Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced Subsidiary

Thursday 27 May 2010 – Morning

Time: 1 hour 30 minutes

Materials required for examination	Items included with question papers
Mathematical Formulae (Pink)	Nil

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

Instructions to Candidates

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.
Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

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

Advice to Candidates

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

H35395A

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:

$$S_{xx} = 1022500, \qquad S_{yy} = 130.9, \qquad S_{xy} = 8825.$$

(a) Calculate the product moment correlation coefficient for these data.

(2)

(b) Interpret the value of the correlation coefficient.

(1)

Helen was given the same data to analyse. In view of the large numbers involved she decided to divide the attendance figures by 100. She then calculated the product moment correlation coefficient between $\frac{x}{100}$ and y.

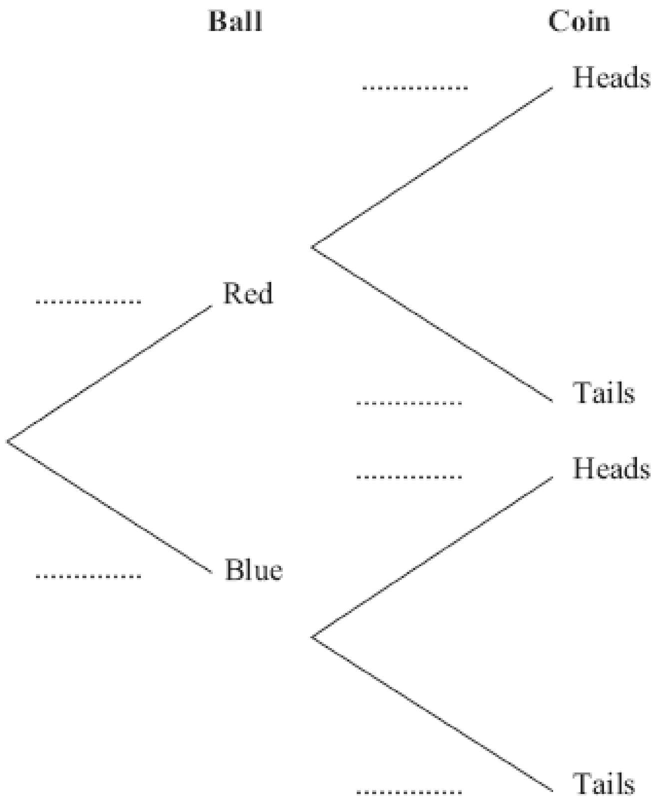
(c) Write down the value Helen should have obtained.

(1)

2. An experiment consists of selecting a ball from a bag and spinning a coin. The bag contains 5 red balls and 7 blue balls. A ball is selected at random from the bag, its colour is noted and then the ball is returned to the bag.

When a red ball is selected, a biased coin with probability $\frac{2}{3}$ of landing heads is spun.
When a blue ball is selected a fair coin is spun.

(*a*) Copy and complete the tree diagram below to show the possible outcomes and associated probabilities.



(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)

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

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

where a is a constant.

(*a*) Find the value of a.

(2)

(*b*) Write down $E(X)$.

(1)

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

(3)

The random variable $Y = 6 - 2X$.

(*d*) Find $\text{Var}(Y)$.

(2)

(*e*) Calculate $P(X \geq Y)$.

(3)

4. The Venn diagram in Figure 1 shows the number of students in a class who read any of 3 popular magazines A, B and C.

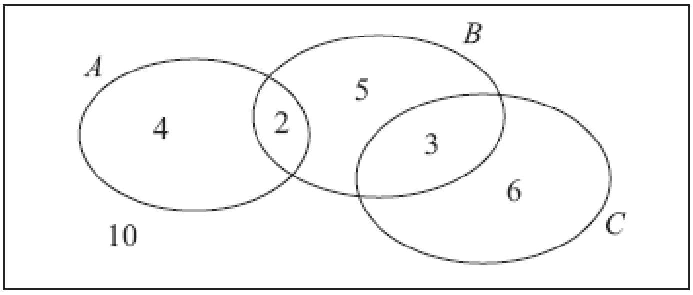


Figure 1

One of these students is selected at random.

(a) Show that the probability that the student reads more than one magazine is $\frac{1}{6}$.

