for the foot lengths of ten-year-old children.

- (d) Using the value found in part (c), comment on Greg's suggestion, giving a reason for your answer. (2)
- 

5. The weight,  $w$  grams, and the length,  $l$  mm, of 10 randomly selected newborn turtles are given in the table below.

$l$	49.0	52.0	53.0	54.5	54.1	53.4	50.0	51.6	49.5	51.2
$w$	29	32	34	39	38	35	30	31	29	30

(You may use  $S_{ll} = 33.381$   $S_{wl} = 59.99$   $S_{ww} = 120.1$ )

- (a) Find the equation of the regression line of  $w$  on  $l$  in the form  $w = a + bl$ . (5)
- (b) Use your regression line to estimate the weight of a newborn turtle of length 60 mm. (2)
- (c) Comment on the reliability of your estimate giving a reason for your answer. (2)
- 
6. The discrete random variable  $X$  has probability function

$$P(X = x) = \begin{cases} a(3-x) & x = 0, 1, 2 \\ b & x = 3 \end{cases}$$

- (a) Find  $P(X = 2)$  and copy and complete the table below.

$x$	0	1	2	3
$P(X=x)$	$3a$	$2a$		$b$

- (1)
- Given that  $E(X) = 1.6$ ,
- (b) find the value of  $a$  and the value of  $b$ . (5)

Find

- (c)  $P(0.5 < X < 3)$ , (2)
- (d)  $E(3X - 2)$ . (2)
- (e) Show that the  $\text{Var}(X) = 1.64$  (3)
- (f) Calculate  $\text{Var}(3X - 2)$ . (2)
-

7. (a) Given that  $P(A) = a$  and  $P(B) = b$  express  $P(A \cup B)$  in terms of  $a$  and  $b$  when

(i)  $A$  and  $B$  are mutually exclusive,

(ii)  $A$  and  $B$  are independent.

(2)

Two events  $R$  and  $Q$  are such that

$$P(R \cap Q) = 0.15, \quad P(Q) = 0.35 \quad \text{and} \quad P(R | Q) = 0.1$$

Find the value of

(b)  $P(R \cup Q)$ ,

(1)

(c)  $P(R \cap Q)$ ,

(2)

(d)  $P(R)$ .

(2)

8. The lifetimes of bulbs used in a lamp are normally distributed.

A company  $X$  sells bulbs with a mean lifetime of 850 hours and a standard deviation of 50 hours.

(a) Find the probability of a bulb, from company  $X$ , having a lifetime of less than 830 hours.

(3)

(b) In a box of 500 bulbs, from company  $X$ , find the expected number having a lifetime of less than 830 hours.

(2)

A rival company  $Y$  sells bulbs with a mean lifetime of 860 hours and 20% of these bulbs have a lifetime of less than 818 hours.

(c) Find the standard deviation of the lifetimes of bulbs from company  $Y$ .

(4)

Both companies sell the bulbs for the same price.

(d) State which company you would recommend. Give reasons for your answer.

(2)

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683/01**

**Edexcel GCE**

**Statistics S1**

**Advanced Subsidiary**

**Wednesday 13 January 2010 – Afternoon**

**Time: 1 hour 30 minutes**

Materials required for examination

Mathematical Formulae (Pink or Green)

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, differentiation and integration, or have retrievable mathematical formulae stored in them.**

#### **Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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 7 questions on this paper. The total mark for this paper is 75.

#### **Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

1. A jar contains 2 red, 1 blue and 1 green bead. Two beads are drawn at random from the jar without replacement.
- (a) Draw a tree diagram to illustrate all the possible outcomes and associated probabilities. State your probabilities clearly. (3)
- (b) Find the probability that a blue bead and a green bead are drawn from the jar. (2)
- 

2. The 19 employees of a company take an aptitude test. The scores out of 40 are illustrated in the stem and leaf diagram below.

	2   6 means a score of 26	
0	7	(1)
1	88	(2)
2	4468	(4)
3	2333459	(7)
4	00000	(5)

Find

- (a) the median score, (1)
- (b) the interquartile range. (3)

The company director decides that any employees whose scores are so low that they are outliers will undergo retraining.

An outlier is an observation whose value is less than the lower quartile minus 1.0 times the interquartile range.

- (c) Explain why there is only one employee who will undergo retraining. (2)
- (d) Draw a box plot to illustrate the employees' scores. (3)
- 

3. The birth weights, in kg, of 1500 babies are summarised in the table below.

Weight (kg)	Midpoint, $x$ kg	Frequency, $f$
0.0 – 1.0	0.50	1
1.0 – 2.0	1.50	6
2.0 – 2.5	2.25	60
2.5 – 3.0		280
3.0 – 3.5	3.25	820
3.5 – 4.0	3.75	320
4.0 – 5.0	4.50	10
5.0 – 6.0		3

[You may use  $\sum fx = 4841$  and  $\sum fx^2 = 15\,889.5$ ]

- (a) Write down the missing midpoints in the table above. (2)
- (b) Calculate an estimate of the mean birth weight. (2)
- (c) Calculate an estimate of the standard deviation of the birth weight. (3)
- (d) Use interpolation to estimate the median birth weight. (2)
- (e) Describe the skewness of the distribution. Give a reason for your answer. (2)
-

4. There are 180 students at a college following a general course in computing. Students on this course can choose to take up to three extra options.

112 take systems support,  
 70 take developing software,  
 81 take networking,  
 35 take developing software and systems support,  
 28 take networking and developing software,  
 40 take systems support and networking,  
 4 take all three extra options.

- (a) Draw a Venn diagram to represent this information. (5)

A student from the course is chosen at random.

Find the probability that the student takes

- (b) none of the three extra options, (1)

- (c) networking only. (1)

Students who want to become technicians take systems support and networking. Given that a randomly chosen student wants to become a technician,

- (d) find the probability that this student takes all three extra options. (2)
- 

5. The probability function of a discrete random variable  $X$  is given by

$$p(x) = kx^2, \quad x = 1, 2, 3.$$

where  $k$  is a positive constant.

- (a) Show that  $k = \frac{1}{14}$ . (2)

Find

- (b)  $P(X \geq 2)$ , (2)

- (c)  $E(X)$ , (2)

- (d)  $\text{Var}(1 - X)$ . (4)
- 

6. The blood pressures,  $p$  mmHg, and the ages,  $t$  years, of 7 hospital patients are shown in the table below.

Patient	A	B	C	D	E	F	G
$t$	42	74	48	35	56	26	60
$P$	98	130	120	88	182	80	135

$$[\sum t = 341, \sum p = 833, \sum t^2 = 18181, \sum p^2 = 106397, \sum tp = 42948]$$

- (a) Find  $S_{pp}$ ,  $S_{tp}$  and  $S_{tt}$  for these data. (4)

- (b) Calculate the product moment correlation coefficient for these data. (3)

- (c) Interpret the correlation coefficient. (1)

- (d) Draw the scatter diagram of blood pressure against age for these 7 patients. (2)

- (e) Find the equation of the regression line of  $p$  on  $t$ . (4)

- (f) Plot your regression line on your scatter diagram. (2)

- (g) Use your regression line to estimate the blood pressure of a 40 year old patient. (2)
- 

7. The heights of a population of women are normally distributed with mean  $\mu$  cm and standard deviation  $\sigma$  cm. It is known that 30% of the women are taller than 172 cm and 5% are shorter than 154 cm.

- (a) Sketch a diagram to show the distribution of heights represented by this information. (3)

- (b) Show that  $\mu = 154 + 1.6449\sigma$ . (3)

- (c) Obtain a second equation and hence find the value of  $\mu$  and the value of  $\sigma$ . (4)

A woman is chosen at random from the population.

- (d) Find the probability that she is taller than 160 cm. (3)
- 

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced Subsidiary****Thursday 27 May 2010 – Morning****Time: 1 hour 30 minutes****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, differentiation and integration, or have retrievable mathematical formulas stored in them.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has 7 questions.

The total mark for this paper is 75.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

**H35395A**

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.  
©2010 Edexcel Limited

1. Gary compared the total attendance,  $x$ , at home matches and the total number of goals,  $y$ , scored at home during a season for each of 12 football teams playing in a league. He correctly calculated:

$$S_{xx} = 1022500, \quad S_{yy} = 130.9, \quad S_{xy} = 8825.$$

- (a) Calculate the product moment correlation coefficient for these data. (2)
- (b) Interpret the value of the correlation coefficient. (1)

Helen was given the same data to analyse. In view of the large numbers involved she decided to divide the attendance figures by 100. She then calculated the product moment correlation coefficient between  $\frac{x}{100}$  and  $y$ .

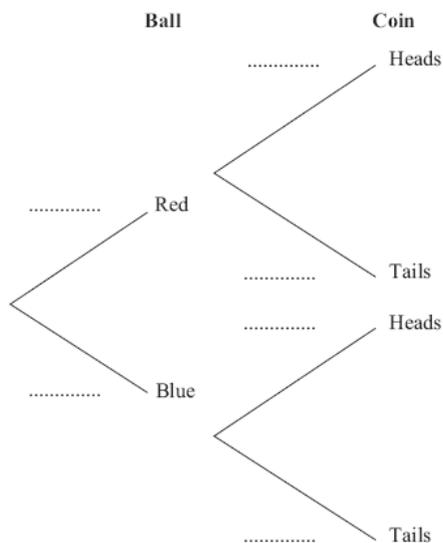
- (c) Write down the value Helen should have obtained. (1)
-

2. An experiment consists of selecting a ball from a bag and spinning a coin. The bag contains 5 red balls and 7 blue balls. A ball is selected at random from the bag, its colour is noted and then the ball is returned to the bag.

When a red ball is selected, a biased coin with probability  $\frac{2}{3}$  of landing heads is spun.

When a blue ball is selected a fair coin is spun.

- (a) Copy and complete the tree diagram below to show the possible outcomes and associated probabilities.



Shivani selects a ball and spins the appropriate coin.

- (b) Find the probability that she obtains a head.

(2)

Given that Tom selected a ball at random and obtained a head when he spun the appropriate coin,

- (c) find the probability that Tom selected a red ball.

(3)

Shivani and Tom each repeat this experiment.

- (d) Find the probability that the colour of the ball Shivani selects is the same as the colour of the ball Tom selects.

(3)

3. The discrete random variable  $X$  has probability distribution given by

$x$	-1	0	1	2	3
$P(X = x)$	$\frac{1}{5}$	$a$	$\frac{1}{10}$	$a$	$\frac{1}{5}$

where  $a$  is a constant.

- (a) Find the value of  $a$ .

(2)

- (b) Write down  $E(X)$ .

(1)

- (c) Find  $\text{Var}(X)$ .

(3)

The random variable  $Y = 6 - 2X$ .

- (d) Find  $\text{Var}(Y)$ .

(2)

- (e) Calculate  $P(X \geq Y)$ .

