Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Friday 5 June 2015 – Morning

Time: 1 hour 30 minutes

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

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

Instructions to Candidates

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

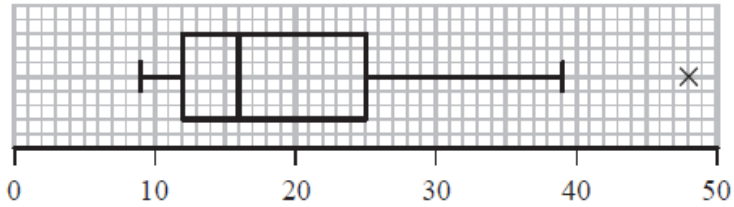
Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
This paper has 6 questions.
The total mark for this paper is 75.

Advice to Candidates

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

1. Each of 60 students was asked to draw a 20° angle without using a protractor. The size of each angle drawn was measured. The results are summarised in the box plot below.



(*a*) Find the range for these data.

(1)

(*b*) Find the interquartile range for these data.

(1)

The students were then asked to draw a 70° angle.
The results are summarised in the table below.

Angle, *a*, (degrees)	**Number of students**
$55 \leq a < 60$	6
$60 \leq a < 65$	15
$65 \leq a < 70$	13
$70 \leq a < 75$	11
$75 \leq a < 80$	8
$80 \leq a < 85$	7

(*c*) Use linear interpolation to estimate the size of the median angle drawn. Give your answer to 1 decimal place.

(2)

(*d*) Show that the lower quartile is 63°.

(2)

For these data, the upper quartile is 75°, the minimum is 55° and the maximum is 84°.

An outlier is an observation that falls either

more than 1.5 × (interquartile range) above the upper quartile or

more than 1.5 × (interquartile range) below the lower quartile.

(*e*) (i) Show that there are no outliers for these data.

(ii) On graph paper, draw a box plot for these data.

(5)

(*f*) State which angle the students were more accurate at drawing. Give reasons for your answer.

(3)

---

2. An estate agent recorded the price per square metre, p £/m^2, for 7 two-bedroom houses.

He then coded the data using the coding $q = \dfrac{p - a}{b}$, where a and b are positive constants.

His results are shown in the table below.

p	1840	1848	1830	1824	1819	1834	1850
q	4.0	4.8	3.0	2.4	1.9	3.4	5.0

(a) Find the value of a and the value of b.

(2)

The estate agent also recorded the distance, d km, of each house from the nearest train station. The results are summarised below.

$$S_{dd} = 1.02 \qquad S_{qq} = 8.22 \qquad S_{dq} = -2.17$$

(b) Calculate the product moment correlation coefficient between d and q.

(2)

(c) Write down the value of the product moment correlation coefficient between d and p.

(1)

The estate agent records the price and size of 2 additional two-bedroom houses, H and J.

House	Price (£)	Size (m^2)
H	156 400	85
J	172 900	95

(d) Suggest which house is most likely to be closer to a train station. Justify your answer.

(3)

3. A college has 80 students in Year 12.

20 students study Biology.
28 students study Chemistry.
30 students study Physics.
7 students study both Biology and Chemistry.
11 students study both Chemistry and Physics.
5 students study both Physics and Biology.
3 students study all 3 of these subjects.

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

(5)

A Year 12 student at the college is selected at random.

(*b*) Find the probability that the student studies Chemistry but not Biology or Physics.

(1)

(*c*) Find the probability that the student studies Chemistry or Physics or both.

(2)

Given that the student studies Chemistry or Physics or both,

(*d*) find the probability that the student does not study Biology.

(2)

(*e*) Determine whether studying Biology and studying Chemistry are statistically independent.

(3)

---

4. Statistical models can provide a cheap and quick way to describe a real world situation.

(*a*) Give two other reasons why statistical models are used.

(2)

A scientist wants to develop a model to describe the relationship between the average daily temperature, x °C, and her household's daily energy consumption, y kWh, in winter.

