Time: 1 hour 30 minutes

Materials required for examination papers	**Items included with question**
Mathematical Formulae (Green)	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 8 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.

Suggested grade boundaries for this paper:

A*	A	B	C	D	E
58	51	44	37	30	23

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 $x = \dfrac{g-60}{4}$ and $y = \dfrac{e}{10}$. 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)

May 2013 (R)

---

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

(1)

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

Find

(*b*) P(*B*),

(4)

(*c*) $P(A' \cap B)$,

(2)

(*d*) $P(B'|A)$.

(2)

January 2012

---

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.

May 2009

---

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)

May 2013 (R)

3

5. 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)

January 2013

6. 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)

January 2009

4

7. 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)

May 2013

---

8. (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

$$P(R \cap Q \,\square\,) = 0.15, \quad P(Q) = 0.35 \quad \text{and} \quad P(R \mid Q) = 0.1$$

Find the value of

(b) $P(R \cup Q)$,

(1)

(c) $P(R \cap Q)$,

(2)

(d) $P(R)$.

(2)

May 2009

TOTAL FOR PAPER: 75 MARKS

END

Question Number	Scheme	Marks
1. (a)	$b = \dfrac{18.35}{312.1}[= 0.058795...]$	M1
	$a = 5.8 - "0.058795..." \times 4.8$	M1
	$a = \text{awrt} \ \underline{\mathbf{5.52}}$	A1
	So $\underline{\boldsymbol{y = 5.52 + 0.0588x}}$	A1 (4)
(b)	$\dfrac{e}{10} = "5.52" + "0.0588" \times \left(\dfrac{g - 60}{4}\right)$	M1
	$4e = 220.71 + 0.588(g - 60)$	dM1
	$\underline{\boldsymbol{e = 46 + 0.15g}}$	A1A1 (4)
(c)	$e = "46" + "0.15" \times 100$	M1
	$= \underline{\mathbf{61}}$	A1 (2)
		[10]
2. (a)	(R and S are mutually) exclusive.	B1 (1)
(b)	$\dfrac{2}{3} = \dfrac{1}{4} + P(B) - P(A \cap B)$ use of Addition Rule	M1
	$\dfrac{2}{3} = \dfrac{1}{4} + P(B) - \dfrac{1}{4} \times P(B)$ use of independence	M1 A1
	$\dfrac{5}{12} = \dfrac{3}{4} P(B)$	
	$P(B) = \dfrac{5}{9}$	A1 (4)
(c)	$P(A' \cap B) = \dfrac{3}{4} \times \dfrac{5}{9} = \dfrac{15}{36} = \dfrac{5}{12}$	M1A1ft (2)
(d)	$P(B' \mid A) = \dfrac{(1 - (b)) \times 0.25}{0.25}$ or $P(B')$ or $\dfrac{\frac{1}{9}}{\frac{1}{4}}$	M1
	$= \dfrac{4}{9}$	A1 (2)
		[9]

Question Number	Scheme	Marks
3. (a)	1(cm) cao	B1 (1)
(b)	10 cm^2 represents 15 10/15 cm^2 represents 1 Therefore frequency of 9 is $\dfrac{10}{15} \times 9$ or $\dfrac{9}{1.5}$ height = 6 (cm)	M1 M1 A1 (2) **[3]**
4. (a)	$\left[\mathrm{P}(M < 145) = \right] \ \mathrm{P}\left(Z < \dfrac{145 - 150}{10} \right)$ $= \mathrm{P}(Z < -0.5)$ or $\mathrm{P}(Z > 0.5)$ = **awrt <u>0.309</u>**	M1 A1 A1 (3)
(b)	$\left[\mathrm{P}(B > 115) = 0.15 \Rightarrow \right] \ \dfrac{115 - 100}{d} = 1.0364$ (Calc gives 1.036433...) <u>**d = 14.5**</u> (Calc gives 14.4727...)	M1B1 A1 A1 (4)
(c)	$[\mathrm{P}(X > \mu + 15 \mid X > \mu - 15) =] \ \dfrac{\mathrm{P}\left(X > \mu + 15\right)}{\mathrm{P}\left(X > \mu - 15\right)}$ $= \dfrac{0.35}{1 - 0.35}$ $= \dfrac{7}{13}$ or **<u>awrt 0.538</u>**	M1 A1 A1 (3) **[10]**

