Edexcel Maths S1

Past Paper Pack

2005-2013

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Thursday 9 June 2005 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

Materials required for examination
Mathematical Formulae (Lilac or 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

In the boxes above, write your centre number, candidate number, your surname, initial(s) and signature.
Check that you have the correct question paper.
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 individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 7 questions in this question paper.
The total for this question paper is 75.
There are 20 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 gain no credit.

Printer's Log. No.
N20910A
W850/R6683/57570 4/7/3/3/2/6/3/

N 2 0 9 1 0 A 0 1 2 0

Turn over

edexcel

Leave blank

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

Diagram *A*

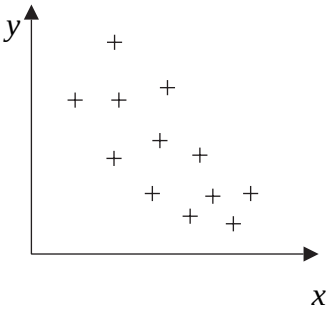


Diagram *B*

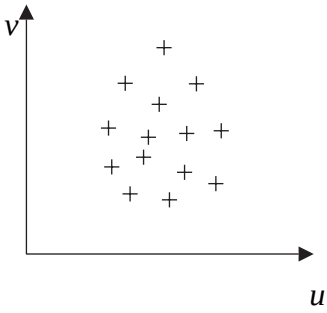
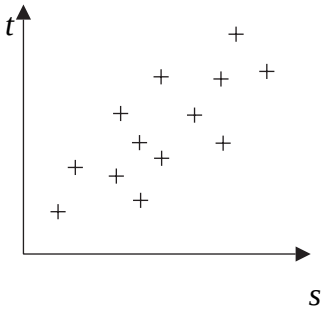


Diagram *C*



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

The values were

$$0.68 \quad -0.79 \quad 0.08$$

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

(6)

Leave blank

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.

(DO NOT DRAW THE HISTOGRAM)

(2)

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

(4)

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

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

(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)

Leave blank

Question 2 continued

Leave blank

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 $y = f - 100$.

$$\Sigma x = 130 \quad \Sigma y = 48 \quad \Sigma xy = 8880 \quad S_{xx} = 20\,487.5$$

(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)

Leave blank

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× (interquartile range) above the upper quartile or 1.5× (interquartile range) below the lower quartile.

(a) On the graph paper opposite draw a box plot to represent these data.

(7)

(b) Comment on the distribution of delays. Justify your answer.

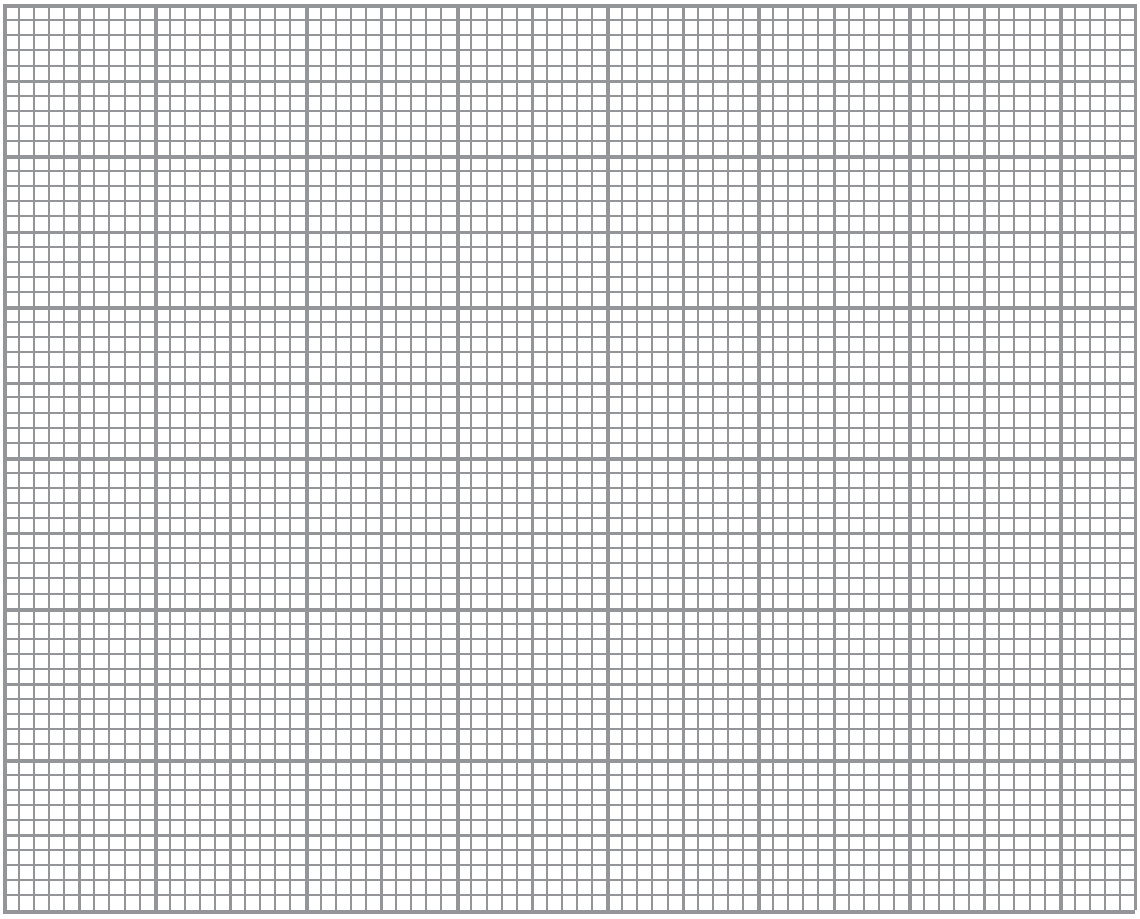
(2)

(c) Suggest how the distribution might be interpreted by a passenger who frequently flies from City *A* to City *B*.

(1)

Leave blank

Question 4 continued



(Total 10 marks)

Q4

Leave blank

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, Var(X) = 1.47.

(4)

(d) Find, to 1 decimal place, Var(4 – 3X).

(2)

Leave blank

Question 5 continued

Leave blank

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($150 \leqslant M \leqslant 157$).

(4)

(c) the value of m, to 1 decimal place, such that P($M \leqslant m$) = 0.30.

(4)

Leave blank

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)

Leave blank

Question 7 continued

Q7

(Total 12 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Wednesday 24 May 2006 – Afternoon

Time: 1 hour 30 minutes

Examiner's use only		

Team Leader's use only		

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
Total	

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

In the boxes above, write your centre number, candidate number, your surname, initial(s) and signature.
Check that you have the correct question paper.
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 individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 6 questions in this question paper.
The total for this question paper is 75.
There are 20 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 gain no credit.

Printer's Log. No.
N22337A
W850/R6683/57570 3/3/3/3/3/

Turn over

edexcel

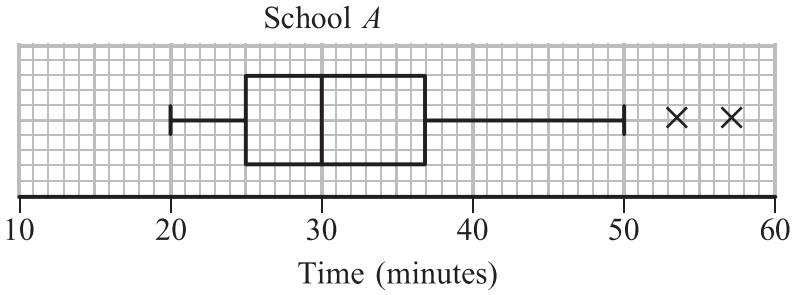
Leave blank

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.

(ii) State the name given to this value.

(2)

(c) Explain what you understand by the two crosses (×) on Figure 1.

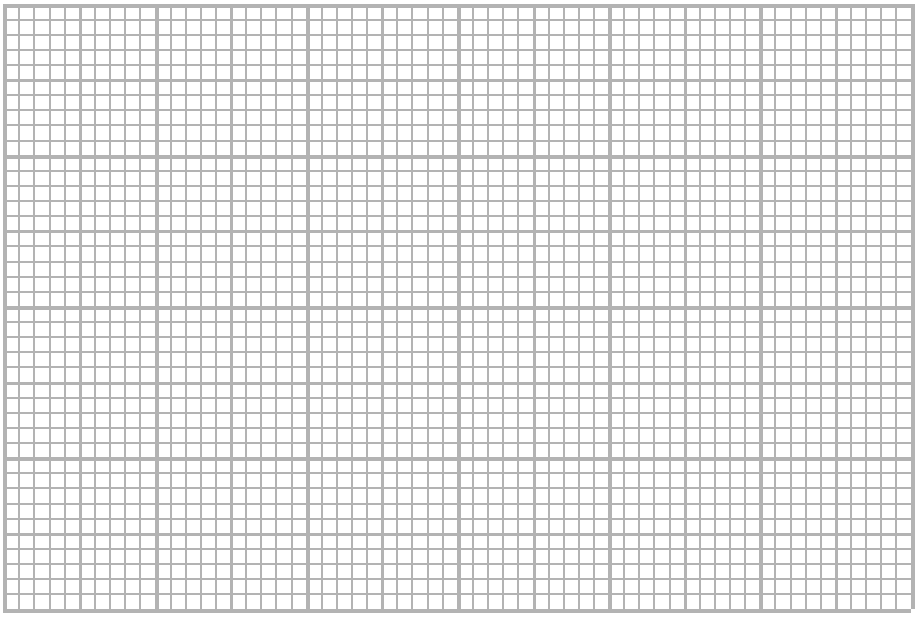
(2)

Leave blank

Question 1 continued

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) Draw a box plot to represent the data from school *B*.



(4)

(e) Compare and contrast these two box plots.

