1. 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)

a) Random Variable $L \sim N(\mu, 0.5^2)$, $\sigma = 0.5$

$P(L > 50.98) = 0.025$ (1)

$\boxed{1}$ Find the mean (μ)

$\boxed{2}$ Find the probability ($P(49 < L < 50.75)$)

Z-Score $= \frac{x - \mu}{\sigma}$ = 1.96 (1)

$1 - 0.025 = 0.9750$

look on table / or do on calculator

$\Rightarrow 1.96 = \frac{50.98 - \mu}{0.5} \Rightarrow 50.98 - \mu = 1.96 \times 0.5$

$\mu = 50.98 - 1.96 \times 0.5$

$\mu = 49.97 = \underline{50}$ cm (1)

(1)

$\Rightarrow L \sim N(50, 0.5^2) \Rightarrow P(49 < L < 50.75) = P(L < 50.75) - P(L < 49)$

$P(L < 50.75) = P\left(Z < \frac{50.75 - 50}{0.5}\right) = P(Z < 1.5) = \Phi(1.5) = 0.9332$

$P(L < 49) = P\left(Z < \frac{49 - 50}{0.5}\right) = P(Z < -2) = 1 - \Phi(2) = 0.0228$

$\Rightarrow P(49 < L < 50.75) = 0.9332 - 0.0228 = \underline{0.910}$ (1)

b) The probability that a strip cannot be used will be equal to
$1-0.910$ * (0.910 was our answer to part a)

Now, if we let X be a random variable which denotes the number of strips that cannot be used then we're going to have that X is binomially distributed, with $n=10$ and $p=0.09$. $(1-0.910=0.09)$ *

$\Rightarrow X \sim B(10, 0.09)$ (1). $\Rightarrow P(X \leq 3) = \underline{0.99}$ (1)

c) $n = 15$, $\sigma = 0.6\text{cm}$, Sample mean $\bar{x} = 50.4\text{cm}$

$H_0 : \mu = 50.1\text{cm}$ v.s. $H_1 : \mu > 50.1\text{cm}$ (1) (one-sided test)

Standard error of the mean : $\sigma/\sqrt{n} = \dfrac{0.6}{\sqrt{15}}$

$\Rightarrow \bar{X} \sim N\left(50.1, \dfrac{0.6^2}{15}\right)$ (1)

$\Rightarrow P(\bar{X} > 50.4) = P\left(Z > \dfrac{50.4-50.1}{0.6/\sqrt{15}}\right) = P(Z > 1.94) = 1 - \Phi(1.94)$
$= \underline{0.026}$ (p-value) (1)

$\Rightarrow 0.026 > 0.01 = \alpha$ (1)

$\Rightarrow$ Do not reject H_0 and we can conclude that there is insufficient evidence that the mean length of the strips is greater than $\underline{50.1\text{cm}}$. (1)

2. 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)

a) If they test all their peas then they will have none to sell. This is not an effective business method, as they can't get any income if they have destroyed all their peas. 1

b) let S is the random variable which is the number of seeds out of 24 that germinate. $\Rightarrow S \sim B(24, 0.55)$.

$n = 24$
$p = 0.55$

let T be the random variable which is the number of trays with at least 15 or more seeds germinating. $= T \sim B(10, q)$ 1

$\Rightarrow q = P(S \geq 15)$ $(n = 24, p = 0.55)$

$q = 0.299$ 1

$\Rightarrow T \sim B(10, 0.299)$

$\Rightarrow P(T \geq 5) = 0.149$ 1

c) • n is large
• P (probability) must be close to $\frac{1}{2}$ or 0.5 (1)

d) $X \sim N(\mu, \sigma^2)$

When this approximation : $\mu = n \cdot p$ and $\sigma^2 = npq$ $n = 240, p = 0.55$
$\Rightarrow \mu = 240 \times 0.55 = 132$ and $\sigma^2 = 240 \times 0.55 \times 0.45 = 59.4$ $q = 0.45$

$\Rightarrow X \sim N(132, 59.4)$ (1) * we must use the continuity correction, so when we work out our probability, we subtract 0.5. *

$\Rightarrow P(X \geqslant 149.5) = P\left(Z \geqslant \frac{149.5 - 132}{\sqrt{59.4}}\right) = P(Z \geqslant 2.27) = \underline{\underline{0.0116}}$ (1) (1)

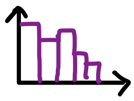
e) Probability from part d was 0.0116, which is a very small number.
This small number tells us that there is evidence to suggest that the company's claim is <u>incorrect</u>. (1)

3. 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 $8 \leqslant y < 11$ group was represented by a bar of width 1.5 cm and height 8 cm.