Question Number	Scheme	Marks
5. (a)	$\dfrac{127-100}{15}$	M1
	So $P(L > 127) = P(Z > 1.8)$ or $1 - P(Z < 1.8)$ o.e.	A1
	$= 1 - 0.9641 = \underline{\mathbf{0.0359}}$ (awrt **0.0359**)	A1 (3)
(b)	$\dfrac{d-100}{15} = -1.2816$ (Calculator gives $-1.2815515\ldots$)	M1, B1
	$d = 80.776$ (awrt **80.8**)	A1 (3)
(c)	Require $P(L > 133 \mid L > 127)$	M1
	$= \left[\dfrac{P(L > 133)}{P(L > 127)} \right] = \dfrac{P(Z > 2.2)}{P(L > 127)}$	dM1
	$= \left[\dfrac{1 - 0.9861}{1 - 0.9641} \right] = \dfrac{0.0139}{[0.0359]}$	A1
	$= 0.3871\ldots$ $=$ awrt **0.39**	A1 (4) **[10]**

Question Number	Scheme	Marks
6. (a)	8-10 hours: width = 10.5 - 7.5 = 3 represented by 1.5cm	
	16-25 hours: width = 25.5 - 15.5 = 10 so represented by <u>5 cm</u>	B1
	8- 10 hours: height = fd = 18/3 = 6 represented by 3 cm	M1
	16-25 hours: height = fd = 15/10 = 1.5 represented by <u>0.75 cm</u>	A1
		(3)
(b)	$Q_2 = 7.5 + \dfrac{(52-36)}{18} \times 3 = 10.2$	M1 A1
	$Q_1 = 5.5 + \dfrac{(26-20)}{16} \times 2 [= 6.25 \text{ or } 6.3]$ or $5.5 + \dfrac{(26.25-20)}{16} \times 2 \ [=6.3]$	A1
	$Q_3 = 10.5 + \dfrac{(78-54)}{25} \times 5 [= 15.3]$ or $10.5 + \dfrac{(78.75-54)}{25} \times 5 \ [=15.45 \backslash 15.5]$	A1
	IQR = (15.3 - 6.3) = <u>9</u>	A1ft
		(5)
(c)	$\sum fx = 1333.5 \Rightarrow \bar{x} = \dfrac{1333.5}{104} = \qquad\qquad\qquad\qquad$ awrt <u>12.8</u>	M1 A1
	$\sum fx^2 = 27254 \Rightarrow \sigma_x = \sqrt{\dfrac{27254}{104} - \bar{x}^2} = \sqrt{262.05 - \bar{x}^2} \qquad$ awrt <u>9.88</u>	M1 A1
		(4)
(d)	$Q_3 - Q_2 [= 5.1] > Q_2 - Q_1 \ [= 3.9]$ or $Q_2 < \bar{x}$	B1ft
	So data is positively skew	dB1
		(2)
(e)	Use median and IQR,	B1
	since data is skewed <u>or</u> not affected by extreme values or outliers	B1
		(2)
		[16]