(4)

Q1

(Total 15 marks)

Leave blank

2. Sunita and Shelley talk to one another 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 $\Sigma fx = 1060$.

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

(2)

During the following 25 weeks they monitored their weekly conversations 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)

Leave blank

Question 2 continued

Leave blank

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 $\Sigma x^2 = 15965.01$ and $\Sigma xy = 757.467$)

(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)

Leave blank

Question 3 continued

Leave blank

4. The random variable X has the discrete uniform distribution

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

(a) Write down the value of E(X) and show that Var(X) = 2.

(3)

Find

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

(2)

(c) Var($4 - 3X$).

(2)

Leave blank

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)

Leave blank

Question 5 continued

Leave blank

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 that they were left handed and had dark hair.

(3)

Leave blank

Question 6 continued

Q6

(Total 13 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Tuesday 16 January 2007 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

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

In the boxes above, write your centre number, candidate number, your surname, initial(s) and signature.
Check that you have the correct question paper.
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 individual questions and parts of questions are shown in round brackets: e.g. **(2)**.
There are 7 questions in this question paper. The total for this question paper is 75.
There are 20 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 gain no credit.

Printer's Log. No.
N23957A
W850/R6683/57570 3/3/3/3/3/3/17,400

Turn over

edexcel
advancing learning, changing lives

Leave blank

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 $\Sigma t^2 = 5478 \quad \Sigma m^2 = 2101 \quad \Sigma tm = 2485$)

(6)

(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)

Leave blank

Question 1 continued

Leave blank

Question 1 continued

Leave blank

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)

Leave blank

Question 2 continued

Q2

(Total 11 marks)

Leave blank

3. The random variable X has probability function

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

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

(3)

Find

(b) $P(2 < X \leqslant 5)$,

(2)

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

(2)

(d) Show that $Var(X) = 1.97$ to 3 significant figures.

(4)

(e) Find $Var(2 - 3X)$.

(2)

Leave blank

Question 3 continued

Q3

(Total 13 marks)

Leave blank

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

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

(c) Estimate the mean and the 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)

Leave blank

(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)

Leave blank

Question 4 continued

Leave blank

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 $0.8\,cm^2$.

(3)

The total area under the histogram was $24\,cm^2$.

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

(2)

Leave blank

Question 5 continued

Q5

(Total 7 marks)

Leave blank

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)

Leave blank

Question 6 continued

Q6

(Total 5 marks)

Leave blank

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)

Leave blank

Question 7 continued

Q7

(Total 10 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Tuesday 5 June 2007 – Afternoon

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

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 formulas stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer 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 individual questions and 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 should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

Printer's Log. No.
N26118A
W850/R6683/57570 3/3/3/3/3/3/3

N 2 6 1 1 8 A 0 1 2 4

Turn over

edexcel
advancing learning, changing lives

Leave blank

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

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

(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)

Leave blank

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.

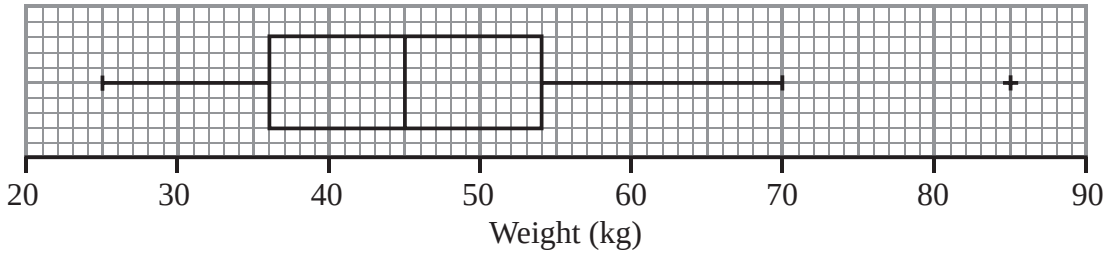


Figure 1

The airline's recommended weight limit for each musician's luggage was 45 kg.
Given that none of the musicians' 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)

Leave blank

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

(2)

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

(e) Find the standard deviation of this normal distribution.

(4)

Leave blank

3. A student is investigating the relationship between the price (y pence) of 100g of chocolate and the percentage (x%) of 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) On the graph paper on page 9 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 scatter diagram.

(2)

The student believes that one brand of chocolate is overpriced.

(e) Use the scatter diagram to

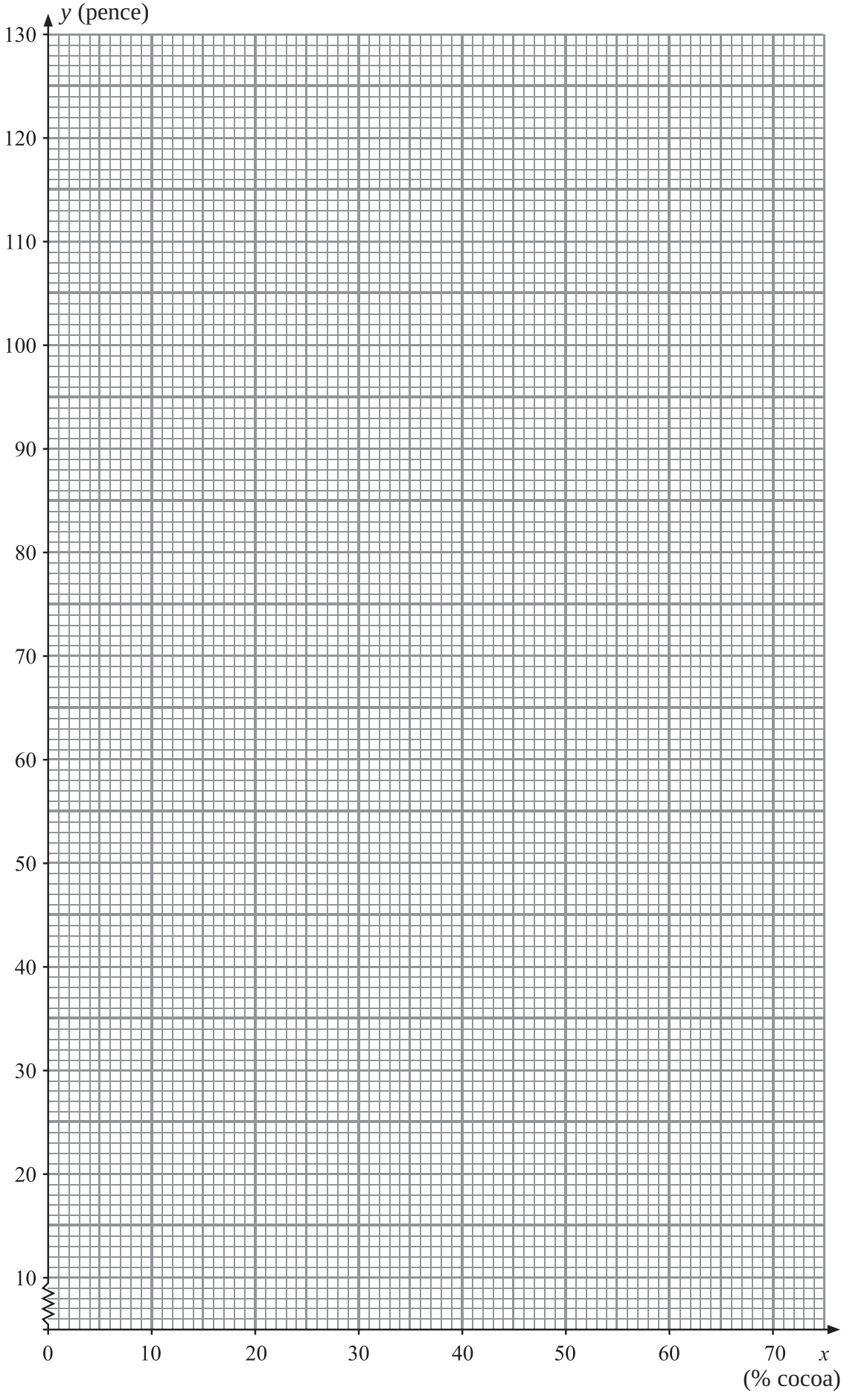
(i) state which brand is overpriced,
(ii) suggest a fair price for this brand.

Give reasons for both your answers.

(4)

Leave blank

Question 3 continued



Leave blank

The data on page 8 has been repeated here to help you

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

Question 3 continued

Leave blank

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) In the space on page 13 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)

Leave blank

Question 4 continued

Q4

(Total 9 marks)

Leave blank

5.

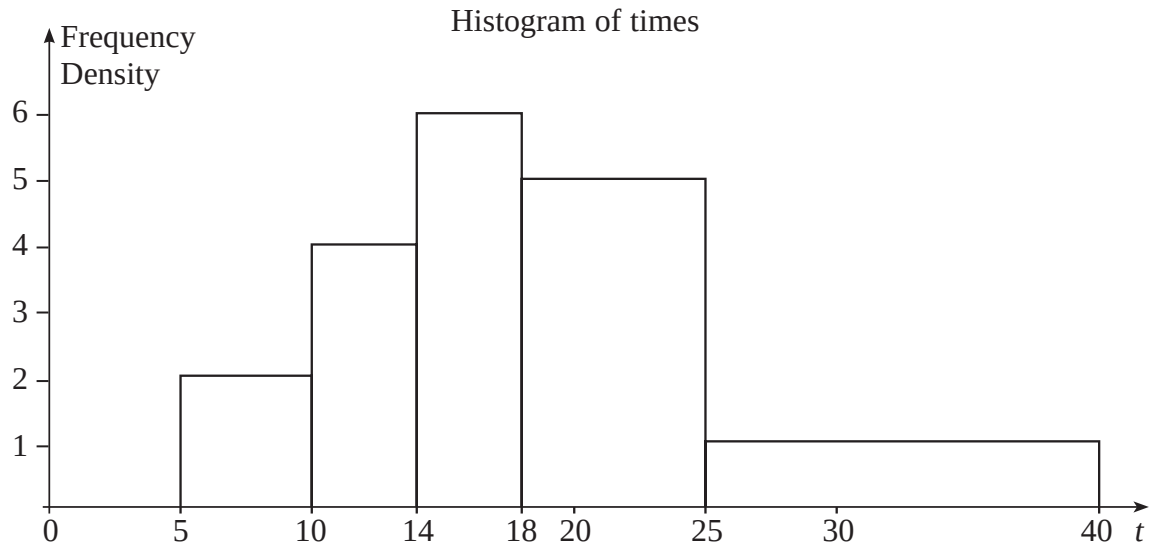


Figure 2

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

(a) 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 500m.

