Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced Subsidiary

Friday 17 May 2013 – 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. A meteorologist believes that there is a relationship between the height above sea level, h m, and the air temperature, t °C. Data is collected at the same time from 9 different places on the same mountain. The data is summarised in the table below.

h	1400	1100	260	840	900	550	1230	100	770
t	3	10	20	9	10	13	5	24	16

[You may assume that $\sum h = 7150$, $\sum t = 110$, $\sum h^2 = 7171500$, $\sum t^2 = 1716$, $\sum th = 64\,980$ and $S_{tt} = 371.56$]

(*a*) Calculate S_{th} and S_{hh}. Give your answers to 3 significant figures.

(3)

(*b*) Calculate the product moment correlation coefficient for this data.

(2)

(*c*) State whether or not your value supports the use of a regression equation to predict the air temperature at different heights on this mountain. Give a reason for your answer.

(1)

(*d*) Find the equation of the regression line of t on h giving your answer in the form $t = a + bh$.

(4)

(*e*) Interpret the value of b.

(1)

(*f*) Estimate the difference in air temperature between a height of 500 m and a height of 1000 m.

(2)

---

2. The marks of a group of female students in a statistics test are summarised in Figure 1.

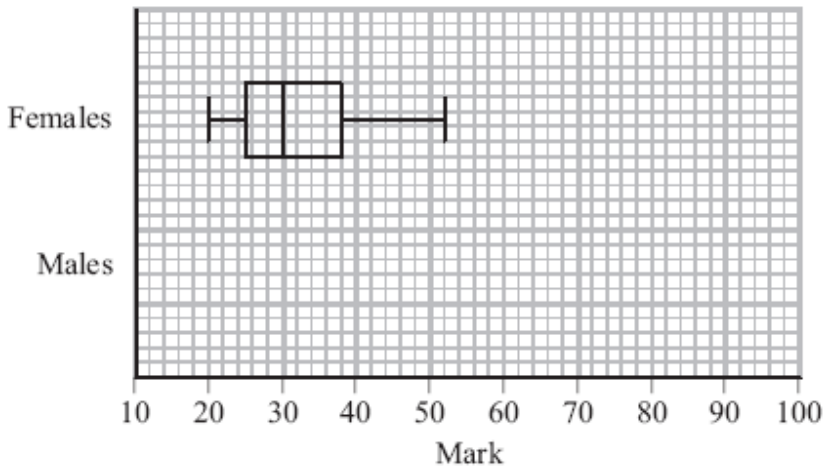


Figure 1

(*a*) Write down the mark which is exceeded by 75% of the female students.

(1)

The marks of a group of male students in the same statistics test are summarised by the stem and leaf diagram below.

Mark	(2|6 means 26)	Totals
1	4	(1)
2	6	(1)
3	4 4 7	(3)
4	0 6 6 7 7 8	(6)
5	0 0 1 1 1 3 6 7 7	(9)
6	2 2 3 3 3 8	(6)
7	0 0 8	(3)
8	5	(1)
9	0	(1)

(*b*) Find the median and interquartile range of the marks of the male students.

(3)

An outlier is a mark that is

either more than 1.5 × interquartile range above the upper quartile

or more than 1.5 × interquartile range below the lower quartile.

(*c*) On graph paper draw a box plot to represent the marks of the male students, indicating clearly any outliers.

(5)

(*d*) Compare and contrast the marks of the male and the female students.

(2)

---

3. In a company the 200 employees are classified as full-time workers, part-time workers or contractors.

The table below shows the number of employees in each category and whether they walk to work or use some form of transport.

	Walk	Transport
Full-time worker	2	8
Part-time worker	35	75
Contractor	30	50

The events F, H and C are that an employee is a full-time worker, part-time worker or contractor respectively. Let W be the event that an employee walks to work.

An employee is selected at random.

Find

(a) P(H)

(2)

(b) P($[F \cap W]'$)

(2)

(c) P($W \mid C$)

(2)

Let B be the event that an employee uses the bus.

Given that 10% of full-time workers use the bus, 30% of part-time workers use the bus and 20% of contractors use the bus,

(d) draw a Venn diagram to represent the events F, H, C and B,

(4)

(e) find the probability that a randomly selected employee uses the bus to travel to work.

