Paper Reference(s) 6683/01 Edexcel GCE

Statistics S1

Advanced Level

Friday 18 January 2013 – Afternoon

Time: 1 hour 30 minutes

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

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

Instructions to Candidates

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

Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information for Candidates

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

Advice to Candidates

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

P41805A

This publication may only be reproduced in accordance with Edexcel Limited copyright policy. ©2013 Edexcel Limited

1. A teacher asked a random sample of 10 students to record the number of hours of television, t, they watched in the week before their mock exam. She then calculated their grade, g, in their mock exam. The results are summarised as follows.

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

(a) Find S_{tt} and S_{gt} .

(3)

(b) Calculate, to 3 significant figures, the product moment correlation coefficient between t and g.

(2)

(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. 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
$$k = 13$$
.

(*b*) Find the probability distribution of *X*.

Given that $\operatorname{Var}(X) = \frac{259}{320}$,

(c) find the exact value of Var(4X - 5).

(2)

(2)

(4)

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
<i>m</i> secs	6.5	4.5	6	5	4	3	2	1

(You may use $\sum tm = 469.5$,	$S_{tt} = 354,$	$S_{mm} = 25.5$)
----------------------------------	-----------------	-------------------

(<i>a</i>)	Show that $S_{tm} = -90.5$.	
		(4)
(<i>b</i>)	Find the equation of the regression line of <i>m</i> on <i>t</i> giving your answer in the form $m = a + a$	bt. (4)
(<i>c</i>)	Use your regression line to estimate the time interval between mating calls when surrounding temperature is 10 $^{\circ}$ C.	the
		(1)
(<i>d</i>)	Comment on the reliability of this estimate, giving a reason for your answer.	(1)
		(1)

4. The length of time, L hours, that a phone will work before it needs charging is normally distributed with a mean of 100 hours and a standard deviation of 15 hours.

(a) Find
$$P(L > 127)$$
. (3)

(b) Find the value of d such that P(L < d) = 0.10.

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)

(3)

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

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

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

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

(b) Use linear interpolation to estimate the median weekly income to the nearest pound.

(c) Estimate the mean and the standard deviation of the weekly income for these data.

(4)

(3)

(2)

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

(d) Use this measure to calculate the skewness for these data and describe its value.

(2)

Katie suggests using the random variable X which has a normal distribution with mean 320 and standard deviation 150 to model the weekly income for these data.

(e) Find
$$P(240 < X < 400)$$
.

(2)

(f) With reference to your calculations in parts (d) and (e) and the data in the table, comment on Katie's suggestion.(2)

- 6. A fair blue die has faces numbered 1, 1, 3, 3, 5 and 5. The random variable *B* represents the score when the blue die is rolled.
 - (*a*) Write down the probability distribution for *B*.
 - (*b*) State the name of this probability distribution.
 - (c) Write down the value of E(B).

(1)

(2)

(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
$\mathbf{P}(R=r)$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

(*d*) Find E(R).

(e) Find Var(R).

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)

(2) (3) 7. Given that

$$P(A) = 0.35$$
, $P(B) = 0.45$ and $P(A \cap B) = 0.13$,

find

(a)
$$P(A \cup B)$$
, (2)

(b)
$$P(A' | B')$$
. (2)

The event *C* has P(C) = 0.20.

The events A and C are mutually exclusive and the events B and C are independent.

END

January 2013 6683 Statistics S1 Mark Scheme

Question Number	Scheme	Marks
1. (a)	$(S_{tt}) = 8702 - \frac{258^2}{10}$ or $(S_{gt}) = 1550.2 - \frac{258 \times 63.6}{10}$ $(S_{tt}) = 2045.6$, $(S_{gt}) = -90.68$ awrt (2046), awrt - 90.7	M1 A1, A1
(b)	$r = \frac{-90.68}{\sqrt{2045.6 \times 7.864}} = -0.714956$ awrt -0.715	(3) M1 A1 (2)
(c)	Positive e.g. high v corresponds to low t and low t corresponds to high g so expect high v to corresponds to high g or expect more revision to result in a better grade	B1 B1 (2) 7
	Notes	
(a)	M1 for at least one correct expression 1^{st}A1 for $S_{tt} = \text{awrt } 2046$ (Condone $S_{xx} = \dots$ or even $S_{yy} = \dots$) 2^{nd}A1 for $S_{gt} = \text{awrt } -90.7$ (Condone $S_{xy} = \dots$)	
	On epen take care that 1^{st} A1 is for S_{tt} and 2^{nd} A1 is for S_{gt}	
	On open take care that 1 AT is for S_{tt} and 2 AT is for S_{gt}	
(b)	M1 for attempt at correct formula. Must have their S_{tt} , S_{gt} and given S_{gg} in the correct places. Condone miss Award M1A0 for awrt -0.71 with no expression seen M0 for $\frac{1550.2}{\sqrt{8702 \times 7.864}}$ Correct answer only is 2/2	ing "—"
(c)		