(3)

4. The Venn diagram in Figure 1 shows the number of students in a class who read any of 3 popular magazines A, B and C.

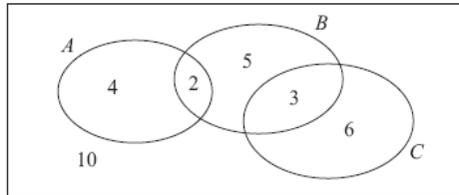


Figure 1

One of these students is selected at random.

- (a) Show that the probability that the student reads more than one magazine is  $\frac{1}{6}$ . (2)
- (b) Find the probability that the student reads A or B (or both). (2)
- (c) Write down the probability that the student reads both A and C. (1)

Given that the student reads at least one of the magazines,

- (d) find the probability that the student reads C. (2)
- (e) Determine whether or not reading magazine B and reading magazine C are statistically independent. (3)

5. A teacher selects a random sample of 56 students and records, to the nearest hour, the time spent watching television in a particular week.

Hours	1–10	11–20	21–25	26–30	31–40	41–59
Frequency	6	15	11	13	8	3
Mid-point	5.5	15.5		28		50

- (a) Find the mid-points of the 21–25 hour and 31–40 hour groups. (2)

A histogram was drawn to represent these data. The 11–20 group was represented by a bar of width 4 cm and height 6 cm.

- (b) Find the width and height of the 26–30 group. (3)
- (c) Estimate the mean and standard deviation of the time spent watching television by these students. (5)
- (d) Use linear interpolation to estimate the median length of time spent watching television by these students. (2)

The teacher estimated the lower quartile and the upper quartile of the time spent watching television to be 15.8 and 29.3 respectively.

- (e) State, giving a reason, the skewness of these data. (2)

6. A travel agent sells flights to different destinations from *Beerow* airport. The distance  $d$ , measured in 100 km, of the destination from the airport and the fare £ $f$  are recorded for a random sample of 6 destinations.

Destination	A	B	C	D	E	F
$d$	2.2	4.0	6.0	2.5	8.0	5.0
$f$	18	20	25	23	32	28

[You may use  $\sum d^2 = 152.09$      $\sum f^2 = 3686$      $\sum fd = 723.1$ ]

- (a) On graph paper, draw a scatter diagram to illustrate this information. (2)
- (b) Explain why a linear regression model may be appropriate to describe the relationship between  $f$  and  $d$ . (1)
- (c) Calculate  $S_{dd}$  and  $S_{fd}$ . (4)
- (d) Calculate the equation of the regression line of  $f$  on  $d$  giving your answer in the form  $f = a + bd$ . (4)
- (e) Give an interpretation of the value of  $b$ . (1)
- Jane is planning her holiday and wishes to fly from *Beerow* airport to a destination  $t$  km away. A rival travel agent charges 5p per km.
- (f) Find the range of values of  $t$  for which the first travel agent is cheaper than the rival. (2)
- 

7. The distances travelled to work,  $D$  km, by the employees at a large company are normally distributed with  $D \sim N(30, 8^2)$ .

- (a) Find the probability that a randomly selected employee has a journey to work of more than 20 km. (3)
- (b) Find the upper quartile,  $Q_3$ , of  $D$ . (3)
- (c) Write down the lower quartile,  $Q_1$ , of  $D$ . (1)

An outlier is defined as any value of  $D$  such that  $D < h$  or  $D > k$  where

$$h = Q_1 - 1.5 \times (Q_3 - Q_1) \text{ and } k = Q_3 + 1.5 \times (Q_3 - Q_1).$$

- (d) Find the value of  $h$  and the value of  $k$ . (2)

An employee is selected at random.

- (e) Find the probability that the distance travelled to work by this employee is an outlier. (3)

---

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced Level****Friday 14 January 2011 – Morning****Time: 1 hour 30 minutes**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, differentiation and integration, or have retrievable mathematical formulas stored in them.

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S2), the paper reference (6684), your surname, other name and signature.

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.

This paper has 8 questions.

The total mark for this paper is 75.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

**H35410A**

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.  
©2011 Edexcel Limited

1. A random sample of 50 salmon was caught by a scientist. He recorded the length  $l$  cm and weight  $w$  kg of each salmon.

The following summary statistics were calculated from these data.

$$\sum l = 4027 \quad \sum l^2 = 327\,754.5 \quad \sum w = 357.1 \quad \sum lw = 29\,330.5 \quad S_{ww} = 289.6$$

- (a) Find  $S_{ll}$  and  $S_{lw}$ . (3)
- (b) Calculate, to 3 significant figures, the product moment correlation coefficient between  $l$  and  $w$ . (2)
- (c) Give an interpretation of your coefficient. (1)
- 

2. Keith records the amount of rainfall, in mm, at his school, each day for a week. The results are given below.

2.8    5.6    2.3    9.4    0.0    0.5    1.8

Jenny then records the amount of rainfall,  $x$  mm, at the school each day for the following 21 days. The results for the 21 days are summarised below.

$$\sum x = 84.6$$

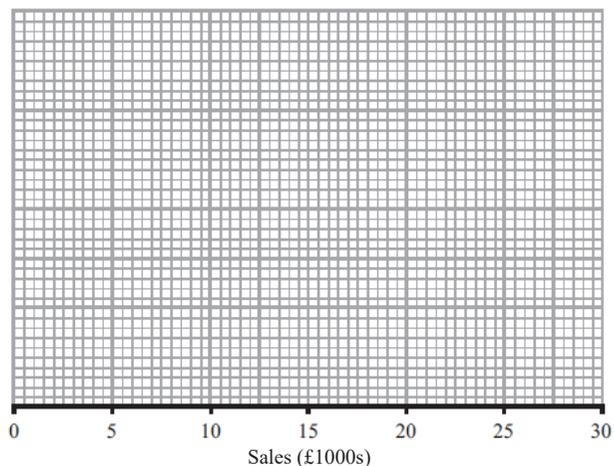
- (a) Calculate the mean amount of rainfall during the whole 28 days. (2)
- Keith realises that he has transposed two of his figures. The number 9.4 should have been 4.9 and the number 0.5 should have been 5.0.
- Keith corrects these figures.
- (b) State, giving your reason, the effect this will have on the mean. (2)
-

3. Over a long period of time a small company recorded the amount it received in sales per month. The results are summarised below.

	Amount received in sales (£1000s)
Two lowest values	3, 4
Lower quartile	7
Median	12
Upper quartile	14
Two highest values	20, 25

An outlier is an observation that falls either  $1.5 \times$  interquartile range above the upper quartile or  $1.5 \times$  interquartile range below the lower quartile.

- (a) On the graph paper below, draw a box plot to represent these data, indicating clearly any outliers. (5)



- (b) State the skewness of the distribution of the amount of sales received. Justify your answer. (2)
- (c) The company claims that for 75 % of the months, the amount received per month is greater than £10 000. Comment on this claim, giving a reason for your answer. (2)

4. A farmer collected data on the annual rainfall,  $x$  cm, and the annual yield of peas,  $p$  tonnes per acre.

The data for annual rainfall was coded using  $v = \frac{x-5}{10}$  and the following statistics were found.

$$S_{vv} = 5.753 \quad S_{pv} = 1.688 \quad S_{pp} = 1.168 \quad \bar{p} = 3.22 \quad \bar{v} = 4.42$$

- (a) Find the equation of the regression line of  $p$  on  $v$  in the form  $p = a + bv$ . (4)
- (b) Using your regression line estimate the annual yield of peas per acre when the annual rainfall is 85 cm. (2)

5. On a randomly chosen day, each of the 32 students in a class recorded the time,  $t$  minutes to the nearest minute, they spent on their homework. The data for the class is summarised in the following table.

Time, $t$	Number of students
10 – 19	2
20 – 29	4
30 – 39	8
40 – 49	11
50 – 69	5
70 – 79	2

- (a) Use interpolation to estimate the value of the median. (2)

Given that

$$\sum t = 1414 \quad \text{and} \quad \sum t^2 = 69\,378,$$

- (b) find the mean and the standard deviation of the times spent by the students on their homework. (3)
- (c) Comment on the skewness of the distribution of the times spent by the students on their homework. Give a reason for your answer. (2)

6. The discrete random variable  $X$  has the probability distribution

$x$	1	2	3	4
$P(X = x)$	$k$	$2k$	$3k$	$4k$

(a) Show that  $k = 0.1$  (1)

Find

(b)  $E(X)$  (2)

(c)  $E(X^2)$  (2)

(d)  $\text{Var}(2 - 5X)$  (3)

Two independent observations  $X_1$  and  $X_2$  are made of  $X$ .

(e) Show that  $P(X_1 + X_2 = 4) = 0.1$  (2)

(f) Complete the probability distribution table for  $X_1 + X_2$ . (2)

$y$	2	3	4	5	6	7	8
$P(X_1 + X_2 = y)$	0.01	0.04	0.10		0.25	0.24	

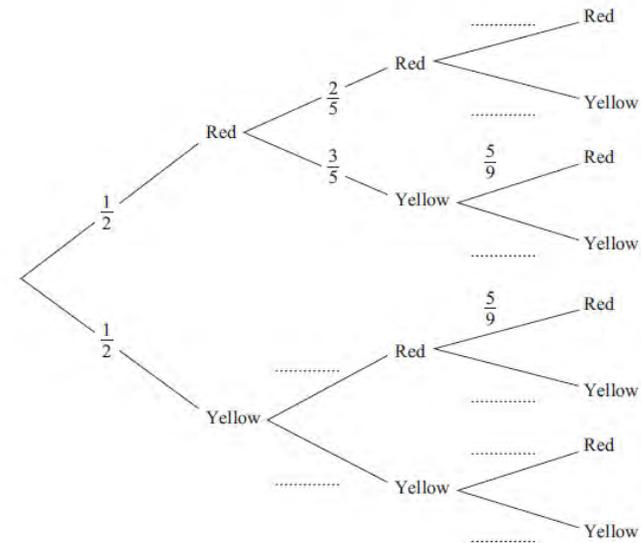
(g) Find  $P(1.5 < X_1 + X_2 \leq 3.5)$  (2)

---

7. The bag  $P$  contains 6 balls of which 3 are red and 3 are yellow.  
 The bag  $Q$  contains 7 balls of which 4 are red and 3 are yellow.  
 A ball is drawn at random from bag  $P$  and placed in bag  $Q$ . A second ball is drawn at random from bag  $P$  and placed in bag  $Q$ . A third ball is then drawn at random from the 9 balls in bag  $Q$ .

The event  $A$  occurs when the 2 balls drawn from bag  $P$  are of the same colour.  
 The event  $B$  occurs when the ball drawn from bag  $Q$  is red.

(a) Copy and complete the tree diagram shown below.



