Symmetric Confidence Intervals from Small Samples Using the t-Distribution

When the true population variance is unknown, which can be the case for real-life examples, the population mean can still be estimated using a **sample variance**, s^2. Instead normal distributions, the ***t*-distribution** is used for the population.

For a t-distribution when the sample size is sufficiently small, the c% symmetric confidence interval for the population mean, μ, is :

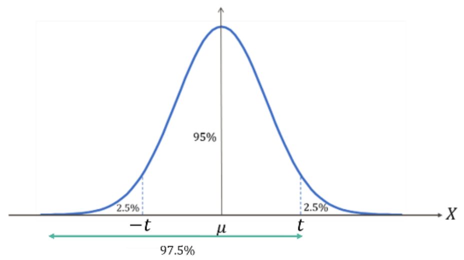
$$\left(\bar{x} - t\frac{s}{\sqrt{n}}, \bar{x} + t\frac{s}{\sqrt{n}}\right)$$

where $\bar{x}$ is the sample mean, n is the sample size and t is a statistic chosen such that $P(t_v < t) = 0.5 + \frac{\frac{1}{2}c}{100}$.

This can also be given as a range:

$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}.$$

When the sample is sufficiently small, the population size should be less than 30. A normal distribution must also be used to obtain the sample. The distribution for t-statistic is shown in the graph below.



It is possible to see that with determining the t value when looking for a 95% confidence interval, the upper bound of the interval can be used, in this case the 97.5% point.

The equation $v = n - 1$ can be used to obtain the number of **degrees of freedom**. This can be used to determine the t-score in conjunction with the tables for the values of v and c%.

Example 1: A sample of shoe sizes is taken for a class: $\{3, 4, 5, 6, 6, 7, 8, 8\}$. Assuming the data is drawn from a normal distribution, find a 90% confidence interval for the mean of the data.

The sample mean and an unbiased estimate for the variance can be determined.	$\bar{x} = \frac{47}{8} = 5.875$ $\sigma^2 = \frac{\sum x^2}{n} - \bar{x}^2 = \frac{245.5}{7} - 5.875^2 \approx 2.86$ $s^2 = \frac{8}{7} \times \sigma^2 \approx 3.27$ $s \approx 1.81$ (3 s.f.)
The number of degrees of freedom are found. The t-score can be found using the tables to find the value at the 90% confidence interval when $v = 7$. It is important to note that for the 90% confidence interval, below the upper bound is 95% of the score. The t_7 at the 95% percentage point can then be used as the t-score.	$v = n - 1 = 7$ 95th percentage point of t_7 is 1.90 to 3 s.f.
The confidence interval for μ can then be determined using the formula for t-statisics.	$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$ $5.875 - 1.90\frac{1.81}{\sqrt{8}} < \mu < 5.875 + 1.90\frac{1.81}{\sqrt{8}}$ $4.79 < \mu < 6.96$ (3 s.f.)

Example 2: The depth of the deep end of a swimming is measured every half hour. The results for the first two hours are displayed in the table below.

Time	9:00	9:30	10:00	10:30	11:00
Water Depth (cm)	136	140	142	139	x

A symmetric confidence interval for the mean change in water depth was found to include values from 0.019m to 0.042m. It can be assumed that the water depths follow a normal distribution.

a) Find the value of x.
b) Find the confidence level of this interval.

a) The two values given for the change in water depth can be used to determine the mean of the sample. The sum of the change in water depth for the first two hours can be assumed to be equal to the mean determined and therefore used to find x. It is important to consider the conversion of units using the information $100cm = 1m$.	$\bar{x} = \frac{0.019 + 0.042}{2} = 0.0305$ m $\bar{x} = \frac{(1.40 - 1.36) + (1.42 - 1.40) + (1.42 - 1.39) + (1.39 - x)}{4} = 3.05$cm $x = 135.8$ cm $x = 136$ cm (3 s.f.)
b) The process for the determining the confidence interval is the inverse of some of the previous calculations. You can determine the value of t in order to find the confidence level. This is found using the number of degrees of freedom and the unbiased estimate for the variance.	$v = n - 1 = 3$ $\sigma^2 = 0.3924$ cm $s^2 = \frac{4}{3} \times 0.003924 = 0.5232$ cm $t\frac{s}{\sqrt{n}} = \frac{4.2 - 1.9}{2} = 1.15$ Therefore, $t = 3.18$ (3 s.f.) Using the t-score table, the confidence level is approximately 95% .