

ANOVA Test

Tushar B. Kute,
<http://tusharkute.com>



Introduction

- Buying a new product or testing a new technique but not sure how it stacks up against the alternatives? It's an all too familiar situation for most of us. Most of the options sound similar to each other so picking the best out of the lot is a challenge.
- Consider a scenario where we have three medical treatments to apply on patients with similar diseases.
- Once we have the test results, one approach is to assume that the treatment which took the least time to cure the patients is the best among them.
- What if some of these patients had already been partially cured, or if any other medication was already working on them?

Introduction

- In order to make a confident and reliable decision, we will need evidence to support our approach.
- This is where the concept of ANOVA comes into play.



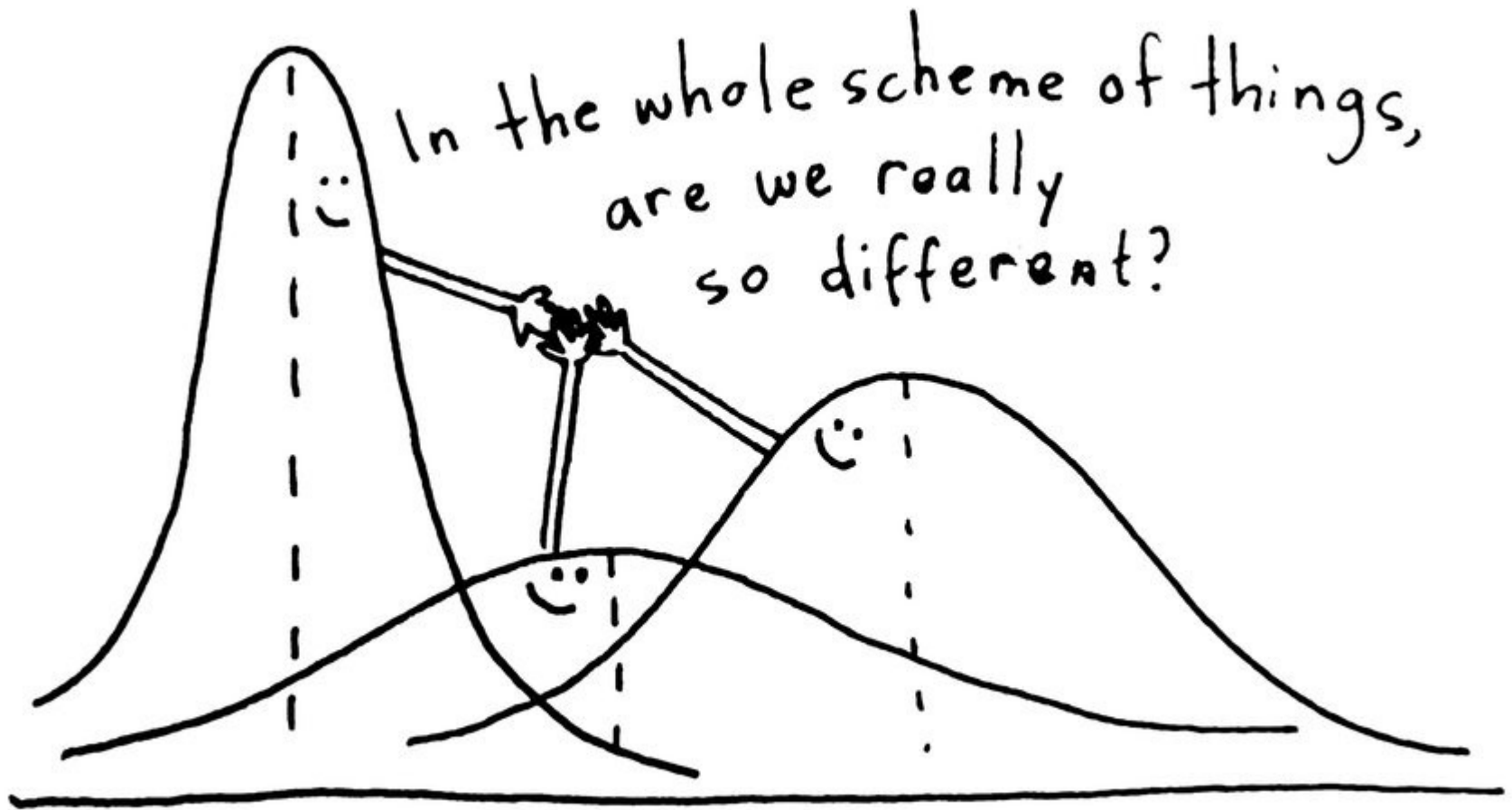
ANOVA

- A common approach to figure out a reliable treatment method would be to analyse the days it took the patients to be cured.
- We can use a statistical technique which can compare these three treatment samples and depict how different these samples are from one another.
- Such a technique, which compares the samples on the basis of their means, is called ANOVA.

