Confidence intervals

So far, the means of the population have been known to us and the variance may have been known or has had to be estimated from the sample.

If the mean and variance are unknown or not given in a question and we wish to say something about the **population mean** based on a sample, we give a confidence interval for the population mean. Our degree of confidence in the interval is usually expressed in probability terms:

A 95% confidence interval is such that the probability that the interval (after it has been calculated) contains the population mean is 0.95.

A 99% confidence interval is such that the probability that the interval contains the population mean μ is 0.99. (we can only be 100% certain that the interval contains the population mean if we take a sample of size infinity – totally impractacle).

The **sample mean** has the distribution $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ **The variance has to be estimated from the sample if all we have is a sample.**

This sample **32, 44, 45, 21, 18, 56, 27, 31, 57, 28, 39, 49, 22, 15, 33** of size 15, has **mean** of **34.5** and a **variance** of **165 (using the divisor n).** The **estimate of the population variance** must have divisor n-1. This is easily obtained by multiplying the sample variance by n (to clear the denominator) and dividing by n-1. $165 \times \frac{15}{14}$ = **176.8 Denote this estimate by $\hat{\sigma}^2$ and the estimate of the population standard deviation by $\hat{\sigma}$ (sigma hat).** $\hat{\sigma} = \sqrt{176.8} = 13.3$ **The standard error** $\frac{\hat{\sigma}}{\sqrt{n}}$ **= 3.43.**

When the sample size n is large we use the Normal Distribution to get the confidence interval $\bar{x} \pm 1.96\frac{\hat{\sigma}}{\sqrt{n}}$ = 34.5 ± **(1.96 x 3.43).**

When the sample size is small as in this case (n = 15) we use the t – distribution in a similar way to the Normal Distribution but with the added extra of "degrees of freedom". The degrees of freedom will be **n-1** and in this case we look at t-tables along the row for **d.f. = 14.**

A 95% confidence interval for μ is now $\bar{x} \pm t_{95\%}(14)\frac{\hat{\sigma}}{\sqrt{n}}$ = 34.5 ± **2.145**x3.43 = **(27.1, 41.9).**

A 99% confidence interval for μ is $\bar{x} \pm t_{99\%}(14)\frac{\hat{\sigma}}{\sqrt{n}}$ = **34.5** ± 2.977**x3.43** = (24.3, 44.7).

Find these values in t-tables.
row 14
columns 0.975 and 0.995