A random sample of the average daily temperature and her household's daily energy consumption are taken from 10 winter days and shown in the table.

x	–0.4	–0.2	0.3	0.8	1.1	1.4	1.8	2.1	2.5	2.6
y	28	30	26	25	26	27	26	24	22	21

[You may use $\sum x^2 = 24.76$ $\sum y = 255$ $\sum\sum xy = 283.8$ $S_{xx} = 10.36$]

(*b*) Find S_{xy} for these data.

(3)

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

Give the value of a and the value of b to 3 significant figures.

(4)

(*d*) Give an interpretation of the value of a.

(1)

(*e*) Estimate her household's daily energy consumption when the average daily temperature is 2°C.

(2)

The scientist wants to use the linear regression model to predict her household's energy consumption in the summer.

(*f*) Discuss the reliability of using this model to predict her household's energy consumption in the summer.

(2)

---

5. In a quiz, a team gains 10 points for every question it answers correctly and loses 5 points for every question it does not answer correctly. The probability of answering a question correctly is 0.6 for each question. One round of the quiz consists of 3 questions.

The discrete random variable X represents the total number of points scored in one round. The table shows the incomplete probability distribution of X.

x	30	15	0	–15
$P(X = x)$	0.216			0.064

(*a*) Show that the probability of scoring 15 points in a round is 0.432.

(2)

(*b*) Find the probability of scoring 0 points in a round.

(1)

(*c*) Find the probability of scoring a total of 30 points in 2 rounds.

(3)

(*d*) Find E(X).

(2)

(*e*) Find Var(X).

(3)

In a bonus round of 3 questions, a team gains 20 points for every question it answers correctly and loses 5 points for every question it does not answer correctly.

(*f*) Find the expected number of points scored in the bonus round.

(3)

---

6. The random variable $Z \sim N(0, 1)$.

A is the event $Z > 1.1$
B is the event $Z > -1.9$
C is the event $-1.5 < Z < 1.5$

(*a*) Find

(i) $P(A)$,

(ii) $P(B)$,

(iii) $P(C)$,

(iv) $P(A \cup C)$.

(6)

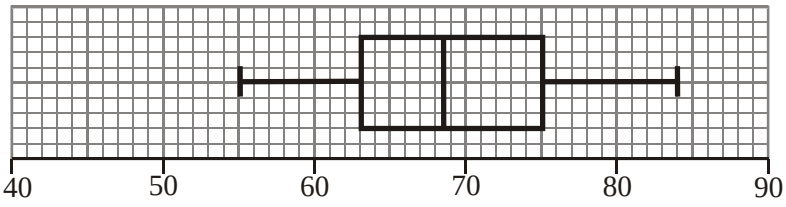
The random variable X has a normal distribution with mean 21 and standard deviation 5.

(*b*) Find the value of w such that $P(X > w \mid X > 28) = 0.625$.

(6)