(2)

4. The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

[You may use $\sum ft^2 = 134281.25$]

(*a*) Estimate the mean and standard deviation of these data.

(5)

(*b*) Use linear interpolation to estimate the value of the median.

(2)

(*c*) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures.

(1)

(*d*) Estimate the interquartile range of this distribution.

(2)

(*e*) Give a reason why the mean and standard deviation are not the most appropriate summary statistics to use with these data.

(1)

The person timing the exam made an error and each student actually took 5 minutes less than the times recorded above. The table below summarises the actual times.

Time (minutes) t	6 – 15	16 – 20	21 – 25	26 – 30	31 – 40	41 – 55
Number of students f	62	88	16	13	11	10

(*f*) Without further calculations, explain the effect this would have on each of the estimates found in parts (*a*), (*b*), (*c*) and (*d*).

(3)

---

5. A biased die with six faces is rolled. The discrete random variable X represents the score on the uppermost face. The probability distribution of X is shown in the table below.

x	1	2	3	4	5	6
$P(X = x)$	a	a	a	b	b	0.3

(a) Given that $E(X) = 4.2$ find the value of a and the value of b.

(5)

(b) Show that $E(X^2) = 20.4$.

(1)

(c) Find $\text{Var}(5 - 3X)$.

(3)

A biased die with five faces is rolled. The discrete random variable Y represents the score which is uppermost. The cumulative distribution function of Y is shown in the table below.

y	1	2	3	4	5
$F(y)$	$\frac{1}{10}$	$\frac{2}{10}$	$3k$	$4k$	$5k$

(d) Find the value of k.

(1)

(e) Find the probability distribution of Y.

(3)

Each die is rolled once. The scores on the two dice are independent.

(f) Find the probability that the sum of the two scores equals 2.

(2)

---

6. 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.

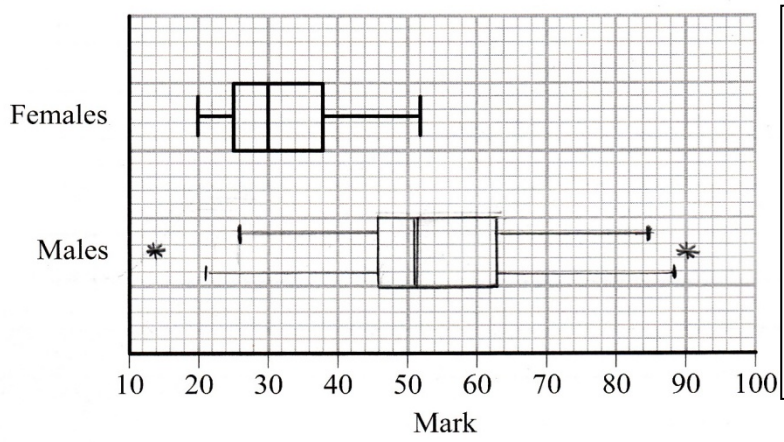
(4)