(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)

Leave blank

Question 5 continued

Leave blank

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)

Leave blank

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($4 < X \leqslant 7$).

(2)

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

(d) Var(X),

(2)

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

(1)

(f) Var($19 - 4X$).

(2)

Leave blank

Question 7 continued

Q7

(Total 13 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Tuesday 15 January 2008 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

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 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 individual questions and 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 should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

Printer's Log. No.
M29283A
W850/R6683/57570 3/3/3/3/3

M 2 9 2 8 3 A 0 1 2 4

Turn over

edexcel
advancing learning, changing lives

Leave blank

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)

Leave blank

Question 1 continued

Leave blank

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 $\dfrac{(Q_1 - 2Q_2 + Q_3)}{(Q_3 - Q_1)}$.

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

(3)

Leave blank

Question 2 continued

Leave blank

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

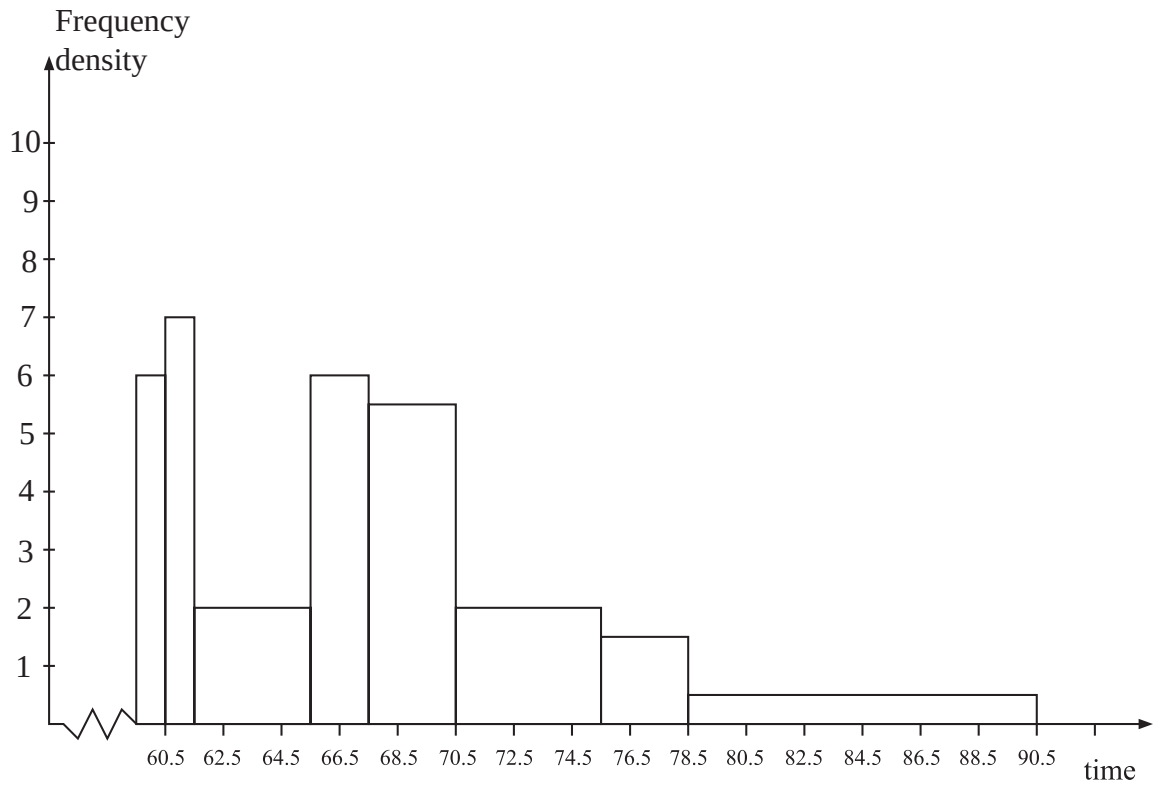


Figure 1

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

(5)

Leave blank

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)

Leave blank

Question 4 continued

Leave blank

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,

(1)

(c) wine *A* but not wine *B*,

(2)

(d) any wine in the survey except wine *C*,

(2)

(e) exactly two of the three kinds of wine.

(2)

Given that a person from the survey likes wine *A*,

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

(3)

Leave blank

Question 5 continued

Leave blank

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)

Leave blank

Question 6 continued

Q6

(Total 9 marks)

Leave blank

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 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.

3				
2		2		
1	0			
0				
B / R	**0**	**1**	**2**	**3**

Sample space diagram of T

(3)

(c) 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	1/8	1/8	c	1/8	d

Find the values of a, b, c and d.

(3)

Find the values of

(d) E(T),

(2)

(e) Var(T).

(4)

Leave blank

Question 7 continued

Q7

(Total 14 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Thursday 15 May 2008 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

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 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 individual questions and 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 28 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

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

Printer's Log. No.
H32582A
W850/R6683/57570 3/3/3/2/

H 3 2 5 8 2 A 0 1 2 8

Turn over

edexcel
advancing learning, changing lives

Leave blank

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)

Leave blank

Question 1 continued

Leave blank

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)	9751	1		
(4)	9831	2	6	(1)
(11)	99997665332	3	447	(3)
(6)	987750	4	005569	(6)
(1)	8	5	000013667	(9)
		6	233457	(6)
		7	015	(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)

Leave blank

Question 2 continued

Leave blank

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)

Leave blank

Question 3 continued

Leave blank

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.

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

(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)

Leave blank

Question 4 continued

Leave blank

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)

Leave blank

(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)

Leave blank

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)

Leave blank

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)

Leave blank

Question 7 continued

Q7

(Total 12 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Monday 19 January 2009 – Afternoon

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
Total	

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 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 individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 6 questions in this question paper. The total mark for this paper is 75.
There are 24 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

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

Printer's Log. No.
N32680A
W850/R6683/57570 3/3/3

N 3 2 6 8 0 A 0 1 2 4

Turn over

edexcel
advancing learning, changing lives

Leave blank

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 $\sum x = 21.4, \quad \sum y = 96, \quad \sum x^2 = 57.22, \quad \sum xy = 313.7$]

(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)

Leave blank

Question 1 continued

Leave blank

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)

Leave blank

Question 2 continued

Leave blank

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 $Var(X) = 1$

(4)

(d) Find $Var(5 - 3X)$.

(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)

Leave blank

Question 3 continued

Leave blank

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) Using the graph paper on page 15 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)

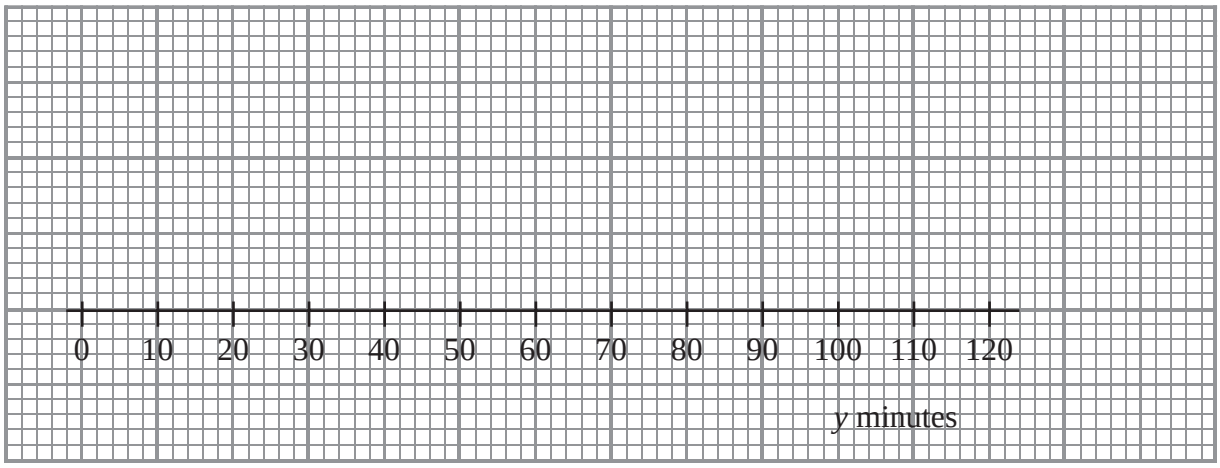
A parent believes that a student who sends a large number of text messages will spend fewer minutes on calls.

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

(1)

Leave blank

Question 4 continued



Leave blank

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)

Leave blank

Question 5 continued

Leave blank

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)

Leave blank

Question 6 continued

Q6

(Total 10 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Wednesday 20 May 2009 – Afternoon

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
8	
Total	

Materials required for examination
Mathematical Formulae (Orange 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 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 individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 8 questions in this question paper. The total mark for this paper is 75.
There are 24 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

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

Printer's Log. No.
H34279A
W850/R6683/57570 3/5/4/3

H 3 4 2 7 9 A 0 1 2 4

Turn over

edexcel
advancing learning, changing lives

Leave blank

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.

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

(a) Find S_{pp} and S_{pt}.

(3)

Given that $S_{tt} = 21760$,

(b) calculate the product moment correlation coefficient.

(2)

(c) Give an interpretation of your answer to part (b).

(1)

Leave blank

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)

Leave blank

Question 2 continued

Leave blank

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.

Leave blank

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**
$10 \leqslant l < 12$	5
$12 \leqslant l < 17$	53
$17 \leqslant l < 19$	29
$19 \leqslant l < 21$	15
$21 \leqslant l < 23$	11
$23 \leqslant l < 25$	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)

Leave blank

Question 4 continued

Leave blank

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)

Leave blank

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 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 $Var(X) = 1.64$

(3)

(f) Calculate $Var(3X - 2)$.

