Questions

Q1.

A machine cuts strips of metal to length L cm, where L is normally distributed with standard deviation 0.5 cm.

Strips with length either less than 49 cm or greater than 50.75 cm **cannot** be used.

Given that 2.5% of the cut lengths exceed 50.98 cm,

(a) find the probability that a randomly chosen strip of metal **can** be used.

(5)

Ten strips of metal are selected at random.

(b) Find the probability fewer than 4 of these strips **cannot** be used.

(2)

A second machine cuts strips of metal of length X cm, where X is normally distributed with standard deviation 0.6 cm

A random sample of 15 strips cut by this second machine was found to have a mean length of 50.4 cm

(c) Stating your hypotheses clearly and using a 1% level of significance, test whether or not the mean length of all the strips, cut by the second machine, is greater than 50.1 cm

(5)

(Total for question = 12 marks)

Q2.

A company sells seeds and claims that 55% of its pea seeds germinate.

(a) Write down a reason why the company should not justify their claim by testing all the pea seeds they produce.

(1)

A random selection of the pea seeds is planted in 10 trays with 24 seeds in each tray.

(b) Assuming that the company's claim is correct, calculate the probability that in at least half of the trays 15 or more of the seeds germinate.

(3)

(c) Write down two conditions under which the normal distribution may be used as an approximation to the binomial distribution.

(1)

A random sample of 240 pea seeds was planted and 150 of these seeds germinated.

(d) Assuming that the company's claim is correct, use a normal approximation to find the probability that at least 150 pea seeds germinate.

(3)

(e) Using your answer to part (d), comment on whether or not the proportion of the company's pea seeds that germinate is different from the company's claim of 55%

(1)

(Total for question = 9 marks)

Q3.

The number of hours of sunshine each day, y, for the month of July at Heathrow are summarised in the table below.

Hours	$0 \leqslant y < 5$	$5 \leqslant y < 8$	$8 \leqslant y < 11$	$11 \leqslant y < 12$	$12 \leqslant y < 14$
Frequency	12	6	8	3	2

A histogram was drawn to represent these data. The $8 \leq y < 11$ group was represented by a bar of width 1.5 cm and height 8 cm.

(a) Find the width and the height of the $0 \leq y < 5$ group.

(3)

(b) Use your calculator to estimate the mean and the standard deviation of the number of hours of sunshine each day, for the month of July at Heathrow.

Give your answers to 3 significant figures.

(3)

The mean and standard deviation for the number of hours of daily sunshine for the same month in Hurn are 5.98 hours and 4.12 hours respectably.

Thomas believes that the further south you are the more consistent should be the number of hours of daily sunshine.

(c) State, giving a reason, whether or not the calculations in part (b) support Thomas' belief.

(2)

(d) Estimate the number of days in July at Heathrow where the number of hours of sunshine is more than 1 standard deviation above the mean.

(2)

Helen models the number of hours of sunshine each day, for the month of July at Heathrow by $N(6.6, 3.7^2)$.

(e) Use Helen's model to predict the number of days in July at Heathrow when the number of hours of sunshine is more than 1 standard deviation above the mean.

(2)

(f) Use your answers to part (d) and part (e) to comment on the suitability of Helen's model.

(1)

(Total for question = 13 marks)

Q4.

The lifetime, L hours, of a battery has a normal distribution with mean 18 hours and standard deviation 4 hours.

Alice's calculator requires 4 batteries and will stop working when any one battery reaches the end of its lifetime.

(a) Find the probability that a randomly selected battery will last for longer than 16 hours.

(1)

At the start of her exams Alice put 4 new batteries in her calculator.
She has used her calculator for 16 hours, but has another 4 hours of exams to sit.

(b) Find the probability that her calculator will not stop working for Alice's remaining exams.

(5)

Alice only has 2 new batteries so, after the first 16 hours of her exams, although her calculator is still working, she randomly selects 2 of the batteries from her calculator and replaces these with the 2 new batteries.

(c) Show that the probability that her calculator will not stop working for the remainder of her exams is 0.199 to 3 significant figures.

