MATH 140
DR. MCLOUGHLIN'S CLASS
STATISTICAL FORMULAE FOR T-TEST
HANDOUT V (REVISED AND EXPANDED)

We are called upon many times to determine if the mean performance of two groups are significantly different. When attempting to determine if the difference between two means is greater than that expected from chance, the "t" test may be the needed statistical technique. If the data is from a normal population, we have a pseudo-random sample, and the data is at least interval[1] in nature, then we are surer that this is the technique to use.

"T" is the difference between two sample means measured in terms of the standard error of those means, or "T" is a comparison between two groups means which takes into account the differences in group variation and group size of the two groups. The statistical hypothesis for the "T" test is stated as the null hypothesis concerning differences.

Let D = $\{X_1, X_2, X_3, \ldots, X_n\}$ be a finite data set from a population of interest.

Let $X_1, X_2, X_3, \ldots, X_n$ be the finite random sample.

Recall these statistical formulae from previous handouts:

$$\overline{X} = \frac{\sum_{k=1}^{n} X_k}{n} \qquad S^2 = \frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1} \qquad s = \sqrt{= \frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{(n-1)}}$$

$$Z_i = \frac{X_i - \mu_X}{\sigma_X} \qquad Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \qquad T = \frac{X - \overline{X}}{S_X} \qquad T_{n-1} = \frac{\overline{X} - \mu_{\overline{X}}}{S_{\overline{X}}}$$

[1] Ordinal data is used but technically should not be used.

Independent t-test Pooled Variance Formula

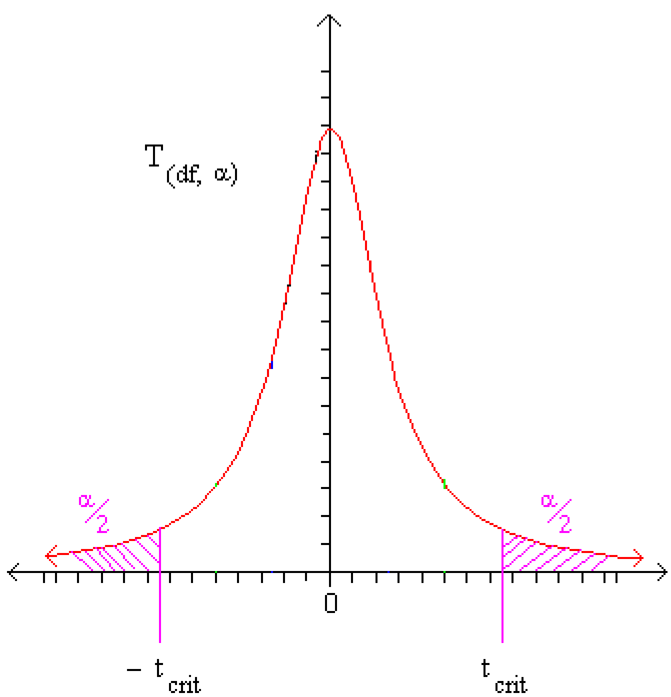
We use the pooled variance formula if:

If $n_1 = n_2$ and $s_1^2 = s_2^2$	$df = n_1 + n_2 - 2$
If $n_1 \neq n_2$ and $s_1^2 = s_2^2$	$df = n_1 + n_2 - 2$
If $n_1 = n_2$ and $s_1^2 \neq s_2^2$	$df = n_1 - 1$ or $n_2 - 1$

$$t_{pool} = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\left(\frac{\sum_{k=1}^{n}(X_k - \overline{X})^2 + \sum_{k=1}^{n}(Y_k - \overline{Y})^2}{n_1 + n_2 - 2}\right) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