ANOVA

- Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other.
- ANOVA checks the impact of one or more factors by comparing the means of different samples.
- We can use ANOVA to prove/disprove if all the medication treatments were equally effective or not.

ANOVA



ANOVA

- Another measure to compare the samples is called a t-test. When we have only two samples, t-test and ANOVA give the same results.
- However, using a t-test would not be reliable in cases where there are more than 2 samples.
- If we conduct multiple t-tests for comparing more than two samples, it will have a compounded effect on the error rate of the result.

Grand Mean

- Mean is a simple or arithmetic average of a range of values. There are two kinds of means that we use in ANOVA calculations, which are separate sample means

$$(\mu_1, \mu_2, \mu_3)$$

- and the grand mean

$$(\mu)$$

- The grand mean is the mean of sample means or the mean of all observations combined, irrespective of the sample.

Hypothesis

- Just like any other kind of hypothesis that you might have studied in statistics, ANOVA also uses a Null hypothesis and an Alternate hypothesis.
- The Null hypothesis in ANOVA is valid when all the sample means are equal, or they don't have any significant difference. Thus, they can be considered as a part of a larger set of the population.
- On the other hand, the alternate hypothesis is valid when at least one of the sample means is different from the rest of the sample means. In mathematical form, they can be represented as:

$$H_0 : \mu_1 = \mu_2 = \mu_L \quad \text{Null Hypothesis}$$

$$H_1 : \mu_1 \neq \mu_m \quad \text{Alternate Hypothesis}$$

Hypothesis

- μ_L and μ_m belong to any two sample means out of all the samples considered for the test.
- the null hypothesis states that all the sample means are equal or the factor did not have any significant effect on the results.
- Whereas, the alternate hypothesis states that at least one of the sample means is different from another. But we still can't tell which one specifically.

Between Group Variability

- Consider the distributions of the below two samples. As these samples overlap, their individual means won't differ by a great margin.
- Hence the difference between their individual means and grand mean won't be significant enough.



Between Group Variability

- Now consider these two sample distributions. As the samples differ from each other by a big margin, their individual means would also differ.
- The difference between the individual means and grand mean would therefore also be significant.

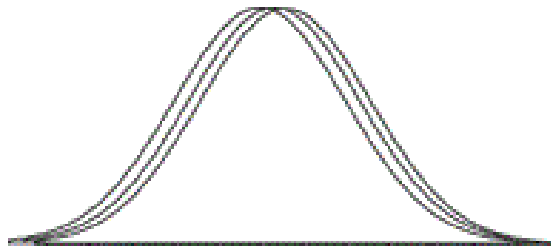


Between Group Variability

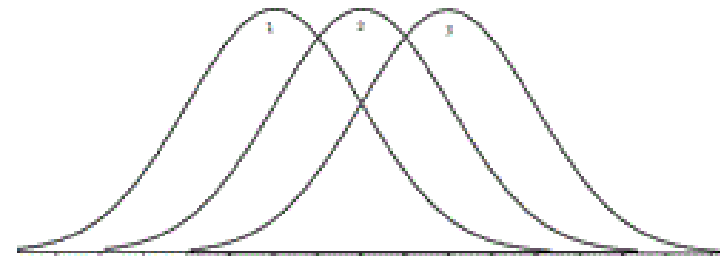
- Each sample is looked at and the difference between its mean and grand mean is calculated to calculate the variability.
- If the distributions overlap or are close, the grand mean will be similar to the individual means whereas if the distributions are far apart, difference between means and grand mean would be large.