(3)

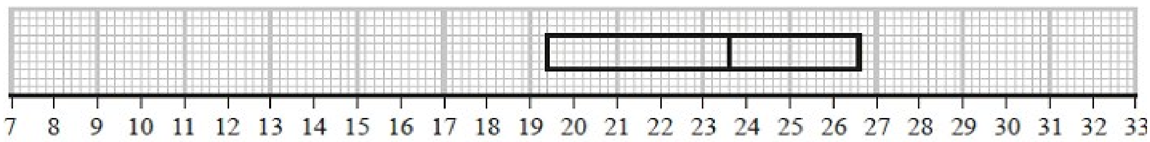
After her exams, Alice believed that the lifetime of the batteries was more than 18 hours.
She took a random sample of 20 of these batteries and found that their mean lifetime was 19.2 hours.

(d) Stating your hypotheses clearly and using a 5% level of significance, test Alice's belief.

(5)

(Total for question = 14 marks)

Q5.



Temperature (°C)

Figure 1

The partially completed box plot in Figure 1 shows the distribution of daily mean air temperatures using the data from the large data set for Beijing in 2015

An outlier is defined as a value

more than 1.5 × IQR below Q_1 or
more than 1.5 × IQR above Q_3

The three lowest air temperatures in the data set are 7.6 °C, 8.1 °C and 9.1 °C
The highest air temperature in the data set is 32.5 °C

(a) Complete the box plot in Figure 1 showing clearly any outliers.

(4)

(b) Using your knowledge of the large data set, suggest from which month the two outliers are likely to have come.

(1)

Using the data from the large data set, Simon produced the following summary statistics for the daily mean air temperature, x °C, for Beijing in 2015

$$n = 184 \qquad \sum x = 4153.6 \qquad S_{xx} = 4952.906$$

(c) Show that, to 3 significant figures, the standard deviation is 5.19 °C

(1)

Simon decides to model the air temperatures with the random variable

$$T \sim N(22.6, 5.19^2)$$

(d) Using Simon's model, calculate the 10^{th} to 90^{th} interpercentile range.

(3)

Simon wants to model another variable from the large data set for Beijing using a normal distribution.

(e) State two variables from the large data set for Beijing that are **not** suitable to be modelled by a normal distribution. Give a reason for each answer.

(2)

(Total for question = 11 marks)

Q6.

A machine puts liquid into bottles of perfume. The amount of liquid put into each bottle, D ml, follows a normal distribution with mean 25 ml

Given that 15% of bottles contain less than 24.63 ml

(a) find, to 2 decimal places, the value of k such that $P(24.63 < D < k) = 0.45$

(5)

A random sample of 200 bottles is taken.

(b) Using a normal approximation, find the probability that fewer than half of these bottles contain between 24.63 ml and k ml

(3)

The machine is adjusted so that the standard deviation of the liquid put in the bottles is now 0.16 ml

Following the adjustments, Hannah believes that the mean amount of liquid put in each bottle is less than 25 ml

She takes a random sample of 20 bottles and finds the mean amount of liquid to be 24.94 ml

(c) Test Hannah's belief at the 5% level of significance.

You should state your hypotheses clearly.

(5)

(Total for question = 13 marks)

Q7.

A health centre claims that the time a doctor spends with a patient can be modelled by a normal distribution with a mean of 10 minutes and a standard deviation of 4 minutes.

(a) Using this model, find the probability that the time spent with a randomly selected patient is more than 15 minutes.

(1)

Some patients complain that the mean time the doctor spends with a patient is more than 10 minutes.

The receptionist takes a random sample of 20 patients and finds that the mean time the doctor spends with a
patient is 11.5 minutes.

(b) Stating your hypotheses clearly and using a 5% significance level, test whether or not there is evidence to support the patients' complaint.

(4)

The health centre also claims that the time a dentist spends with a patient during a routine appointment,
T minutes, can be modelled by the normal distribution where $T \sim N(5, 3.5^2)$

(c) Using this model,

(i) find the probability that a routine appointment with the dentist takes less than 2 minutes

(1)

(ii) find $P(T < 2 \mid T > 0)$

(3)

(iii) hence explain why this normal distribution may not be a good model for *T*.

(1)

The dentist believes that she cannot complete a routine appointment in less than 2 minutes.

She suggests that the health centre should use a refined model only including values of $T > 2$

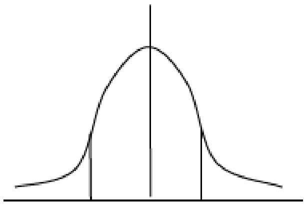
(d) Find the median time for a routine appointment using this new model, giving your answer correct to one decimal place.

(5)

(Total for question = 15 marks)

Mark Scheme

Q1.

