

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Tuesday 5 June 2007 – Afternoon

Time: 1 hour 30 minutes

Materials required for examination

Mathematical Formulae (Green)

Items included with question papers

Nil

Candidates may use any calculator 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

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

Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.

Full marks may be obtained for answers to ALL questions.

There are 7 questions in this question paper. The total mark for this paper is 75.

Advice to Candidates

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

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

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

$$S_{xx} = 113573, \quad S_{yy} = 8.657, \quad S_{xy} = -808.917$$

- (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)
-

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.

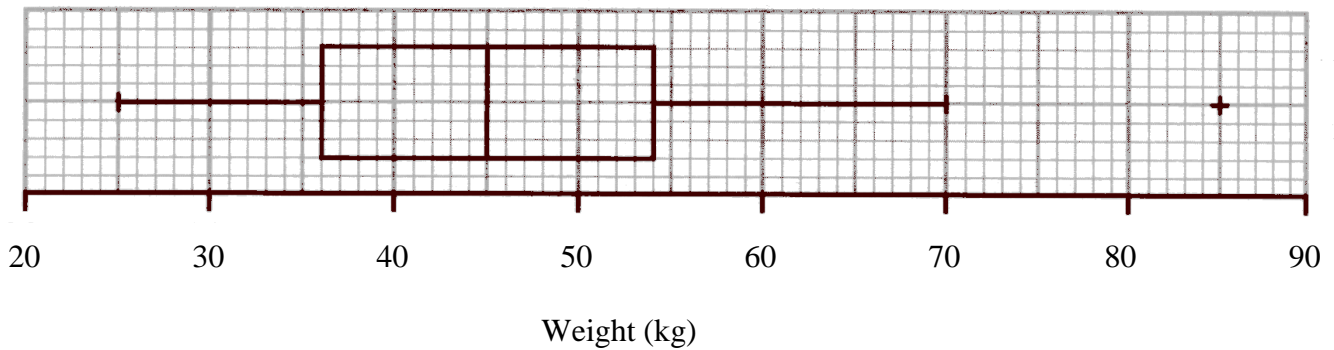


Figure 1

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)

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

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

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

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

The following data is obtained

Chocolate brand	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>
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 graph paper, 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)

4. A survey of the reading habits of some students revealed that, on a regular basis, 25% read quality newspapers, 45% read tabloid newspapers and 40% do not read newspapers at all.
- (a) Find the proportion of students who read both quality and tabloid newspapers. **(3)**

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

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

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

5.

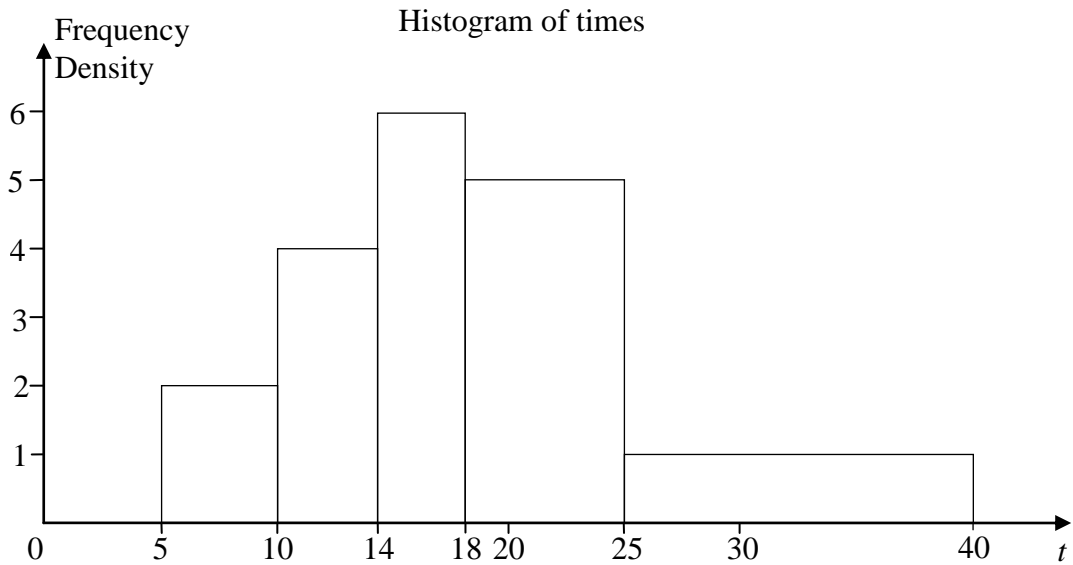


Figure 2

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

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

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

(2)

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

(2)

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

(4)

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

(3)

(e) Find the median and quartiles for t .