TOTAL FOR PAPER: 75 MARKS

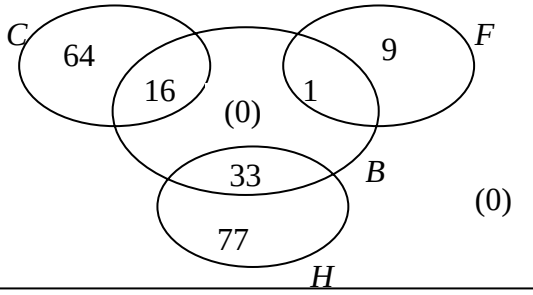
END

Question	Scheme	Marks
1. (a)	$(S_{th}) = 64980 - \frac{7150\times110}{9} = -22408.9\ldots$ $\quad -\underline{\mathbf{22\ 400}}$	M1 A1
	$(S_{hh}) = 7171500 - \frac{7150^2}{9} = 1491222.2\ldots$ $\quad \underline{\mathbf{1\ 490\ 000}}$	A1
		(3)
(b)	$r = \frac{-22408.9}{\sqrt{1491222\times371.56}} = -0.95200068\ldots$ $\quad$ awrt $\underline{\mathbf{-0.952}}$	M1A1
		(2)
(c)	Yes as r is close to -1 (if $-1 < r < -0.5$) or Yes as r is close to 1 (if $1 > r > 0.5$) [If $-0.5 \leq r \leq 0.5$ allow "no since r is close to 0"] [If $\lvert r\rvert > 1$ award B0]	B1ft
		(1)
(d)	$b = \frac{-22408.9}{1491222.2} = -0.015027...$ $\quad (\text{allow } \frac{-56}{3725})$ $\quad$ awrt -0.015	M1 A1
	$a = \frac{110}{9} - \text{"their } b\text{"}\times\frac{7150}{9} = (12.2 - -0.015\times794.4), = 24.1604...$ so $\boldsymbol{t = 24.2 - 0.015h}$	M1, A1
		(4)
(e)	0.015 is the drop in temp, (in ^{0}C), for every 1(m) increase in height above sea level.	B1
		(1)
(f)	Change = ("24.2 – 0.015"$\times$500) – ("24.2 – 0.015"$\times$1000) or 500 $\times$"0.015" $= \pm 7.5$ (awrt $\pm$ 7.5) (only ft a value < 100)	M1 A1ft **(2)**
		(13 marks)

	Notes
(a)	M1 for at least one correct expression (condone transcription error) 1st A1 for S_{hh} = awrt 1 490 000 or S_{th} = awrt $-22\ 400$ (Condone S_{xx} or S_{xy} =… or even S_{yy} =...) 2nd A1 for S_{th} = $-22\ 400$ and S_{hh} = 1 490 000 only. [This mark is assessing correct rounding] (Allow no labels but mis-labelling S_{th} as S_{hh} etc loses the final A1)
(b)	M1 for attempt at correct formula. Allow minor transcription errors of 2 or 3 digits. Must have their S_{hh}, S_{th} and given S_{tt} (3sf or better) in the correct places. Condone missing "$-$" Award M1A0 for awrt -0.95 with no expression seen. M0 for $\frac{64980}{\sqrt{7171500\times7.864}}$
(c)	B1ft must comment on supporting **and** state: high/strong/clear (negative or positive) correlation "points lie close to a straight line" is B0 since there is no evidence of this.
(d)	1st M1 for a correct expression for b. Follow through their S_{hh} & S_{th}. Condone missing "$-$" 1st A1 for awrt -0.015 or allow exact fraction from rounded values. 2nd M1 for a correct method for a. Follow through their value of b 2nd A1 for a correct equation for t and h with a = awrt 24.2 and b = awrt -0.015 No fractions
(e)	B1 Must mention h (or height) and t (or temperature) and their (1 sf) value of b in a correct comment
(f)	M1 for a correct expression seen based on their equation. Allow transcription error of 1 digit. If answer is 500$\times$ their b to 2sf and < 100 (M1A1), If answer is 500$\times$ their b to 2sf and $\geq$100 (M1A0)

Question	Scheme	Marks
2. (a)	25 (allow any x where $24 < x < 26$)	B1 **(1)**
(b)	**Q_2** (or median or m) = **51** **IQR** = 63 – 46 ,= **17** (or $Q_3 - Q_1 = 17$)	B1 M1, A1 **(3)**
(c)	Outliers given by $46 - 1.5 \times 17 = 20.5$ or $63 + 1.5 \times 17 = 88.5$ Outliers limits are **20.5** and **88.5**  Allow lower whisker to 20.5 and upper whisker to 88.5 Do **not** allow a mix of whiskers e.g 20.5 and 85 Do **not** allow both sets of whiskers	M1 A1 M1 A1ft B1 **(5)**
(d)	**Medians:** Median for females lower than males **IQR:** IQR for females smaller than males. Allow "lower/higher" but not "wider" **Range:** Range of females is less than males **Skewness:** Male and female marks are both positively skew Ignore other statements about average, spread, mean, st. Dev, variation, outliers etc	B1ft B1ft **(2)** **(11 marks)**