Question Number	Scheme	Marks
2.		
(a)	F(3) = 1 gives $\frac{3^3 + k}{40} = 1$	M1
	So $k = \underline{13}$	A1cso (2)
(b)	$P(X=1) = \frac{14}{40}$ or 0.35 (o.e.)	(2) B1
		M1
	Use of $P(X = 2) = F(2) - F(1)$ or $P(X = 3) = F(3) - F(2)$ $P(X = 2) = \frac{7}{40}$ or 0.175, $P(X = 3) = \frac{19}{40}$ or 0.475	A1, A1
(c)	$Var(4X-5) = 4^2 Var(X)$	(4) M1
(0)	So $Var(4X - 5) = 4$ Var(X) So $Var(4X - 5) = \frac{259}{20}$ or 12.95	
	$30 \text{ Val}(4x-3) = \frac{1}{20} 0112.93$	A1 (2)
	Notes	8
(a)	M1 for use of F(3) = 1 Attempt at $\frac{3^3 + k}{40} = 1$ must be seen	
	27+k = 40 without reference to F(3) = 1 is M0 A1cso for no incorrect working seen and M1 scored.	
<u>Verify</u>	Allow M1 for $\frac{3^3 + 13}{40} = 1$ but the A1 requires an <u>explicit</u> comment such as "so	o <i>k</i> = 13"
	40	
	If a table such as this is seen then award B1M1A1A1. Ignore labels on 2^{nd} row	
(b)	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	
	$\frac{7}{20} \text{ or } 0.35 \frac{7}{40} \text{ or } 0.175 \frac{19}{40} \text{ or } 0.475$	
	Otherwise apply the following:	
	B1 for $\frac{14}{40}$ or 0.35 or any exact equivalent. Can be labelled F(1), P(X = 1) or p	p(x) and
	associated with $x = 1$ or given in a table but must have <u>a</u> label.	
	M1 for clear method showing how to obtain $P(X =)$ from $F(x)$ M1 can be implied if either $P(X = 2)$ or $P(X = 3)$ is correct	
	1 st A1 for P(X = 2) = $\frac{7}{40}$ or 0.175 or exact equivalent	
	2^{nd} A1 for P(X = 3) = $\frac{19}{40}$ or 0.475 or exact equivalent	
	NB on epen make sure that the 1^{st} A1 is for P(X = 2) and 2^{nd} A1 is for	$\mathbf{P}(X=3)$
(c)	M1 for correct use of the variance formula $(4^2 \text{Var}(X) \text{ alone secures M1})$ A value for Var(X) is not required for this M1	
	A1 for any exact equivalent to 12.95 Correct answer only is 2/2	