Question Number	Scheme	Marks
7. (a)	[Let X be the amount of beans in a tin. $P(X < 200) = 0.1$]	
	$\dfrac{200 - \mu}{7.8} = -1.2816$ [calc gives 1.28155156…]	M1 B1
	$\mu = 209.996\ldots$ awrt 210	A1 (3)
(b)	$P(X > 225) = P\left(Z > \dfrac{225 - "210"}{7.8} \right)$	M1
	$= P(Z > 1.92)$ <u>or</u> $1 - P(Z < 1.92)$ (allow 1.93)	A1
	$= 1 - 0.9726 = 0.0274$ (or better) [calc gives 0.0272037…]	
	$= 0.0274$	
	$=$ awrt **2.7%** allow **0.027**	A1 (3)
(c)	[Let Y be the new amount of beans in a tin]	
	$\dfrac{210 - 205}{\sigma} = 2.3263$ <u>or</u> $\dfrac{200 - 205}{\sigma} = -2.3263$ [calc gives 2.3263478…]	M1 B1
	$\sigma = \dfrac{5}{2.3263}$	dM1
	$\sigma = 2.15$ $(2.14933\ldots)$	A1 (4) **[10]**
8. (a)(i)	$P(A \cup B) = a + b$ cao	B1
(ii)	$P(A \cup B) = a + b - ab$ or equivalent	B1 (2)
(b)	$P(R \cup Q) = 0.15 + 0.35$	
	$= 0.5$	B1 (1)
(c)	$P(R \cap Q) = P(R|Q) \times P(Q)$	
	$= 0.1 \times 0.35$	M1
	$= 0.035$	A1 (2)
(d)	$P(R \cup Q) = P(R) + P(Q) - P(R \cap Q)$ or $P(R) = P(R \cap Q') + P(R \cap Q)$	
	$= 0.15 +$ their (c)	M1
	$0.5 \ = P(R) + 0.35 - 0.035$ $= 0.15 + 0.035$	
	$P(R) = 0.185$ $= 0.185$	A1 (2) **[7]**

Examiner reports

Question 1

Part (a) was answered very well although some only rounded their value of *b* to 3 decimal places rather than 3 significant figures as required. Most knew how to start part (b) and made a correct substitution but errors often occurred when simplifying to the required form. The final part was answered well with most substituting $g = 100$ into their answer for part (b) however earlier errors meant that the correct answer was less frequently seen.

Question 2

Despite the question using *R* and *S* in part (a) and *A* and *B* for the rest of the question, candidates assumed *A* and *B* were mutually exclusive and made no use of independence. In part (a) candidates were let down by their inability to express in unambiguous English "mutually exclusive". A number of candidates just restated the question, writing that it meant the probability of the intersection was 0 rather than describing the relationship between *R* and *S*.

Part (b) was not as well done as it ought to have been by the majority of candidates. Many didn't realise that the letters *R* and *S* were replaced from part (a) by *A* and *B* and so confused independence with mutually exclusive. Many candidates did write the full formula and substituted at least one probability correctly, although far fewer candidates realised the "independent" statement in the question meant that $P(A \cap B)$ could be replaced with $P(A) \times P(B)$. Of those who successfully used the addition rule and independence, it was very disappointing to see some who could not handle the resulting linear equation because it had fractions in it. Those who did not start by quoting a formula and assumed exclusivity scored no marks.

Part (c) was answered well, with either a correct answer (even if part (b) incorrect) or a correct follow through.

In part (d) most knew that they had to use conditional probability, with only a few dividing by $P(B')$ by mistake. The ability to find $P(B' \cap A)$ for the numerator from previous working was often lacking and very few candidates used the fact that *A* and *B* were independent to simply state $P(B'|A) = P(B')$.

Question 3

Although there were more correct solutions than in previous papers for this type of question the process required to answer this question was not applied successfully by a large number of candidates. The most common error in part (a) was to give an answer of 0.8. In tandem with this was an answer to part (b) of 7.5 where candidates recognised that the answer to part (a) times the answer to part (b) should be 6. Many candidates divided 9 by 3 in part (b) but failed to multiply by 2. Other candidates however produced two correct answers but nothing else. The variety of approaches may suggest some logical thinking rather than a taught approach to this type of problem.

Question 4

The normal distribution was handled well by most candidates on this paper. Part (a) caused few problems although some candidates failed to subtract their tables' value from 1 and most made good progress in part (b) too although some failed to use the table of percentage points of the normal distribution and had a z value of 1.04 or 1.03 rather than 1.0364. Part (c) was more challenging requiring the identification of a conditional probability and then the correct evaluation of the numerator using the symmetry of the distribution but there were a good number of correct responses seen.

Question 5

A small minority were still unsure whether the final answer was 0.9641 or 1 – 0.9641. There were 3 common sources of error in part (b). Some candidates simply set their standardised expression equal to 0.1 or 0.5398 and lost all 3 marks. Others realised that the standardised expression should be set equal to a z value but did not use the percentage points table and lost a mark. The final problem was choosing the correct sign on their z value and a number of answers of 119 were seen. Some candidates gave an answer of 80.776… from their calculators and gained all 3 marks.