(b) Find  $P(A)$ . (3)

(c) Show that  $P(B) = \frac{5}{9}$ . (3)

(d) Show that  $P(A \cap B) = \frac{2}{9}$ . (2)

(e) Hence find  $P(A \cup B)$ . (2)

(f) Given that all three balls drawn are the same colour, find the probability that they are all red. (3)

---

8. The weight,  $X$  grams, of soup put in a tin by machine  $A$  is normally distributed with a mean of 160 g and a standard deviation of 5 g.

A tin is selected at random.

- (a) Find the probability that this tin contains more than 168 g. (3)

The weight stated on the tin is  $w$  grams.

- (b) Find  $w$  such that  $P(X < w) = 0.01$ . (3)

The weight,  $Y$  grams, of soup put into a carton by machine  $B$  is normally distributed with mean  $\mu$  grams and standard deviation  $\sigma$  grams.

- (c) Given that  $P(Y < 160) = 0.99$  and  $P(Y > 152) = 0.90$ , find the value of  $\mu$  and the value of  $\sigma$ . (6)

END

TOTAL FOR PAPER: 75 MARKS

Paper Reference(s)

6683/01

**Edexcel GCE**

**Statistics S1**

**Advanced Level**

**Friday 20 May 2011 – Afternoon**

**Time: 1 hour 30 minutes**

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, differentiation and integration, or have retrievable mathematical formulas stored in them.

#### Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has 8 questions.

The total mark for this paper is 75.

#### Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

**P38164A**

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.  
©2011 Edexcel Limited

1. On a particular day the height above sea level,  $x$  metres, and the mid-day temperature,  $y$  °C, were recorded in 8 north European towns. These data are summarised below

$$S_{xx} = 3\,535\,237.5 \quad \sum y = 181 \quad \sum y^2 = 4305 \quad S_{xy} = -23\,726.25$$

- (a) Find  $S_{yy}$ . (2)
- (b) Calculate, to 3 significant figures, the product moment correlation coefficient for these data. (2)
- (c) Give an interpretation of your coefficient. (1)

A student thought that the calculations would be simpler if the height above sea level,  $h$ , was measured in kilometres and used the variable  $h = \frac{x}{1000}$  instead of  $x$ .

- (d) Write down the value of  $S_{hh}$ . (1)
- (e) Write down the value of the correlation coefficient between  $h$  and  $y$ . (1)
- 

2. The random variable  $X \sim N(\mu, 5^2)$  and  $P(X < 23) = 0.9192$ .

- (a) Find the value of  $\mu$ . (4)
- (b) Write down the value of  $P(\mu < X < 23)$ . (1)
- 

3. The discrete random variable  $Y$  has the probability distribution

$y$	1	2	3	4
$P(Y = y)$	$a$	$b$	0.3	$c$

where  $a$ ,  $b$  and  $c$  are constants.

The cumulative distribution function  $F(y)$  of  $Y$  is given in the following table.

$y$	1	2	3	4
$F(y)$	0.1	0.5	$d$	1.0

where  $d$  is a constant.

- (a) Find the value of  $a$ , the value of  $b$ , the value of  $c$  and the value of  $d$ . (5)
- (b) Find  $P(3Y + 2 \geq 8)$ . (2)
- 

4. Past records show that the times, in seconds, taken to run 100 m by children at a school can be modelled by a normal distribution with a mean of 16.12 and a standard deviation of 1.60.

A child from the school is selected at random.

- (a) Find the probability that this child runs 100 m in less than 15 s. (3)

On sports day the school awards certificates to the fastest 30% of the children in the 100 m race.

- (b) Estimate, to 2 decimal places, the slowest time taken to run 100 m for which a child will be awarded a certificate. (4)
-

5. A class of students had a sudoku competition. The time taken for each student to complete the sudoku was recorded to the nearest minute and the results are summarised in the table below.

Time	Mid-point, $x$	Frequency, $f$
2 – 8	5	2
9 – 12		7
13 – 15	14	5
16 – 18	17	8
19 – 22	20.5	4
23 – 30	26.5	4

(You may use  $\sum fx^2 = 8603.75$ )

- (a) Write down the mid-point for the 9 – 12 interval. (1)
- (b) Use linear interpolation to estimate the median time taken by the students. (2)
- (c) Estimate the mean and standard deviation of the times taken by the students. (5)

The teacher suggested that a normal distribution could be used to model the times taken by the students to complete the sudoku.

- (d) Give a reason to support the use of a normal distribution in this case. (1)

On another occasion the teacher calculated the quartiles for the times taken by the students to complete a different sudoku and found

$$Q_1 = 8.5 \quad Q_2 = 13.0 \quad Q_3 = 21.0$$

- (e) Describe, giving a reason, the skewness of the times on this occasion. (2)
- 

6. Jake and Kamil are sometimes late for school. The events  $J$  and  $K$  are defined as follows

$J$  = the event that Jake is late for school,  
 $K$  = the event that Kamil is late for school.

$$P(J) = 0.25, \quad P(J \cap K) = 0.15 \quad \text{and} \quad P(J' \cap K') = 0.7.$$

On a randomly selected day, find the probability that

- (a) at least one of Jake or Kamil are late for school, (1)
- (b) Kamil is late for school. (2)

Given that Jake is late for school,

- (c) find the probability that Kamil is late. (3)

The teacher suspects that Jake being late for school and Kamil being late for school are linked in some way.

- (d) Determine whether or not  $J$  and  $K$  are statistically independent. (2)
- (e) Comment on the teacher's suspicion in the light of your calculation in part (d). (1)
-

7. A teacher took a random sample of 8 children from a class. For each child the teacher recorded the length of their left foot,  $f$  cm, and their height,  $h$  cm. The results are given in the table below.

$f$	23	26	23	22	27	24	20	21
$h$	135	144	134	136	140	134	130	132

(You may use  $\sum f = 186$     $\sum h = 1085$     $S_{ff} = 39.5$     $S_{hh} = 139.875$     $\sum fh = 25\,291$ )

- (a) Calculate  $S_{fh}$ . (2)
- (b) Find the equation of the regression line of  $h$  on  $f$  in the form  $h = a + bf$ .  
Give the value of  $a$  and the value of  $b$  correct to 3 significant figures. (5)
- (c) Use your equation to estimate the height of a child with a left foot length of 25 cm. (2)
- (d) Comment on the reliability of your estimate in part (c), giving a reason for your answer. (2)

The left foot length of the teacher is 25 cm.

- (e) Give a reason why the equation in part (b) should not be used to estimate the teacher's height. (1)
- 

8. A spinner is designed so that the score  $S$  is given by the following probability distribution.

$s$	0	1	2	4	5
$P(S = s)$	$p$	0.25	0.25	0.20	0.20

- (a) Find the value of  $p$ . (2)
- (b) Find  $E(S)$ . (2)
- (c) Show that  $E(S^2) = 9.45$ . (2)
- (d) Find  $\text{Var}(S)$ . (2)
- Tom and Jess play a game with this spinner. The spinner is spun repeatedly and  $S$  counters are awarded on the outcome of each spin. If  $S$  is even then Tom receives the counters and if  $S$  is odd then Jess receives them. The first player to collect 10 or more counters is the winner.
- (e) Find the probability that Jess wins after 2 spins. (2)
- (f) Find the probability that Tom wins after exactly 3 spins. (4)
- (g) Find the probability that Jess wins after exactly 3 spins. (3)

---

END

TOTAL FOR PAPER: 75 MARKS

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced Level****Tuesday 17 January 2012 – Morning****Time: 1 hour 30 minutes****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, differentiation and integration, or have retrievable mathematical formulas stored in them.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has 7 questions.

The total mark for this paper is 75.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

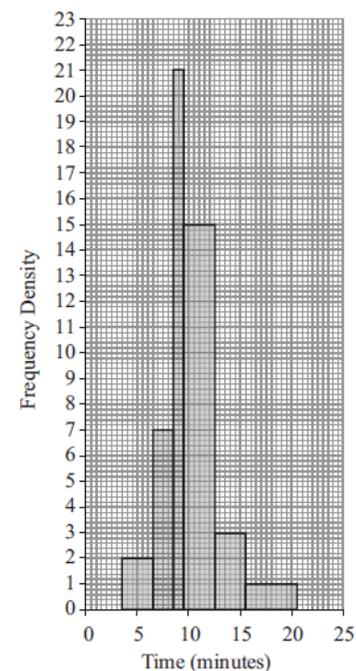
You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

P44699A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.  
©2012 Edexcel Limited

1. The histogram in Figure 1 shows the time, to the nearest minute, that a random sample of 100 motorists were delayed by roadworks on a stretch of motorway.

**Figure 1**

- (a) Complete the table.

Delay (minutes)	Number of motorists
4 – 6	6
7 – 8	
9	21
10 – 12	45
13 – 15	9
16 – 20	

(2)

- (b) Estimate the number of motorists who were delayed between 8.5 and 13.5 minutes by the roadworks.

(2)

2. (a) State in words the relationship between two events  $R$  and  $S$  when  $P(R \cap S) = 0$ . (1)

The events  $A$  and  $B$  are independent with  $P(A) = \frac{1}{4}$  and  $P(A \cup B) = \frac{2}{3}$ .

Find

- (b)  $P(B)$ , (4)  
 (c)  $P(A' \cap B)$ , (2)  
 (d)  $P(B'|A)$ . (2)

3. The discrete random variable  $X$  can take only the values 2, 3, 4 or 6. For these values the probability distribution function is given by

$x$	2	3	4	6
$P(X=x)$	$\frac{5}{21}$	$\frac{2k}{21}$	$\frac{7}{21}$	$\frac{k}{21}$

where  $k$  is a positive integer.

- (a) Show that  $k = 3$ . (2)  
 Find  
 (b)  $F(3)$ , (1)  
 (c)  $E(X)$ , (2)  
 (d)  $E(X^2)$ , (2)  
 (e)  $\text{Var}(7X - 5)$ . (4)

4. The marks,  $x$ , of 45 students randomly selected from those students who sat a mathematics examination are shown in the stem and leaf diagram below.

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

Key	(3   6 means 36)
-----	------------------

- (a) Write down the modal mark of these students. (1)  
 (b) Find the values of the lower quartile, the median and the upper quartile. (3)

For these students  $\sum x = 2497$  and  $\sum x^2 = 143\,369$ .

- (c) Find the mean and the standard deviation of the marks of these students. (3)  
 (d) Describe the skewness of the marks of these students, giving a reason for your answer. (2)

The mean and standard deviation of the marks of all the students who sat the examination were 55 and 10 respectively. The examiners decided that the total mark of each student should be scaled by subtracting 5 marks and then reducing the mark by a further 10 %.