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



a) $\text{Frequency Density} = \dfrac{\text{Frequency}}{\text{Width}}$ and $\dfrac{\text{Freq Density 1}}{\text{Freq Density 2}} = \dfrac{\text{Height 1}}{\text{Height 2}}$

Freq. Density 1: $8 \leq y < 11$, Freq : 8, Width = 11 − 8 = 3

Freq. Density 2 : $0 \leq y < 5$, Freq : 12, Width = 5 − 0 = 5

Freq. Density 1 = $\dfrac{8}{3}$ and Freq. Density 2 = $\dfrac{12}{5}$

$\Rightarrow \dfrac{8/3}{12/5} = \dfrac{8}{H_2} \Rightarrow H_2 = \dfrac{12/5 \cdot 8}{8/3} = 7.2\text{ cm}$ (1) (1)

Width of $8 \leq y < 11$ = 1.5cm : 3

$\Rightarrow 1.5/3 : 1$

$0 \leq y < 5 \Rightarrow \frac{1.5}{3} \times 5 = \frac{5}{2} : 5$ $\Rightarrow$ Width is 2.5cm (1)

$\frac{5}{2} = 2.5$

(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)

b) Mean : Standard Deviation :

We need to find the midpoint of each interval :

2.5, 6.5, 9.5, 11.5, 13 (midpoints for calculator) (1)

12, 6, 8, 3, 2 (frequencies) Mean = 6.63 cm (1) and Standard deviation = 3.69 cm (1)

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)

c)

Heathrow : $\mu = 6.63$ Hurn : $\mu = 5.98$
$\sigma = 3.69$ $\sigma = 4.12$

For a lower standard deviation, data will be more consistent.
$\Rightarrow$ The number of hours of daily sunshine is more consistent at (1) Heathor, but Hurn is further South than Heathrow, so therefore Thomas' belief is not Supported. (1)

(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)

d) For Heathrow, $\mu = 6.63$ and $\sigma = 3.69$.
1 Standard deviation above the mean : $\mu + \sigma = 6.63 + 3.69 = 10.32$.

Since $11 > 10.32$, all the observations in the $11 \leq y < 12$ and $12 < y \leq 14$ must be greater than 10.32 (3 and 2 observations respectively).

• $8 \leq y < 11$, we need to estimate how many observations are in this group and are greater than 10.32.

(1)
$\frac{11 - 10.32}{3} \times 8 = 1.8$ $\Rightarrow$ we estimate that there is 1.8 observations which are greater than 10.32. $\Rightarrow 1.8 + 3 + 2 = 6.8 = 7$ days (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)

e) $N(6.6, 3.7^2)$, $\mu = 6.6$, $\sigma = 3.7$
$\Rightarrow Z = \frac{X - \mu}{\sigma} = \frac{10.32 - 6.6}{3.7} = 1.0054$, $P(X > 10.32) = P(X > 1)$
$= 1 - P(X \leq 1) = 1 - 0.841...$
$\Rightarrow$ Number of days $= 31 \times 0.159 = 4.9$ days (1) $= 0.159.$ (1)

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

(1) (1)

f) Part d: $\mu = 6.8$ days, Part e : $\mu = 4.9$ days. $6.8 \neq 4.9 \Rightarrow$ The model is not Suitable.

4. 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)

a) $L \sim N(18, 4^2)$

$P(L>16) = 0.691$ (3s.f.) - 1

b) $P(L>20 \mid L>16) = \dfrac{P(L>20)}{P(L>16)}$ - 1

$= \dfrac{0.3085...}{0.6915...}$ - 1

(for one battery) $\rightarrow = 0.4462$ (4d.p) - 1

(for all four)

$(0.4462)^4 = 0.0396$ - 2

c) $(P(L>4))^2 \times (P(L>20 \mid L>16))^2$ - 1

$= (0.9998...)^2 \times (0.4462...)^2$ - 1

$= 0.199$ (3s.f.) - 1

d) $H_0: \mu = 18$ $H_1: \mu > 18$ - ①

$\bar{L} \sim N\left(18, \frac{4^2}{20}\right)$ $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

- ①

$\bar{L} \sim N\left(18, \left(\frac{4}{\sqrt{20}}\right)^2\right)$

$P(\bar{L} > 19.2) = 0.0899$ (4 d.p.) - ①

$0.0899 > 0.05$

Insufficient evidence to reject H_0. - ①

Therefore there's not enough evidence to support Alice's belief. - ①

5.