(2)

(b) Find the probability that the student reads A or B (or both).

(2)

(c) Write down the probability that the student reads both A and C.

(1)

Given that the student reads at least one of the magazines,

(d) find the probability that the student reads C.

(2)

(e) Determine whether or not reading magazine B and reading magazine C are statistically independent.

(3)

5. A teacher selects a random sample of 56 students and records, to the nearest hour, the time spent watching television in a particular week.

Hours	1–10	11–20	21–25	26–30	31–40	41–59
Frequency	6	15	11	13	8	3
Mid-point	5.5	15.5		28		50

(*a*) Find the mid-points of the 21−25 hour and 31−40 hour groups.

(2)

A histogram was drawn to represent these data. The 11−20 group was represented by a bar of width 4 cm and height 6 cm.

(*b*) Find the width and height of the 26−30 group.

(3)

(*c*) Estimate the mean and standard deviation of the time spent watching television by these students.

(5)

(*d*) Use linear interpolation to estimate the median length of time spent watching television by these students.

(2)

The teacher estimated the lower quartile and the upper quartile of the time spent watching television to be 15.8 and 29.3 respectively.

(*e*) State, giving a reason, the skewness of these data.

(2)

6. A travel agent sells flights to different destinations from *Beerow* airport. The distance d, measured in 100 km, of the destination from the airport and the fare £f are recorded for a random sample of 6 destinations.

Destination	*A*	*B*	*C*	*D*	*E*	*F*
d	2.2	4.0	6.0	2.5	8.0	5.0
f	18	20	25	23	32	28

[You may use $\sum d^2 = 152.09$ $\sum f^2 = 3686$ $\sum fd = 723.1$]

(*a*) On graph paper, draw a scatter diagram to illustrate this information.

(2)

(*b*) Explain why a linear regression model may be appropriate to describe the relationship between f and d.

(1)

(*c*) Calculate S_{dd} and S_{fd}.

(4)

(*d*) Calculate the equation of the regression line of f on d giving your answer in the form $f = a + bd$.

(4)

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

(1)

Jane is planning her holiday and wishes to fly from *Beerow* airport to a destination t km away. A rival travel agent charges 5p per km.

(*f*) Find the range of values of t for which the first travel agent is cheaper than the rival.

(2)

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

(*a*) Find the probability that a randomly selected employee has a journey to work of more than 20 km.

(3)

(*b*) Find the upper quartile, Q_3, of D.

(3)

(*c*) Write down the lower quartile, Q_1, of D.

(1)

An outlier is defined as any value of D such that $D < h$ or $D > k$ where

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

(*d*) Find the value of h and the value of k.

(2)

An employee is selected at random.

(*e*) Find the probability that the distance travelled to work by this employee is an outlier.

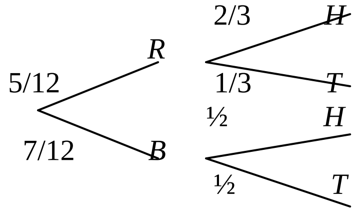
(3)

TOTAL FOR PAPER: 75 MARKS

END

June 2010
Statistics S1 6683
Mark Scheme

Question Number	Scheme	Marks
Q1 (a)	$r = \frac{8825}{\sqrt{1022500 \times 130.9}}$, = awrt **<u>0.763</u>**	M1 A1 (2)
(b)	Teams with high attendance scored more goals (oe, statement in context)	B1 (1)
(c)	0.76(3)	B1ft (1)
		Total 4
(a)	M1 for a correct expression, square root required Correct answer award 2/2	
(b)	Context required (attendance and goals). Condone causality. B0 for 'strong positive correlation between attendance and goals' on its own oe	
(c)	Value required. Must be a correlation coefficient between -1 and +1 inclusive. B1ft for 0.76 or better or same answer as their value from part (a) to at least 2 d.p.	