Between Group Variability

Little discrimination



Some Discrimination



Discrimination between Two Groups,
but not the third

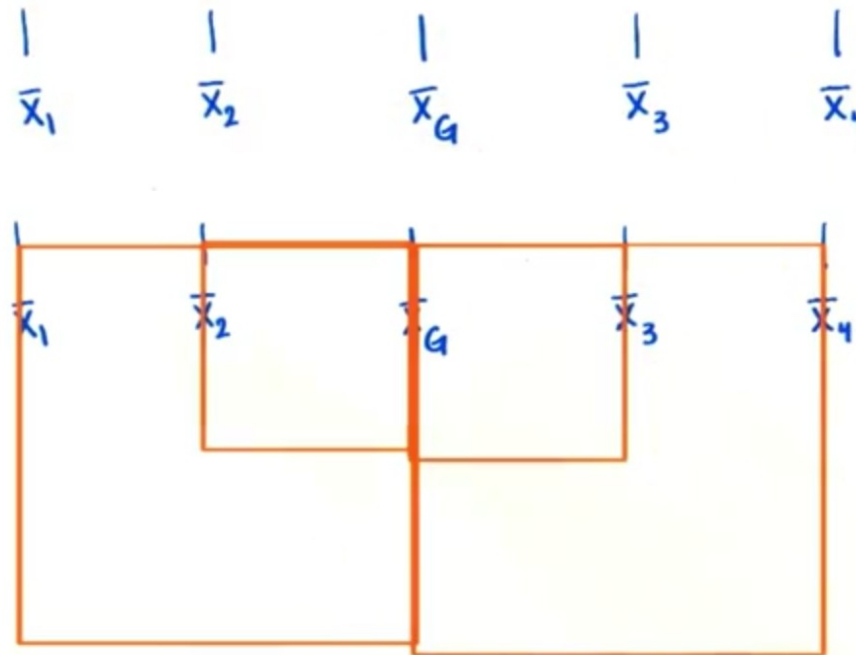


Large Discrimination



Between Group Variability

- We will calculate Between Group Variability just as we calculate the standard deviation. Given the sample means and Grand mean, we can calculate it as:



Between Group Variability

- We also want to weigh each squared deviation by the size of the sample.
- In other words, a deviation is given greater weight if it's from a larger sample.
- Hence, we'll multiply each squared deviation by each sample size and add them up. This is called the sum-of-squares for between-group variability

$$SS_{\text{between}} = n_1(\bar{x}_1 - \bar{x}_G)^2 + n_2(\bar{x}_2 - \bar{x}_G)^2 + n_3(\bar{x}_3 - \bar{x}_G)^2 + \dots + n_k(\bar{x}_k - \bar{x}_G)^2$$

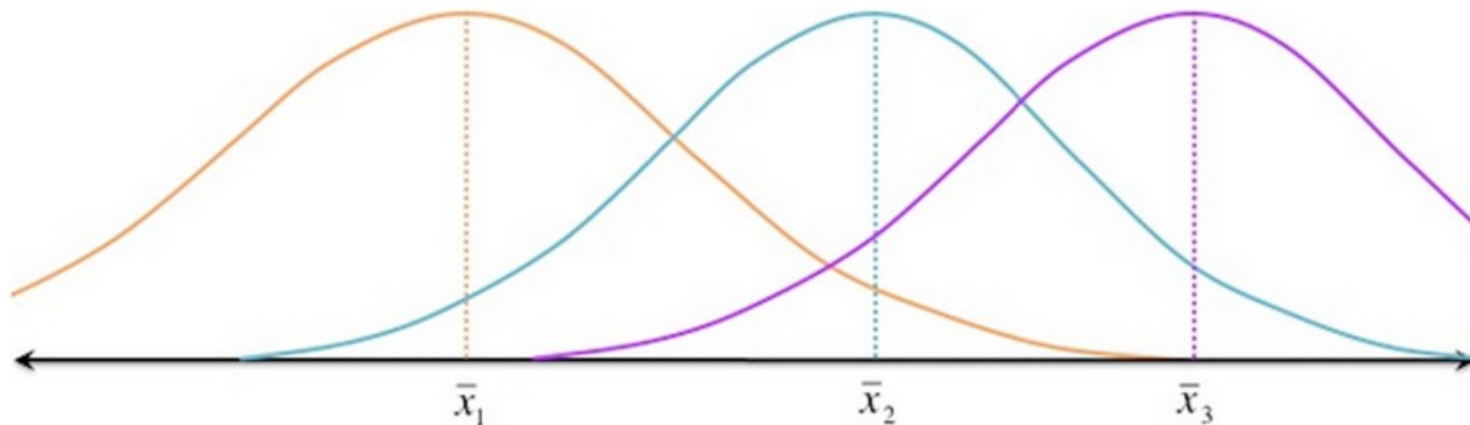
Between Group Variability

- There's one more thing we have to do to derive a good measure of between-group variability. Again, recall how we calculate the sample standard deviation.
- We find the sum of each squared deviation and divide it by the degrees of freedom.
- For our between-group variability, we will find each squared deviation, weigh them by their sample size, sum them up, and divide by the degrees of freedom (), which in the case of between-group variability is the number of sample means (k) minus 1.