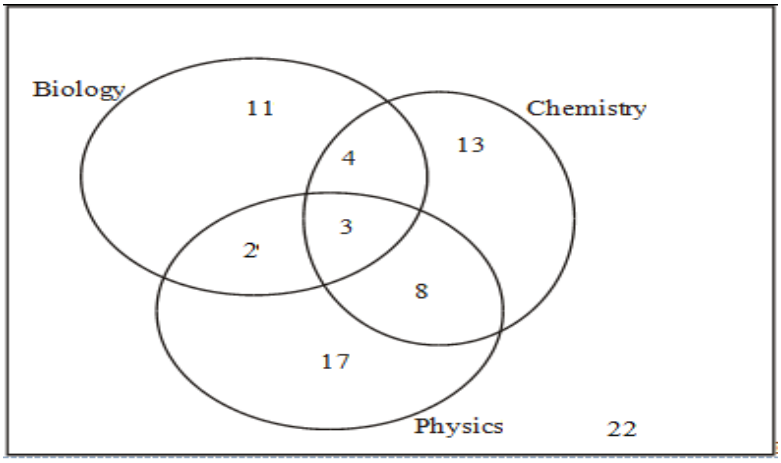
TOTAL FOR PAPER: 75 MARKS

END

Question	Scheme	Marks
1.		
(a)	[Range = 48 – 9] = **39**	B1 (1)
(b)	[IQR = 25 – 12]= **13**	B1 (1)
(c)	Median = $65+\frac{[9]}{13}\times 5 = \frac{890}{13}$ = awrt **68.5**° $\left[\text{Condone: } 65+\frac{[9.5]}{13}\times 5 = 68.7\right]$	M1 A1 (2)
(d)	Lower Quartile = $60+\frac{9}{15}\times 5 =$ **63** **(*)**	M1 A1cso (2)
(e)(i)	$63-1.5\times(75-63)=45$ $75+1.5\times(75-63)=93$ No data above 93 and no data below 45 or 55>45 etc or there are no outliers.	M1A1 A1
(ii)		M1 A1ft (5)
(f)	Median for the 70°angle is closer (to 70°)[than the 20° median is to 20°] The range/IQR for the 70° angle box plot is smaller/shorter Therefore, students were more accurate at drawing the 70° angle.	B1 B1 dB1 (3)
		(14 marks)
	Notes	
(c)	M1 for an attempt (should have 65 or 70, 13 and 5)NB working down: $70-\frac{[4]}{13}\times 5$ Allow any correct method leading to $\frac{890}{13}$, the "5" may be implied by 65 and 70 seen A1 awrt 68.5 (condone 68.7 if (n+1) is used). Ans only of 68.5 is 2/2 but 68.7 needs M1	
(d)	M1 for correct expression for the lower quartile (condone 9.25 if (n+1) used) Watch out for working down e.g. $65-\frac{6}{15}\times 5$ (M1) but e.g. $\frac{60+65}{2}=62.5=63$ is M0 A1 for correct solution with no incorrect working seen (condone (n+1) giving 63.08..)	
(e)(i)	M1 for either correct calculation (may be implied by one correct limit) A1 for either 45 or 93 A1 for 45 and 93 and conclusion	
(ii)	M1 for a box with 1 whisker drawn on each side (must see the line drawn) A1ft their median $63<Q_2<75$ but quartiles (63 and 75), 55 and 84 must be correct.	
Accuracy	Use 0.5 sq. accuracy so condone median on 68 or 69 if 68.5 seen	
(f)	1st B1 for correct comparison of their **medians** (63 < (c) < 75) to true value 2nd B1 for correct comparison of their **range** or **IQR** ("spread" is B0) Allow saying IQRs of 12 and 13 are similar. Ignore mention of "skewness" or "outliers" 3rd dB1 dependent upon at least one previous B1 being scored for choosing 70°	

Question	Scheme	Marks
2. (a)	$\frac{1840-a}{b}=4.0$ $\quad$ $\frac{1848-a}{b}=4.8$ a = **<u>1800</u>** $\quad$ b = **<u>10</u>**	M1 A1 (2)
(b)	$r=\frac{-2.17}{\sqrt{1.02\times 8.22}}=-0.749417343....$ $\quad$ awrt $-$ **<u>0.749</u>**	M1A1 (2)
(c)	-0.749	B1ft (1)
(d)	House H: 156 400/85 = [£1840/m^2 or q = 4] House J: 172 900/95 = [£1820/m^2 or q = 2] Since (r = -0.749,) there is negative correlation. <u>or</u> The higher the price (per square metre), the lower the distance from the train station. Therefore…..House H is likely to be closer.	M1 dM1 A1 (3) (8 marks)

	Notes
(a)	M1 for setting up two suitable equations which could lead to a and b (may be implied by one correct answer) A1 for a = 1800 <u>and</u> b = 10 (a = 10 and b = 1800 is A0) Correct answer only is 2/2
(b)	M1 for a correct expression (condone missing $-$) A1 for awrt -0.749 (-0.75 <u>or</u> awrt 0.749 with no working scores M1 A0).
(c)	B1ft for -0.749 or ft their answer to (b) to at least 2sf. Must be in the range $-1<'(b)'<1$
(d)	M1 for calculating price/square metre for <u>both</u> H and J. Can be implied by sight of 1840 and 1820 (so OK if not labelled or mis-labelled) These may be seen in the table in the question. Allow comment like "H is £20/square metre more than J" dM1 dependent on 1st M1 for a statement that correlation is negative <u>or</u> a contextualised interpretation of the negative correlation. **$r > 0$** If $r > 0$ allow equivalent statements about positive correlation A1 (dependent on both Ms) for House H is likely to be closer (No ft if $r > 0$)