(4)

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

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

(2)

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

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

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

7. The random variable X has probability distribution

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

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

Find

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

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

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

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

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

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

TOTAL FOR PAPER: 75 MARKS

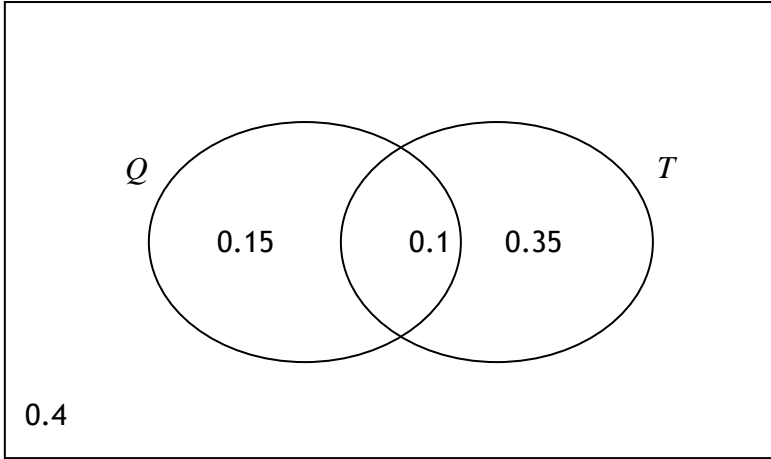
END

**June 2007
6683 Statistics S1
Mark Scheme**

Question Number	Scheme	Marks
1. (a)	$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{-808.917}{\sqrt{113573 \times 8.657}}$ $= -0.81579\dots$	<p style="text-align: center;">M1</p> <p style="text-align: center;">A1 (2)</p>
(b)	Houses are <u>cheaper</u> further away from the station or equivalent statement	<p style="text-align: center;">B1 (1)</p>
(c)	-0.816	<p style="text-align: center;">B1 ∫ (1)</p> <p style="text-align: right;">Total 4 marks</p>
Notes:		
1(a)	<p>M1 for knowing formula and clear attempt to sub in correct values from question. Root required for method. Anything that rounds to -0.82 for A1. Correct answer with no working award 2/2</p>	
(b)	<p>Context based on negative correlation only required. Accept <u>Houses are more expensive</u> closer to the <u>station</u> or equivalent statement. Require 'house prices' or 'station' and a clear correct comparison.</p>	
(c)	<p>Accept anything that rounds to -0.82 or 'the same' or 'unchanged' or equivalent. Award B1 if value quoted same as answer to (a).</p>	

Question Number	Scheme	Marks
2(a)	$\frac{1}{2}$	<p>B1 (1)</p>
(b)	54	<p>B1 (1)</p>
(c)	<p>+ is an 'outlier' or 'extreme value' Any heavy musical instrument or a statement that the instrument is heavy</p>	<p>B1 B1 (2)</p>
(d)	<p>$Q_3 - Q_2 = Q_2 - Q_1$ so symmetrical or no skew</p>	<p>B1 Dependent - only award if B1 above B1 (2)</p>
(e)	<p>$P(W < 54) = 0.75$ (or $P(W > 54) = 0.25$) or correctly labelled and shaded diagram</p> $\frac{54 - 45}{\sigma} = 0.67$ <p>$\sigma = 13.43\dots$</p>	<p>M1 M1B1 A1 (4)</p>
Total 10 marks		
Notes 2(a)	<p>Accept 50% or half or 0.5. Units not required.</p>	
(b)	<p>Correct answer only. Units not required.</p>	
(c)	<p>'Anomaly' only award B0 Accept '85kg was heaviest instrument on the trip' or equivalent for second B1. Examples of common acceptable instruments; double bass, cello, harp, piano, drums, tuba Examples of common unacceptable instruments: violin, viola, trombone, trumpet, french horn, guitar</p>	
(d)	<p>'Quartiles equidistant from median' or equivalent award B1 then symmetrical or no skew for B1 Alternative: 'Positive tail is longer than negative tail' or 'median closer to lowest value' or equivalent so slight positive skew. B0 for 'evenly' etc. instead of 'symmetrical' B0 for 'normal' only</p>	
(e)	<p>Please note that B mark appears first on ePEN First line might be missing so first M1 can be implied by second. Second M1 for standardising with sigma and equating to z value NB Using 0.7734 should not be awarded second M1 Anything which rounds to 0.67 for B1. Accept 0.675 if to 3sf obtained by interpolation Anything that rounds to 13.3 - 13.4 for A1.</p>	