(2)

Leave blank

Question 6 continued

Leave blank

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, P(Q) = 0.35 and 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)

Leave blank

Question 7 continued

Q7

(Total 7 marks)

Leave blank

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)

Leave blank

Question 8 continued

Q8

(Total 11 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Wednesday 13 January 2010 – Afternoon

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

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 above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 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 should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

Printer's Log. No.
N35711A
W850/R6683/57570 5/5/5/3

N 3 5 7 1 1 A 0 1 2 4

Turn over

edexcel
advancing learning, changing lives

Leave blank

1. A jar contains 2 red, 1 blue and 1 green bead. Two beads are drawn at random from the jar without replacement.

(a) In the space below, 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)

Leave blank

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.

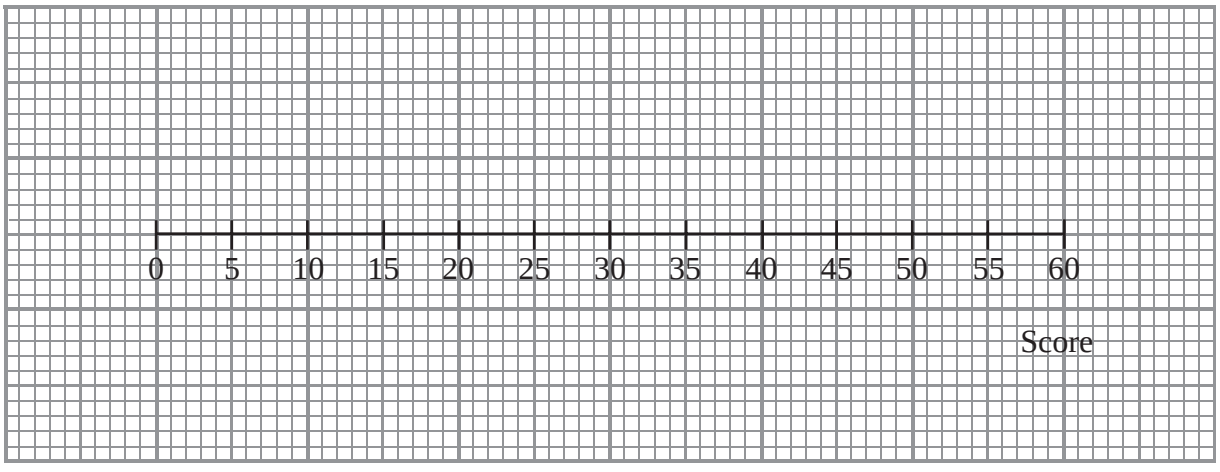
(2)

(d) On the graph paper on page 5, draw a box plot to illustrate the employees' scores.

(3)

Leave blank

Question 2 continued



Q2

(Total 9 marks)

Leave blank

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)

Leave blank

Question 3 continued

Leave blank

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) In the space below, draw a Venn diagram to represent this information.

(5)

A student from the course is chosen at random.

Find the probability that this 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)

Leave blank

Question 4 continued

Leave blank

5. The probability function of a discrete random variable X is given by

$$p(x) = kx^2 \qquad x = 1, 2, 3$$

where k is a positive constant.

(a) Show that $k = \frac{1}{14}$

(2)

Find

(b) $P(X \geqslant 2)$

(2)

(c) $E(X)$

(2)

(d) $\text{Var}(1 - X)$

(4)

Leave blank

Question 5 continued

Q5

(Total 10 marks)

Leave blank

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

$$[\sum t = 341, \sum p = 833, \sum t^2 = 18\,181, \sum p^2 = 106\,397, \sum tp = 42\,948]$$

(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) On the graph paper on page 17, 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)

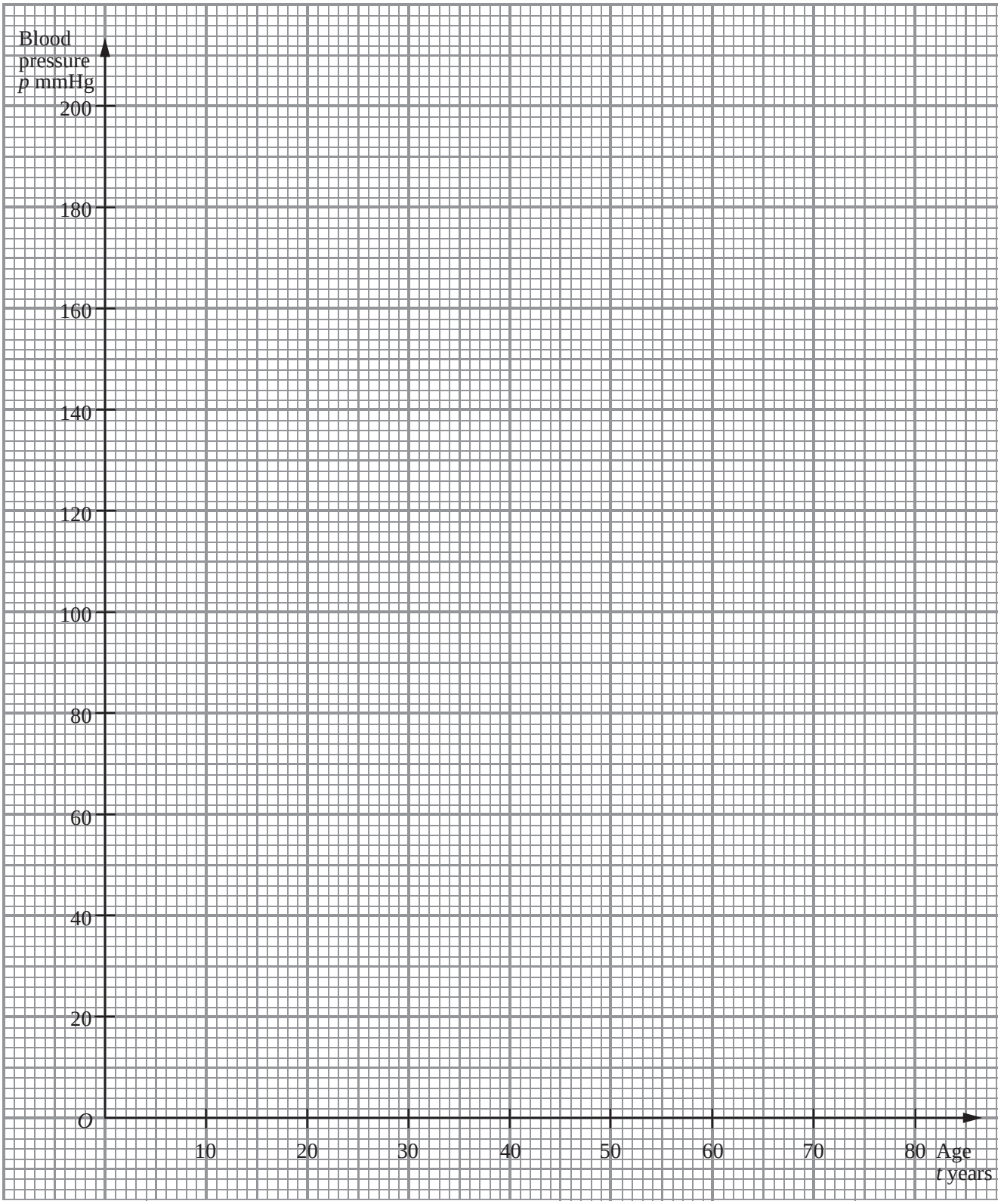
(f) Plot your regression line on your scatter diagram.

(2)

(g) Use your regression line to estimate the blood pressure of a 40 year old patient.

(2)

Question 6 continued



Leave blank

Question 6 continued

Leave blank

7. The heights of a population of women are normally distributed with mean μ cm and standard deviation σ 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 μ and the value of σ.

(4)

A woman is chosen at random from the population.

(d) Find the probability that she is taller than 160 cm.

(3)

Leave blank

Question 7 continued

Q7

(Total 13 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Thursday 27 May 2010 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

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 formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 7 questions in this question paper. The total mark for this paper is 75.
There are 28 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

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

Printer's Log. No.
H35395A
W850/R6683/57570 4/5/4/4

H 3 5 3 9 5 A 0 1 2 8

Turn over

edexcel
advancing learning, changing lives

Leave blank

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)

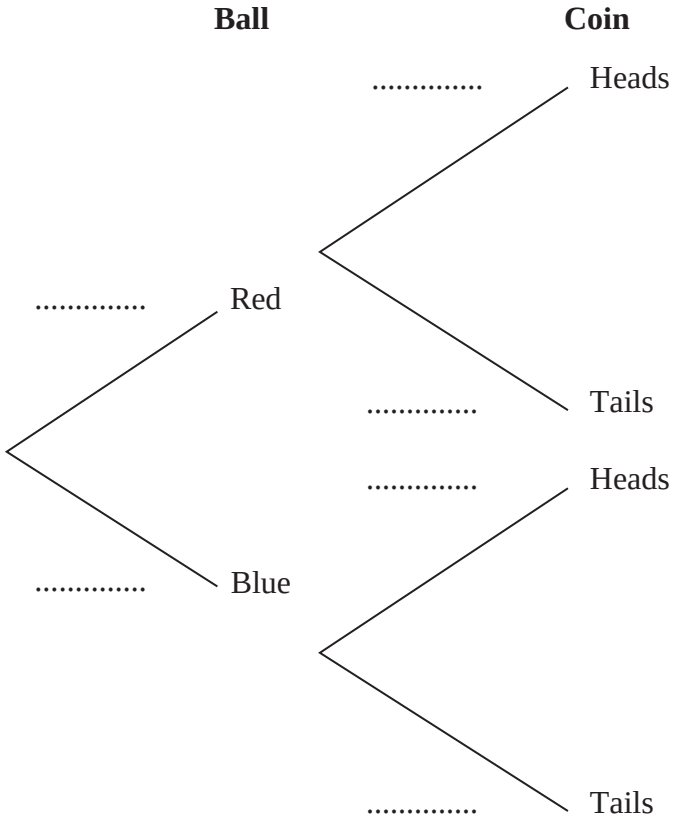
Leave blank

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) Complete the tree diagram below to show the possible outcomes and associated probabilities.