- (e) Find the mean and standard deviation of the scaled marks of all the students. (4)

5. The age,  $t$  years, and weight,  $w$  grams, of each of 10 coins were recorded. These data are summarised below.

$$\sum t^2 = 2688 \quad \sum tw = 1760.62 \quad \sum t = 158 \quad \sum w = 111.75 \quad S_{ww} = 0.16$$

- (a) Find  $S_{tt}$  and  $S_{tw}$  for these data. (3)
- (b) Calculate, to 3 significant figures, the product moment correlation coefficient between  $t$  and  $w$ . (2)
- (c) Find the equation of the regression line of  $w$  on  $t$  in the form  $w = a + bt$ . (4)
- (d) State, with a reason, which variable is the explanatory variable. (2)
- (e) Using this model, estimate
- (i) the weight of a coin which is 5 years old,
  - (ii) the effect of an increase of 4 years in age on the weight of a coin.
- (2)

It was discovered that a coin in the original sample, which was 5 years old and weighed 20 grams, was a fake.

- (f) State, without any further calculations, whether the exclusion of this coin would increase or decrease the value of the product moment correlation coefficient. Give a reason for your answer. (2)
- 

6. The following shows the results of a survey on the types of exercise taken by a group of 100 people.

65 run  
 48 swim  
 60 cycle  
 40 run and swim  
 30 swim and cycle  
 35 run and cycle  
 25 do all three

- (a) Draw a Venn Diagram to represent these data. (4)
- Find the probability that a randomly selected person from the survey
- (b) takes none of these types of exercise, (2)
  - (c) swims but does not run, (2)
  - (d) takes at least two of these types of exercise. (2)

Jason is one of the above group.

Given that Jason runs,

- (e) find the probability that he swims but does not cycle. (3)
- 

7. A manufacturer fills jars with coffee. The weight of coffee,  $W$  grams, in a jar can be modelled by a normal distribution with mean 232 grams and standard deviation 5 grams.

- (a) Find  $P(W < 224)$ . (3)
- (b) Find the value of  $w$  such that  $P(232 < W < w) = 0.20$ . (4)

Two jars of coffee are selected at random.

- (c) Find the probability that only one of the jars contains between 232 grams and  $w$  grams of coffee. (3)
- 

END

TOTAL FOR PAPER: 75 MARKS

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced Level****Friday 18 May 2012 – Afternoon****Time: 1 hour 30 minutes****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, differentiation and integration, or have retrievable mathematical formulas stored in them.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has 7 questions.

The total mark for this paper is 75.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

**P40105XA**

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.  
©2012 Edexcel Limited

1. A discrete random variable  $X$  has the probability function

$$P(X = x) = \begin{cases} k(1-x)^2 & x = -1, 0, 1 \text{ and } 2 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that  $k = \frac{1}{6}$ . (3)

(b) Find  $E(X)$ . (2)

(c) Show that  $E(X^2) = \frac{4}{3}$ . (2)

(d) Find  $\text{Var}(1 - 3X)$ . (3)

2. A bank reviews its customer records at the end of each month to find out how many customers have become unemployed,  $u$ , and how many have had their house repossessed,  $h$ , during that month. The bank codes the data using variables  $x = \frac{u-100}{3}$  and  $y = \frac{h-20}{7}$ .

The results for the 12 months of 2009 are summarised below.

$$\sum x = 477 \quad S_{xx} = 5606.25 \quad \sum y = 480 \quad S_{yy} = 4244 \quad \sum xy = 23\,070$$

(a) Calculate the value of the product moment correlation coefficient for  $x$  and  $y$ . (3)

(b) Write down the product moment correlation coefficient for  $u$  and  $h$ . (1)

The bank claims that an increase in unemployment among its customers is associated with an increase in house repossessions.

(c) State, with a reason, whether or not the bank's claim is supported by these data. (2)

3. A scientist is researching whether or not birds of prey exposed to pollutants lay eggs with thinner shells. He collects a random sample of egg shells from each of 6 different nests and tests for pollutant level,  $p$ , and measures the thinning of the shell,  $t$ . The results are shown in the table below.

$p$	3	8	30	25	15	12
$t$	1	3	9	10	5	6

[You may use  $\sum p^2 = 1967$  and  $\sum pt = 694$ ]

- (a) On graph paper, draw a scatter diagram to represent these data. (2)
- (b) Explain why a linear regression model may be appropriate to describe the relationship between  $p$  and  $t$ . (1)
- (c) Calculate the value of  $S_{pt}$  and the value of  $S_{pp}$ . (4)
- (d) Find the equation of the regression line of  $t$  on  $p$ , giving your answer in the form  $t = a + bp$ . (4)
- (e) Plot the point  $(\bar{p}, \bar{t})$  and draw the regression line on your scatter diagram. (2)

The scientist reviews similar studies and finds that pollutant levels above 16 are likely to result in the death of a chick soon after hatching.

- (f) Estimate the minimum thinning of the shell that is likely to result in the death of a chick. (2)

4.

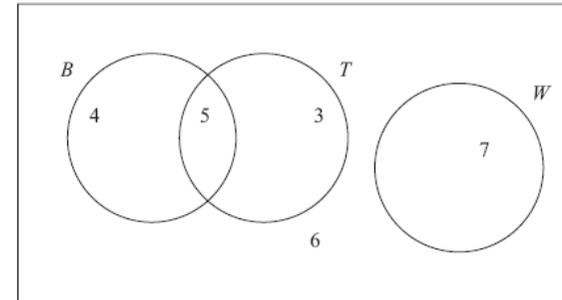


Figure 1

Figure 1 shows how 25 people travelled to work.

Their travel to work is represented by the events

- $B$  bicycle
- $T$  train
- $W$  walk

- (a) Write down 2 of these events that are mutually exclusive. Give a reason for your answer. (2)
- (b) Determine whether or not  $B$  and  $T$  are independent events. (3)

One person is chosen at random.

Find the probability that this person

- (c) walks to work, (1)
- (d) travels to work by bicycle and train. (1)

Given that this person travels to work by bicycle,

- (e) find the probability that they will also take the train. (2)

5.

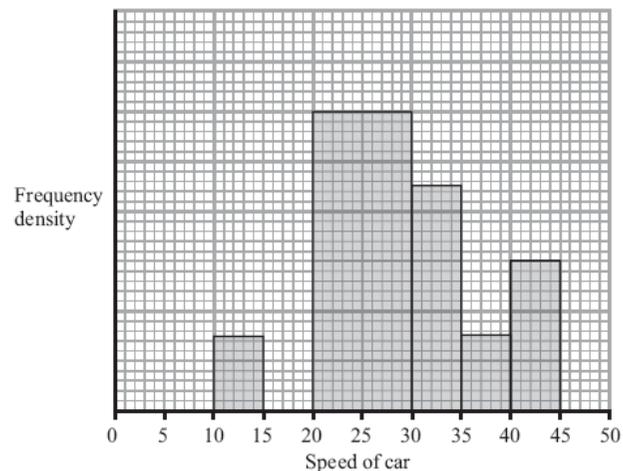


Figure 2

A policeman records the speed of the traffic on a busy road with a 30 mph speed limit.

He records the speeds of a sample of 450 cars. The histogram in Figure 2 represents the results.

- Calculate the number of cars that were exceeding the speed limit by at least 5 mph in the sample. (4)
- Estimate the value of the mean speed of the cars in the sample. (3)
- Estimate, to 1 decimal place, the value of the median speed of the cars in the sample. (2)
- Comment on the shape of the distribution. Give a reason for your answer. (2)
- State, with a reason, whether the estimate of the mean or the median is a better representation of the average speed of the traffic on the road. (2)

- The heights of an adult female population are normally distributed with mean 162 cm and standard deviation 7.5 cm.

- Find the probability that a randomly chosen adult female is taller than 150 cm. (3)

Sarah is a young girl. She visits her doctor and is told that she is at the 60th percentile for height.

- Assuming that Sarah remains at the 60th percentile, estimate her height as an adult. (3)

The heights of an adult male population are normally distributed with standard deviation 9.0 cm.

Given that 90% of adult males are taller than the mean height of adult females,

- find the mean height of an adult male. (4)

- A manufacturer carried out a survey of the defects in their soft toys. It is found that the probability of a toy having poor stitching is 0.03 and that a toy with poor stitching has a probability of 0.7 of splitting open. A toy without poor stitching has a probability of 0.02 of splitting open.

- Draw a tree diagram to represent this information. (3)
- Find the probability that a randomly chosen soft toy has exactly one of the two defects, poor stitching or splitting open. (3)

The manufacturer also finds that soft toys can become faded with probability 0.05 and that this defect is independent of poor stitching or splitting open. A soft toy is chosen at random.

- Find the probability that the soft toy has none of these 3 defects. (2)
- Find the probability that the soft toy has exactly one of these 3 defects. (4)

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced Level****Friday 18 January 2013 – Afternoon****Time: 1 hour 30 minutes****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, differentiation and integration, or have retrievable mathematical formulas stored in them.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has 7 questions.

The total mark for this paper is 75.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

**P41805A**

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.  
©2013 Edexcel Limited

1. A teacher asked a random sample of 10 students to record the number of hours of television,  $t$ , they watched in the week before their mock exam. She then calculated their grade,  $g$ , in their mock exam. The results are summarised as follows.

$$\sum t = 258 \quad \sum t^2 = 8702 \quad \sum g = 63.6 \quad S_{gg} = 7.864 \quad \sum gt = 1550.2$$

- (a) Find  $S_{tt}$  and  $S_{gt}$ . (3)

- (b) Calculate, to 3 significant figures, the product moment correlation coefficient between  $t$  and  $g$ . (2)

The teacher also recorded the number of hours of revision,  $v$ , these 10 students completed during the week before their mock exam. The correlation coefficient between  $t$  and  $v$  was  $-0.753$ .

- (c) Describe, giving a reason, the nature of the correlation you would expect to find between  $v$  and  $g$ . (2)

2. The discrete random variable  $X$  can take only the values 1, 2 and 3. For these values the cumulative distribution function is defined by

$$F(x) = \frac{x^3 + k}{40}, \quad x = 1, 2, 3.$$

- (a) Show that  $k = 13$ . (2)

- (b) Find the probability distribution of  $X$ . (4)

Given that  $\text{Var}(X) = \frac{259}{320}$ ,

- (c) find the exact value of  $\text{Var}(4X - 5)$ . (2)

3. A biologist is comparing the intervals ( $m$  seconds) between the mating calls of a certain species of tree frog and the surrounding temperature ( $t$  °C). The following results were obtained.

$t$ °C	8	13	14	15	15	20	25	30
$m$ secs	6.5	4.5	6	5	4	3	2	1

(You may use  $\sum tm = 469.5$ ,  $S_{tt} = 354$ ,  $S_{mm} = 25.5$ )

- (a) Show that  $S_{tm} = -90.5$ . (4)
- (b) Find the equation of the regression line of  $m$  on  $t$  giving your answer in the form  $m = a + bt$ . (4)
- (c) Use your regression line to estimate the time interval between mating calls when the surrounding temperature is 10 °C. (1)
- (d) Comment on the reliability of this estimate, giving a reason for your answer. (1)

4. The length of time,  $L$  hours, that a phone will work before it needs charging is normally distributed with a mean of 100 hours and a standard deviation of 15 hours.