Question Number	Scheme		Marks
Q2 (a)		P(R) and P(B) 2^{nd} set of probabilities	B1 B1 (2)
(b)	$\text{P}(H) = \frac{5}{12}\times\frac{2}{3}+\frac{7}{12}\times\frac{1}{2}, = \frac{41}{72}$ or awrt 0.569		M1 A1 (2)
(c)	$\text{P}(R\mid H) = \frac{\frac{5}{12}\times\frac{2}{3}}{"\frac{41}{72}"}, = \frac{20}{41}$ or awrt 0.488		M1 A1ft A1 (3)
(d)	$\left(\frac{5}{12}\right)^2+\left(\frac{7}{12}\right)^2$ $=\frac{25}{144}+\frac{49}{144}=\frac{74}{144}$ or $\frac{37}{72}$ or awrt 0.514		M1 A1ft A1 (3) **Total 10**
(a)	1^{st} B1 for the probabilities on the first 2 branches. Accept $0.41\dot{6}$ and $0.58\dot{3}$ 2^{nd} B1 for probabilities on the second set of branches. Accept $0.\dot{6}$, $0.\dot{3}$, 0.5 and $\frac{1.5}{3}$ Allow exact decimal equivalents using clear recurring notation if required.		
(b)	M1 for an expression for P(H) that follows through their sum of two products of **probabilities** from their tree diagram		
(c) Formula seen Formula not seen	M1 for $\frac{\text{P}(R\cap H)}{\text{P}(H)}$ with denominator their (b) substituted e.g. $\frac{\text{P}(R\cap H)}{\text{P}(H)} = \frac{\frac{5}{12}}{\text{(their (b))}}$ award M1. M1 for $\frac{\text{probability}\times\text{probability}}{\text{their } b}$ but M0 if fraction repeated e.g. $\frac{\frac{5}{12}\times\frac{2}{3}}{\frac{2}{3}}$. 1^{st} A1ft for a fully correct expression or correct follow through 2^{nd} A1 for $\frac{20}{41}$ o.e.		
(d)	M1 for $\left(\frac{5}{12}\right)^2$ or $\left(\frac{7}{12}\right)^2$ can follow through their equivalent values from tree diagram 1^{st} A1 for both values correct or follow through from their original tree and + 2^{nd} A1 for a correct answer Special Case $\frac{5}{12}\times\frac{4}{11}$ or $\frac{7}{12}\times\frac{6}{11}$ seen award M1A0A0		

Question Number	Scheme	Marks
Q3 (a)	$2a+\frac{2}{5}+\frac{1}{10}=1$ (or equivalent)	M1
	$a=\frac{1}{4}$ **or 0.25**	A1 (2)
(b)	E(X) = **1**	B1 (1)
(c)	$\text{E}(X^2)=1\times\frac{1}{5}+1\times\frac{1}{10}+4\times\frac{1}{4}+9\times\frac{1}{5}$ (= 3.1)	M1
	Var(X) = $3.1-1^2$, **= 2.1 or $\frac{21}{10}$ oe**	M1 A1 (3)
(d)	Var(Y) = $(-2)^2$ Var(X) , = **8.4 or $\frac{42}{5}$ oe**	M1 A1 (2)
(e)	$X \geq Y$ when X = 3 or 2, so probability = $"\frac{1}{4}"+\frac{1}{5}$	M1 A1ft
	= $\frac{9}{20}$ **oe**	A1 (3)
		Total 11
(a)	M1for a clear attempt to use $\sum \text{P}(X=x)=1$ Correct answer only 2/2. **NB Division by 5 in parts (b), (c) and (d) seen scores 0. Do not apply ISW.**	
(b)	B1 for 1	
(c)	1st M1 for attempting $\sum x^2\text{P}(X=x)$ at least two terms correct. Can follow through. 2nd M1 for attempting $\text{E}(X^2)-[\text{E}(X)]^2$ or allow subtracting 1 from their attempt at $\text{E}(X^2)$ provided no incorrect formula seen. Correct answer only 3/3.	
(d)	M1 for $(-2)^2\text{Var}(X)$ or 4Var(X) Condone missing brackets provided final answer correct for their Var(X) . Correct answer only 2/2.	
(e)	Allow M1 for distribution of $Y=6-2X$ and correct attempt at $\text{E}(Y^2)-[\text{E}(Y)]^2$ M1 for identifying X = 2, 3 1st A1ft for attempting to find their P(X=2) + P(X = 3) 2nd A1 for $\frac{9}{20}$ or 0.45	