Question	Scheme	Marks
3. (a)		B1 M1 A1 A1 B1 (5)
(b)	$\frac{'13'}{80}$ or 0.1625	B1ft (1)
(c)	$\frac{28+30-11}{80}$ or $\frac{2+3+4+8+13+17}{80}$ or $1-\frac{(11+22)}{80}=\frac{47}{80}$ or 0.5875	M1 A1 (2)
(d)	$\frac{"17+8+13"}{"47"}$ or $\frac{\frac{"38"}{80}}{\frac{"47"}{80}}$ or $1-\frac{"2+3+4"}{"47"}=\frac{38}{47}$ (condone awrt 0.809)	M1 A1cao (2)
(e)	$P(B\|C)=\frac{7}{28}$, $P(B)=\frac{20}{80}$ $P(C\|B)=\frac{7}{20}$, $P(C)=\frac{28}{80}$ $P(B\cap C)=\frac{7}{80}$, $P(B)=\frac{20}{80}P(C)=\frac{28}{80}$ $P(B\|C)=P(B)$, $P(C\|B)=P(C)$ these may be implied by correct conclusion $P(B\cap C)=P(B)\times P(C)$ this approach requires the product to be seen So, they are independent.	M1 M1 A1 (3) (13 marks)

	Notes
(a)	B1 for 3 intersecting circles with 3 in the centre. **Allow probs. or integers in diagram.** M1 for some correct subtraction e.g. at least one of 2, 4, 8 or for *B*: 20 – their(2+3+4) etc A1 for 2, 4 and 8 (ignore labels) A1 for 11, 13 and 17 (must be in compatible regions with 2, 4, 8 if no labels) B1 for correct labels and 22 and box (Do not treat "blank" as 0 so can't use 0 for ft in (c))
(c)	M1 for a correct expression seen in (c) (or ft their diagram). Correct ans M1A1
(d)	M1 for denominator of 47 or ft their numerator from part (c) and numerator of 38 or their (17 + 8 + 13) or (their 47) – their (2 + 3 + 4). Correct ans M1A1
(e)	M1 for stating at least the required probs.& labelled for a correct test (can ft their diagram) M1 for use of a correct test with *B* and *C* Must see product attempted for $P(B\cap C)$ test. A1 for a correct test with all probabilities correct and a correct concluding statement. NB M0M1A0 should be possible but A1 requires both Ms

Question	Scheme	Marks
4. (a)	To simplify (or represent) a real world problem (o.e.) To improve understanding (o.e.) To analyse a real world problem or can change variables/replicate easily (oe) To make predictions or find estimates (o.e.)	B1g B1h (2)
(b)	$\sum x = 12$ $S_{xy} = 283.8 - \frac{'12' \times 255}{10}$, $= -\underline{\mathbf{22.2}}$	B1 M1,A1cao (3)
(c)	$b = \frac{'-22.2'}{10.36} =, -2.142857....$ (A1 for awrt -2.1) $\left[a = \bar{y} - b\bar{x} \quad \Rightarrow\right] \; a = \frac{255}{10} - 'b' \times \frac{"12"}{10} = 28.07143$ $y = 28.1 - 2.14x$ $\left[\text{Condone: } y = 28.1 + -2.14x\right]$	M1A1 M1 A1 (4)
(d)	(28.1 kWh) of energy are used when the temperature is 0[°C]	B1 (1)
(e)	$y = 28.1 - 2.14(2) =$ awrt **$\underline{\mathbf{23.8}}$**	M1 A1 (2)
(f)	The regression model is based on temperatures from the winter, so not reliable in the summer. Stating it **is** reliable (whatever the reason) is B0B0	B1 dB1 (2) (14 marks)