Question Number	Scheme	Marks	
2	$\sum t = 140$ (or $\overline{t} = 17.5$) and $\sum m = 32$ (or $\overline{m} = 4$)	B1 B1	
	$\sum t = 140$ (or $\overline{t} = 17.5$) and $\sum m = 32$ (or $\overline{m} = 4$) (S _{tm}) = 469.5 - $\frac{"140" \times "32"}{8}$	M1	
	$(S_{tm} =) - 90.5$	A1cso (4)	
(b)	$b = \frac{S_{tm}}{S_{tt}} = \frac{-90.5}{354}$ $b = -0.255649 \text{ (allow } \frac{181}{708}\text{)} -0.25 \text{ or awrt } -0.26$ $a = \frac{"32"}{8} - b \times \frac{"140"}{8}$	M1	
	$b = -0.255649$ (allow $\frac{181}{708}$) -0.25 or awrt -0.26	A1	
	$a = \frac{"32"}{8} - b \times \frac{"140"}{8}$	M1	
	So equation of the line is $\underline{m = 8.47 - 0.256t}$ (allow $m = \frac{11999}{1416} - \frac{181}{708}t$)	A1 (4)	
(c)	$(8.47 - 0.256 \times 10 =) 5.9$ awrt <u>5.9</u>	B1 (1)	
(d)	Should be reliable since 10 is in the range (of the data)	B1 (1) 10	
	Notes	10	
(a)	1 st B1 for 140 seen in correct context or correctly labelled 2 nd B1 for 32 seen in correct context or correctly labelled. (allow a fully correct expression – not "++") 4480 used correctly is H M1 for attempting a correct expression. Follow through their 140 and their 32 You may see attempt at $\sum (t-\overline{t})(m-\overline{m})$. This must have all the product A1cso requires a correct expression seen and no incorrect working leading to -90	cts seen.	
(b)	1 st M1 for a correct expression for <i>b</i> . Follow through their S_{tm} . Condone miss	ing "—"	
	1 st A1for a correct expression for b . Follow through their S_{tm} . Condone missing1st A1for a wrt -0.26 or condone -0.25 2nd M1for a correct method for a . Follow through their sums from part (a) and their value of b 2nd A1for a correct equation for m and t with $a = awrt 8.47$ and $b = awrt -0.256$ Must be an equation in m and t , use of x or y scores A0 here.		
(c)	B1 for awrt 5.9 Accept 6 if the correct expression (awrt $8.47 - 10 \times awrt 0.2$	256) is seen	
(d)	B1 for suggesting it is reliable and mentioning 10 within the range (of the data or suggesting it is reliable since interpolating or not extrapolating	.)	
	 <u>or</u> suggesting it is reliable since <u>it</u> is in the range" is B0 since "<u>it</u>" is not explicit enough Condone extra irrelevant comment but penalise contradictory comments. e.g. "near the extreme so <u>not reliable</u> but not extrapolated so reliable" is B0 since <u>contradicts</u> "reliable since 10 is within the range (of temps) <u>and 5.9 within range of times</u>" is B1 since <u>irrelevant</u> 		

Question Number	Scheme	Marks
4. (a)	$\frac{127 - 100}{15}$ So P(L > 127) = P(Z > 1.8) or 1-P(Z < 1.8) o.e. = 1 - 0.9641 = <u>0.0359</u> (awrt <u>0.0359</u>)	M1 A1 A1 (3)
(b)	$\frac{d-100}{15} = -1.2816 \text{(Calculator gives } -1.2815515)}d = 80.776 \text{(awrt } \underline{80.8}\text{)}$	M1, B1 A1
(c)	Require $P(L > 133 L > 127)$ $= \left[\frac{P(L > 133)}{P(L > 127)} \right] = \frac{P(Z > 2.2)}{P(L > 127)}$ $= \left[\frac{1 - 0.9861}{1 - 0.9641} \right] = \frac{0.0139}{[0.0359]}$ $= 0.3871 = awrt 0.39$	(3) M1 dM1 A1 A1
S.C.	An attempt at P($L < 133 L > 127$) that leads to awrt 0.61 (M0M1A0A0)	(4) 10
(a) (b)	B1 for $z = \pm 1.2816$ (or better) seen anywhere [May be implied by 80.776(72 A1 for awrt 80.8 (can be scored for using 1.28 but then they get M1B0A1)	1 not poss.
Calc	The 80.8 must follow from correct workingfudging – sign scores A0 If answer is awrt 80.8 and awrt 80.777 or 80.776 or better seen then award M If answer is awrt 80.8 or 80.77 then award M1B0A1 (unless of course $z = 1.2816$	
(c)	1^{st} M1for clear indication of correct conditional probability or attempt at correct So clear attempt at $\frac{P(L>133)}{P(L>127)}$ is sufficient for the 1^{st} M1 2^{nd} dM1dependent on 1^{st} M1 for $P(L>133)$ leading to $P(Z>2.2)$. 1^{st} A1for 0.0139 or better seen coming from $P(Z>2.20)$. 2^{nd} A1for awrt 0.39. Both Ms required	
ALT	If they assume Alice did not check that the phone was working you may see: [P(L < 127).0] + P(L > 127).P(L > 133 L > 127) Provided the <u>conditional probabilities</u> as part of this calculation the 1 st M1 can be scored and their final answer will be 0.0 An answer of 0.0139 without sight of the conditional probability is 0/4 (If in doubt	139(4/4)