Question Number	Scheme	Marks
Q4 (a)	$\frac{2+3}{\text{their total}} = \frac{5}{\text{their total}} = \frac{1}{6}$ **(** given answer**)**	M1 A1cso (2)
(b)	$\frac{4+2+5+3}{\text{total}}, \quad = \frac{14}{30}$ or $\frac{7}{15}$ or $0.4\dot{6}$	M1 A1 (2)
(c)	$\mathrm{P}(A \cap C) = 0$	B1 (1)
(d)	$\mathrm{P}(\mathrm{C} \mid \text{reads at least one magazine}) = \frac{6+3}{20} = \frac{9}{20}$	M1 A1 (2)
(e)	$\mathrm{P}(B) = \frac{10}{30} = \frac{1}{3}, \quad \mathrm{P}(C) = \frac{9}{30} = \frac{3}{10}, \quad \mathrm{P}(B \cap C) = \frac{3}{30} = \frac{1}{10}$ or $\mathrm{P}(B \mid C) = \frac{3}{9}$	M1
	$\mathrm{P}(B) \times \mathrm{P}(C) = \frac{1}{3} \times \frac{3}{10} = \frac{1}{10} = \mathrm{P}(B \cap C)$ or $\mathrm{P}(B \mid C) = \frac{3}{9} = \frac{1}{3} = \mathrm{P}(B)$	M1
	So yes they are statistically independent	A1cso (3)
		Total 10
(a)	M1 for $\frac{2+3}{\text{their total}}$ or $\frac{5}{30}$	
(b)	M1 for adding at least 3 of "4, 2, 5, 3" and dividing by their total to give a probability Can be written as separate fractions substituted into the completely correct Addition Rule	
(c)	B1 for 0 or 0/30	
(d)	M1 for a **denominator of 20** or $\frac{20}{30}$ leading to an answer with denominator of 20 $\frac{9}{20}$ only, 2/2	
(e)	1st M1 for attempting all the required probabilities for a suitable test 2nd M1 for use of a correct test - must have attempted all the correct probabilities. Equality can be implied in line 2. A1 for fully correct test carried out with a comment	

Question Number	Scheme	Marks
Q5 (a)	23, 35.5 (may be in the table)	B1 B1 (2)
(b)	Width of 10 units is 4 cm so width of 5 units is **<u>2 cm</u>**	B1
	Height = 2.6×4 =**<u>10.4 cm</u>**	M1 A1 (3)
(c)	$\sum \text{f}x = 1316.5 \Rightarrow \bar{x} = \frac{1316.5}{56} =$ awrt **<u>23.5</u>**	M1 A1
	$\sum \text{f}x^2 = 37378.25$ can be implied	B1
	So $\sigma = \sqrt{\frac{37378.25}{56} - \bar{x}^2} =$ awrt**<u>10.7</u>** allow s = 10.8	M1 A1 (5)
(d)	$Q_2 = (20.5) + \frac{(28-21)}{11}\times 5$ = 23.68… **awrt <u>23.7 or 23.9</u>**	M1 A1 (2)
(e)	$Q_3 - Q_2 = 5.6$, $Q_2 - Q_1 = 7.9$ (or $\bar{x} < Q_2$)	M1
	[7.9 >5.6 so] **<u>negative skew</u>**	A1 (2)
		Total 14

(b)	M1 for their width x their height=20.8. Without labels assume width first, height second and award marks accordingly.
(c)	1st M1 for reasonable attempt at $\sum x$ and /56 2nd M1 for a method for σ or s , $\sqrt{\ }$ is required Typical errors $\sum(\text{f}x)^2 = 354806.3$ M0, $\sum \text{f}^2 x = 13922.5$ M0 and $(\sum \text{f}x)^2 = 1733172$ M0 Correct answers only, award full marks.
(d)	Use of $\sum \text{f}(x-\bar{x})^2 =$ awrt 6428.75 for B1 lcb can be 20, 20.5 or 21, width can be 4 or 5 and the fraction part of the formula correct for M1 - Allow 28.5 in fraction that gives awrt 23.9 for M1A1
(e)	M1for attempting a test for skewness using quartiles or mean and median. Provided median greater than 22.55 and less than 29.3 award for M1 for $Q_3 - Q_2 < Q_2 - Q_1$ without values as a valid reason. SC Accept mean close to median and no skew oe for M1A1