	Notes
(a)	Make sure reasons refer to **models** and not **tests** 1st B1g (be fairly generous) for a sensible reason not using "quick", "cheap" or "describe" 2nd B1h (be slightly harder) for two convincing reasons (both based on the list above) Use professional judgement and mark as B0B0 or B1B0 or B1B1 do **not** use B0B1
(b)	B1 for $\sum x = 12$ (May be by the table) (Can be implied by 3060 seen or the next line) M1 for attempt at correct formula (ft their $\sum x$ where $10 < \sum x < 14$) A1 for -22.2 only
(c)	M1 for a correct expression for b (ft their $S_{xy} \neq 283.8$) A1 for awrt -2.1 (allow $-15/7$) M1 for a correct expression for a and ft their 12 (allow use of a letter b) A1 for $y = 28.1 - 2.14x$ (awrt 28.1 and awrt -2.14) Must be y and x and no fractions
(d)	B1 for a contextualised interpretation e.g. the amount of <u>energy</u> used when <u>temperature</u> is <u>0</u>[°C] **or** [28.1] <u>kWh</u> used when <u>temp. is 0</u>[°C] [Can ft their 28.1]Need <u>temp</u> or °<u>sign</u> [B0 for "value of y when $x = 0$" since no context in words]
(e)	M1 for substituting $x = 2$ into their equation
(f)	B1 for reasoning to suggest that temperatures are different in summer or the model was based only on data from the winter. Allow mention of <u>extrapolation</u> (o.e.) dB1 so not reliable.

Question	Scheme	Marks
5. (a)	To score 15 points, 2 correct and 1 not correct $[0.6\times0.6\times0.4]+[0.6\times0.4\times0.6]+[0.4\times0.6\times0.6]$ <u>or</u> $3\times(0.6\times0.6\times0.4)$ $= 0.432$ **(*)**	M1 A1cso (2)
(b)	$1 - (0.216 + 0.432 + 0.064) =$ **<u>0.288</u>** or $3\times0.6\times(0.4)^2$	B1 (1)
(c)	[(30, 0), (0, 30) or (15, 15)] $0.216\times'0.288'+'0.288'\times0.216+0.432\times0.432$ **awrt <u>0.311</u>**	M1 A1ft A1 (3)
(d)	$E(X)=[30\times0.216]+[15\times0.432]+[0\times0.288]+[(-15)\times0.064]$ $E(X) = 12$ **<u>12</u>** (only)	M1 A1 (2)
(e)	$E(X^2)=30^2\times0.216+15^2\times0.432+0^2\times0.288+(-15)^2\times0.064$ (= 306) $Var(X)=E(X^2)-[E(X)]^2$ ='306'–'12'^{2} =, **<u>162</u>**	M1 M1, A1 (3)
(f)	Let Y = number of points scored in bonus round [y] | 60 | 35 | 10 | –15 [P($Y = y$)] | 0.216 | 0.432 | 0.288 | 0.064 $E(Y) = 60\times0.216+35\times0.432+10\times0.288+(-15)\times0.064$ $=$ **<u>30</u>**	M1 dM1 A1 (3) (14 marks)
	Notes	
(a)	M1 for $0.6^2\times0.4$ may be $\Rightarrow$ by tree diagram with 0.6 & 0.4 but just 3×0.144 or 2×0.216 is M0 A1 cso for $3\times0.6^2\times0.4$ (seen) and no incorrect working seen	
(b)	0.288 or $\frac{36}{125}$ answer may be seen in table. [NB Fractions: $\frac{27}{125},\frac{54}{125},\frac{36}{125}$ and $\frac{8}{125}$]	
	Correct answers to (c), (d) and (e) score full marks for these parts.	
(c)	M1 for either $0.216\times'0.288' = (0.062208)$ <u>or</u> $0.432\times0.432 = 0.186624$ (ft (b) provided their (b) is a probability) 1st A1ft for a fully correct expression 2nd A1 for awrt 0.311 or $\frac{972}{3125}$	
SC	**6 questions 4 correct** Award M1&1st A1 for $6C4\times0.6^4\times0.4^2$ or $15\times0.6^4\times0.4^2$	
(d)	M1 for a correct expression for E(X) (0 term not required, ft their (b)) NB alt: $3\times(10\times0.6+(-5)\times0.4)$. E($X$) = 12 scores M1A1 if (b) is a probability.	
(e)	1st M1 for correct expres' for $E(X^2)$ (0 term not required, ft their(b))Condone -15^2 Ignore label so Var(X) = [E(X^2)] = 306 can score M1M0A0 2nd M1 for correct expression for Var(X) (may follow through their values)	
ALT	1st M1 for $[10^2\times0.6+(-5)^2\times0.4 = 70]$ 2nd M1 for $3\times(70-4^2) = 54$ and A1 for 162	
(f)	1st M1 for correct distribution for Y (ft(b)) <u>or</u> $20\times0.6+(-5)\times0.4$ <u>or</u> $Y=\frac{5}{3}X+10$ 2nd dM1 for correct expres' for E(Y) <u>or</u> $3\times(20\times0.6+(-5)\times0.4)$ <u>or</u> $E(Y)=\frac{5}{3}E(X)+10$ Dep. on 1st M1 but can ft their (b) or their E(X). Correct expres' (line 2) scores M1M1 A1 for 30 with at least 1 M mark scored. Answer only is 0/3 but 30 after M1 is 3/3	