Question	Scheme	Marks	AOs
(a)			
	$\text{P}(L > 50.98) = 0.025$	B1cao	3.4
	$\therefore \frac{50.98 - \mu}{0.5} = 1.96$	M1	1.1b
	$\therefore \quad \mu = 50$	A1cao	1.1b
	$\text{P}(49 < L < 50.75)$	M1	3.4
	$= 0.9104\ldots$ awrt **<u>0.910</u>**	A1ft	1.1b
		(5)	
(b)	S = number of strips that cannot be used so $S \sim \text{B}(10, 0.090)$	M1	3.3
	$= \text{P}(S \leqslant 3) = 0.991166\ldots$ awrt 0.991	A1	1.1b
		(2)	
(c)	$\text{H}_0 : \mu = 50.1$ $\quad \text{H}_1 : \mu > 50.1$	B1	2.5
	$\bar{X} \sim \text{N}\left(50.1, \frac{0.6^2}{15}\right)$ and $\bar{X} > 50.4$	M1	3.3
	$\text{P}(\bar{X} > 50.4) = 0.0264$	A1	3.4
	$p = 0.0264 > 0.01$ <u>or</u> $z = 1.936\ldots < 2.3263$ and not significant	A1	1.1b
	There is insufficient evidence that the **<u>mean length</u>** of strips is **<u>greater than 50.1</u>**	A1	2.2b
		(5)	
			(12 marks)

Notes:
(a) **1st M1:** for standardizing with μ and 0.5 and setting equal to a z value ($\lvert z \rvert > 1$) **2nd M1:** for attempting the correct probability for strips that can be used **2nd A1ft:** awrt 0.910 (allow ft of their μ)
(b) **M1:** for identifying a suitable binomial distribution **A1:** awrt 0.991 (from calculator)
(c) **B1:** hypotheses stated correctly **M1:** for selecting a correct model (stated or implied) **1st A1:** for use of the correct model to find p = awrt 0.0264 (allow z = awrt 1.94) **2nd A1:** for a correct calculation, comparison and correct statement **3rd A1:** for a correct conclusion in context mentioning "mean length" and 50.1

Q2.

Question	Scheme	Marks	AOs
(a)	The seeds would be destroyed in the process so they would have none to sell	B1	2.4
		(1)	
(b)	[S = no. of seeds out of 24 that germinate, $S \sim B(24, 0.55)$]		
	T = no. of trays with at least 15 germinating. $T \sim B(10, p)$	M1	3.3
	$p = P(S \geqslant 15) = 0.299126...$	A1	1.1b
	So $P(T \geqslant 5) = 0.1487...$ awrt <u>**0.149**</u>	A1	1.1b
		(3)	
(c)	n is large and p close to 0.5	B1	1.2
		(1)	
(d)	$X \sim N(132, 59.4)$	B1	3.4
	$P(X \geqslant 149.5) = P\left(Z \geqslant \frac{149.5 - 132}{\sqrt{59.4}}\right)$	M1	1.1b
	$= 0.01158...$ awrt <u>**0.0116**</u>	A1cso	1.1b
		(3)	
(e)	e.g The probability is very small therefore there is evidence that the company's claim is incorrect.	B1	2.2b
		(1)	
			(9 marks)

Notes:

(a)
B1: cao

(b)
M1: for selection of an appropriate model for T
1st A1: for a correct value of the parameter p (accept 0.3 or better)
2nd A1: for awrt 0.149

(c)
B1: both correct conditions

(d)
B1: for correct normal distribution
M1: for correct use of continuity correction
A1: cso

(e)
B1: correct statement

Q3.

Question	Scheme	Marks	AOs
(a)	**Area** = $8\times1.5=12$ cm^2 **Frequency** = 8 so 1 cm^2 = $\frac{2}{3}$ **hour (o.e.)**	M1	3.1a
	Frequency of 12 corresponds to area of 18 so height = $18\div2.5$ = **7.2 (cm)**	A1	1.1b
	Width = 5×0.5 = **2.5 (cm)**	B1cao	1.1b
		(3)	
(b)	$[\bar{y}=]\ \frac{205.5}{31}$ = awrt 6.63	B1cao	1.1b
	$[\sigma_y=]\ \sqrt{\frac{1785.25}{31}-\bar{y}^2}$ = $\sqrt{13.644641}$ = **awrt 3.69** allow $[s=]\ \sqrt{\frac{1785.25-31\bar{y}^2}{30}}$ = **awrt 3.75**	M1 A1	1.1a 1.1b
		(3)	
(c)	Mean of Heathrow is higher than Hurn and standard deviation smaller suggesting Heathrow is more reliable	M1	2.4
	Hurn is South of Heathrow so does <u>not</u> support his belief	A1	2.2b
		(2)	
(d)	$\bar{x}+\sigma\approx10.3$ so number of days is e.g. $\frac{(11-\text{"}10.3\text{"})}{3}\times8\ (+5)$	M1	1.1b
	= 6.86 so **7 days**	A1	1.1b
		(2)	
(e)	[H = no. of hours] P(H > 10.3) or P(Z > 1) = [0.15865…]	M1	3.4
	Predict $31\times0.15865...$ = <u>**4.9 or 5 days**</u>	A1	1.1b
		(2)	
(f)	(5 or) 4.9 days < (7 or) 6.9 days so model may **not** be suitable	B1	3.5a
		(1)	
			(13 marks)