	Notes
	Mark (b) and (c) together BUT must see clear statement that median (or m or Q_2) = 51 and IQR = 17
(b)	M1 for 2 quartiles (at least one correct) and attempt to find the difference. Must see their 63 – their 46 A1 for 17 only. [Answer only of IQR= 17 scores M1A1]
(c)	**A fully correct box-plot (either version) with no supporting work scores 5/5. Otherwise:** 1st M1 for correct attempt to calc' at least one limit for outliers, ft their quartiles or IQR or award for sight of 20.5 or 88.5 1st A1 for identifying both limits of 20.5 and 88.5 2nd M1 for a box with an upper and a lower whisker(s) with at least 2 correct values (or correct ft) (condone no median marked) (condone 2 upper or 2 lower whiskers) 2nd A1ft for their 20.5 or 26 ,46, 51, 63 and 85 or their 88.5 in appropriate places and readable off their scale. Follow through their 20.5 and their 88.5 only, other values need to be correct If there are 2 upper or 2 lower whiskers A0 B1 for only 2 outliers appropriately marked at 14 and 90 Do not award if whiskers go beyond these values. **Apply ± 0.5 square accuracy for diagram** A box plot not on the graph paper can only score the 1st M1A1
(d)	**In (d) ft from their diagrams (if no diagram then use their values)** 1st B1ft for one correct comment comparing median, IQR , range or skewness 2nd B1ft for a second correct comment comparing median, IQR, range or skewness Do not allow contradictory statements

Question	Scheme	Marks
3. (a)	$\frac{35+75}{200}=0.55$	M1 A1 **(2)**
(b)	$\frac{200-2}{200}=0.99$	M1 A1 **(2)**
(c)	$\left[\mathrm{P}(W \mid C)\right]=\frac{\mathrm{P}(W \cap C)}{\mathrm{P}(C)}=\frac{30/200}{80/200}=\frac{30}{80}=0.375$	M1 A1 **(2)**
(d)	 Allow diagrams with intersections between *F*, *C* and *H* provided these are marked with 0. If their diagram indicates extra empty regions do not treat a blank as 0.	M1 B1 for 9, 1 B1 for 77,33 B1 for 64,16 **(4)**
(e)	$\frac{1+16+33}{200}=0.25$	M1 A1 **(2)**
		(12 marks)

	Notes
	Correct answers only score full marks for each part **If a probability is not in [0, 1] award M0**
(a)	M1 for denominator of 200 and attempt to add 2 + 8 or 35 + 75 or 30 + 50 A1 for 0.55 or exact equivalent fraction e.g. $\frac{11}{20}$
(b)	M1 for a fully correct expression (e.g. $1-0.01$) A1 for 0.99 or an exact equivalent fraction
(c)	M1 for a correct ratio or a correct formula and at least one correct prob (i.e. a correct num or denom). BUT award M0 if num is $\mathrm{P}(W)\times\mathrm{P}(C)=\frac{67}{200}\times\frac{80}{200}$ or if num>denom A1 for 0.375 or 3/8 or any exact equivalent.
(d)	M1 for a box and the 3 regions *F, C* and *H* labelled or <u>implied</u> and single set *B* labelled. There should be no intersections between *F, C* and *H* unless marked by zeros. They may have 3 circles for *F, C* and *B* with $H=F'\cap C'$ etc. Condone lack of zero in the <u>given</u> diagram. *F* 1st B1 for the 9 and 1 <u>or</u> 0.045 and 0.005 (o.e.) in the correct regions *H* 2nd B1 for the 77 and 33 <u>or</u> 0.385 and 0.165 (o.e.) in the correct regions *C* 3rd B1 for the 64 and 16 <u>or</u> 0.32 and 0.08 (o.e.) in the correct regions. May have *B* in 3 bits that are disconnected.
(e)	M1 for a numerator made up of their 1 + their 16 + their 33 and a denom of 200 and num < 200 Also allow sum of their probabilities (provided sum < 1) A1 for 0.25 or any exact equivalent