$$MS_{\text{between}} = \frac{n_1(\bar{x}_1 - \bar{x}_G)^2 + n_2(\bar{x}_2 - \bar{x}_G)^2 + n_3(\bar{x}_3 - \bar{x}_G)^2 + \dots + n_k(\bar{x}_k - \bar{x}_G)^2}{k-1}$$

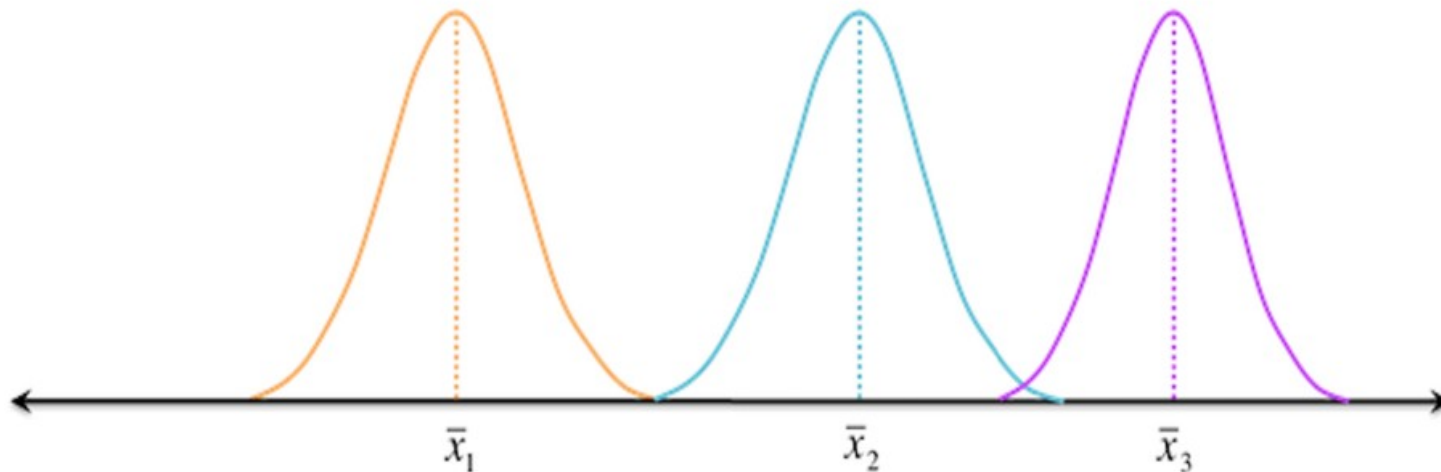
Within Group Variability

- Consider the given distributions of three samples. As the spread (variability) of each sample is increased, their distributions overlap and they become part of a big population.



Within Group Variability

- Now consider another distribution of the same three samples but with less variability.
- Although the means of samples are similar to the samples in the above image, they seem to belong to different populations.