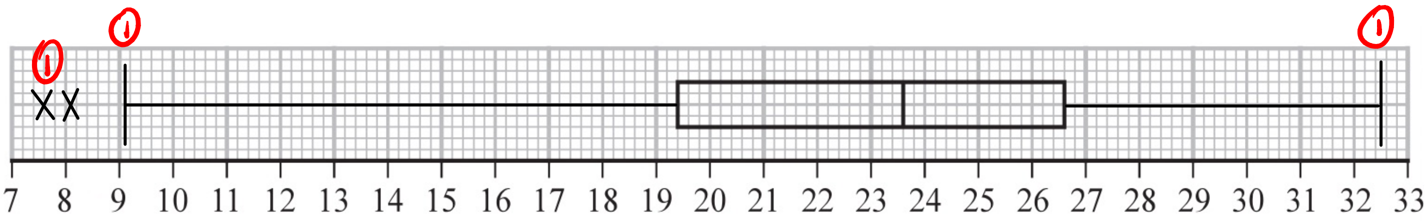


Figure 1

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

$Q_1 = 19.4$ $\quad$ $IQR = 26.6 - 19.4 = 7.2$ 1

$Q_3 = 26.6$ get values from Figure 1

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)

$1.5 \times 7.2 = 10.8$ $\quad$ $19.4 - 10.8 = 8.6$ 1 $\therefore$ since $7.6, 8.1 < 8.6$
we know 7.6°C and 8.1°C are outliers
$26.6 + 10.8 = 37.4$ $\quad$ $32.5 < 37.4$ so NOT outlier

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

(1)

October 1 (Since it's the month with the coldest temperatures between May and October in Beijing)

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

$\Sigma(x-\bar{x})^2$

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

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

(1)

$\sigma = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n}}$ $\quad$ 1 $\quad$ $\sigma = \sqrt{\dfrac{4952.906}{184}} = \sqrt{26.917967\ldots}$

$= 5.188\ldots = 5.19$ (3sf)

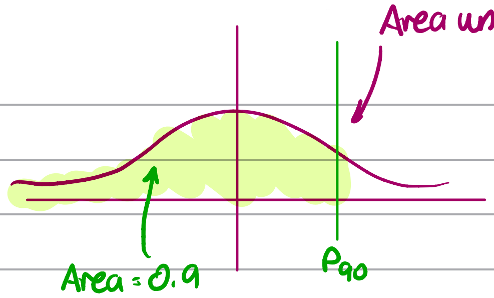
Simon decides to model the air temperatures with the random variable

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

μ σ Normal distribution

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

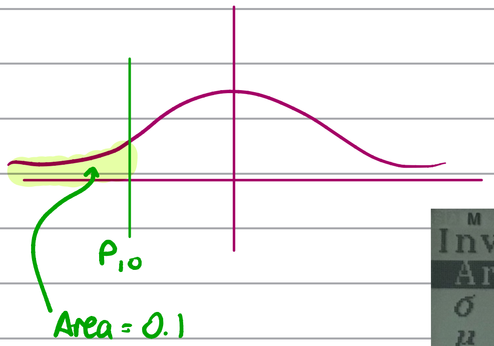
(3)

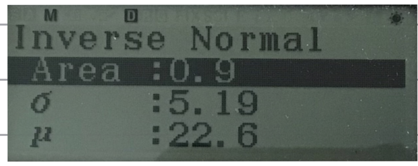


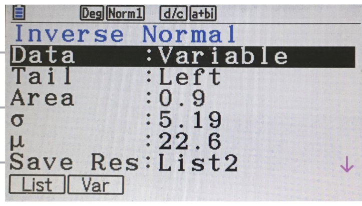
$P(T < P_{90}) = 0.9$ $\quad \mu = 22.6$

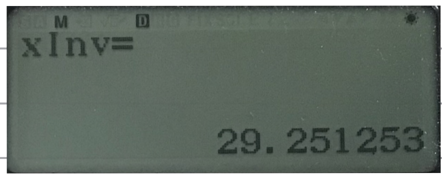
$P(T < P_{10}) = 0.1$ $\quad \sigma = 5.19$

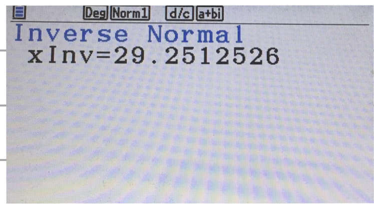
using inverse normal mode on the calculator











$P_{90} = 29.251\ldots$ (1)

Setting Area = 0.1 on Calculator

$P_{10} = 15.948\ldots$

Interpercentile range $= 29.251\ldots - 15.948\ldots = 13.3$ (1dp)