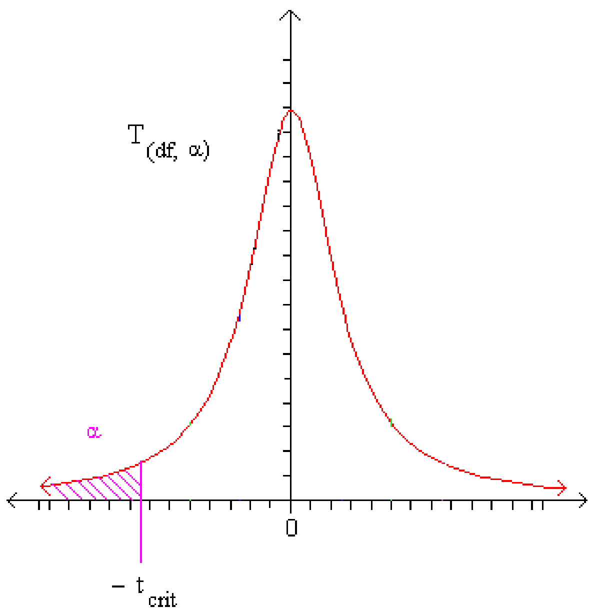
2-tailed $H_0 : \mu_1 = \mu_2$ also stated as $H_0 : \mu_1 - \mu_2 = 0$

$H_A : \mu_1 \neq \mu_2$ $H_A : \mu_1 - \mu_2 \neq 0$



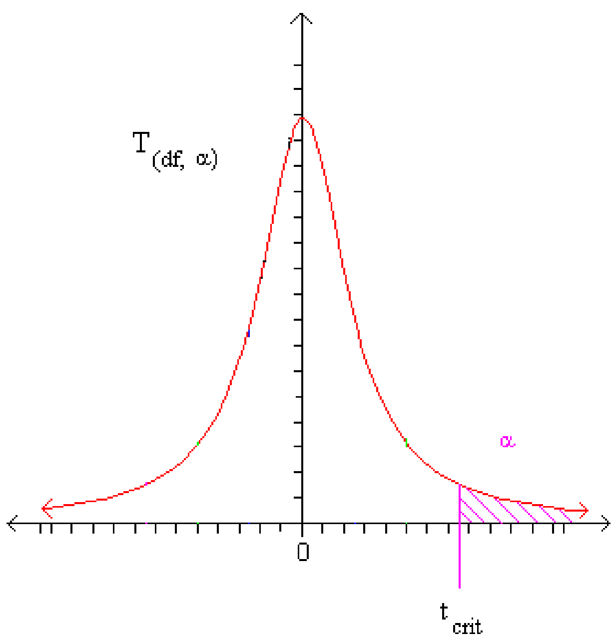
1-tailed low $H_0: \mu_1 \geq \mu_2$ also stated as $H_0: \mu_1 - \mu_2 \geq 0$

$H_A: \mu_1 < \mu_2$ $H_A: \mu_1 - \mu_2 < 0$



1-tailed high $H_0: \mu_1 \leq \mu_2$ also stated as $H_0: \mu_1 - \mu_2 \leq 0$

$H_A: \mu_1 > \mu_2$ $H_A: \mu_1 - \mu_2 > 0$



Correlated Data Formula (Paired Sample t-test)

First compute r_{xy}

$$r_{xy} = \frac{\sum_{k=1}^{n}(X_k - \overline{X})(Y_k - \overline{Y})}{\sqrt{\sum_{k=1}^{n}(X_k - \overline{X})^2 \sum_{k=1}^{n}(Y_k - \overline{Y})^2}} = \frac{\sum_{k=1}^{n}(X_k - \overline{X})(Y_k - \overline{Y})}{n\sqrt{(S_X)(S_Y)}} = \frac{\sum_{k=1}^{n}(z_{x_k})(z_{y_k})}{n}$$

Since f the samples are related (two measures from the same subject or matched pairs), the correlated data formula is used.

$$t_{paired} = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\left(\frac{S_X^2}{n}\right) + \left(\frac{S_Y^2}{n}\right) - 2r_{XY} \cdot (\frac{S_X \cdot S_Y}{n})}}$$

Recall the assumptions are:

1. Representative sample (Random)

2. Normal distribution for population

3. At least ordinal measures