Question	Scheme	Marks
4. (a)	$\sum ft = 4837.5$ (allow 4838 or 4840)	B1
	Mean $= \frac{"4837.5"}{200} = 24.1875$ awrt **<u>24.2</u>** or $\frac{387}{16}$	M1 A1
	$\sigma = \sqrt{\frac{134281.25}{200} - \left(\frac{4837.5}{200}\right)^2}$	M1
	$= 9.293$ …..... (accept s =9.32) awrt **<u>9.29</u>**	A1 **(5)**
(b)	$Q_2 = [20.5] + \frac{(100/100.5 - 62)}{88} \times 5 = 22.659...$ awrt **<u>22.7</u>**	M1 A1 **(2)**
(c)	$Q_1 = 10.5 + \frac{(50/50.25)}{62} \times 10 [= 18.56]$ **(*)** (n + 1 gives 18.604…)	B1 cso **(1)**
(d)	Q_3= 25.5 (Use of n + 1 gives 25.734…)	B1
	IQR = 6.9 (Use of n + 1 gives 7.1)	B1 ft **(2)**
(e)	The data is skewed (condone "negative skew")	B1 **(1)**
(f)	Mean decreases and st. dev. remains the same. [Must mention mean and st. dev.] (from(a))	B1
	The median and quartiles would decrease. [Must refer to median <u>and</u> at least Q_1.] ((b)(c))	B1
	The IQR would remain unchanged (from (d))	B1 **(3)** **(14 marks)**

	Notes
	Correct answers only score full marks in each part except (c)
(a)	B1 for 4837.5 or 4838 or 4840 seen.
	If no $\sum ft$ seen (or attempt at $\sum ft$ seen), B1 can be implied by a correct mean of awrt 24.2
	1st M1 for attempt at their $\frac{\sum ft}{\sum f}$ allow 1sf so $\sum f$ = awrt 200 and $\sum ft$ = awrt 5000.
	<u>Or</u> award M1 for a clear attempt at mean where at least 4 correct products of $\sum ft$ are seen
	2nd M1 for correct expression including square root seen. Follow through their mean. Allow a transcription error in 134281.25 but not an incorrect re-calculation.
(b)	M1 for a correct fraction $\times 5$. Ignore end point but must be +. Allow use of (n + 1) giving 100.5…
(c)	B1cso for a fully correct expression including end point. NB Answer is given. Allow use of (n + 1) giving 50.25…but use of 50.5 scores B0
(d)	1st B1 for 25.5 (or awrt 25.7 using n + 1)
	2nd B1ft for their Q_3 – their Q_1 (or 18.6) (provided > 0) Accept awrt 2sf. Correct ans. only scores 2/2
(e)	B1 Must mention that the data is skewed or not symmetrical. Do not award for "outliers"
(f)	1st B1 for one correct comment from the above. May refer to parts (a), (b), (c) or (d)
	2nd B1 for two correct comments from the above
	3rd B1 for all 3 correct comments from the above

Question	Scheme	Marks
5. (a)	$3a + 2b = 0.7$	M1
	$a + 2a + 3a + 4b + 5b + 1.8 = 4.2$ or $6a + 9b = 2.4$	M1
	$5b = 1$ Attempt to solve	M1
	$b =$ **0.2** cao	B1
	$a =$ **0.1** cao	B1
		(5)
(b)	$E(X^2) = 1\times0.1 + 2^2\times0.1 + 3^2\times0.1 + 4^2\times0.2 + 5^2\times0.2 + 6^2\times0.3$ (= 20.4) (*)	B1cso
		(1)
(c)	[Var (X) =] $20.4 - 4.2^2$ [= 2.76]	M1
	Var$(5 - 3X)$ = 9 Var (X)	M1
	= **24.84** or **24.8** (allow $\frac{621}{25}$) cao	A1
		(3)
(d)	[$5k = 1$ so] $k =$ **0.2**	B1
		(1)
(e)	P($Y = 1$) = 0.1	B1
	e.g. P($Y = 2$) = F(2) – F(1) = 0.1	M1
	y | 1 | 2 | 3 | 4 | 5 P($Y = y$) | 0.1 | 0.1 | 0.4 | 0.2 | 0.2 Condone use of $X(x)$ instead of $Y(y)$ Ignore incorrect or no label if table fully correct	A1
		(3)
(f)	P($X = 1$) × P($Y = 1$) ,= **0.01** cao	M1, A1 **(2)**
		(15 marks)

