Analysis of Variance

(ANOVA)

What is **ANOVA**?

It is a method used whenever we wish to test for the equality of at least three population means simultaneously.

How do we set up H_0 and H_1 for this method?

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$
$$H_1 : \text{ At least one population mean is different.}$$

This method uses F-distribution and it is always a Right-Tail Test.

What does an **ANOVA Table** consist of?

Here is a typical ANOVA table:

Source	DF	SS	MS	F	P
Factor					
Error					
Total					

Where

- **Factor** also referred to as between groups or samples.
- **Error** also referred to as within groups or samples.
- **DF** stands for degrees of freedom.
- **SS** stands for the sum of squares.
- **MS** stands for the mean sum of squares.
- **F** is the value of the computed test statistic.
- **P** is the corresponding p-value for the computed test statistic.

How do we find **DF** for **Factor** and **Error**?

$$\text{DF(Factor)} = k - 1$$
$$\text{DF(Error)} = n - k$$
$$\text{DF(Total)} = n - 1$$

Where

- k is the number of groups or samples.
- n is the total sample size of all groups or samples.

How do we compute **SS** for **Factor**, **Error** and **Total**?

$$SS(\text{Factor}) = \sum n_i(\bar{x}_i - \bar{x})^2$$

$$SS(\text{Error}) = \sum (n_i - 1)s_i^2$$

$$SS(\text{Total}) = \sum (x_{ij} - \bar{x})^2$$

$$SS(\text{Total}) = SS(\text{Factor}) + SS(\text{Error})$$

Where

- $\bar{x}$ is the grand mean, the mean of all sample values combined.
- n_i is the number of values in the ith sample.
- x_{ij} is the ith value of the jth sample.
- $\bar{x}_i$ is the mean of values in the ith sample.
- s_i^2 is the variance of values in the ith sample.

How do we compute **MS** for **Factor** and **Error**?

$$\text{MS(Factor)} = \frac{\textbf{SS(Factor)}}{\textbf{DF(Factor)}}$$

$$\text{MS(Error)} = \frac{\textbf{SS(Error)}}{\textbf{DF(Error)}}$$

How do we compute **F** ?

$$F = \frac{\textbf{Variance Between Samples}}{\textbf{Variance Within Samples}} = \frac{\dfrac{\sum n_i(\bar{x}_i - \bar{x})^2}{k-1}}{\dfrac{\sum (n_i - 1)s_i^2}{n-k}} = \frac{\textbf{MS(Factor)}}{\textbf{MS(Error)}}$$

Example:

In an assembly plant it is suspected that some models are assembled with greater care than others.
To investigate whether there is any basis for this feeling, the number of defects for several of the three models were recorded and are displayed in the table below.

Model 1:	5	7	6	6		
Model 2:	5	4	3	5	3	4
Model 3:	7	6	8	9	5	

Construct the ANOVA table using the data provided in this table.

Solution:

We begin by finding the following

- $k = 3$ & $DF(Factor) = 2$
- $n = 15$ & $DF(Error) = 12$
- $\bar{x}_1 = 6$ & $s_1^2 = 0.\bar{6} = \dfrac{2}{3}$
- $\bar{x}_2 = 4$ & $s_2^2 = 0.8 = \dfrac{4}{5}$
- $\bar{x}_3 = 7$ & $s_3^2 = 2.5 = \dfrac{5}{2}$
- $\bar{x} = 5.5\bar{3} = \frac{83}{15}$

Now we compute **SS(Factor)** and **SS(Error)**.

Solution Continued:

$$
\begin{aligned}
SS(Factor) &= \sum n_i(\bar{x}_i - \bar{x})^2 \\
&= 4\left(6 - \frac{83}{15}\right)^2 + 6\left(4 - \frac{83}{15}\right)^2 + 5\left(7 - \frac{83}{15}\right)^2 \\
&= 25.7\bar{3} = \frac{386}{15}
\end{aligned}
$$

$$
\begin{aligned}
SS(Error) &= \sum (n_i - 1)s_i^2 \\
&= (4-1)\cdot\frac{2}{3} + (6-1)\cdot\frac{4}{5} + (5-1)\cdot\frac{5}{2} \\
&= 16
\end{aligned}
$$

---

it is recommended to readers to use the formula SS(Total)= $\sum (x_{ij} - \bar{x})^2$ to find SS(Total) and then verify that SS(Total)=SS(Factor) + SS(Error).

Solution Continued:

Now we compute **MS(Factor)** and **MS(Error)**.

$$\begin{aligned} MS(Factor) &= \frac{SS(Factor)}{DF(Factor)} \\ &= \frac{25.7\overline{3}}{2} \\ &\approx 12.867 \end{aligned}$$

$$\begin{aligned} MS(Error) &= \frac{SS(Error)}{DF(Error)} \\ &= \frac{16}{12} \\ &\approx 1.333 \end{aligned}$$

Solution Continued:

We are finally ready to compute **Computed Test Statistic F**.

$$F = \frac{MS(Factor)}{MS(Error)} = \frac{12.867}{1.333} = 9.65$$

We can use technology to find the corresponding $p-$value which is 0.003 in this case. We now have all the information to make the **ANOVA** table.

