0.0.1 Small samples

- We indicated that use of the normal distribution in estimating a population mean is warranted for any large sample ($n > 30$).
- For a small sample ($n \leq 30$) only if the population is normally distributed **and** σ is known, the standard normal distribution can be used compute quantiles. In practice, this case is unusual.
- Now we consider the situation in which the sample is small and the population is normally distributed, but σ is not known.

0.0.2 Independent one-sample t-test

In testing the null hypothesis that the population mean is equal to a specified value μ_0, one uses the statistic

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \tag{1}$$

where s is the sample standard deviation and n is the sample size. The degrees of freedom used in this test is $n - 1$.

later, remember that a confidence interval is a **two-tailed** procedure, i.e. $k = 2$.

0.1 Student-t Distribution

0.1.1 Student's $t-$distribution (1)

(Revision from last class, but very important).

- We use the ***Student's*** $t-$***distribution*** for small samples.
- The Student $t-$distribution is the appropriate basis for determining the standardized test statistic when the sampling distribution of the mean is normally distributed but s is not known.
- The sampling distribution can be assumed to be normal either because the population is normal or because the sample is large enough to invoke the central limit theorem.
- **[IMPORTANT]** The t distribution is required when the sample is small ($n < 30$). For larger samples, where $n \geq 30$, normal approximation can be used.
- For the critical value approach, the procedure is identical to that described for the normal distribution, except for the use of t instead of z as the test statistic.

0.1.2 Using the Student $t-$ distribution in statistical software

- Student-t distribution is used for the data that follow the normal model with unknown standard deviation specially when sample sizes are small.
- For Students t distribution, statistical tables such (e.g. Murdoch Barnes and State Examinations Commision tables) only tabulate quantiles with degrees of freedom of less than 30. (*Some other tables go as far as 50*).

- This constraint has given rise to the convention that a sample of size greater than 30 is a large sample and in this case the standard normal distribution should be used.

$$n > 30 \rightarrow \text{Large Sample} \rightarrow \text{Use Z distribution}$$

$$n \leq 30 \rightarrow \text{Small Sample} \rightarrow \text{Use t-distribution}$$

- However there is a disparity between the Z value and the correct t value. For a sample size of 61 (i.e. degrees of freedom =60), the 97.5% t-quantiles of Student's t distribution is 2.003, and not 1.96.
- However, statistical software is free from this restraint. The correct distribution will be automatically used. The Students t distribution can be used in all appropriate cases. As the sample size increases the Student t distribution converges with the standard normal distribution
- This is worth remembering when doing analyses with statistical software.

.

0.2 Student-t Distribution

We use the student's $t-$ distribution when the population variance σ is not known and either one or both of these conditions are met.

- The Population is normally distributed.
- The sample size n is greater than 30.

N.B. The student t distribution is different for different sample sizes.

- The Student t distribution has approximately the same shape as the normal distribution, but has longer tails.
- That is to say, reflects the higher variability associated with small samples.
- As the sample size gets progressively larger, the student t distribution becomes more and more like the normal distribution.

0.2.1 degrees of freedom

- Degrees of freedom corresponds to the number of sample values that can vary after certain restrictions have been imposed on all data values.
- df $= n - 1$

Summary

IMPORTANT : Using Murdoch Barnes Table 7

Size	Sample Size (n)	degrees of freedom
Small	30 or less	n-1
Large	more than 30	∞

0.2.2 The Student $t-$distribution

- A similar distribution to standard normal is student-t. The student-t is defined in part by degree of freedom in the formula n-1, where n is sample size. This means that were variables from the distribution to be picked one by one, all but the last one could be chosen freely.
- There is no choice but to take the very last one and no freedom to choose any other variable at that point. Therefore one variable is not free; it's like having to pick the last tile out of a bag during a Scrabble game where there is no choice but to choose that letter.
- Different distributions like the F and the chi-square have different definitions of degree of freedom, and some even use more than one df in definition. The issue gets confusing because df definition is linked to type of test performed and isnt the same with the various parametric (based on parameters) and non-parametric (not based on parameters) tests.
- Essentially, it won't always be n-1. Goodness of fit or contingency table testing may use the chi-square distribution with different df than that which evaluates single variable hypothesis testing of the variance or standard deviation.

What is important to remember is that each time degree of freedom is used to define a distribution, it changes it. It still may have certain characteristics that are unchanging, but size and appearance vary. When people are drawing representations of distributions, particularly two of the same distributions that have a different df, theyre advised to make them look different in size to convey that df is not the same.