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?	$$\frac{(\bar{x}-\mu)}{\left(\frac{S}{\sqrt{n}}\right)}$$ Where S^2 is an unbiased estimator of σ^2 which is found using… $$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$
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, $n \leq 30$). *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?	Key point The standard deviation of the sample mean, $\frac{s}{\sqrt{n}}$, is called the **standard error**. Which is…

	$$\bar{X} \approx \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$
How can t-distributions be used to generate a p% confidence interval? And when?	$$\bar{x} - t \times \frac{s}{\sqrt{n}} < \mu < \bar{x} + t \times \frac{s}{\sqrt{n}}$$ Where t is from a t-distribution of n-1 degrees of freedom: $$t = t_{n-1}^{-1}\left(\frac{1+p}{2}\right),$$ When sample size is small and population variance is unknown. Example: A sample of size 16 is taken, whose mean is 13.6 and whose standard deviation is 20.4, in order to generate a 95% confidence interval for the mean of the population. Find the confidence interval. For $p = 95\%$ and with $16-1=15$ degrees of freedom, $t_{15}^{-1}(0.975) = 2.13$ The interval $13.6 - 2.13 \times \frac{20.4}{\sqrt{16}} < \mu < 13.6 + 2.13 \times \frac{20.4}{\sqrt{16}}$ simplifies to $2.74 < \mu < 24.46$
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?	$$\bar{x} - z \times \frac{s}{\sqrt{n}} < \mu < \bar{x} + z \times \frac{s}{\sqrt{n}}$$ Where z is calculated from p.