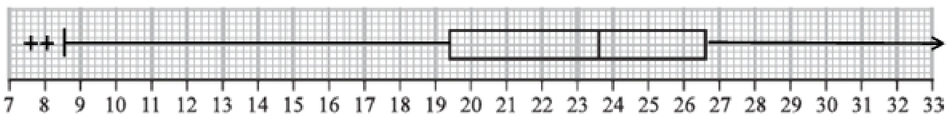
Notes:	
(a)	
M1:	for clear attempt to relate the area to frequency. Can also award if their height $\times$ their width = 18
A1:	for height = 7.2 (cm)
(b)	
M1:	for a correct expression for σ or s, can ft their value for mean
A1:	awrt 3.69 (allow $s = 3.75$)
(c)	
M1:	for a suitable comparison of standard deviations to comment on reliability.
A1:	for stating Hurn is south of Heathrow and a correct conclusion
(d)	
M1:	for a correct expression – ft their $\bar{x} + \sigma \approx 10.3$
A1:	for 7 days but accept 6 (rounding down) following a correct expression
(e)	
M1:	for a correct probability attempted
A1:	for a correct prediction
(f)	
B1:	for a suitable comparison and a compatible conclusion

Q4.

Qu	Scheme	Marks	AO
(a)	$P(L > 16) = 0.69146...$ **awrt 0.691**	B1 (1)	1.1b
(b)	$P(L > 20 \mid L > 16) = \dfrac{P(L > 20)}{P(L > 16)}$	M1	3.1b
	$= \dfrac{0.308537...}{(a)}$ or $\dfrac{1-(a)}{(a)}$, $= 0.44621...$	A1ft, A1	1.1b 1.1b
	For calc to work require $(0.44621...)^4 = 0.03964...$ **awrt <u>0.0396</u>**	dM1 A1 (5)	2.1 1.1b
(c)	Require: $[P(L > 4)]^2 \times [P(L > 20 \mid L > 16)]^2$	M1	1.1a
	$= (0.99976...)^2 \times (\text{"}0.44621...\text{"})^2$	A1ft	1.1b
	$= 0.19901...$ awrt <u>0.199</u> (*)	A1cso* (3)	1.1b
(d)	$H_0: \mu = 18 \quad H_1: \mu > 18$	B1	2.5
	$\bar{L} \sim N\left(18, \left(\dfrac{4}{\sqrt{20}}\right)^2\right)$	M1	3.3
	$P(\bar{L} > 19.2) = P(Z > 1.3416...) = 0.089856...$	A1	3.4
	$(0.0899 > 5\%)$ or $(19.2 < 19.5)$ or $1.34 < 1.6449$ so not significant	A1	1.1b
	Insufficient evidence to support Alice's claim (or belief)	A1 (5)	3.5a
		(14 marks)	