Part (c) was not answered well and most attempts did not notice the usual prompt (the wording "given that…") and thus did not attempt a conditional probability. Common solutions were simply a calculation of P($L > 133$) or P($127 < L < 133$).

Question 6

Part (a) was not answered well. Many candidates attempted to calculate frequency densities but they often forgot to deal with the scale factor and the widths of the classes were frequently incorrect. There are a variety of different routes to a successful answer here but few candidates gave any explanation to accompany their working and it was therefore difficult for the examiners to give them much credit. The linear interpolation in part (b) was tackled with more success but a number missed the request for the Inter Quartile Range. Whilst the examiners did allow the use of ($n + 1$) here, candidates should remember that the data is being treated as continuous and it is therefore not appropriate to "round" up or down their point on the cumulative frequency axis. Although the mean was often found correctly the usual problems arose in part (c) with the standard deviation. Apart from those who rounded

prematurely, some forgot the square root and others used $\sum f^2 x$ or $\left(\sum fx\right)^2$ or $\dfrac{\sum fx^2}{\sum fx}$

instead of the correct first term in their expression and there was the usual crop of candidates who used $n = 6$ instead of 104. The majority were able to propose and utilise a correct test for the skewness in part (d) with most preferring the quartiles rather than the mean and median. Few scored both marks in part (e) as, even if they chose the median, they missed mentioning the Inter Quartile Range. A number of candidates gave the mean and standard deviation without considering the implications of their previous result.

Question 7

As usual the normal distribution posed serious problems for some candidates. There were two major reasons for lost marks. The first was a failure to use the table of percentage points of the normal distribution where appropriate and many candidates lost marks for using z values of 1.28 or 2.32 or even 2.33 rather than the 4dp values available. The second problem is a fundamental one of understanding where candidates confuse probabilities (areas) with z values (points on the horizontal axis).

In part (a) many could standardize correctly and often set their expression equal to a suitable z value but often there was a sign error and this led to an answer of 190 for the mean. Other candidates stated that the mean was 210 which was correct but this didn't follow from their equation and accuracy marks will not be awarded in such cases.

Part (b) was more straightforward and provided a correct mean was found in part (a) full marks were usually obtained. Many left their answer as a probability rather than the percentage asked for in the question but this was condoned on this occasion. The final part proved quite challenging. Some drew a diagram but were unable to represent the information in the question in a useful way. Others tried subtracting two standardizations and ended up with $\dfrac{10}{\sigma} = z$ which was of little use to them. Those who realized that just using the value of 210 along with the mean of 205 and $z = 2.3263$ was all that was required usually formed a simple equation for σ and were able to solve it successfully.

Question 8

Generally this question was not well answered by a large number of candidates. The terms and properties relating to probability do not seem to be fully understood, especially by weaker candidates. Part (a) was done surprising badly, with often the rest of the question fully correct. Part (c) was often correct when all else was wrong, demonstrating that candidates can use the conditional probability formula even if they do not understand it. Too few candidates write down the formula they are trying to use, which in part (d) was helpful in ascertaining if they were trying to use the correct method.

Statistics for S1 Practice Paper Gold Level G3

Qu	Max Score	Modal score	Mean %	Mean score for students achieving grade:							
				ALL	A*	A	B	C	D	E	U
1	10		70	7.02	8.94	8.33	6.78	5.85	5.23	4.72	3.34
2	9		46	4.15	7.92	6.66	3.97	2.98	2.28	1.81	1.09
3	3		47	1.40		2.27	1.57	1.17	0.92	0.64	0.45
4	10		70	6.99	9.01	8.21	7.01	6.02	5.02	3.62	2.08
5	10	6	45	4.53	6.79	5.70	4.80	3.99	3.27	2.63	1.43
6	16		46	7.28		11.14	7.55	5.26	3.51	2.42	0.92
7	10	0	43	4.32	8.55	7.78	5.59	4.04	2.74	1.72	0.65
8	7		47	3.28		5.54	3.78	2.71	1.88	1.37	0.69
	75		52	38.97		55.63	41.05	32.02	24.85	18.93	10.65