Question Number	Scheme	Marks
5. (a)		B1
	Area of 14 cm^2 represents frequency 28 and area of $4h$ represents 18	M1
	Or $\frac{4h}{18} = \frac{14}{28}$ (o.e.) $h = \underline{2.25}$ (cm)	A1
		(3)
(b)	$m = (240) + \frac{10}{22} \times 80$ (o.e.)	M1
	$= 276.36 (\frac{3040}{11}) \qquad ((\pounds)\underline{276} \le m < (\pounds)276.5)$	A1 (2)
(c)	$\sum fy = 31600$ leading to $\overline{y} = 316$	(2) M1A1
	$\sigma_y = \sqrt{\frac{12452800}{100} - (\bar{y})^2} = 157.07 \text{ (awrt } \underline{157}\text{)} \text{ Allow } s = 157.86$	M1A1
(d)	Skewness = 0.764 (awrt 0.76 or 0.75)	(4) B1
	[If $n+1$ used in (b) and $m = \pounds 278$ accept awrt 0.73 or 0.72] <u>Positive</u> skew	B1ft (2)
(e)	$z = \pm \frac{80}{150}$	M1
	$P(240 < X < 400) = 0.40 \sim 0.41$	A1 (2)
(f)	(e) suggests a reasonable fit for this range BUT(d) since skew it will not be a good fit overall	B2/1/0 (2) (2) 15
	Notes	13
(a)	 B1 for width (ignore units) M1 for clear method using area and frequency <u>or</u> their width × their height = 9 e.g. seeing both fd of 0.7 and 0.225 (may see fd in the table) [Must use correction] 	
(b)	M1 for $\frac{10}{22} \times 80$ or $\frac{10.5}{22} \times 80$ (o.e.). Allow use of $(n + 1)$ leading to £278.18 A1 Do not award if incorrect end-point seen but answer only is 2/2	8 or [278, 278.5)
(c)	1^{st} M1 attempt at $\sum fy$ with at least 3 correct products or ans. that rounds to 30 0	000 (to 1 sf) &/100
	2 nd M1 for correct expression including $$. Follow through \overline{y} . Need $\sum fy^2$ condone a minor transcription error e.g. 12458200.	
(d)	1 st B1 for awrt 0.76/0.75 for $m = \pounds 276$ or awrt 0.73/0.72 for $m = \pounds 278$ 2 nd B1ft for a correct description of their skew based on their measure <u>or</u> if no m based on their values of mean and median. (correlation is B0)	easure given
(e)		
(f)	For B2 we need 2 comments that make reference to each of part (e) and p. One comment should suggest it is <u>not</u> good since <u>skew</u> . The other it <u>is</u> since match 1 st B1 for one relevant comment 2 nd B1 for both comments NB Do not use B0B1	