Question	Scheme	Marks
6. (a)(i)	$P(A) = P(Z > 1.1) = 1 - 0.8643 =$ **0.1357** (accept awrt 0.136)	B1
(ii)	$P(B) = P(Z > -1.9) =$ **0.9713** (accept awrt 0.971)	B1
(iii)	$P(C) = [P(-1.5 < Z < 1.5)] = 0.9332 - (1 - 0.9332)$ or $(0.9332 - 0.5) \times 2$ $=$ **0.8664** (accept awrt 0.866)	M1 A1
(iv)	$P(A \cup C) = P(Z > -1.5)$ or $P(Z < 1.5)$ or $= P(A) + P(C) - P(A \cap C) = "0.1357" + "0.8664" - (0.9332 - 0.8643)$ $=$ **0.9332** (accept awrt 0.933)	M1 A1 (6)
(b)	$[P(X > w \mid X > 28) =] \dfrac{P(X > w)}{P(X > 28)} = [0.625]$	M1
	$P(X > 28) = P\left(Z > \dfrac{28-21}{5}\right) = P(Z > 1.4) = [0.0808$ calc: 0.80756..$]$	M1
	$P(X > w) = 0.0808 \times 0.625$ (= 0.0505) or ($P(X < w) = 0.9495$)	A1
	$\dfrac{w-21}{5} = 1.64$	M1 B1
	$w =$ awrt **29.2**	A1 (6) (12 marks)
	Notes	
	Mark final answer here so in (ii) 0.9713 followed by 1 – 0.9713 is B0 but for rounding errors e.g. 29.245 followed by 29.3 apply ISW and award for 29.245	
(a)(iii)	M1 for correct expression with probability values . Correct ans implies M1A1	
(iv)	M1 for a correct addition formula with some correct substitution (or correct ft) or $P(Z > -1.5)$ (o.e) or for a fully correct expression with correct probabilities A1 for 0.9332 (accept 0.933) Correct answer only is M1A1	
(b)	M1 for correct expression for conditional probability- must have $P(X > w)$ as num' May be implied by $P(X > w) = 0.625 \times$ (any probability) M1 for standardising 28 with 21 and 5 Allow $\pm$ (May be implied by 0.0808 [or awrt 0.081] seen in correct position) A1 for $P(X > w) = 0.0808 \times 0.625$ or $P(X > w) = 0.0505$ or $P(X < w) = 0.9495$) This A1 depends on both Ms but seeing $P(X > w) = 0.0808 \times 0.625$ scores M1M1A1	
1st 3 marks	Allow $P\left(Z > \dfrac{w-21}{5}\right)$ instead of $P(X > w)$ for these first 3 marks	
	M1 for standardising w with 21 and 5 (allow $\pm$) and setting equal to a z-value $\lvert z \rvert > 1$ Allow any letter instead of w B1 for 1.64 (or better) used correctly. [Calculator gives: 1.6402851...] A1 allow awrt 29.2	

Greg Attwood 13th June 2015