(1) (1)

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)

- Daily mean wind speed / Beaufort conversion since it's qualitative
- Rainfall since it's not symetric (lots of days with 0 rainfall)

6. 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)

a) $D \sim N(25, \sigma^2)$ μ

Standard normal distribution

$P(D < 24.63) = 0.15$

using standard normal distribution $Z \sim N(0, 1^2)$ μ σ

$P\left(Z < \frac{24.63 - 25}{\sigma}\right) = 0.15$

$\frac{24.63-25}{\sigma} = -1.0364$ (1)

$z = \frac{x - \mu}{\sigma}$

work out using inverse normal mode on calculator with $\mu = 0$ $\sigma = 1$

$\sigma = \frac{24.63-25}{-1.0364}$

$= 0.357$ (1)

$\therefore D \sim N(25, 0.357^2)$

$P(24.63 < D < k) = 0.45$

24.63 k

$P(D<k) - P(D<24.63) = 0.45$

$P(D<k) - 0.15 = 0.45$

(1) $P(D<k) = 0.6$

inverse normal $\mu = 25$ $\sigma = 0.357$

(2) $k = 25.09$ (2dp)

b) $Y \sim B(200, 0.45)$ Mean = np = 90
Variance = $np(1-p)$ = 49.5

$[Y \sim B(200, 0.45) \rightarrow]$ $W \sim N(90, 49.5)$ (1)

$P(Y < 100) \approx P(W < 99.5)$ (1)

$P(W < 99.5) = 0.9115... = 0.912$ (3dp) (1)

c) $H_0: \mu = 25$ $n = 20$ Sample mean = 24.94 SL = 5%
(1)
$H_1: \mu < 25$ Assume H_0 is true $D \sim N(25, 0.16^2)$

For a random sample of size n taken from a random variable $X \sim N(\mu, \sigma^2)$, the sample mean, $\bar{X}$, is normally distributed with $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$\bar{D} \sim N(25, \frac{0.16^2}{20})$ (1) $P(\bar{D} < 24.94) = 0.046766...$ (1)

$0.047 < 0.05$ (1) $\therefore$ Reject H_0, there is sufficient evidence to support Hannah's belief (1)

7. 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)

a) let X = time spent $X \sim N(10, 4^2)$ μ σ^2

$P(X > 15) = 0.105649... = 0.106$ (3dp) (1)

b) $H_0: \mu = 10$ (1)
$H_1: \mu > 10$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ where n = sample size

Significance level = 5%

$\bar{X} \sim N(10, \frac{4^2}{20})$ (2) $\frac{4^2}{20} = \sigma^2 \Rightarrow \frac{4}{\sqrt{20}} = \sigma$ (1)

$P(\bar{X} > 11.5) = 0.046766...$ 0.0468 (4dp) < 0.05 so there is evidence to support the patients complaint

c) i) $P(T<2) = 0.1956\ldots = 0.196$ (3dp) 1

$T \sim N(5, 3.5^2)$

ii) $P(T<2 \mid T>0)$ $\quad P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

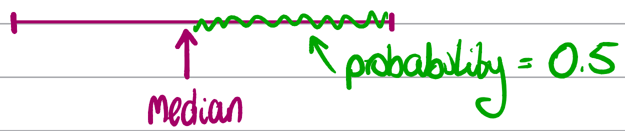
'given that'

3

$$P(T<2 \mid T>0) = \frac{P(0<T<2)}{P(T>0)} = \frac{0.119119\ldots}{0.923436\ldots} = 0.1289955\ldots = 0.129 \text{ (3dp)}$$

$P(T<0) \approx 0.077$

iii) The current model suggests non-negligible probability of T values <0 which is impossible 1 (length of time can't be negative so suggests model unsuitable)

d) 

$P(T>t \mid T>2) = 0.5$ 1

median

using conditional probability formula

$$P(T>t \mid T>2) = \frac{P(T>t)}{P(T>2)} = 0.5 \quad 1$$

$T \sim N(5, 3.5^2)$

Note: $P(T>t \cap T>2) = P(T>t)$ since new model only includes $T>2$ so median definetly more than 2

$$\frac{P(T>t)}{0.8043\ldots} = 0.5$$

$P(T>t) = 0.40215\ldots$ 1

$P(T<t) = 1 - 0.40215\ldots = 0.5978\ldots$ 1

Need to do this step for calculators that can only test the 'lower tail'

$t = 5.867\ldots$

$= 5.9$ (1dp) 1

use inverse normal on calculator