- (a) Find  $P(L > 127)$ . (3)
- (b) Find the value of  $d$  such that  $P(L < d) = 0.10$ . (3)

Alice is about to go on a 6 hour journey. Given that it is 127 hours since Alice last charged her phone,

- (c) find the probability that her phone will not need charging before her journey is completed. (4)

5. A survey of 100 households gave the following results for weekly income £ $y$ .

Income $y$ (£)	Mid-point	Frequency $f$
$0 \leq y < 200$	100	12
$200 \leq y < 240$	220	28
$240 \leq y < 320$	280	22
$320 \leq y < 400$	360	18
$400 \leq y < 600$	500	12
$600 \leq y < 800$	700	8

(You may use  $\sum fy^2 = 12\,452\,800$ )

A histogram was drawn and the class  $200 \leq y < 240$  was represented by a rectangle of width 2 cm and height 7 cm.

- (a) Calculate the width and the height of the rectangle representing the class  $320 \leq y < 400$ . (3)
- (b) Use linear interpolation to estimate the median weekly income to the nearest pound. (2)
- (c) Estimate the mean and the standard deviation of the weekly income for these data. (4)

One measure of skewness is  $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$ .

- (d) Use this measure to calculate the skewness for these data and describe its value. (2)

Katie suggests using the random variable  $X$  which has a normal distribution with mean 320 and standard deviation 150 to model the weekly income for these data.

- (e) Find  $P(240 < X < 400)$ . (2)
- (f) With reference to your calculations in parts (d) and (e) and the data in the table, comment on Katie's suggestion. (2)

6. A fair blue die has faces numbered 1, 1, 3, 3, 5 and 5. The random variable  $B$  represents the score when the blue die is rolled.

(a) Write down the probability distribution for  $B$ . (2)

(b) State the name of this probability distribution. (1)

(c) Write down the value of  $E(B)$ . (1)

A second die is red and the random variable  $R$  represents the score when the red die is rolled.

The probability distribution of  $R$  is

$r$	2	4	6
$P(R = r)$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

(d) Find  $E(R)$ . (2)

(e) Find  $\text{Var}(R)$ . (3)

Tom invites Avisha to play a game with these dice.

Tom spins a fair coin with one side labelled 2 and the other side labelled 5. When Avisha sees the number showing on the coin she then chooses one of the dice and rolls it. If the number showing on the die is **greater** than the number showing on the coin, Avisha wins, otherwise Tom wins.

Avisha chooses the die which gives her the best chance of winning each time Tom spins the coin.

(f) Find the probability that Avisha wins the game, stating clearly which die she should use in each case. (4)

---

7. Given that

$$P(A) = 0.35, P(B) = 0.45 \text{ and } P(A \cap B) = 0.13,$$

find

(a)  $P(A \cup B)$ , (2)

(b)  $P(A' | B')$ . (2)

The event  $C$  has  $P(C) = 0.20$ .

The events  $A$  and  $C$  are mutually exclusive and the events  $B$  and  $C$  are independent.

(c) Find  $P(B \cap C)$ . (2)

(d) Draw a Venn diagram to illustrate the events  $A$ ,  $B$  and  $C$  and the probabilities for each region. (4)

(e) Find  $P([B \cup C]')$ . (2)

---

**END** **TOTAL FOR PAPER: 75 MARKS**

Paper Reference(s)

**6683/01****Edexcel GCE****Statistics S1****Advanced Subsidiary****Friday 17 May 2013 – Morning****Time: 1 hour 30 minutes****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, differentiation and integration, or have retrievable mathematical formulas stored in them.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has 6 questions.

The total mark for this paper is 75.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

P42831A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.  
©2013 Edexcel Limited

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  $\sum h = 7150$ ,  $\sum t = 110$ ,  $\sum h^2 = 7171500$ ,  $\sum t^2 = 1716$ ,  $\sum th = 64\,980$  and  $S_{tt} = 371.56$ ]

- (a) Calculate  $S_{hh}$  and  $S_{th}$ . 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)

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

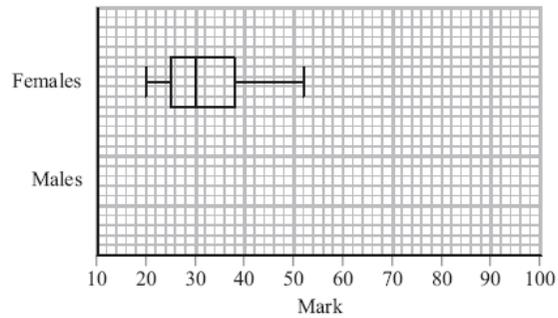


Figure 1

(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  $1.5 \times$  interquartile range above the upper quartile

or more than  $1.5 \times$  interquartile range below the lower quartile.

(c) On graph paper 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)

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)
- (b)  $P([F \cap W])$  (2)
- (c)  $P(W|C)$  (2)

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)
  - (e) find the probability that a randomly selected employee uses the bus to travel to work. (2)
- 

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)
-

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  $\text{Var}(5 - 3X)$ . (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)
- 

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

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

- (a) the value of  $\mu$ , (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  $\sigma$ .

- (c) Given that 98% of tins contain between 200 g and 210 g find the value of  $\sigma$ . (4)

---

**END** **TOTAL FOR PAPER: 75 MARKS**

Paper Reference(s)

**6683/01R****Edexcel GCE****Statistics S1 (R)****Advanced/Advanced Subsidiary****Friday 17 May 2013 – Morning****Time: 1 hour 30 minutes****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.

This paper is strictly for students outside the UK.

**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 for 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 the parts of questions are shown in round brackets, e.g. (2). There are 7 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 must show sufficient working to make your methods clear to the Examiner. Answers without working may not gain full credit.

**P43956A**

This publication may only be reproduced in accordance with Edexcel Limited copyright policy. ©2013 Edexcel Limited.

1. Sammy is studying the number of units of gas,  $g$ , and the number of units of electricity,  $e$ , used in her house each week. A random sample of 10 weeks use was recorded and the data for each week were coded so that  $x = \frac{g-60}{4}$  and  $y = \frac{e}{10}$ . The results for the coded data are summarised below

$$\sum x = 48.0, \quad \sum y = 58.0, \quad S_{xx} = 312.1, \quad S_{yy} = 2.10, \quad S_{xy} = 18.35$$

- (a) Find the equation of the regression line of  $y$  on  $x$  in the form  $y = a + bx$ .

Give the values of  $a$  and  $b$  correct to 3 significant figures.

(4)

- (b) Hence find the equation of the regression line of  $e$  on  $g$  in the form  $e = c + dg$ .

Give the values of  $c$  and  $d$  correct to 2 significant figures.

(4)

- (c) Use your regression equation to estimate the number of units of electricity used in a week when 100 units of gas were used.

(2)

2. The discrete random variable  $X$  takes the values 1, 2 and 3 and has cumulative distribution function  $F(x)$  given by

$x$	1	2	3
$F(x)$	0.4	0.65	1

- (a) Find the probability distribution of  $X$ .

(3)

- (b) Write down the value of  $F(1.8)$ .

(1)

3. An agriculturalist is studying the yields,  $y$  kg, from tomato plants. The data from a random sample of 70 tomato plants are summarised below.

Yield ( $y$ kg)	Frequency ( $f$ )	Yield midpoint ( $x$ kg)
$0 \leq y < 5$	16	2.5
$5 \leq y < 10$	24	7.5
$10 \leq y < 15$	14	12.5
$15 \leq y < 25$	12	20
$25 \leq y < 35$	4	30

(You may use  $\sum fx = 755$  and  $\sum fx^2 = 12\,037.5$ )

A histogram has been drawn to represent these data.

The bar representing the yield  $5 \leq y < 10$  has a width of 1.5 cm and a height of 8 cm.

- (a) Calculate the width and the height of the bar representing the yield  $15 \leq y < 25$ . (3)
- (b) Use linear interpolation to estimate the median yield of the tomato plants. (2)
- (c) Estimate the mean and the standard deviation of the yields of the tomato plants. (4)
- (d) Describe, giving a reason, the skewness of the data. (2)
- (e) Estimate the number of tomato plants in the sample that have a yield of more than 1 standard deviation above the mean. (2)
- 

4. The time, in minutes, taken to fly from London to Malaga has a normal distribution with mean 150 minutes and standard deviation 10 minutes.

- (a) Find the probability that the next flight from London to Malaga takes less than 145 minutes. (3)

The time taken to fly from London to Berlin has a normal distribution with mean 100 minutes and standard deviation  $d$  minutes.

Given that 15% of the flights from London to Berlin take longer than 115 minutes,

- (b) find the value of the standard deviation  $d$ . (4)

The time,  $X$  minutes, taken to fly from London to another city has a normal distribution with mean  $\mu$  minutes.

Given that  $P(X < \mu - 15) = 0.35$

- (c) find  $P(X > \mu + 15 \mid X > \mu - 15)$ . (3)
-

5. A researcher believes that parents with a short family name tended to give their children a long first name. A random sample of 10 children was selected and the number of letters in their family name,  $x$ , and the number of letters in their first name,  $y$ , were recorded.

The data are summarised as:

$$\sum x = 60, \quad \sum y = 61, \quad \sum y^2 = 393, \quad \sum xy = 382, \quad S_{xx} = 28$$

- (a) Find  $S_{yy}$  and  $S_{xy}$ . (3)
- (b) Calculate the product moment correlation coefficient,  $r$ , between  $x$  and  $y$ . (2)
- (c) State, giving a reason, whether or not these data support the researcher's belief. (2)

The researcher decides to add a child with family name "Turner" to the sample.

- (d) Using the definition  $S_{xx} = \sum (x - \bar{x})^2$ , state the new value of  $S_{xx}$  giving a reason for your answer. (2)

Given that the addition of the child with family name "Turner" to the sample leads to an increase in  $S_{yy}$

- (e) use the definition  $S_{xy} = \sum (x - \bar{x})(y - \bar{y})$  to determine whether or not the value of  $r$  will increase, decrease or stay the same. Give a reason for your answer. (2)
- 

6.

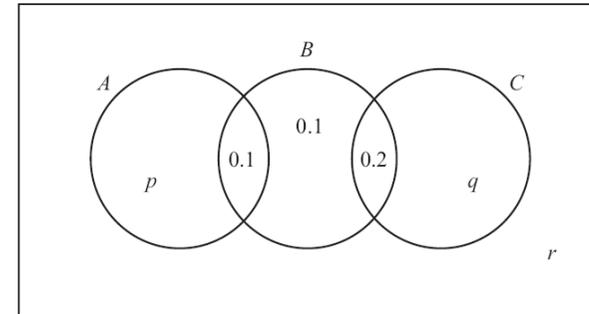


Figure 1

The Venn diagram in Figure 1 shows three events  $A$ ,  $B$  and  $C$  and the probabilities associated with each region of  $B$ . The constants  $p$ ,  $q$  and  $r$  each represent probabilities associated with the three separate regions outside  $B$ .