	Notes
	Probabilities outside [0, 1] should be awarded M0
(a)	1st M1 for an attempt at a linear equation in a and b based on sum of probs. = 1
	2nd M1 for an attempt at a second linear equation in a and b based on E(X) = 4.2 Allow one slip.
	3rd M1 for an attempt to solve their 2 linear equations based on sum of probs and E(X) . Must reduce to a linear equation in one variable. 1st B1 for b and 2nd B1 for a. Answers only score B1B1 only
	The 3rd M1 may be implied if M2 is scored and both correct answers are given.
ALT	B1B1 for stating b and a.
	1st M1 for showing that sum of probs. = 1
	2nd M1 for showing that E(X) = 4.2
	3rd M1 for an overall comment "(therefore) a = …and b = …" No comment loses this mark.
(b)	B1cso for a fully correct expression (no incorrect work seen). E.g. allow $14\times0.1 + 41\times0.2 + 36\times0.3$
	Or $0.1 + 0.4 + 0.9 + 3.2 + 5 + 10.8$. Allow in a table (with 20.4) but without "+" explicitly seen.
(c)	1st M1 for a correct expression for Var(X). Must see -4.2^2
	2nd M1 for $(-3)^2$ Var(X) or better, no need for a value. Accept -3^2 if it clearly is used as +9 later.
(e)	B1 for P($Y = 1$) = 0.1
	M1 for correct use of F(y) to find one other prob. Can ft their k if finding P($Y = y$) for y >2
	Can be implied by one other prob. correct or correct ft Look out for P(3) = $3k$ – 0.2 or P(4) = P(5) = k.
	A1 for a fully correct probability distribution. Correct table only is 3/3
(f)	M1 for a correct expression or answer ft their P($Y = 1$) and their P($X = 1$)
	A1 for 0.01 or exact equivalent only
	Don't ISW here e.g. $0.1\times0.1 + 0.1\times0.1$ or $2\times0.1\times0.1$ are M0A0

Question	Scheme	Marks
6. (a)	[Let X be the amount of beans in a tin. P(X < 200) = 0.1]	
	$\frac{200-\mu}{7.8} = -1.2816$ [calc gives 1.28155156…]	M1 B1
	μ = 209.996…. awrt 210	A1
		(3)
(b)	$P(X > 225) = P\left(Z > \frac{225-"210"}{7.8}\right)$	M1
	$= P(Z > 1.92)$ or $1 - P(Z < 1.92)$ (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]	
	$\frac{210-205}{\sigma} = 2.3263$ or $\frac{200-205}{\sigma} = -2.3263$ [calc gives 2.3263478…]	M1 B1
	$\sigma = \frac{5}{2.3263}$	dM1
	$\sigma = 2.15$ (2.14933…)	A1
		(4)
		(10 marks)

	Notes
	Condone poor handling of notation if answers are correct but A marks must have correct working.
(a)	M1 for an attempt to standardise (allow ±) with 200 and 7.8 and set = ± any z value ($\|z\| > 1$)
	B1 for z = ± 1.2816 (or better used as a z)[May be implied by 209.996(102…) or better seen]
	A1 for awrt 210 (can be scored for using 1.28 but then they get M1B0A1)
	The 210 must follow from correct working – sign scores A0
	If answer is awrt 210 **and** 209.996… or better seen then award M1B1A1
	z = 1.28 gives 209.984 and z = 1.282 gives 209.9996 and both score M1B0A1
	If answer is awrt 210 or awrt 209.996 then award M1B0A1 (unless of course z = 1.2816 is seen)
(b)	M1 for attempting to standardise with 225, their mean and 7.8 . Allow ±
	1st A1 for Z > awrt 1.92/3. Allow a diagram but must have 1.92/3 and correct area indicated.
	Must have the Z so P(X > 225) with or without a diagram is not sufficient.
	Award for 1 – 0.9726 or 1 – 0.9732
	2nd A1 for 2.7 % or better (calculator gives 2.72…) Allow awrt 0.027. Correct ans scores 3/3
(c)	1st M1 for an attempt to standardise with 200 or 210, 205 and σ and set = ± any z value ($\|z\| > 2$)
	B1 for z = 2.3263 (or better) **and** compatible signs.
	If B0 in (a) for using a value in [1.28, 1.29) but not using 1.2816: allow awrt 2.33 here
	2nd dM1 **Dependent on the first M1** for correctly rearranging to make σ = ... May be implied
	e.g. $\frac{5}{\sigma} = 2.32 \rightarrow \sigma = 2.16$ (M1A0) BUT must have $\sigma > 0$
	A1 for awrt 2.15 . Must follow from correct working but a range of possible z values will do.
	NB $2.320 < z \leq 2.331$ will give an answer of awrt 2.15