Question Number	Scheme	Marks
Q6 (a)	See overlay	B1 B1 (2)
(b)	The **points** lie reasonably close to a straight **line** (o.e.)	B1 (1)
(c)	$\sum d = 27.7, \quad \sum f = 146$ (both, may be implied)	B1
	$S_{dd} = 152.09 - \frac{(27.7)^2}{6} = 24.208\ldots$ **awrt <u>24.2</u>**	M1 A1
	$S_{fd} = 723.1 - \frac{27.7 \times 146}{6} = 49.06\ldots$ **awrt <u>49.1</u>**	A1 (4)
(d)	$b = \frac{S_{fd}}{S_{dd}} = 2.026\ldots$ **awrt <u>2.03</u>**	M1 A1
	$a = \frac{146}{6} - b \times \frac{27.7}{6} = 14.97\ldots$ so <u>***f* = 15.0 + 2.03*d***</u>	M1 A1 (4)
(e)	A flight costs **£2.03 (or about £2)** for every extra **100km** or about **2p** per **km**.	B1ft (1)
(f)	$15.0 + 2.03d < 5d$ so $d > \frac{15.0}{(5-2.03)} = 5.00 \sim 5.05$	M1
	So $t > 500 \sim 505$	A1 (2)
		Total 14

(a)	1st B1 for at least 4 points correct (allow ± one 2mm square) 2nd B1 for all points correct (allow ± one 2 mm square
(b)	Ignore extra points and lines Require reference to points and line for B1.
(c)	M1 for a correct method seen for either - a correct expression 1st A1 for S_{dd} awrt 24.2 2nd A1 for S_{fd} awrt 49.1
(d)	1st M1 for a correct expression for *b* - can follow through their answers from (c) 2nd M1 for a correct method to find *a* - follow through their *b* and their means 2nd A1 for *f* =.... in terms of *d* and all values awrt given expressions. Accept 15 as rounding from correct answer only.
(e)	Context of cost and distance required. Follow through their value of *b*
(f)	M1 for an attempt to find the intersection of the 2 lines. Value of *t* in range 500 to 505 seen award M1. Value of *d* in range 5 to 5.05 award M1. Accept *t* greater than *500* to 505 inclusive to include graphical solution for M 1A1

Question Number	Scheme	Marks
Q7 (a)	$\text{P}(D > 20) = \text{P}\left(Z > \frac{20-30}{8}\right)$	M1
	$= \text{P}(Z > -1.25)$	A1
	$=$ **<u>0.8944</u>** **<u>awrt 0.894</u>**	A1 (3)
(b)	$\text{P}(D < Q_3) = 0.75$ so $\frac{Q_3 - 30}{8} = 0.67$	M1 B1
	$Q_3 =$ **awrt <u>35.4</u>**	A1 (3)
(c)	35.4 - 30= 5.4 so $Q_1 = 30 - 5.4$ = **awrt <u>24.6</u>**	B1ft (1)
(d)	$Q_3 - Q_1 = 10.8$ so $1.5(Q_3 - Q_1) = 16.2$ so $Q_1 - 16.2 = h$ or $Q_3 + 16.2 = k$	M1
	h=**<u>8.4 to 8.6</u>** and k= **<u>51.4 to 51.6</u>** both	A1 (2)
(e)	$2\text{P}(D > 51.6) = 2\text{P}(Z > 2.7)$	M1
	$= 2[1 - 0.9965]$ = awrt **<u>0.007</u>**	M1 A1 (3)
		Total 12

(a) M1 for an attempt to standardise 20 or 40 using 30 and 8.
1st A1 for $z = \pm 1.25$
2nd A1 for awrt 0.894

(b) M1 for $\frac{Q_3 - 30}{8}$ = to a z value
M0 for 0.7734 on RHS.
B1 for (z value) between 0.67~0.675 seen.
M1B0A1 for use of $z = 0.68$ in correct expression with awrt 35.4

(c) Follow through using their of quartile values.

(d) M1 for an attempt to calculate 1.5(IQR) and attempt to add or subtract using one of the formulae given in the question - follow through their quartiles

(e) 1st M1 for attempting 2P(D > their k) or (P(D >their k)+ P(D < their h))
2nd M1 for standardising their h or k (may have missed the 2) so allow for standardising P(D > 51.6) or P(D < 8.4)
Require boths Ms to award A mark.