The events  $A$  and  $B$  are independent.

- (a) Find the value of  $p$ . (3)

Given that  $P(B|C) = \frac{5}{11}$ ,

- (b) find the value of  $q$  and the value of  $r$ . (4)
- (c) Find  $P(A \cup C | B)$ . (2)
-

7. The score  $S$  when a spinner is spun has the following probability distribution.

$s$	0	1	2	4	5
$P(S = s)$	0.2	0.2	0.1	0.3	0.2

- (a) Find  $E(S)$ . (2)
- (b) Show that  $E(S^2) = 10.4$ . (2)
- (c) Hence find  $\text{Var}(S)$ . (2)
- (d) Find
- (i)  $E(5S - 3)$ ,
- (ii)  $\text{Var}(5S - 3)$ . (4)
- (e) Find  $P(5S - 3 > S + 3)$ . (3)

The spinner is spun twice.

The score from the first spin is  $S_1$  and the score from the second spin is  $S_2$ .

The random variables  $S_1$  and  $S_2$  are independent and the random variable  $X = S_1 \times S_2$ .

- (f) Show that  $P(\{S_1 = 1\} \cap X < 5) = 0.16$ . (2)
- (g) Find  $P(X < 5)$ . (3)

---

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683/01**

**Edexcel GCE**

**Statistics S1**

**Advanced Subsidiary**

**Friday 17 May 2013 – Morning**

**Time: 1 hour 30 minutes**

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, differentiation and integration, or have retrievable mathematical formulas stored in them.

**Instructions to Candidates**

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has 6 questions.

The total mark for this paper is 75.

**Advice to Candidates**

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

**P42831A**

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.  
©2013 Edexcel Limited

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  $\sum h = 7150$ ,  $\sum t = 110$ ,  $\sum h^2 = 7171500$ ,  $\sum t^2 = 1716$ ,  $\sum th = 64\,980$  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)

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

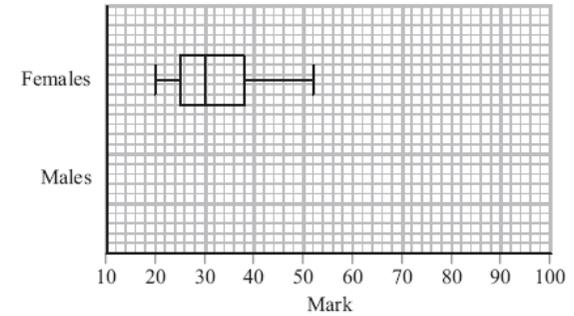


Figure 1

- (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  $1.5 \times$  interquartile range above the upper quartile

or more than  $1.5 \times$  interquartile range below the lower quartile.

(c) On graph paper 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)

---

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)

(b)  $P([F \cap W]')$  (2)

(c)  $P(W|C)$  (2)

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)

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

---

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)

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  $\text{Var}(5 - 3X)$ . (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)

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

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

- (a) the value of  $\mu$ , (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  $\sigma$ .

- (c) Given that 98% of tins contain between 200 g and 210 g find the value of  $\sigma$ . (4)

---

END TOTAL FOR PAPER: 75 MARKS

Paper Reference(s)

**WST01/01**

# Pearson Edexcel International Advanced Level

## Statistics S1

### Advanced/Advanced Subsidiary

**Friday 17 January 2014 – Afternoon**

**Time: 1 hour 30 minutes**

**Materials required for examination**

Mathematical Formulae (Blue)

**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, 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.

**P43140A**

©2014 Edexcel Limited.

This publication may only be reproduced in accordance with Edexcel Limited copyright policy.

1. A price comparison website publishes data on the cost per month, £ $c$ , and the level of satisfaction,  $s$ , of a random sample of six internet service providers. A low value of  $s$  corresponds to a low level of satisfaction. The data are given in the table below.

Internet service provider	A	B	C	D	E	F
$c$	20	15	12	30	9	25
$s$	5	3	4	2	3	4

(You may use  $\Sigma c = 111$ ,  $\Sigma c^2 = 2375$ ,  $\Sigma s = 21$ ,  $\Sigma s^2 = 79$ ,  $\Sigma cs = 380$ ,  $S_{cc} = 321.5$ .)

- (a) Calculate the value of  $S_{cs}$  and the value of  $S_{ss}$ . (3)
- (b) Calculate the product moment correlation coefficient for these data. (2)

Brad is not satisfied with his current internet service and decides to change his provider. He decides to pay a lot more for his new internet service.

- (c) On the basis of your calculation in part (b), comment on Brad's decision. Give a reason for your answer. (2)
- 

2. A rugby club coach uses club records to take a random sample of 15 players from 1990 and an independent random sample of 15 players from 2010. The body weight of each player was recorded to the nearest kg and the results from 2010 are summarised in the table below.

Body weight (kg)	75–79	80–84	85–89	90–94	95–99	100–104	105–109
Number of Players (2010)	1	2	2	4	3	2	1

- (a) Find the estimated values in kg of the summary statistics  $a$ ,  $b$  and  $c$  in the table below.

	Estimate in 1990	Estimate in 2010
Mean	83.0	$a$
Median	82.0	$b$
Variance	44.0	$c$

Give your answers to 3 significant figures.

(6)

The rugby coach claims that players' body weight increased between 1990 and 2010.

- (b) Using the table in part (a), comment on the rugby coach's claim. (2)
-

3. Jean works for an insurance company. She randomly selects 8 people and records the price of their car insurance, £ $p$ , and the time,  $t$  years, since they passed their driving test. The data is shown in the table below.

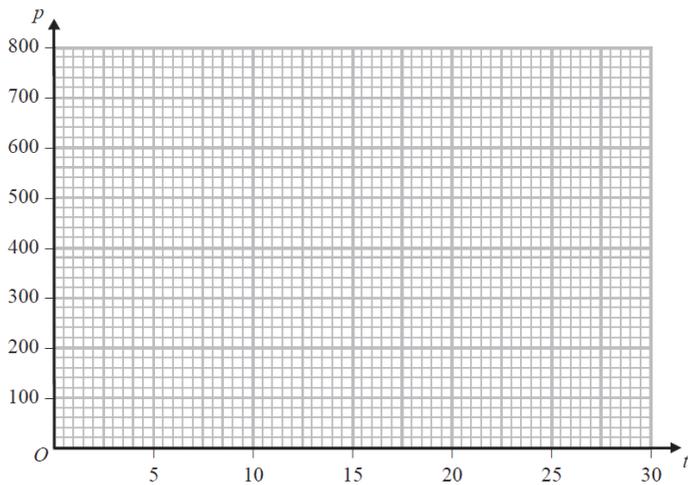
$t$	10	13	17	18	22	24	25	27
$p$	720	650	430	490	500	390	280	300

(You may use  $\bar{t} = 19.5$ ,  $\bar{p} = 470$ ,  $S_p = -6080$ ,  $S_t = 254$ ,  $S_{pp} = 169\,200$ .)

- (a) On the graph below draw a scatter diagram for these data. (2)
- (b) Comment on the relationship between  $p$  and  $t$ . (1)
- (c) Find the equation of the regression line of  $p$  on  $t$ . (4)
- (d) Use your regression equation to estimate the price of car insurance for someone who passed their driving test 20 years ago. (2)

Jack passed his test 39 years ago and decides to use Jean's data to predict the price of his car insurance.

- (e) Comment on Jack's decision. Give a reason for your answer. (2)



4. A discrete random variable  $X$  has the probability distribution given in the table below, where  $a$  and  $b$  are constants.

$x$	-1	0	1	2	3
$P(X = x)$	$a$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{10}$	$b$

Given  $E(X) = \frac{9}{5}$

- (a) (i) find two simultaneous equations for  $a$  and  $b$ ,  
 (ii) show that  $a = \frac{1}{20}$  and find the value of  $b$ . (4)
- (b) Specify the cumulative distribution function  $F(x)$  for  $x = -1, 0, 1, 2$  and  $3$ . (2)
- (c) Find  $P(X < 2.5)$ . (1)
- (d) Find  $\text{Var}(3 - 2X)$ . (4)

5. A group of 100 students are asked if they like folk music, rock music or soul music.

All students who like folk music also like rock music

No students like both rock music and soul music

75 students do not like soul music

12 students who like rock music do not like folk music

30 students like folk music

- (a) Draw a Venn diagram to illustrate this information.

(4)

- (b) State two of these types of music that are mutually exclusive.

(1)

Find the probability that a randomly chosen student

- (c) does not like folk music, rock music or soul music,

(1)

- (d) likes rock music,

(1)

- (e) likes folk music or soul music.

(1)

Given that a randomly chosen student likes rock music,

- (f) find the probability that he or she also likes folk music.

(2)

---

6. A manufacturer has a machine that fills bags with flour such that the weight of flour in a bag is normally distributed. A label states that each bag should contain 1 kg of flour.

- (a) The machine is set so that the weight of flour in a bag has mean 1.04 kg and standard deviation 0.17 kg. Find the proportion of bags that weigh less than the stated weight of 1 kg.

(3)

The manufacturer wants to reduce the number of bags which contain less than the stated weight of 1 kg. At first she decides to adjust the mean but not the standard deviation so that only 5% of the bags filled are below the stated weight of 1 kg.

- (b) Find the adjusted mean.

(3)

The manufacturer finds that a lot of the bags are overflowing with flour when the mean is adjusted, so decides to adjust the standard deviation instead to make the machine more accurate. The machine is set back to a mean of 1.04 kg. The manufacturer wants 1% of bags to be under 1 kg.

- (c) Find the adjusted standard deviation. Give your answer to 3 significant figures.

(3)

---

7. In a large college,  $\frac{3}{5}$  of the students are male,  $\frac{3}{10}$  of the students are left handed and  $\frac{1}{5}$  of the male students are left handed.

A student is chosen at random.

- (a) Given that the student is left handed, find the probability that the student is male.

(2)

- (b) Given that the student is female, find the probability that she is left handed.

(3)

- (c) Find the probability that the randomly chosen student is male and right handed.

(2)

Two students are chosen at random.

- (d) Find the probability that one student is left handed and one is right handed.

(2)

---

8. A manager records the number of hours of overtime claimed by 40 staff in a month.

The histogram in Figure 1 represents the results.