	Notes
(a)	B1 for evaluating probability using their calculator (awrt 0.691) Accept 0.6915
(b)	1st M1 for a first step of identifying a suitable conditional probability (either form) 1st A1ft for a ratio of probabilities with numerator = awrt 0.309 or 1 – (a) and denom = their (a) 2nd A1 for awrt 0.446 (o.e.) Accept 0.4465 (from $\frac{0.3085}{0.691} = 0.44645...$) NB $\frac{P(16<L<20)}{P(L>16)} = 0.5538...$ scores M1A1A1 when they do $1 - 0.5538 = 0.4462...$ 2nd M1 (dep on 1st M1) for 2nd correct step i.e. (their $0.446...)^4$ or $X \sim B(4, \text{"}0.446\text{"})$ and $P(X = 4)$ 3rd A1 for awrt 0.0396
(c)	1st M1 for a correct approach to solving the problem (May be implied by A1ft) 1st A1ft for $P(L > 4)$ = awrt 0.9998 used <u>and</u> ft their 0.44621 in correct expression If use $P(L > 20) = 0.3085..$ as 0.446.. in (b) then M1 for $(0.3085..)^2 \times [P(L > 4)]^2$; A1ft as above
*	2nd A1cso for 0.199 or better with clear evidence of M1 [NB $(0.4662..)^2 = 0.199...$ is M0A0A0] **Must see M1 scored by correct expression in symbols or values (M1A1ft)**
(d)	B1 for both hypotheses in terms of μ M1 for selecting a suitable model. Sight of <u>normal</u>, <u>mean</u> 18, <u>sd</u> $\frac{4}{\sqrt{20}}$ (o.e.) or <u>variance</u> = 0.8 1st A1 for using the model correctly. Allow awrt 0.0899 <u>or</u> 0.09 from correct prob. statement
ALT	**CR** $(\bar{L}) > 19.471...$ (accept awrt 19.5) <u>or</u> **CV** of 1.6449 (or better: calc 1.6448536..)
	2nd A1 for correct non-contextual conclusion. Wrong comparison or contradictions A0 Error giving 2nd A0 implies 3rd A0 but just a correct contextual conclusion can score A1A1 3rd A1 dep on M1 and 1st A1 for a correct contextual conclusion mentioning <u>Alice's claim</u> /<u>belief</u> <u>or</u> there is insufficient evidence that the mean <u>lifetime</u> is more than 18 hours

Q5.

Part	Working or answer an examiner might expect to see	Mark	Notes
(a)	IQR = 2.6. − 19.4 = 7.2	B1	This mark is given for finding the interquartile range
	$19.4 - (1.5 \times 7.2) = 8.6$ $19.4 + (1.5 \times 7.2) = 37.4$	M1	This mark is given for a method find the values for the whiskers of the boxplot
			
		A1	This mark is given for plotting the correct whisker (8.6) on the boxplot
		A1	This mark is given for plotting the two correct outliers 7.6 °C and 8.1 °C
(b)	October (since it is the month with the coldest temperatures between May and October in Beijing)	B1	This mark is given for a correct suggestion with a supporting reason.
(c)	$\sigma = \sqrt{\frac{S_{xx}}{n}} = \sqrt{\frac{4952.906}{184}} = \sqrt{26.92} = 5.19$	B1	This mark is given for showing the calculation for the standard deviation to three significant figures
(d)	z = (±) 1.2816	B1	This mark is given for identifying the z-value for the 10th and 90th percentiles (from tables or calculator)
	$2 \times z \times 5.19$	M1	This mark is given for a method to find the interpercentile range between the 10th and 90th value
	= 13.303	A1	This mark is given for finding a correct interpercentile range between the 10th and 90th value
(e)	Daily wind speed (Beaufort) since it is qualitative data	B1	This mark is given for stating a correct variable with a supporting reason
	Rainfall (since it is not symmetric)	B1	This mark is given for stating a correct variable with a supporting reason
			(Total 11 marks)

Q6.

Part	Working or answer an examiner might expect to see	Mark	Notes
(a)	$\frac{24.63-25}{\sigma} = -1.0364$	M1	This mark is given for standardising as part of a method to find σ
	$\sigma = 0.357$	A1	This mark is given for a correct value of σ
	$P(D > K) = 0.4$ or $P(D < K) = 0.6$	B1	This mark is given for
	$\frac{k-25}{\sigma} = \frac{k-25}{0.357} = 0.2533$	M1	This mark is given for using a normal model to find the probability
	$k = 25.09$	A1	This mark is given for a correct value for k
(b)	$Y \sim B(200, 0.45)$ so $W \sim N(90, 49.5)$	B1	This mark is given for setting up the normal distribution approximation of the binomial
	$P(Y < 100) \approx P(W < 99.5) = P\left(Z < \frac{99.5-90}{\sqrt{49.5}}\right)$	M1	This mark is given for using the normal model with a continuity correction
	$= 0.912$	A1	This mark is given for finding a correct value of the probability
(c)	$H_0 : \mu = 25$ $H_1 : \mu < 25$	B1	This mark is given for both hypotheses in terms of μ found correctly
	$\overline{D} \sim N\left(25, \frac{0.16^2}{20}\right)$	M1	This mark is given for a method to set up the normal distribution
	$P(\overline{D} < 24.94) = 0.0468$	A1	This mark is given for using the model to find a correct p-value
	$p = 0.0468 < 0.05$, so reject H_0	M1	This mark is given for a correct comparison and non-contextual conclusion
	There is sufficient evidence to support Hannah's belief	A1	This mark is given for a correct conclusion in context stated
			(Total 13 marks)

Q7.