(2)

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)

Leave blank

Question 2 continued

Leave blank

3. The discrete random variable X has probability distribution given by

x	-1	0	1	2	3
$\mathrm{P}(X = x)$	$\frac{1}{5}$	a	$\frac{1}{10}$	a	$\frac{1}{5}$

where a is a constant.

(a) Find the value of a.

(2)

(b) Write down $\mathrm{E}(X)$.

(1)

(c) Find $\mathrm{Var}(X)$.

(3)

The random variable $Y = 6 - 2X$

(d) Find $\mathrm{Var}(Y)$.

(2)

(e) Calculate $\mathrm{P}(X \geqslant Y)$.

(3)

Leave blank

Question 3 continued

Leave blank

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*.

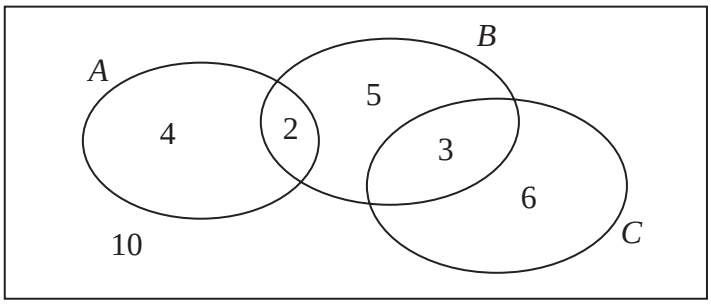


Figure 1

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)

Leave blank

Question 4 continued

Leave blank

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)

Leave blank

Question 5 continued

Leave blank

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) Using the axes below, complete 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)

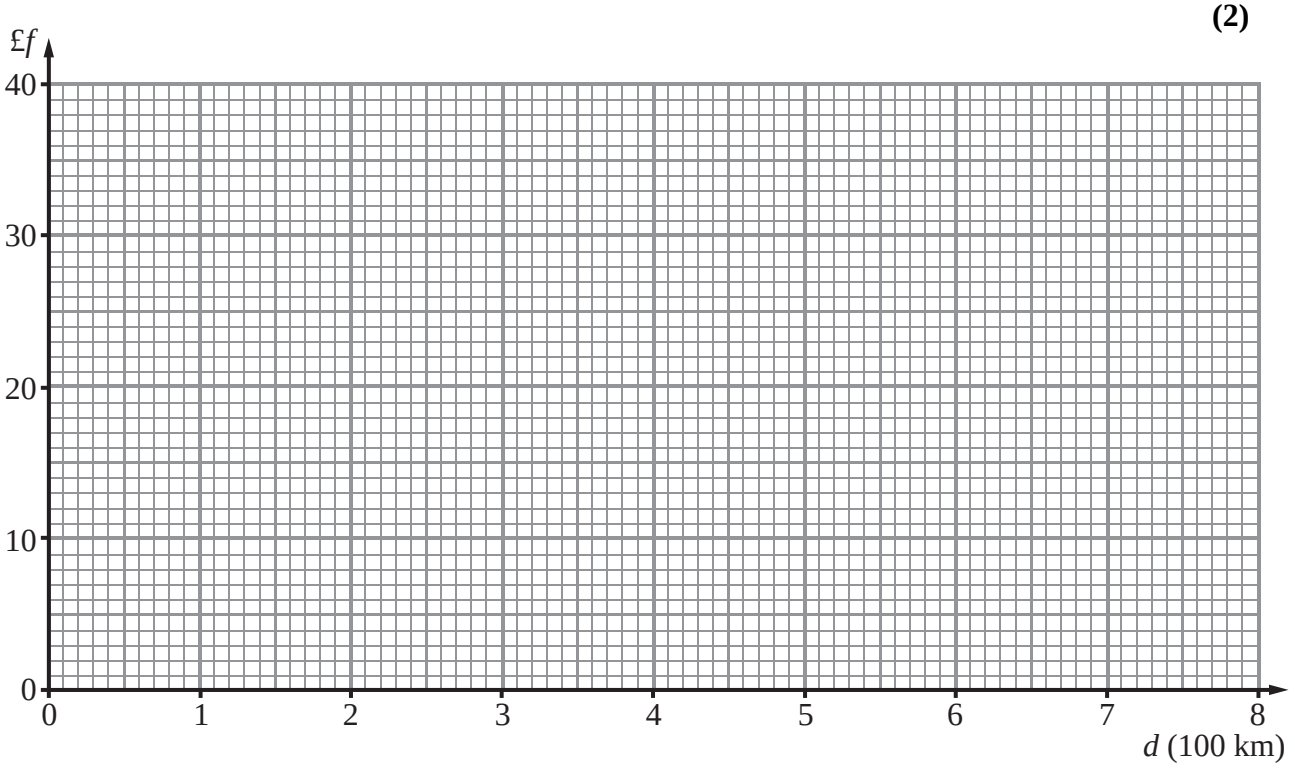
(e) Give an interpretation of the value of b.

(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)



Leave blank

Question 6 continued

Leave blank

7. The distances travelled to work, D km, by the employees at a large company are normally distributed with $D \sim \mathrm{N}(30, 8^2)$.

(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

$$h = Q_1 - 1.5 \times (Q_3 - Q_1) \quad \text{and} \quad k = Q_3 + 1.5 \times (Q_3 - Q_1)$$

(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)

Leave blank

Question 7 continued

Q7

(Total 12 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 14 January 2011 – Afternoon

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
8	
Total	

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 formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 8 questions in this question paper. The total mark for this paper is 75.
There are 24 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

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

Printer's Log. No.
H35410A
W850/R6683/57570 5/5/3

H 3 5 4 1 0 A 0 1 2 4

Turn over

edexcel
advancing learning, changing lives

Leave blank

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.

$\sum l = 4027$ $\sum l^2 = 327\,754.5$ $\sum w = 357.1$ $\sum lw = 29\,330.5$ $S_{ww} = 289.6$

(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)

Leave blank

Question 1 continued

Q1

(Total 6 marks)

Leave blank

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)

Leave blank

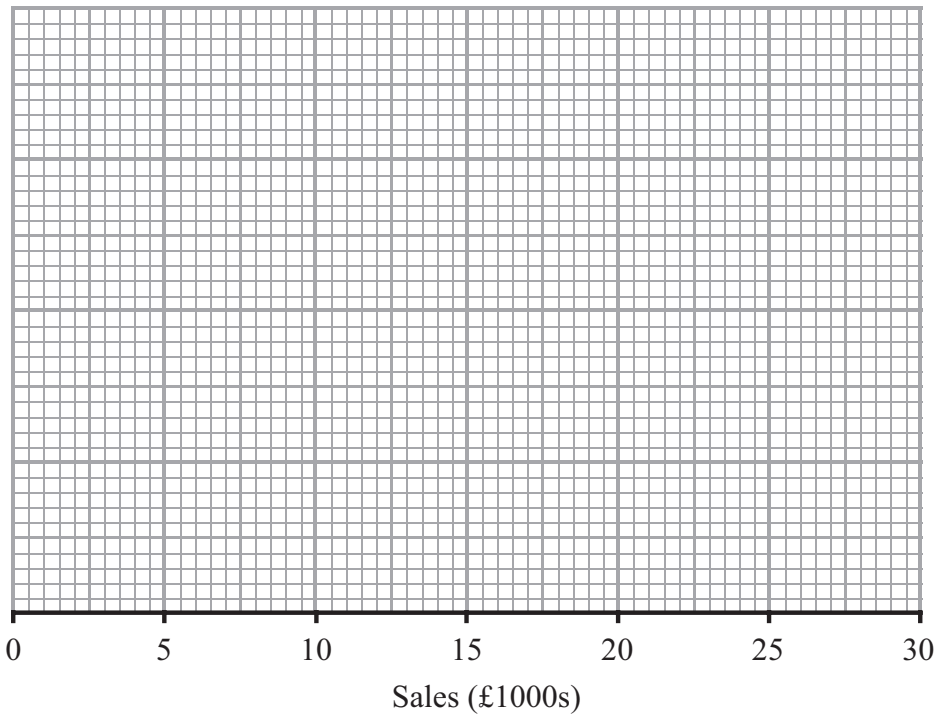
3. Over a long period of time a small company recorded the amount it received in sales per month. The results are summarised below.

	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 × interquartile range above the upper quartile
or 1.5 × 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)

Leave blank

Question 3 continued

Q3

(Total 9 marks)

Leave blank

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.

$$S_{vv} = 5.753 \qquad S_{pv} = 1.688 \qquad S_{pp} = 1.168 \qquad \overline{p} = 3.22 \qquad \overline{v} = 4.42$$

(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)

Leave blank

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

$$\sum t = 1414 \qquad \text{and} \qquad \sum t^2 = 69\,378$$

(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)

Leave blank

Question 5 continued

Q5

(Total 7 marks)

Leave blank

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 $P(1.5 < X_1 + X_2 \leqslant 3.5)$

(2)