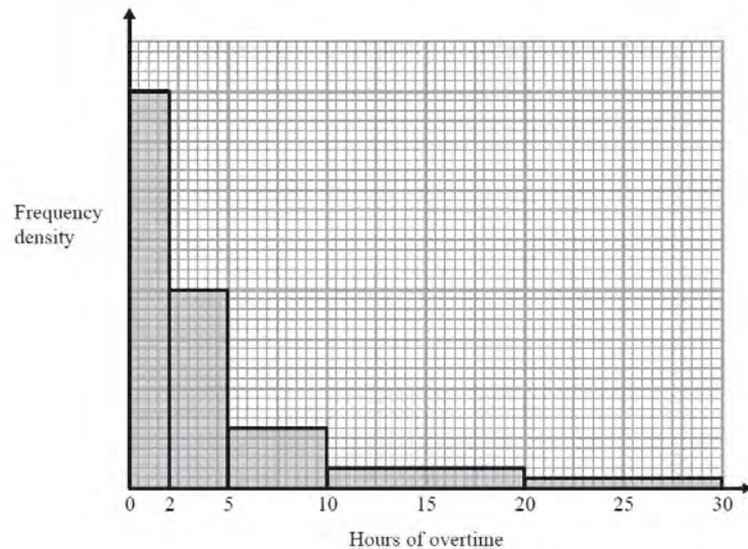


Figure 1

- (a) Calculate the number of staff who have claimed less than 10 hours of overtime in the month. (4)
- (b) Estimate the median number of hours of overtime claimed by these 40 staff in the month. (2)
- (c) Estimate the mean number of hours of overtime claimed by these 40 staff in the month. (2)

The manager wants to compare these data with overtime data he collected earlier to find out if the overtime claimed by staff has decreased.

- (d) State, giving a reason, whether the manager should use the median or the mean to compare the overtime claimed by staff. (2)

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**WST01/01**

# Pearson Edexcel International Advanced Level

**Statistics S1**

**Advanced/Advanced Subsidiary**

**Tuesday 10 June 2014 – Morning**

**Time: 1 hour 30 minutes**

**Materials required for examination**

Mathematical Formulae (Blue)

**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, 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.

**P43144A**

This publication may only be reproduced in accordance with Pearson Education Limited copyright policy. ©2014 Pearson Education Limited.

1. A medical researcher is studying the relationship between age ( $x$  years) and volume of blood ( $y$  ml) pumped by each contraction of the heart. The researcher obtained the following data from a random sample of 8 patients.

<b>Age (<math>x</math>)</b>	20	25	30	45	55	60	65	70
<b>Volume (<math>y</math>)</b>	74	76	77	72	68	67	64	62

[You may use  $\Sigma x = 370$ ,  $S_{xx} = 2587.5$ ,  $\Sigma y = 560$ ,  $\Sigma y^2 = 39\,418$ ,  $S_{xy} = -710$ ]

- (a) Calculate  $S_{yy}$ . (2)
- (b) Calculate the product moment correlation coefficient for these data. (2)
- (c) Interpret your value of the correlation coefficient. (1)

The researcher 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 researcher's belief. (1)
- (e) Find the equation of the regression line of  $y$  on  $x$ , giving your answer in the form  $y = a + bx$ . (4)

Jack is a 40-year-old patient.

- (f) (i) Use your regression line to estimate the volume of blood pumped by each contraction of Jack's heart. (2)
- (ii) Comment, giving a reason, on the reliability of your estimate.

2. The table below shows the distances (to the nearest km) travelled to work by the 50 employees in an office.

Distance (km)	Frequency (f)	Distance midpoint ( $x$ )
0 – 2	16	1.25
3 – 5	12	4
6 – 10	10	8
11 – 20	8	15.5
21 – 40	4	30.5

[You may use  $\Sigma fx = 394$ ,  $\Sigma fx^2 = 6500$ ]

A histogram has been drawn to represent these data.

The bar representing the distance of 3 – 5 has a width of 1.5 cm and a height of 6 cm.

- (a) Calculate the width and height of the bar representing the distance of 6 – 10. (3)
- (b) Use linear interpolation to estimate the median distance travelled to work. (2)
- (c) (i) Show that an estimate of the mean distance travelled to work is 7.88 km. (4)
- (ii) Estimate the standard deviation of the distances travelled to work. (4)
- (d) Describe, giving a reason, the skewness of these data. (2)

Peng starts to work in this office as the 51st employee.

She travels a distance of 7.88 km to work.

- (e) Without carrying out any further calculations, state, giving a reason, what effect Peng's addition to the workforce would have on your estimates of the
- (i) mean,
- (ii) median,
- (iii) standard deviation
- of the distances travelled to work. (3)

3. A biased four-sided die has faces marked 1, 3, 5 and 7. The random variable  $X$  represents the score on the die when it is rolled. The cumulative distribution function of  $X$ ,  $F(x)$ , is given in the table below.

$x$	1	3	5	7
$F(x)$	0.2	0.5	0.9	1

- (a) Find the probability distribution of  $X$ . (4)
- (b) Find  $P(2 < X \leq 6)$ . (2)
- (c) Write down the value of  $F(4)$ . (1)
- 

4. The random variable  $Y \sim N(\mu, \sigma^2)$ .

Given that  $P(Y < 17) = 0.6$  find

- (a)  $P(Y > 17)$  (1)
- (b)  $P(\mu < Y < 17)$  (2)
- (c)  $P(Y < \mu \mid Y < 17)$  (2)
- 

5. The discrete random variable  $X$  has the following probability distribution

$x$	-2	0	2	4
$P(X = x)$	$a$	$b$	$a$	$c$

where  $a$ ,  $b$  and  $c$  are probabilities.

Given that  $E(X) = 0.8$

- (a) find the value of  $c$ . (2)

Given also that  $E(X^2) = 5$  find

- (b) the value of  $a$  and the value of  $b$ , (4)
- (c)  $\text{Var}(X)$  (2)

The random variable  $Y = 5 - 3X$

Find

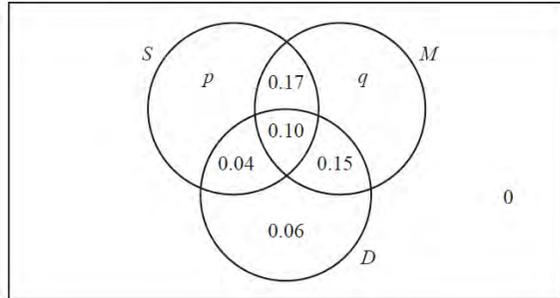
- (d)  $E(Y)$  (1)
- (e)  $\text{Var}(Y)$  (2)
- (f)  $P(Y \geq 0)$  (4)
-

6. The Venn diagram below shows the probabilities of customers having various combinations of a starter, main course or dessert at Polly's restaurant.

$S$  = the event a customer has a starter.

$M$  = the event a customer has a main course.

$D$  = the event a customer has a dessert.



Given that the events  $S$  and  $D$  are statistically independent

- (a) find the value of  $p$ . (4)
- (b) Hence find the value of  $q$ . (2)
- (c) Find
- (i)  $P(D | M \cap S)$
- (ii)  $P(D | M \cap S')$  (4)

One evening 63 customers are booked into Polly's restaurant for an office party. Polly has asked for their starter and main course orders before they arrive.

Of these 63 customers

27 ordered a main course and a starter,

36 ordered a main course without a starter.

- (d) Estimate the number of desserts that these 63 customers will have. (2)

7. One event at *Pentor* sports day is throwing a tennis ball. The distance a child throws a tennis ball is modelled by a normal distribution with mean 32 m and standard deviation 12 m. Any child who throws the tennis ball more than 50 m is awarded a gold certificate.

- (a) Show that, to 3 significant figures, 6.68% of children are awarded a gold certificate. (3)

A silver certificate is awarded to any child who throws the tennis ball more than  $d$  metres but less than 50 m.

Given that 19.1% of the children are awarded a silver certificate,

- (b) find the value of  $d$ . (4)

Three children are selected at random from those who take part in the throwing a tennis ball event.

- (c) Find the probability that 1 is awarded a gold certificate and 2 are awarded silver certificates. Give your answer to 2 significant figures. (4)

---

TOTAL FOR PAPER: 75 MARKS

END

Paper Reference(s)

**6683/01R****Edexcel GCE****Statistics S1 (R)****Advanced/Advanced Subsidiary****Tuesday 10 June 2014 – Morning****Time: 1 hour 30 minutes****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.**

**This paper is strictly for students outside the UK.**

**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 for 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 the parts of questions are shown in round brackets, e.g. (2).

There are 7 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 must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

**P43148A**

This publication may only be reproduced in accordance with Pearson Education Limited copyright policy.  
©2014 Pearson Education Limited.

1. The discrete random variable  $X$  has probability distribution

$x$	-4	-2	1	3	5
$P(X = x)$	0.4	$p$	0.05	0.15	$p$

- (a) Show that  $p = 0.2$ .

(2)

Find

- (b)  $E(X)$

(2)

- (c)  $F(0)$

(1)

- (d)  $P(3X + 2 > 5)$

(2)

Given that  $\text{Var}(X) = 13.35$ ,

- (e) find the possible values of  $a$  such that  $\text{Var}(aX + 3) = 53.4$ .

(2)

2. The discrete random variable  $X$  has probability distribution

$$P(X = x) = \frac{1}{10} \quad x = 1, 2, 3, \dots, 10$$

- (a) Write down the name given to this distribution.

(1)

- (b) Write down the value of

(i)  $P(X = 10)$

(ii)  $P(X < 10)$

(2)

The continuous random variable  $Y$  has the normal distribution  $N(10, 2^2)$ .

- (c) Write down the value of

(i)  $P(Y = 10)$

(ii)  $P(Y < 10)$

(2)

3. A large company is analysing how much money it spends on paper in its offices every year. The number of employees,  $x$ , and the amount of money spent on paper,  $p$  (£ hundreds), in 8 randomly selected offices are given in the table below.

$x$	8	9	12	14	7	3	16	19
$p$ (£ hundreds)	40.5	36.1	30.4	39.4	32.6	31.1	43.4	45.7

(You may use  $\sum x^2 = 1160$      $\sum p = 299.2$      $\sum p^2 = 11\,422$      $\sum xp = 3449.5$ )

- (a) Show that  $S_{pp} = 231.92$  and find the value of  $S_{xx}$  and the value of  $S_{xp}$ . (5)
- (b) Calculate the product moment correlation coefficient between  $x$  and  $p$ . (2)
- The equation of the regression line of  $p$  on  $x$  is given in the form  $p = a + bx$ .
- (c) Show that, to 3 significant figures,  $b = 0.824$  and find the value of  $a$ . (4)
- (d) Estimate the amount of money spent on paper in an office with 10 employees. (2)
- (e) Explain the effect each additional employee has on the amount of money spent on paper. (1)

Later the company realised it had made a mistake in adding up its costs,  $p$ . The true costs were actually half of the values recorded. The product moment correlation coefficient and the equation of the linear regression line are recalculated using this information.