	Scheme	Marks	AO
(a)	{Let X = time spent, $P(X > 15) =$} 0.105649… awrt **0.106**	B1	1.1b
		(1)	
(b)	$H_0 : \mu = 10 \quad H_1 : \mu > 10$	B1	2.5
	$\bar{X} \sim N\left(10, \left(\frac{4}{\sqrt{20}}\right)^2\right)$; $P(\bar{X} > 11.5) = 0.046766\ldots$ [Condone 0.9532…]	M1;A1	3.3;3.4
	[This is significant (< 5%) so] there is evidence to support the complaint	A1	2.2b
		(4)	
(c)(i)	[$P(T < 2) =$] 0.1956… awrt **0.196**	B1	1.1b
		(1)	
(ii)	Require $\frac{P(0 < T < 2)}{P(T > 0)} = \frac{0.119119...}{0.923436...}$; $= 0.1289955\ldots$ awrt **0.129**	M1 A1;A1	3.4 1.1bx2
		(3)	
(iii)	The current model suggests **non-negligible** probability of T values < 0 which is impossible	B1	3.5b
		(1)	
(d)	Require t such that $P(T > t \mid T > 2) = 0.5$ or $P(T < t \mid T > 2) = 0.5$	M1	3.1b
	e.g. $\frac{P(T > t)}{P(T > 2)} = 0.5$; so $P(T > t) = 0.5 \times [1 - \text{(c)(i)}]$ or $P(T > t) = 0.5 \times 0.8043..$	M1; A1ft	1.1b 3.4
	[i.e. $P(T > t) = 0.40\ldots$ implies] $\frac{t-5}{3.5} = 0.2533$ or $P(T < t) =$ "0.5978.."	M1	1.1b
	$t = 5.886\ldots$ or from calculator 5.867… so awrt **5.9**	A1 **(5)**	1.1b
		(15 marks)	

	Notes
(a)	B1 for awrt 0.106 (from calculator) [Allow 10.6%]
(b)	B1 for both hypotheses correct in terms of μ.
	M1 for selection of a correct model (sight or use of correct normal- may not have label $\bar{X}$)
	1st A1 for use of this model to get probability allow 0.046~0.047 [Condone awrt 0.953]
ALT	**OR** test statistic $z = 1.677\ldots$ (awrt 1.68) and cv of 1.64 (or better) **or** CR $\bar{X} > 11.47..$
	2nd A1 (dep on 1st A1 or at least $P(\bar{X} > 11.5) < 0.05$ (o.e.))
	for a correct conclusion in context -must mention **complaint**/claim or **time**/mins is > 10
SC	**(M0 for $\bar{X}$ ~N(11.5, …)** for correct probability **and** conclusion (score M0A0A1 on epen)
(c)(i)	B1 for awrt 0.196 (from calculator) [Allow 19.6%]
(ii)	M1 for a correct probability ratio expression (may be implied by 1st A1 scored)
	1st A1 for a correct ratio of probabilities (both correct or truncated to 2 dp)
	2nd A1 for awrt 0.129
(iii)	B1 for a suitable explanation of why model is not suitable based on negative T values
	Must say that a **significant** proportion of values < 0 (o.e.) e.g. $P(T > 0)$ should be **closer** to 1
	or Difference between $P(T < 2 \mid T > 0)$ and $P(T < 2)$ is **too big** (o.e.)
(d)	1st M1 for a correct conditional probability statement to start the problem or $0.5 \times P(T > 2)$
	2nd M1 for correct ratio of probability expressions [Must have $P(T > t)$ or $P(2 < T < t)$]
	1st A1ft for a correct equation for $P(T > t)$ (o.e.) ft their answer to part (c)[May be in a diagram]
	3rd M1 for attempt to find t (standardising and sight of 0.2533) or prepare to use calc (ft)
	Arriving at $P(T <$ median$) = 1 - 0.5 \times$ "their 0.8043" will score 1st 4 marks
	2nd A1 for awrt 5.9
	Sight of awrt 5.9 and at least one M mark scores 5/5 [Answer only send to review]