Two Samples Tests

Test for the Difference of Means

1. NULL HYPOTHESIS
   Is the value for the difference that we want to test (δ_0 = difference value to test)

$$H_0 : \mu_1 - \mu_2 = \delta_0$$

2. ALTERNATIVE HYPOTHESIS
   Corresponds to what would be true if the null hypothesis is false. Depends on what information we have regarding the population proportion. There are 4 cases

	INFORMATION REGARDING μ_1 AND μ_2	TEST TYPE
$H_1 : \mu_1 - \mu_2 \neq \delta_0$	General case. We have no information for the population means. Hence, if the difference is not equal to δ_0 the only thing we can say is that it is different	Two Tails Test
$H_1 : \mu_1 - \mu_2 > \delta_0$	We have some information about the means indicating that if the difference is not equal to δ_0 then it must be greater	Right-Tail Test
$H_1 : \mu_1 - \mu_2 < \delta_0$	We have some information about the means indicating that if the difference is not equal to δ_0 then it must be smaller	Left-Tail Test
$H_1 : \mu_1 - \mu_2 = \delta_1$	We have some information about the means indicating that if the difference is not equal to δ_0 then it must be equal to a known alternative value δ_1	Right-Tail Test if $\delta_1 > \delta_0$ and Left-Tail Test if $\delta_1 < \delta_0$

3. TEST STATISTIC
   The Test Statistic (TS) to use in this case depends on whether the population variances σ_1^2 i σ_2^2 are both known or not.

σ_1^2 and σ_2^2 KNOWN	σ_1^2 or σ_2^2 UNKNOWN
TS=$\frac{(\bar{X}_1-\bar{X}_2)-\delta_o}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	TS=$\frac{(\bar{X}_1-\bar{X}_2)-\delta_o}{\sqrt{\frac{S^2}{n_1}+\frac{S^2}{n_2}}}$

   In any of these cases, the Ovserved Value of the Test Statistic (OVTS) is obtained by replacing the correspondig values in the formula, where

$\bar{X}_1$ i $\bar{X}_2$	Sample Means
δ_0	Null Hypothesis Value
σ_1^2 i σ_2^2	Population variances (if known)
S^2	Common Sample Variance (if σ_1^2 or σ_2^2 are not known
n_1 i n$_2$	Sample sizes

   In the formulae above the common value for the Sample Variance, S^2(that we use if we do NOT know either σ_1^2 ór σ_2^2) (or any of them) is computed as

$$S^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

   where S_1^2 i S_2^2 are the Sample Variances of the first and the second sample respectively. The rason for using this *common estimation of the sample variance* is that for the test to make sense the two populations must be somehow "homogeneous". Tecnichally, this is equivalent to requiring that the two populations have a similar population variance.

4. DISTRIBUTION OF THE TEST STATISTIC WHEN THE NULL HYPOTHESIS IS TRUE
   If it is true that $\mu_1 - \mu_2 = \delta_0$ then

σ_1^2 and σ_2^2 KNOWN	σ_1^2 or σ_2^2 UNKNOWN
$\frac{(\bar{X}_1-\bar{X}_2)-\delta_o}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$\frac{(\bar{X}_1-\bar{X}_2)-\delta_o}{\sqrt{\frac{S^2}{n_1}+\frac{S^2}{n_2}}} \sim t_{n_1+n_2-2}$

5. REJECTION AREA of size α
   The Rejection Area depends on whether we have a Two Tails Test, a Right-Tail Test, or a Left-Tail Test. This, in turn, depends on what is the specification of the Alternative Hypothesis.