Within Group Variability

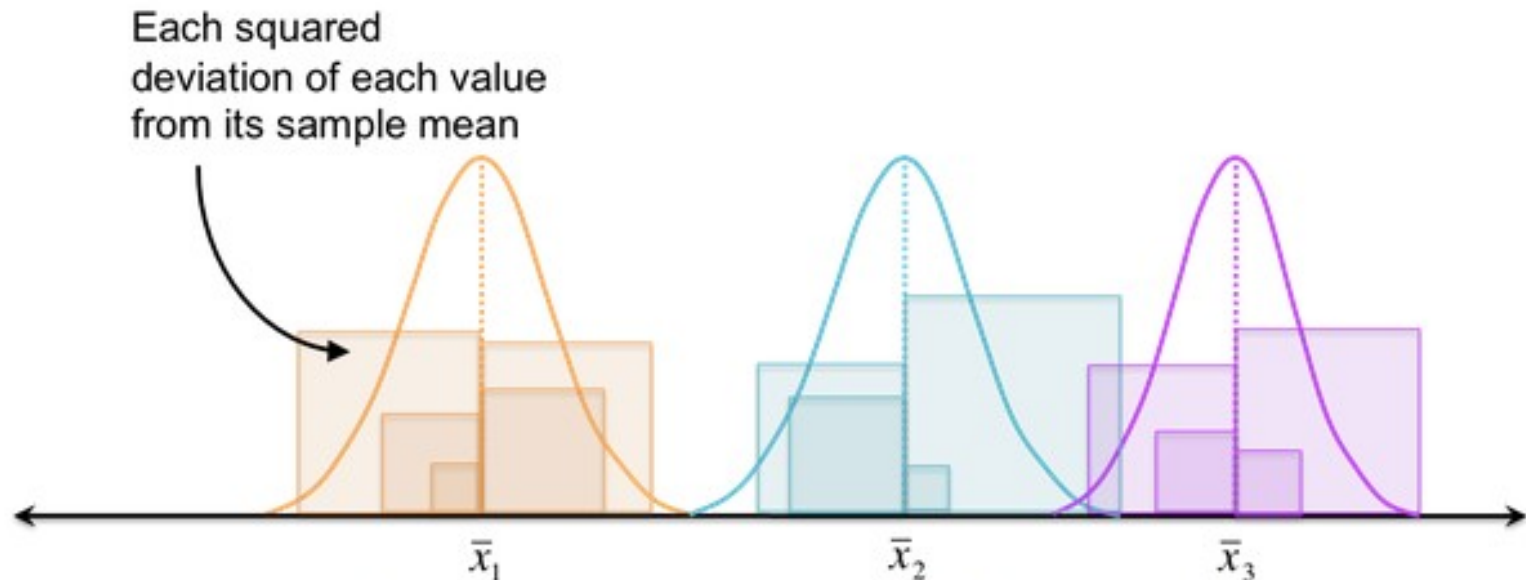
- Such variations within a sample are denoted by Within-group variation.
- It refers to variations caused by differences within individual groups (or levels) as not all the values within each group are the same.
Each sample is looked at on its own and variability between the individual points in the sample is calculated. In other words, no interactions between samples are considered.

Within Group Variability

- We can measure Within-group variability by looking at how much each value in each sample differs from its respective sample mean.
- So first, we'll take the squared deviation of each value from its respective sample mean and add them up.
- This is the sum of squares for within-group variability.

Within Group Variability

Note: x_{i1} is the i th value from the first sample, x_{i2} is the i th value from the second sample, and so on all the way to x_{ik} , the i th value from the k th sample. x_{ij} is therefore the i th value from the j th sample.



With within-group variability, SS_{within} is the sum of each squared deviation of each value from its respective sample mean (the total area of all the squares in the figure above). MS_{within} is the average-sized square.

Degree of freedom

- the degrees of freedom is the sum of the sample sizes (N) minus the number of samples (k).
- Another way to look at degrees of freedom is that we have the total number of values (N), and subtract 1 for each sample:

$$df_{\text{within}} = (n_1 - 1) + (n_2 - 1) + \dots + (n_k - 1) = n_1 + n_2 + n_3 + \dots + n_k - k(1) = N - k$$