Leave blank

Question 6 continued

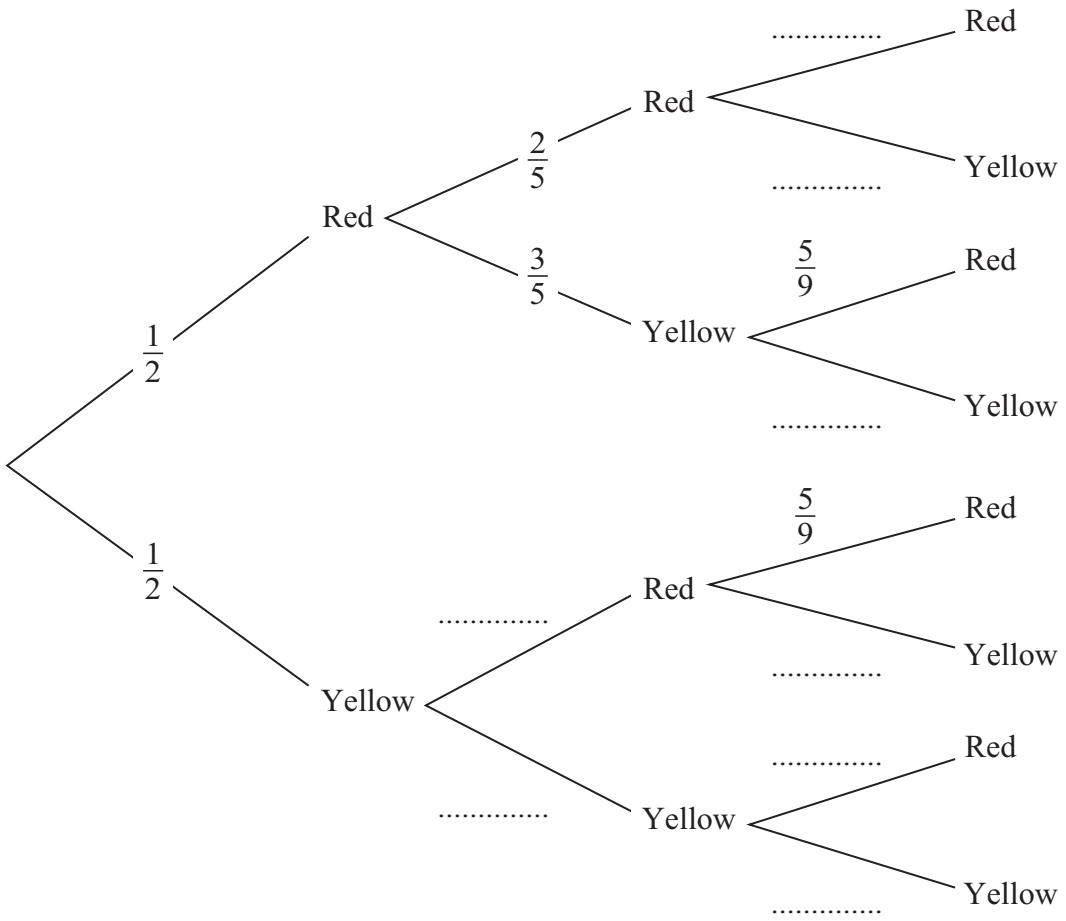
Leave blank

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) Complete the tree diagram shown below.

(4)



(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)

Leave blank

Question 7 continued

Leave blank

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 μ grams and standard deviation σ grams.

(c) Given that $P(Y < 160) = 0.99$ and $P(Y > 152) = 0.90$ find the value of μ and the value of σ.

(6)

Leave blank

Question 8 continued

Q8

(Total 12 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 20 May 2011 – Afternoon

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
8	
Total	

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 8 questions in this question paper. The total mark for this paper is 75.
There are 24 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

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

Printer's Log. No.
P38164A
W850/R6683/57570 3/3/3

P 3 8 1 6 4 A 0 1 2 4

Turn over

edexcel
advancing learning, changing lives

Leave blank

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)

Leave blank

Question 1 continued

Q1

(Total 7 marks)

Leave blank

2. The random variable $X \sim \mathrm{N}(\mu, 5^2)$ and $\mathrm{P}(X < 23) = 0.9192$

(a) Find the value of μ.

(4)

(b) Write down the value of $\mathrm{P}(\mu < X < 23)$.

(1)

Leave blank

3. The discrete random variable Y has 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 $P(3Y + 2 \geqslant 8)$.

(2)

Leave blank

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)

Leave blank

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

$$Q_1 = 8.5 \qquad Q_2 = 13.0 \qquad Q_3 = 21.0$$

(e) Describe, giving a reason, the skewness of the times on this occasion.

(2)

Leave blank

Question 5 continued

Leave blank

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

$\mathrm{P}(J) = 0.25$, $\mathrm{P}(J \cap K) = 0.15$ and $\mathrm{P}(J' \cap K') = 0.7$

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 (d).

(1)

Leave blank

Question 6 continued

Leave blank

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 $\sum f = 186$ $\sum h = 1085$ $S_{ff} = 39.5$ $S_{hh} = 139.875$ $\sum fh = 25291$)

(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 (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 (b) should not be used to estimate the teacher's height.

(1)

Leave blank

Question 7 continued

Leave blank

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 $Var(S)$.

(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)

Leave blank

Question 8 continued

Q8

(Total 17 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Tuesday 17 January 2012 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 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 should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

Printer's Log. No.
P40699A
W850/R6683/57570 5/4/4/5/5/4

P 4 0 6 9 9 A 0 1 2 4

Turn over

PEARSON

Leave blank

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.

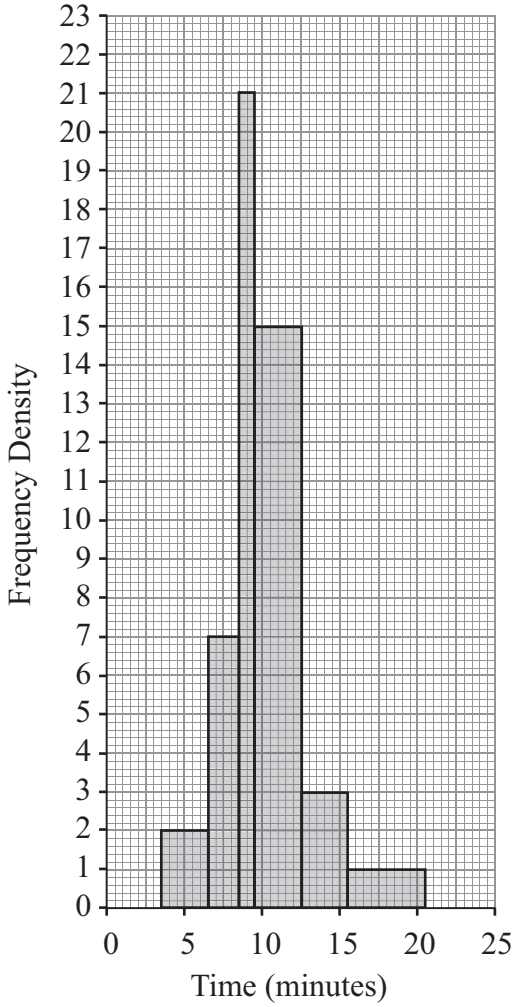


Figure 1

(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)

Leave blank

Question 1 continued

Q1

(Total 4 marks)

Leave blank

2. (a) State in words the relationship between two events R and S when $\mathrm{P}(R \cap S) = 0$

(1)

The events A and B are independent with $\mathrm{P}(A) = \frac{1}{4}$ and $\mathrm{P}(A \cup B) = \frac{2}{3}$

Find

(b) $\mathrm{P}(B)$

(4)

(c) $\mathrm{P}(A' \cap B)$

(2)

(d) $\mathrm{P}(B' \mid A)$

(2)

Leave blank

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) Var($7X - 5$)

(4)

Leave blank

Question 3 continued

Leave blank

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)

Leave blank

Question 4 continued

Leave blank

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)

Leave blank

Question 5 continued

Leave blank

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)

Leave blank

Question 6 continued

Leave blank

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)

Leave blank

Question 7 continued

Q7

(Total 10 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 18 May 2012 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 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 should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

Printer's Log. No.
P40105XA
W850/R6638/57570 5/5/2

P 4 0 1 0 5 X A 0 1 2 4

Turn over

PEARSON

Leave blank

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)

Leave blank

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)

Leave blank

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) Draw a scatter diagram on the axes on page 7 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 $(\overline{p}, \overline{t})$ and draw the regression line on your scatter diagram.

(2)

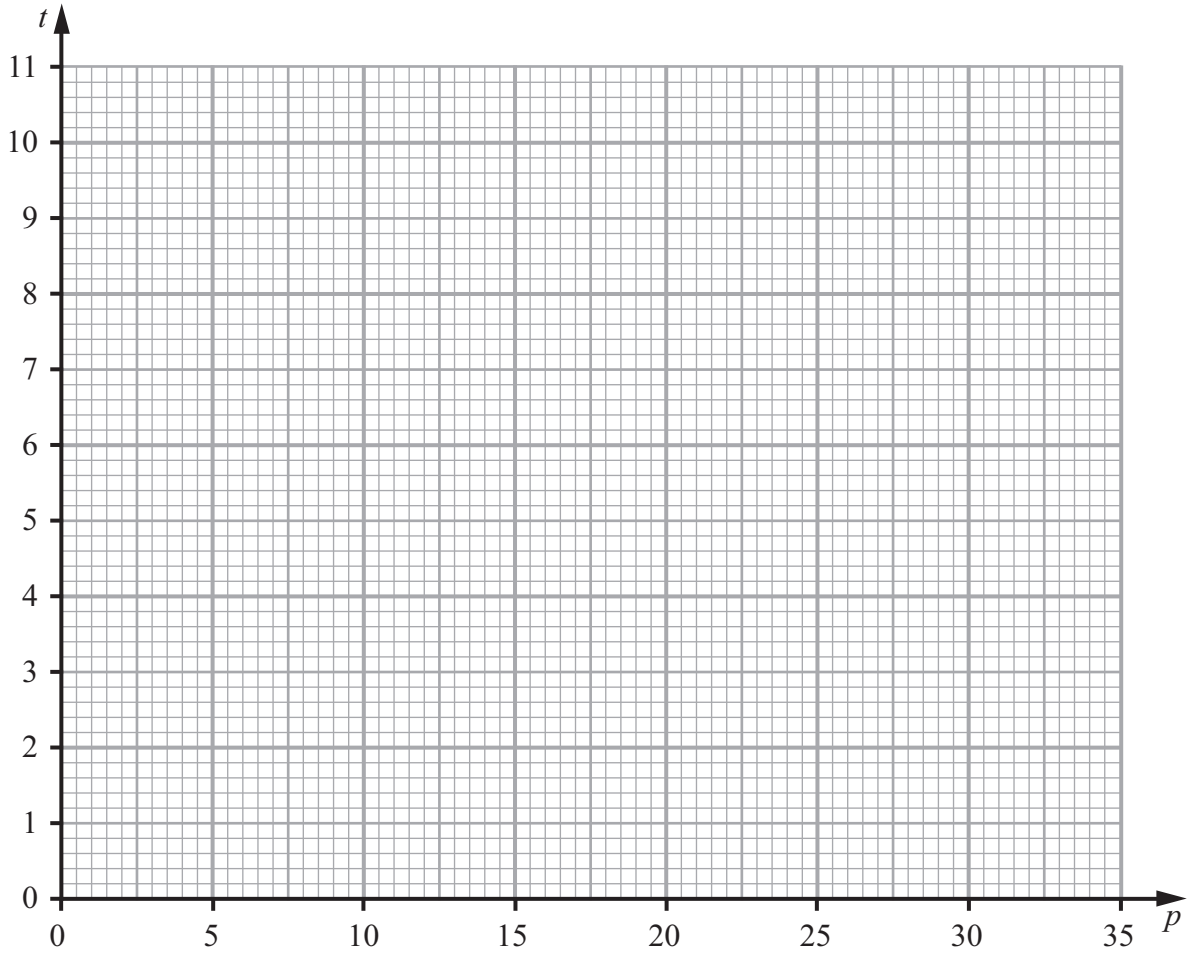
The scientist reviews similar studies and finds that pollutant levels above 16 are likely to result in the death of a chick soon after hatching.