   (a) TWO TAILS TEST. Corresponds to the case when the Alternative Hypothesis is of the type $H_1 : \mu_1 - \mu_2 \neq \delta_0$

The limit values of the Rejection Area, $Z_{1-\frac{\alpha}{2}}$ and $t_{1-\frac{\alpha}{2}}$ can be found in the tables of a $N(0,1)$ or a $t-student$ with n_1+n_2-2 degrees of freedom respectively, depending on whether we know the two population variances or not as explained above (See Figure 3.10)

(b) RIGHT-TAIL TEST. Corresponds to the case when the Alternative Hypothesis is of the type $H_1: \mu_1-\mu_2 > \delta_0$ (or $H_1: \mu_1-\mu_2 = \delta_1$ and $\delta_1 > \delta_0$)

The limit value of the Rejection Area, $Z_{1-\alpha}$ or $t_{1-\alpha}$ can be found in the tables of the $N(0,1)$ or the $t-student$ with n_1+n_2-2 degrees of freedom respectively depending on whether we know the two population variances or not as explained before. (See Figure 3.11)

(c) LEFT-TAIL TEST. Corresponds to the case when the Alternative Hypothesis is of the type $H_1: \mu_1-\mu_2 < \delta_0$ (or $H_1: \mu_1-\mu_2 = \delta_1$ and $\delta_1 < \delta_0$)

The limit value of the Rejection Area, $Z_{1-\alpha}$ or $t_{1-\alpha}$ can be found in the tables of the $N(0,1)$ or $t-student$ with n_1-n_2-2 degrees of freedom

respectively depending on whether we know the two population variances or not. (See Figure 3.12)

6. TEST CONCLUSION
Finally, we have to check if the OBSEVED VALUE OF THE TEST STATISTIC (OVTS) falls, or not, inside the REJECTION AREA. If it does, we then say that the test rejects the NULL HYPOTHESIS. If it does not belong to the rejection area, then we say that the test DOES NOT REJECT THE NULL HYPOTHESIS.

Test for difference between two population variances

0. EXTRA STEP

We change the "denomination" of our two samples so that **ALWAYS** the sample with the highest Sample Variance is the Sample 1, being the Sample 2 the one with the lowest variance. This way, once we have followed this rule, we will always have:

$$S_1^2 > S_2^2$$

1. NULL HYOPTHESIS
   Is always the same and, as said before, it consists of testing whether the two variences are the same or not. Beacuse of the special structure of this test, the correct way to specify this hypothesis is:

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

2. ALTERNATIVE HYPOTHESIS
   As usual, it represents what is true when the Null Hypothesis is false. I this specific case, there are only two possible specifications for this hypothesis (once more, this is so because of the special structure of this test)

	INFORMATION REGARDING σ_1^2 AND σ_2^2	TEST TYPE
$H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$	General Case. We have no information on σ_1^2 nor about σ_2^2. Hence, if they are not equal, the only thing we can say is that they are different	Two Tails Test
$H_1 : \frac{\sigma_1^2}{\sigma_2^2} > 1$	We have some information about σ_1^2 and σ_2^2 indicating that if they are not equal then one of them is greater. Because of the "denomination" in Step 0, the greater will always be σ_1^2	Right-Tail Test

3. TEST STATISTIC

   In this case, the Test Statistic (TC) to use is:

$$textrmTE = \frac{S_1^2}{S_2^2}$$

The observed value of the Test Statistic (OVTS) is easily obtained replacing the corresponding sample variances in the formula, where

S_1^2	Sample Variance of Sample 1
S_2^2	Sample Variance of Sample 2

Notice that, because of Step 0 whe have that $S_1^2 > S_2^2$, and hence we will always find that VOEC> 1

4. DISTRIBUTION OF THE TEST STATISTIC when the Null Hypothesis is true
   In this case, the Test Statistic follows a distribution that is known as a F of Snedecor. This distribution, the same as the $t - student$ or the χ^2 is also characterized by its "degrees of freedom". Unlike those cases, though, the $F - snedecor$ has a "pair" of degrees of freedom, those corresponding to the numerator and those corresponding to the denominator. Hence, the notation:

$$\frac{S_1^2}{S_2^2} \sim F_{(n_1-1,n_2-1)}$$

   indicates that the Test Statistic $\frac{S_1^2}{S_2^2}$ follows a $F - snedecor$ distribution with $n_1 - 1$ degrees of freedom in the numerator (that is, the size of the sample that corresponds to S_1^2 in the numerator minus 1) and $n_2 - 1$ degrees of freedom in the denominator (that is, the size of the sample that corresponds to S_2^2 in the denominator minus 1).
   Remember that it is very important to keep the order stablished in Step 0, that is, sample 1 corresponds to to the sample that has the highest sample variance. In this sense, the "degrees of freedom in the numerator" is the size of such sample minus 1: $n_1 - 1$. This is important whem looking at the tables of the $F - snedecor$ in order to determine the Rejection Area.