- (f) Write down the new value of
- (i) the product moment correlation coefficient,
- (ii) the gradient of the regression line. (2)
- 

4.  $A$  and  $B$  are two events such that

$$P(B) = \frac{1}{2} \qquad P(A | B) = \frac{2}{5} \qquad P(A \cup B) = \frac{13}{20}$$

- (a) Find  $P(A \cap B)$ . (2)
- (b) Draw a Venn diagram to show the events  $A$ ,  $B$  and all the associated probabilities. (3)
- Find
- (c)  $P(A)$  (1)
- (d)  $P(B | A)$  (2)
- (e)  $P(A' \cap B)$  (1)
-

5. The table shows the time, to the nearest minute, spent waiting for a taxi by each of 80 people one Sunday afternoon.

Waiting time (in minutes)	Frequency
2–4	15
5–6	9
7	6
8	24
9–10	14
11–15	12

- (a) Write down the upper class boundary for the 2–4 minute interval. (1)

A histogram is drawn to represent these data. The height of the tallest bar is 6 cm.

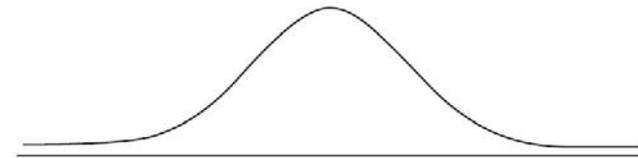
- (b) Calculate the height of the second tallest bar. (3)
- (c) Estimate the number of people with a waiting time between 3.5 minutes and 7 minutes. (2)
- (d) Use linear interpolation to estimate the median, the lower quartile and the upper quartile of the waiting times. (4)
- (e) Describe the skewness of these data, giving a reason for your answer. (2)
- 

6. The time taken, in minutes, by children to complete a mathematical puzzle is assumed to be normally distributed with mean  $\mu$  and standard deviation  $\sigma$ . The puzzle can be completed in less than 24 minutes by 80% of the children. For 5% of the children it takes more than 28 minutes to complete the puzzle.

- (a) Show this information on the Normal curve below. (2)
- (b) Write down the percentage of children who take between 24 minutes and 28 minutes to complete the puzzle. (1)
- (c) (i) Find two equations in  $\mu$  and  $\sigma$ . (7)
- (ii) Hence find, to 3 significant figures, the value of  $\mu$  and the value of  $\sigma$ . (7)

A child is selected at random.

- (d) Find the probability that the child takes less than 12 minutes to complete the puzzle. (3)



7. In a large company,
- 78% of employees are car owners,  
30% of these car owners are also bike owners,  
85% of those who are not car owners are bike owners.
- (a) Draw a tree diagram to represent this information. (3)
- An employee is selected at random.
- (b) Find the probability that the employee is a car owner or a bike owner but not both. (2)
- Another employee is selected at random.
- Given that this employee is a bike owner,
- (c) find the probability that the employee is a car owner. (3)
- Two employees are selected at random.
- (d) Find the probability that only one of them is a bike owner. (3)

---

**TOTAL FOR PAPER: 75 MARKS**

**END**

Paper Reference(s)

**6683/01**

**Edexcel GCE**

**Statistics S1**

**Advanced/Advanced Subsidiary**

**Tuesday 10 June 2014 – Morning**

**Time: 1 hour 30 minutes**

**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, differentiation and integration, or have retrievable mathematical formulas stored in them.**

**Instructions to Candidates**

---

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.

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.

This paper has 8 questions.

The total mark for this paper is 75.

**Advice to Candidates**

---

You must ensure that your answers to parts of questions are clearly labelled.

You must show sufficient working to make your methods clear to the Examiner.

Answers without working may not gain full credit.

1. A random sample of 35 homeowners was taken from each of the villages Greenslax and Penville and their ages were recorded. The results are summarised in the back-to-back stem and leaf diagram below.

Totals	Greenslax		Penville	Totals
(2)	8 7	2	5 5 6 7 8 8 9	(7)
(3)	9 8 7	3	1 1 1 2 3 4 4 5 6 7 9	(11)
(4)	4 4 4 0	4	0 1 2 4 7	(5)
(5)	6 6 5 2 2	5	0 0 5 5 5	(5)
(7)	8 6 5 4 2 1 1	6	2 5 6 6	(4)
(8)	8 6 6 6 4 3 1 1	7	0 5	(2)
(5)	9 8 4 3 2	8		(0)
(1)	4	9	9	(1)

Key: 7 | 3 | 1 means 37 years for Greenslax and 31 years for Penville

Some of the quartiles for these two distributions are given in the table below.

	Greenslax	Penville
Lower quartile, $Q_1$	$a$	31
Median, $Q_2$	64	39
Upper quartile, $Q_3$	$b$	55

- (a) Find the value of  $a$  and the value of  $b$ .

(2)

An outlier is a value that falls either

more than  $1.5 \times (Q_3 - Q_1)$  above  $Q_3$

or more than  $1.5 \times (Q_3 - Q_1)$  below  $Q_1$

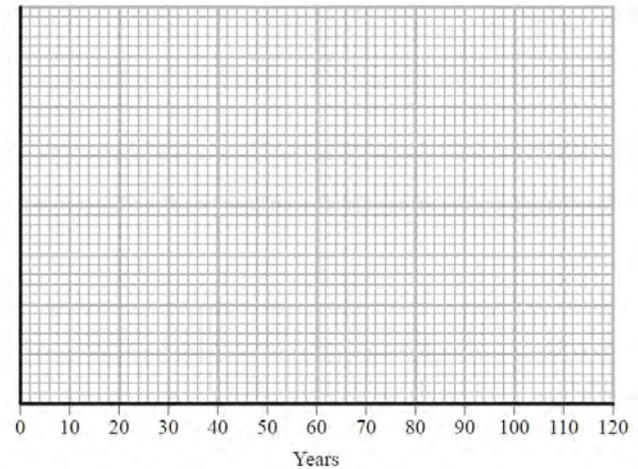
- (b) On the graph paper on the next page draw a box plot to represent the data from Penville. Show clearly any outliers.

(4)

- (c) State the skewness of each distribution. Justify your answers.

(3)

Question 1(b) graph paper



2. The mark,  $x$ , scored by each student who sat a statistics examination is coded using

$$y = 1.4x - 20$$

The coded marks have mean 60.8 and standard deviation 6.60.

Find the mean and the standard deviation of  $x$ .

(4)

3. The table shows data on the number of visitors to the UK in a month,  $v$  (1000s), and the amount of money they spent,  $m$  (£ millions), for each of 8 months.

Number of visitors $v$ (1000s)	2450	2480	2540	2420	2350	2290	2400	2460
Amount of money spent $m$ (£ millions)	1370	1350	1400	1330	1270	1210	1330	1350

You may use

$$S_{vv} = 42587.5 \quad S_{vm} = 31512.5 \quad S_{mm} = 25187.5 \quad \Sigma v = 19390 \quad \Sigma m = 10610$$

- (a) Find the product moment correlation coefficient between  $m$  and  $v$ . (2)
- (b) Give a reason to support fitting a regression model of the form  $m = a + bv$  to these data. (1)
- (c) Find the value of  $b$  correct to 3 decimal places. (2)
- (d) Find the equation of the regression line of  $m$  on  $v$ . (2)
- (e) Interpret your value of  $b$ . (2)
- (f) Use your answer to part (d) to estimate the amount of money spent when the number of visitors to the UK in a month is 2 500 000. (2)
- (g) Comment on the reliability of your estimate in part (f). Give a reason for your answer. (2)
- 

4. In a factory, three machines,  $J$ ,  $K$  and  $L$ , are used to make biscuits.

Machine  $J$  makes 25% of the biscuits.

Machine  $K$  makes 45% of the biscuits.

The rest of the biscuits are made by machine  $L$ .

It is known that 2% of the biscuits made by machine  $J$  are broken, 3% of the biscuits made by machine  $K$  are broken and 5% of the biscuits made by machine  $L$  are broken.

- (a) Draw a tree diagram to illustrate all the possible outcomes and associated probabilities. (2)
- A biscuit is selected at random.
- (b) Calculate the probability that the biscuit is made by machine  $J$  and is not broken. (2)
- (c) Calculate the probability that the biscuit is broken. (2)
- (d) Given that the biscuit is broken, find the probability that it was not made by machine  $K$ . (3)
- 

5. The discrete random variable  $X$  has the probability function

$$P(X = x) = \begin{cases} kx & x = 2, 4, 6 \\ k(x-2) & x = 8 \\ 0 & \text{otherwise} \end{cases}$$

where  $k$  is a constant.

- (a) Show that  $k = \frac{1}{18}$ . (2)
- (b) Find the exact value of  $F(5)$ . (1)
- (c) Find the exact value of  $E(X)$ . (2)
- (d) Find the exact value of  $E(X^2)$ . (2)
- (e) Calculate  $\text{Var}(3 - 4X)$  giving your answer to 3 significant figures. (3)
-

6. The times, in seconds, spent in a queue at a supermarket by 85 randomly selected customers, are summarised in the table below.

Time (seconds)	Number of customers, $f$
0 – 30	2
30 – 60	10
60 – 70	17
70 – 80	25
80 – 100	25
100 – 150	6

A histogram was drawn to represent these data. The 30 – 60 group was represented by a bar of width 1.5 cm and height 1 cm.

- (a) Find the width and the height of the 70 – 80 group. (3)
- (b) Use linear interpolation to estimate the median of this distribution. (2)

Given that  $x$  denotes the midpoint of each group in the table and

$$\sum fx = 6460 \quad \sum fx^2 = 529\,400$$

- (c) calculate an estimate for
- (i) the mean,
- (ii) the standard deviation,
- for the above data. (3)

One measure of skewness is given by

$$\text{coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

- (d) Evaluate this coefficient and comment on the skewness of these data. (3)
- 

7. The heights of adult females are normally distributed with mean 160 cm and standard deviation 8 cm.

- (a) Find the probability that a randomly selected adult female has a height greater than 170 cm. (3)

Any adult female whose height is greater than 170 cm is defined as tall.

An adult female is chosen at random. Given that she is tall,

- (b) find the probability that she has a height greater than 180 cm. (4)

Half of tall adult females have a height greater than  $h$  cm.

- (c) Find the value of  $h$ . (5)
- 

8. For the events  $A$  and  $B$ ,

$$P(A' \cap B) = 0.22 \quad \text{and} \quad P(A' \cap B') = 0.18$$

- (a) Find  $P(A)$ . (1)
- (b) Find  $P(A \cup B)$ . (1)

Given that  $P(A | B) = 0.6$ ,

- (c) find  $P(A \cap B)$ . (3)
- (d) Determine whether or not  $A$  and  $B$  are independent. (2)

---

**TOTAL FOR PAPER: 75 MARKS**

END