<p>3(a)</p> <p>(b)</p> <p>(c)</p> <p>(d)</p> <p>(e)</p>	<p>Use overlay</p> $S_{xy} = 28750 - \frac{315 \times 620}{8} = 4337.5$ <p>**answer given** so award for method</p> $S_{xx} = 15225 - \frac{315^2}{8} = 2821.875$ $b = \frac{4377.5}{S_{xx}} = 1.537... = 1.5$ $a = \bar{y} - b\bar{x} = \frac{620}{8} - b \frac{315}{8} = 16.97... = 17.0$ <p>Use overlay</p> <p>Brand D, since a long way above / from the line Using line: $y = 17 + 35 \times 1.5 = 69.5$</p> <p style="text-align: right;">dependent upon 'Brand D' above</p>	<p>B2 (2)</p> <p>M1</p> <p>M1A1 (3)</p> <p>M1,A1</p> <p>M1,A1 (4)</p> <p>B1 B1 (2)</p> <p>B1 B1 M1A1 (4)</p> <p style="text-align: right;">Total 15 marks</p>
<p>Notes:</p> <p>3(a)</p> <p>(b)</p> <p>(c)</p> <p>(d)</p> <p>(e)</p>	<p>Points B2, within 1 small square of correct point, subtract 1 mark each error minimum 0.</p> <p>Anything that rounds to 2820 for A1</p> <p>Anything that rounds to 1.5 and 17.0 (accept 17)</p> <p>Follow through for the intercept for first B1.. Correct slope of straight line for second B1.</p> <p>Anything that rounds to 69p-71p for final A1. Reading from graph is acceptable for M1A1. If value read from graph at $x = 35$ is answer given but out of range, then award M1A0.</p>	

<p>4(a)</p> <p>(b)</p> <p>(c)</p>	$P(Q \cup T) = 0.6$ $P(Q) + P(T) - P(Q \cap T) = 0.6$ $P(Q \cap T) = 0.1$  $P(Q \cap T' Q \cup T) = \frac{0.15}{0.60} = \frac{1}{4} \text{ or } 0.25 \text{ or } 25\%$	<p>B1 M1 A1 (3)</p> <p>Venn 0.15, 0.35 0.4 and box M1 A1 B1 (3)</p> <p>M1A1 A1 (3)</p> <p>Total 9 marks</p>
<p>Notes:</p> <p>4(a)</p> <p>(b)</p> <p>(c)</p>	<p>B1 for 0.6 M1 for use of $P(Q) + P(T) - P(Q \cap T) = P(Q \cup T)$ 0.1 Correct answer only for A1 Alternative: (25+45+40=)110% B1 110-100=10% M1A1 0.1 stated clearly as the final answer with no working gets 3/3</p> <p>Two intersecting closed curves for M1, no box required. At least one label (Q or T) required for first A1. Follow through (0.25-‘their 0.1’) and (0.45-‘their 0.1’) for A1. 0.4 and box required, correct answer only for B1 Using %, whole numbers in Venn diagram without % sign, whole numbers in correct ratio all OK</p> <p>Require fraction with denominator 0.6 or their equivalent from Venn diagram for M1 Follow through their values in fraction for A1 Final A1 is correct answer only. <u>No working no marks.</u></p>	

