# SG & SH - t-distribution & confidence intervals

|  |  |
| --- | --- |
| **What assumptions are required for the validity of a t-test?** | 1. The sample is random. 2. The sample is taken from a normally distributed population. |
| **How can you calculate the test statistic for a t-test?** | Where S2 is an unbiased estimator of σ2 which is found using… |
| **How are degrees of freedom calculated for t-tests?** | It has (n - 1) degrees of freedom where n is number of datapoints.  *The intuitive understanding is, if we have 6 numbers (of which we know 5) and know the mean then we can easily find the 6th. Ultimately, only 5 of these numbers contribute to the standard error.* |
| **When is a t-test suitable?** | 1. If the population variance is unknown, t-test is suitable (if it is then use normal distribution). 2. If the sample size is small (ie, n ≤ 30).   *Unless explicitly stated otherwise, you can always use a t-test. It’s just a lack of technology that made it the case of using z-distribution at this cut-off as convention early on.* |
| **What happens as the sample size of a t-test increase?** | The t-distribution approaches the standard normal distribution. |
| **What is a standard error?** | Which is… |
| **How can t-distributions be used to generate a p% confidence interval? And when?** | Where t is from a t-distribution of n-1 degrees of freedom:    When sample size is small and population variance is unknown.  Example: |
| **What is the correct interpretation of a p%-confidence interval? What is it generated from?** | * It is expected, ***before generation***, the population mean μ will fall into this interval with probability p%. * If you take repeated samples and form many confidence intervals, you expect p% of them to contain μ. * It’s generated from a sample. |
| **How is a p%-confidence interval generated by a sample size n?** | Where z is calculated from p. |