Source	DF	SS	MS	F	P
Factor	2	25.733	12.867	9.65	0.003
Error	12	16	1.333		
Total	14	41.733			

Example:

Use the ANOVA table we constructed in the last example to test the claim at the 0.05 level of significance that the average number of defects is not the same for the three models.

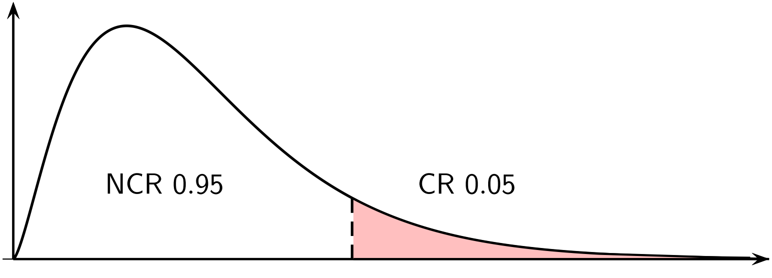
Solution:

We begin by setting up H_0 and H_1.

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : At least one mean is different. Claim & RTT

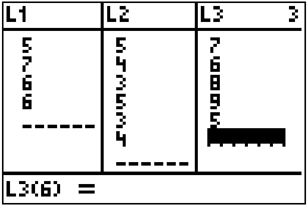
Using technology or F-Distribution table, we find the F critical value to be 3.885.

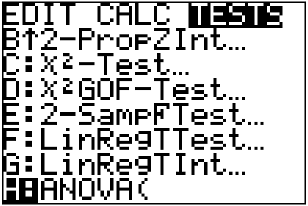


- **CTS** falls in the critical region, therefore H_1 is valid, we support the claim.
- p**−value** is less than the significance level, therefore H_1 is valid, we support the claim.

ANOVA & **TI**

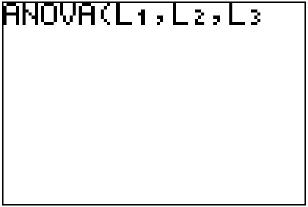
Use L_1, L_2, and L_3 to store the data that belong to each of the 3 samples. Now press **STAT**, go to **TESTS**, uparrow once to find **ANOVA**.

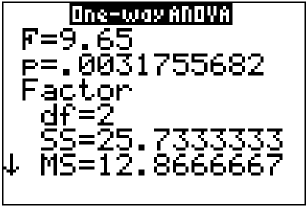




ANOVA & TI

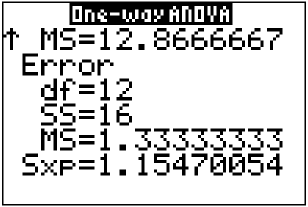
Now press **ENTER** to select **ANOVA**, then enter L_1, L_2, and L_3 followed by **ENTER** to execute **ANOVA**.





ANOVA & TI

Use the down arrow several times to display more results.



Example:

In a biological experiment 4 mixtures of a certain chemical are used to enhance the growth of a certain type of plant over a specific time period. The following growth data, in centimeters, are recorded for the plants that survive.

Mixture 1:	8.2	8.7	9.4	9.2		
Mixture 2:	7.7	8.4	8.6	8.1	8.0	
Mixture 3:	6.9	5.8	7.2	6.8	7.4	6.1
Mixture 4:	6.8	7.3	6.3	6.9	7.1	

Use **TI** to find all information needed to construct the **ANOVA** table, and then test the claim that there is no significant difference in the average growth of these plants for the different mixtures of the chemical. Use a 0.01 level of significance.

Solution:

We begin by setting up H_0 and H_1.

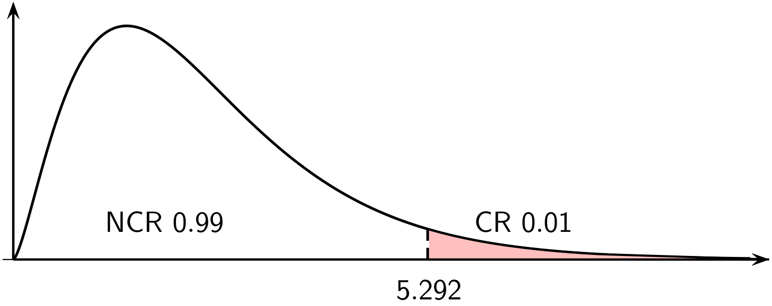
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ Claim

$H_1 :$ At least one mean is different. RTT

Using **TI**, we can find all information needed to construct the **ANOVA** table.

Source	DF	SS	MS	F	P
Factor	3	15.462	5.154	21.213	8.03×10^{-6}
Error	16	3.8875	0.242		
Total	19	19.3495			

Using technology or F-Distribution table, we find the F critical value to be 5.292.



- **CTS** falls in the critical region, therefore H_0 is invalid, we reject the claim.
- $p-$**value** is less than the significance level, therefore H_0 is invalid, we reject the claim.