(f) Estimate the minimum thinning of the shell that is likely to result in the death of a chick.

(2)

Leave blank

Question 3 continued



Leave blank

4.

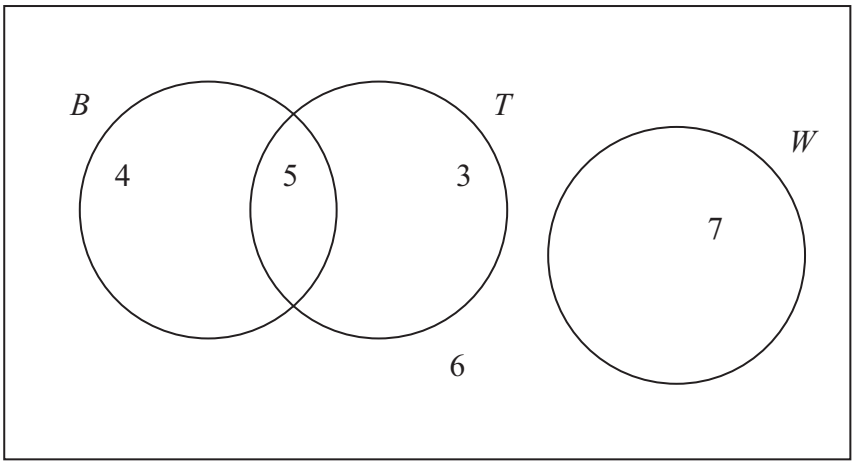


Figure 1

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)

(e) Given that this person travels to work by bicycle, find the probability that they will also take the train.

(2)

Leave blank

Question 4 continued

Leave blank

5.

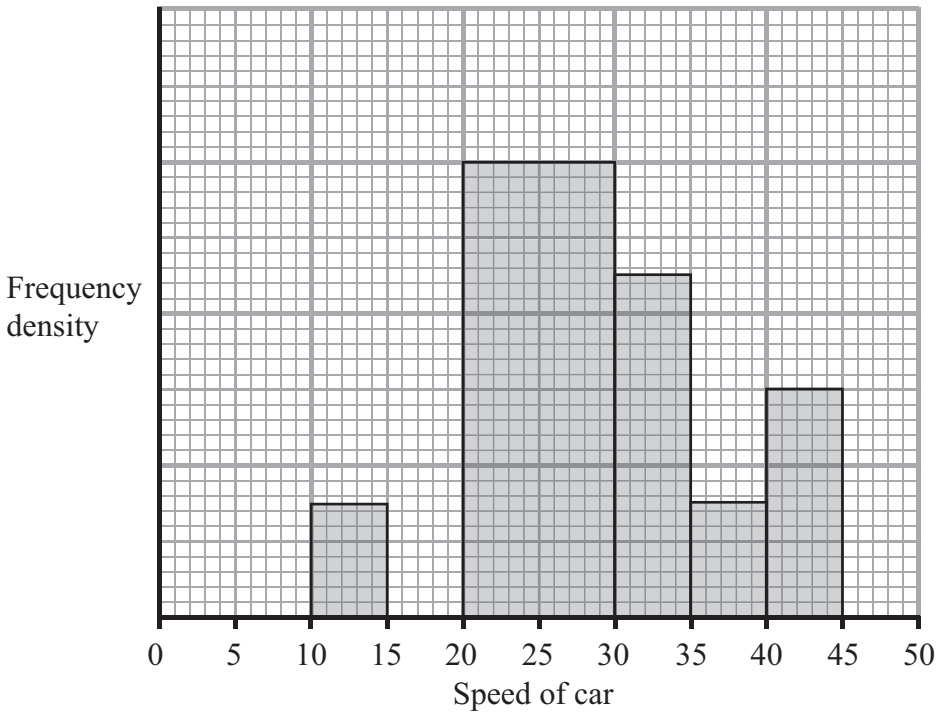


Figure 2

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.

(a) Calculate the number of cars that were exceeding the speed limit by at least 5 mph in the sample.

(4)

(b) Estimate the value of the mean speed of the cars in the sample.

(3)

(c) Estimate, to 1 decimal place, the value of the median speed of the cars in the sample.

(2)

(d) Comment on the shape of the distribution. Give a reason for your answer.

(2)

(e) 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)

Leave blank

Question 5 continued

Leave blank

6. The heights of an adult female population are normally distributed with mean 162 cm and standard deviation 7.5 cm.

(a) 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.

(b) 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,

(c) find the mean height of an adult male.

(4)

Leave blank

Question 6 continued

Leave blank

7. 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.

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

(3)

(b) 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.

(c) Find the probability that the soft toy has none of these 3 defects.

(2)

(d) Find the probability that the soft toy has exactly one of these 3 defects.

(4)

Leave blank

Question 7 continued

Q7

(Total 12 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 18 January 2013 – Afternoon

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 7 questions in this question paper. The total mark for this paper is 75.
There are 20 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

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

Printer's Log. No.
P41805A
W850/R6683/57570 5/5/5/6/

P 4 1 8 0 5 A 0 1 2 0

Turn over

PEARSON

Leave blank

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.

$$\sum t = 258 \qquad \sum t^2 = 8702 \qquad \sum g = 63.6 \qquad S_{gg} = 7.864 \qquad \sum gt = 1550.2$$

(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)

Leave blank

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

$$F(x) = \frac{x^3 + k}{40} \qquad x = 1, 2, 3$$

(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)

Leave blank

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)

Leave blank

Question 3 continued

Q3

(Total 10 marks)

Leave blank

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)

Leave blank

Question 4 continued

Q4

(Total 10 marks)

Leave blank

5. A survey of 100 households gave the following results for weekly income £y.

Income y (£)	Mid-point	Frequency f
$0 \leqslant y < 200$	100	12
$200 \leqslant y < 240$	220	28
$240 \leqslant y < 320$	280	22
$320 \leqslant y < 400$	360	18
$400 \leqslant y < 600$	500	12
$600 \leqslant y < 800$	700	8

(You may use $\sum fy^2 = 12\,452\,800$)

A histogram was drawn and the class $200 \leqslant y < 240$ 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 $320 \leqslant y < 400$

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

(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)

Leave blank

Question 5 continued

Leave blank

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 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)

Leave blank

Question 6 continued

Leave blank

7. Given that

$$P(A) = 0.35 , \quad P(B) = 0.45 \quad \text{and} \quad P(A \cap B) = 0.13$$

find

(a) $P(A \cup B)$

(2)

(b) $P(A' \mid B')$

(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)

Leave blank

Question 7 continued

Q7

(Total 12 marks)

TOTAL FOR PAPER: 75 MARKS

END

Centre No.						Paper Reference							Surname	Initial(s)
Candidate No.						6	6	8	3	/	0	1R	Signature	

Paper Reference(s)

6683/01R

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 17 May 2013 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
Total	

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 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 should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

Printer's Log. No.
P43956A
W850/R6638/57570 5/5/

P 4 3 9 5 6 A 0 1 2 4

Turn over

PEARSON

Leave blank

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)

Leave blank

Question 1 continued

Leave blank

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)

Q2

(Total 4 marks)

Leave blank

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)
$0 \leqslant y < 5$	16	2.5
$5 \leqslant y < 10$	24	7.5
$10 \leqslant y < 15$	14	12.5
$15 \leqslant y < 25$	12	20
$25 \leqslant y < 35$	4	30

(You may use $\sum fx = 755$ and $\sum fx^2 = 12037.5$)

A histogram has been drawn to represent these data.

The bar representing the yield $5 \leqslant y < 10$ 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 $15 \leqslant y < 25$

(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)

Leave blank

Question 3 continued

Leave blank

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 μ minutes.

Given that $P(X < \mu - 15) = 0.35$

(c) find $P(X > \mu + 15 \mid X > \mu - 15)$.

(3)

Leave blank

Question 4 continued

Leave blank

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)

Leave blank

Question 5 continued

Leave blank

6.

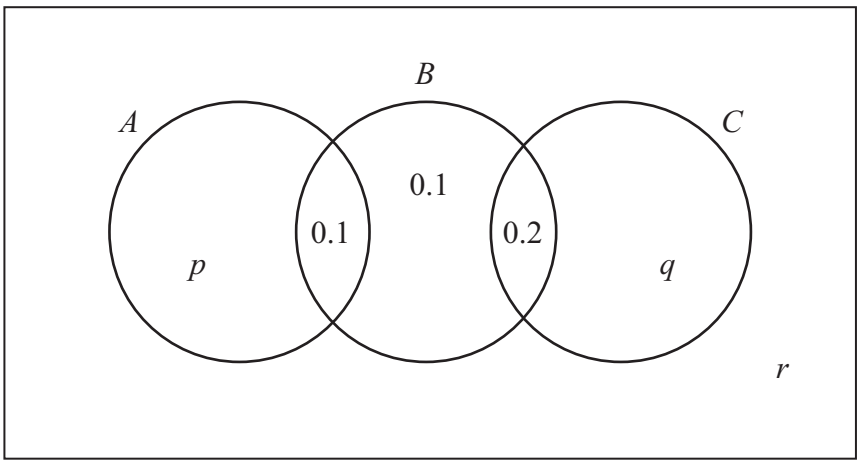


Figure 1

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)

Leave blank

Question 6 continued

Leave blank

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 $Var(S)$.