5. REJECTION AREA of size α
   The Rejection Area depends on whether the test has one or two tails, as given by the Alternative Hypothesis. In this special test, the tail that "matters" will always be the Right-Tail, even if the test is a "Two Tails Test".

   (a) TWO TAILS TEST. Corresponds to thec case when we have an Alternative Hypothesis of the type $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$

   For the limit values of the Rejection Area, $F_{1-\frac{\alpha}{2}}$ and $F_{\frac{\alpha}{2}}$, we only need to find $F_{1-\frac{\alpha}{2}}$ in the tables of the F with $n_1 - 1$ degrees of freedom in the numerator and $n_2 - 1$ degrees of freedom in the denominator. The other value, $F_{\frac{\alpha}{2}}$, is not needed in any case since the OVTS is always > 1, Hence, if it falls into the Rejection Area, it will be on the Right-Tail. BECAUSE OF WHAT IS DONE IN STEP 0 (THE "DENOMINATION"OF THE SAMPLES), THE OBSERVED VALUE OF THE TEST STATISTIC <u>WILL NEVER</u> BE IN THE LEFT-TAIL. (See Figure 3.13)

   (b) RIGHT-TAIL TEST. Corresponds to the case when the Alternative Hypothesis is of the type $H_1 : \frac{\sigma_1^2}{\sigma_2^2} > 1$

   The limit value of the Rejection Area, $F_{1-\alpha}$, can be foundin the tables of a F with $n_1 - 1$ degrees of freedom in the numerator and $n_2 - 1$ degrees of freedom in the denominator. (See Figure 3.14)

6. TEST CONCLUSION
   Finally, we have to check if the OBSEVED VALUE OF THE TEST STATISTIC (OVTS) falls, or not, inside the REJECTION AREA. If it does, we then say that the test rejects the NULL HYPOTHESIS. If it does not belong to the rejection area, then we say that the test DOES NOT REJECT THE NULL HYPOTHESIS.

Exercises

1. (HT for Difference Between Two Population Means)

An airline wants to test the claim that the mean lifetimes of two types of aircraft radios are identical. It installs 800 radios of each type in its current fleet and later selects two simple random samples of radios of each type. After taking two samples of n = 36 each, the statistician finds mean lifetimes a sample standard deviations of $\overline{X_A} = 4120$ *hours*, $S_A = 80$ *hours and* $\overline{X_B} = 4910$ *hours*, $S_B = 120$ hours. Assume that the significance level is 10%.

2. An orchardist wants to compare the mean yield of fruit trees sprayed with gypsy-moth parasites (A) with that of fruit trees sprayed with traditional pesticides (B). Some 250 trees are sampled in each of two large orchards that were given one treatment or the other. The orchardist selects a significance level of 0.005. After taking the samples, the orchardist finds mean fruit yields and sample standard deviations of $\overline{X_A} = 10.5$ bushels, $S_A = 8$ bushels and $\overline{X_B} = 8.9$ bushels, $S_B = 3$ bushels.

3. The American Dental Association wants to determine whether there is a difference in the cavity-fighting ability of two toothpastes, A and B. A simple random sample of 21 users of each type is taken, and the mean number of cavities over a decade is counted. The investigator selects a significance level of 1%. After taking the samples, the investigator finds sample mean numbers of cavities and associated sample standard deviations of $\overline{X_A} = 27$ and $\overline{X_B} = 23$, with $S_A = 6$ and $S_B = 2$.

An investor wants to compare the risks associated to two different stock markets, A and B. Market risk is measured using the variance of the daily changes in stock prices. The investor believes that the risk in market A is lower than the risk in market A. Two random samples are selected, consisting of 21 observations on the changes i prices in market A and 16 observations on the changes of prices in market B. The results are:

Market A	Market B
$\bar{X}_A = 0.3$	$\bar{X}_B = 0.4$
$s_A = 0.25$	$s_B = 0.45$

Assuming that both samples come from two Normal and independent populations, does the data suport the investor's belief ? ($\alpha = 0.5$)