Question Number	Scheme	Marks
6. (a)	b135P(B = b) $\frac{1}{3}$ $\frac{1}{3}$ $\frac{1}{3}$ P(B = b) $\frac{1}{3}$ $\frac{1}{3}$ $\frac{1}{3}$	B1 B1
(b)	Discrete Uniform {distribution}	(2) B1
(c)	[E(B) =] 3 (by symmetry)	(1) B1 (1)
(d)	$[E(R) =] 2 \times \frac{2}{3} + 4 \times \frac{1}{6} + 6 \times \frac{1}{6}$	M1
	= <u>3</u>	A1 (2)
(e)	$[E(R^{2}) =] 2^{2} \times \frac{2}{3} + 4^{2} \times \frac{1}{6} + 6^{2} \times \frac{1}{6} \qquad \left[= \frac{34}{3} \right]$	M1
	$[Var(R) =]\frac{34}{3} - 3^2 = ,\frac{7}{3}$ (or any exact equivalent. NB 2.33 is A0)	dM1, A1
(f)	Coin lands on 2, choose blue die; coin lands on 5 choose red die	(3) B2/1/0
	P(Avisha wins) = $\frac{1}{2} \times \left(\frac{1}{3} + \frac{1}{3}\right) + \frac{1}{2} \times \frac{1}{6}$	M1
	$=\frac{5}{12} (\text{allow awrt } 0.417)$	A1 (4)
		13
(a)	Notes 1^{st} B1 for correctly identifying values of b as 1, 3, 5 or 1,1,3,3,5,5	
(u)	2^{nd} B1 for probabilities all $=\frac{1}{3}$ or exact equivalent (or of course 6 cases of $\frac{1}{6}$)	
(b)	Any correct probability distribution or probability function is 2/2. Must be in part ((a)
(c)	B1 for answer of 3 o.e. Accept $E(X) = 3$	
(d)	M1 for an attempt at correct formula. At least 2 correct products seen. If later div A1 for an answer of 3. Correct answer only scores both marks.	ide by $n(\neq 1)$ M0
(e)	1 st M1 for a correct attempt at $E(R^2)$. At least 2 correct products seen. Condone May be implied by sight of $\frac{34}{3}$ or 11.3 or better.	$e \operatorname{Var}(R) = \operatorname{etc}$
	2^{nd} dM1 Dep. on 1^{st} M1 for clear attempt at $E(R^2) - [E(R)]^2$ Must see their value	s <u>used</u> .
	NB Var(R) = E(R ²) - [E(R)] ² = " $\frac{34}{3}$ " - "3" is M1M0A0 since do not <u>use</u> their	
 (f) B2/1/0 Both correct B1B1, one correct B1B0. Do not use B0B1[e.g. alway NB Allow other descriptions of the die e.g. 1st or fair for blue, 2nd for red a M1 for evaluating correct probabilities i.e. only 1/3, 1/2 seen or if incorrect 		are clear.
	M1 for an answer of : if choose RR $(\frac{1}{4})$, if choose BB $(\frac{1}{3})$, if choose RB $(\frac{1}{6})$	
	NB $\frac{5}{12}$ as answer scores M1A1. Need to see choices of die stated for B marks.	

Question Number	Scheme	Marks
7. (a)	$P(A \cup B) = 0.35 + 0.45 - 0.13 = \underbrace{\frac{\text{or}}{0.67}}_{0.22 + 0.13 + 0.32}$	M1 A1
(b)	$P(A' B') = \frac{P(A' \cap B')}{P(B')} \text{ or } \frac{0.33}{0.55}$	(2) M1
	$=\frac{3}{5}$ or 0.6	A1
(c)	$P(B \cap C) = 0.45 \times 0.2$ $= \underline{0.09}$	(2) M1 A1
		(2)
(d)	A C C Allow 1st B1 for 3 intersecting circles in a box with zeros in the regions for $A \cap C$ Do not accept "blank" for zero	B1 B1ft B1 B1
(e)	$P(B \cup C)' = 0.22 + 0.22$ or $1 - [0.56]$ or $1 - [0.13 + 0.23 + 0.09 + 0.11]$ o.e. = 0.44	(4) M1 A1 (2)
	Notes	12
(a)	NB May see Venn diagram for <i>A</i> and <i>B</i> only used for (a) and (b) but M marks are <u>correct expressions only</u> . No ft from an incorrect diagram for M marks.	
(b)	M1 for a correct ratio of probabilities or a correct formula and at least one correct For a correct formula allow "1 – their (a)" instead of 0.33 but not for correct Do not award for assuming independence i.e. $\frac{P(A' \cap B')}{P(B')} = \frac{0.65 \times 0.55}{0.55}$ is M0. M0 if	ratio case.
(c)	 A1 for 3/5 or any exact equivalent. M1 for correct expression. Need correct values for P(B) and P(C) seen. A1 for 0.09 or any exact equivalent. Correct answer only is 2/2 	
(d)	No labels <i>A</i> , <i>B</i> , <i>C</i> in (d) loses 1 st B1 but can score the other 3 by imp for box with <i>B</i> intersecting <i>A</i> and <i>C</i> but <i>C</i> not intersecting <i>A</i> . No box is B0 B1ft for 0.13 and their 0.09 in correct places. [ft $P(B \cap C)$ from (c)] B1 for any 2 of 0.22, <u>0.22</u> , 0.11 and 0.23 correct B1 for all 4 values correct	
(e)	M1 for a correct expression or follow through from their Venn diagram NB $P(B') \times P(C') = 0.55 \times 0.8$ is OK. Do not ft "blank" for zero and M0 for neg A1 for 0.44 only. Correct answer only is 2/2	gative probs.