(2)

(d) Find

(i) $E(5S - 3)$,

(ii) $Var(5S - 3)$.

(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)

Leave blank

Question 7 continued

Q7

(Total 18 marks)

TOTAL FOR PAPER: 75 MARKS

END

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

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 17 May 2013 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
Total	

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

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

Advice to Candidates

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

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

P 4 2 8 3 1 A 0 1 2 4

Turn over

PEARSON

Leave blank

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

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

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

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

(3)

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

(2)

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

(1)

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

(4)

(e) Interpret the value of b.

(1)

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

(2)

Leave blank

Question 1 continued

Leave blank

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

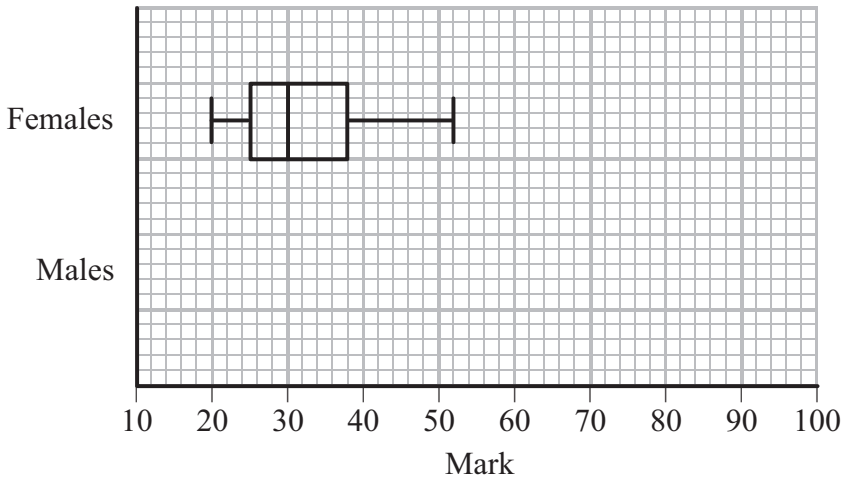


Figure 1

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

(1)

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

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

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

(3)

An outlier is a mark that is

either more than 1.5 × interquartile range above the upper quartile

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

Leave blank

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

(5)

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

(2)

Leave blank

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

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

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

An employee is selected at random.

Find

(a) P(H)

(2)

(b) $\mathrm{P}\left([F \cap W]'\right)$

(2)

(c) P($W \mid C$)

(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)

Leave blank

Question 3 continued

Leave blank

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

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

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

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

(5)

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

(2)

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

(1)

(d) Estimate the interquartile range of this distribution.

(2)

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

(1)

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

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

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

(3)

Leave blank

Question 4 continued

Leave blank

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

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

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

(5)

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

(1)

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

(3)

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

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

(d) Find the value of k.

(1)

(e) Find the probability distribution of Y.

(3)

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

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

(2)

Leave blank

Question 5 continued

Leave blank

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

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

(a) the value of μ

(3)

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

(3)

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

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

(4)

Leave blank

Question 6 continued

Q6

(Total 10 marks)

TOTAL FOR PAPER: 75 MARKS

END

Statistics S1

Probability

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cap B) = P(A)\,P(B \mid A)$$

$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B \mid A)\,P(A) + P(B \mid A')\,P(A')}$$

Discrete distributions

For a discrete random variable X taking values x_i with probabilities $P(X = x_i)$

Expectation (mean): $E(X) = \mu = \Sigma x_i\, P(X = x_i)$

Variance: $\mathrm{Var}(X) = \sigma^2 = \Sigma(x_i - \mu)^2\, P(X = x_i) = \Sigma x_i^2\, P(X = x_i) - \mu^2$

For a function $g(X)$: $E(g(X)) = \Sigma g(x_i)\, P(X = x_i)$

Continuous distributions

Standard continuous distribution:

Distribution of X	P.D.F.	Mean	Variance
Normal $N(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2

Correlation and regression

For a set of n pairs of values (x_i, y_i)

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

$$S_{yy} = \Sigma(y_i - \bar{y})^2 = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}$$

$$S_{xy} = \Sigma(x_i - \bar{x})(y_i - \bar{y}) = \Sigma x_i y_i - \frac{(\Sigma x_i)(\Sigma y_i)}{n}$$

The product moment correlation coefficient is

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\{\Sigma(x_i - \bar{x})^2\}\{\Sigma(y_i - \bar{y})^2\}}} = \frac{\Sigma x_i y_i - \frac{(\Sigma x_i)(\Sigma y_i)}{n}}{\sqrt{\left(\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right)\left(\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}\right)}}$$

The regression coefficient of y on x is $b = \frac{S_{xy}}{S_{xx}} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$

Least squares regression line of y on x is $y = a + bx$ where $a = \bar{y} - b\bar{x}$

THE NORMAL DISTRIBUTION FUNCTION

The function tabulated below is $\Phi(z)$, defined as $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} \, dt$.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.02	0.9783
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.04	0.9793
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.06	0.9803
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.08	0.9812
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.10	0.9821
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.12	0.9830
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.14	0.9838
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.16	0.9846
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.18	0.9854
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	2.20	0.9861
0.11	0.5438	0.61	0.7291	1.11	0.8665	1.61	0.9463	2.22	0.9868
0.12	0.5478	0.62	0.7324	1.12	0.8686	1.62	0.9474	2.24	0.9875
0.13	0.5517	0.63	0.7357	1.13	0.8708	1.63	0.9484	2.26	0.9881
0.14	0.5557	0.64	0.7389	1.14	0.8729	1.64	0.9495	2.28	0.9887
0.15	0.5596	0.65	0.7422	1.15	0.8749	1.65	0.9505	2.30	0.9893
0.16	0.5636	0.66	0.7454	1.16	0.8770	1.66	0.9515	2.32	0.9898
0.17	0.5675	0.67	0.7486	1.17	0.8790	1.67	0.9525	2.34	0.9904
0.18	0.5714	0.68	0.7517	1.18	0.8810	1.68	0.9535	2.36	0.9909
0.19	0.5753	0.69	0.7549	1.19	0.8830	1.69	0.9545	2.38	0.9913
0.20	0.5793	0.70	0.7580	1.20	0.8849	1.70	0.9554	2.40	0.9918
0.21	0.5832	0.71	0.7611	1.21	0.8869	1.71	0.9564	2.42	0.9922
0.22	0.5871	0.72	0.7642	1.22	0.8888	1.72	0.9573	2.44	0.9927
0.23	0.5910	0.73	0.7673	1.23	0.8907	1.73	0.9582	2.46	0.9931
0.24	0.5948	0.74	0.7704	1.24	0.8925	1.74	0.9591	2.48	0.9934
0.25	0.5987	0.75	0.7734	1.25	0.8944	1.75	0.9599	2.50	0.9938
0.26	0.6026	0.76	0.7764	1.26	0.8962	1.76	0.9608	2.55	0.9946
0.27	0.6064	0.77	0.7794	1.27	0.8980	1.77	0.9616	2.60	0.9953
0.28	0.6103	0.78	0.7823	1.28	0.8997	1.78	0.9625	2.65	0.9960
0.29	0.6141	0.79	0.7852	1.29	0.9015	1.79	0.9633	2.70	0.9965
0.30	0.6179	0.80	0.7881	1.30	0.9032	1.80	0.9641	2.75	0.9970
0.31	0.6217	0.81	0.7910	1.31	0.9049	1.81	0.9649	2.80	0.9974
0.32	0.6255	0.82	0.7939	1.32	0.9066	1.82	0.9656	2.85	0.9978
0.33	0.6293	0.83	0.7967	1.33	0.9082	1.83	0.9664	2.90	0.9981
0.34	0.6331	0.84	0.7995	1.34	0.9099	1.84	0.9671	2.95	0.9984
0.35	0.6368	0.85	0.8023	1.35	0.9115	1.85	0.9678	3.00	0.9987
0.36	0.6406	0.86	0.8051	1.36	0.9131	1.86	0.9686	3.05	0.9989
0.37	0.6443	0.87	0.8078	1.37	0.9147	1.87	0.9693	3.10	0.9990
0.38	0.6480	0.88	0.8106	1.38	0.9162	1.88	0.9699	3.15	0.9992
0.39	0.6517	0.89	0.8133	1.39	0.9177	1.89	0.9706	3.20	0.9993
0.40	0.6554	0.90	0.8159	1.40	0.9192	1.90	0.9713	3.25	0.9994
0.41	0.6591	0.91	0.8186	1.41	0.9207	1.91	0.9719	3.30	0.9995
0.42	0.6628	0.92	0.8212	1.42	0.9222	1.92	0.9726	3.35	0.9996
0.43	0.6664	0.93	0.8238	1.43	0.9236	1.93	0.9732	3.40	0.9997
0.44	0.6700	0.94	0.8264	1.44	0.9251	1.94	0.9738	3.50	0.9998
0.45	0.6736	0.95	0.8289	1.45	0.9265	1.95	0.9744	3.60	0.9998
0.46	0.6772	0.96	0.8315	1.46	0.9279	1.96	0.9750	3.70	0.9999
0.47	0.6808	0.97	0.8340	1.47	0.9292	1.97	0.9756	3.80	0.9999
0.48	0.6844	0.98	0.8365	1.48	0.9306	1.98	0.9761	3.90	1.0000
0.49	0.6879	0.99	0.8389	1.49	0.9319	1.99	0.9767	4.00	1.0000
0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772		

PERCENTAGE POINTS OF THE NORMAL DISTRIBUTION

The values z in the table are those which a random variable $Z \sim N(0, 1)$ exceeds with probability p; that is, $P(Z > z) = 1 - \Phi(z) = p$.

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905