5(a)	18-25 group, area=7x5=35 25-40 group, area=15x1=15	B1 B1 (2)
(b)	$(25-20) \times 5 + (40-25) \times 1 = 40$	M1A1 (2)
(c)	Mid points are 7.5, 12, 16, 21.5, 32.5 $\sum f = 100$ $\frac{\sum ft}{\sum f} = \frac{1891}{100} = 18.91$	M1 B1 M1A1 (4)
(d)	$\sigma_t = \sqrt{\frac{41033}{100} - \bar{t}^2}$ $\sqrt{\frac{n}{n-1} \left(\frac{41033}{100} - \bar{t}^2 \right)}$ alternative OK $\sigma_t = \sqrt{52.74\dots} = 7.26$	M1 M1 A1 (3)
(e)	$Q_2 = 18$ or 18.1 if (n+1) used $Q_1 = 10 + \frac{15}{16} \times 4 = 13.75$ or 15.25 numerator gives 13.8125 $Q_3 = 18 + \frac{25}{35} \times 7 = 23$ or 25.75 numerator gives 23.15	B1 M1A1 A1 (4)
(f)	0.376... Positive skew	B1 B1 (2)
Total 17 marks		
Notes: 5(b)	5x5 is enough evidence of method for M1. Condone 19.5, 20.5 instead of 20 etc. Award 2 if 40 seen.	
(c)	Look for working for this question in part (d) too. Use of some mid-points, at least 3 correct for M1. These may be tabulated in (d). Their $\frac{\sum ft}{\sum f}$ for M1 and anything that rounds to 18.9 for A1.	
(d)	Clear attempt at $\frac{41033}{100} - \bar{t}^2$ or $\frac{n}{n-1} \left(\frac{41033}{100} - \bar{t}^2 \right)$ alternative for first M1. They may use their \bar{t} and gain the method mark. Square root of above for second M1 Anything that rounds to 7.3 for A1.	
(e)	Clear attempt at either quartile for M1 These will take the form 'their lower limit'+ correct fraction x 'their class width'. Anything that rounds to 13.8 for lower quartile.	
(f)	23 or anything that rounds to 23.2 dependent upon method used. Anything that rounds to 0.38 for B1 or 0.33 for B1 if (n+1) used. Correct answer or correct statement that follows from their value for B1.	

6(a)	$P(X > 25) = P\left(Z > \frac{25-20}{4}\right)$ $= P(Z > 1.25)$ $= 1 - 0.8944$ $= 0.1056$	M1 M1 A1 (3)
(b)	$P(X < 20) = 0.5 \text{ so } P(X < d) = 0.5 + 0.4641 = 0.9641$ $P(Z < z) = 0.9641, z = 1.80$ $\frac{d-20}{4} = 1.80$ $d = 27.2$	B1 B1 M1 A1 (4)
Total 7 marks		
Notes:		
(a)	Standardise with 20 and 4 for M1, allow numerator 20-25 1- probability for second M1 Anything that rounds to 0.106 for A1. Correct answer with no working award 3/3	
(b)	0.9641 seen or implied by 1.80 for B1 1.80 seen for B1 Standardise with 20 and 4 and equate to z value for M1 Z=0.8315 is M0 Anything that rounds to 27.2 for final A1. Correct answer with no working 4/4	

<p>7(a)</p> <p>(b)</p> <p>(c)</p> <p>(d)</p> <p>(e)</p> <p>(f)</p>	$p + q = 0.45$ $\sum xP(X = x) = 4.5$ $3p + 7q = 1.95$ <p>Attempt to solve equations in (a)</p> $q = 0.15$ $p = 0.30$ $P(4 < X < 7) = P(5) + P(7)$ $= 0.2 + q = 0.35$ $\text{Var}(X) = E(X^2) - [E(X)]^2 = 27.4 - 4.5^2$ $= 7.15$ $E(19 - 4X) = 19 - 4 \times 4.5 = 1$ $\text{Var}(19 - 4X) = 16\text{Var}(X)$ $= 16 \times 7.15 = 114.4$	<p>B1</p> <p>M1</p> <p>A1</p> <p>(3)</p> <p>M1</p> <p>A1</p> <p>A1</p> <p>(3)</p> <p>M1</p> <p>A1</p> <p>(2)</p> <p>M1</p> <p>A1</p> <p>(2)</p> <p>B1</p> <p>(1)</p> <p>M1</p> <p>A1</p> <p>(2)</p> <p>Total 13 marks</p>
<p>Notes:</p> <p>7(a)</p> <p>(b)</p> <p>(c)</p> <p>(d)</p> <p>(f)</p>	<p>$0.55 + p + q = 1$ award B1. Not seen award B0.</p> <p>$0.2 + 3p + 1 + 7q + 1.35 = 4.5$ or equivalent award M1A1</p> <p>$3p + 7q + k = 4.5$ award M1.</p> <p>Attempt to solve must involve 2 linear equations in 2 unknowns Correct answers only for accuracy. Correct answers with no working award 3/3</p> <p>Follow through accuracy mark for their q, $0 < q < 0.8$</p> <p>Attempt to substitute <u>given</u> values <u>only</u> into correct formula for M1. 7.15 only for A1 7.15 seen award 2/2</p> <p>Accept 'invisible brackets' i.e. $-4^2 \text{Var}(X)$ provided answer positive. Anything that rounds to 114 for A1.</p>	