$$MS_{\text{within}} = \sum (x_{ij} - \bar{x}_j)^2 / (N - k)$$

F-Statistic

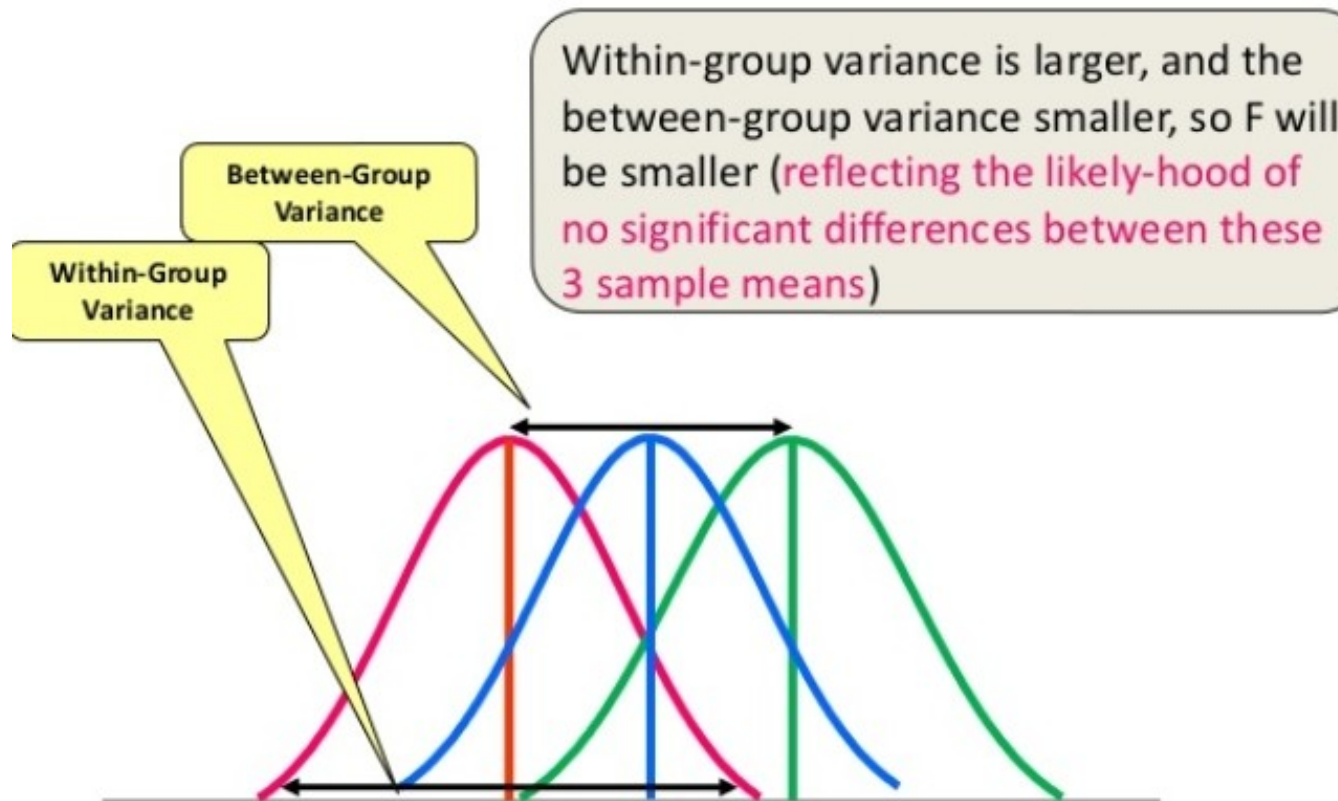
- The statistic which measures if the means of different samples are significantly different or not is called the F-Ratio.
- Lower the F-Ratio, more similar are the sample means. In that case, we cannot reject the null hypothesis.

$$F = \frac{\text{Between group variability}}{\text{Within group variability}}$$

F-Statistic

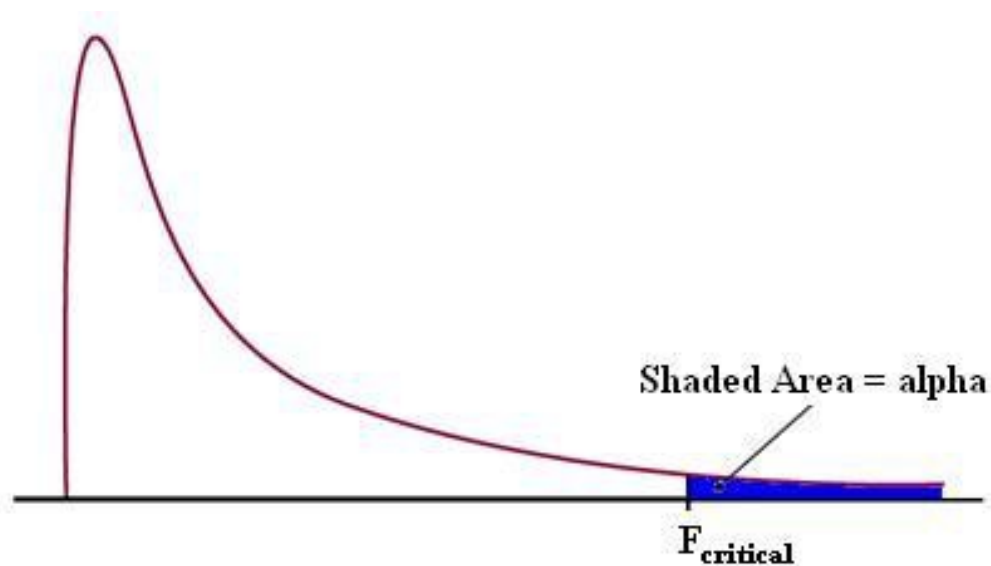
- This above formula is pretty intuitive. The numerator term in the F-statistic calculation defines the between-group variability.
- As between group variability increases, sample means grow further apart from each other. In other words, the samples are more probable to be belonging to totally different populations.
- This F-statistic calculated here is compared with the F-critical value for making a conclusion.
- In terms of our medication example, if the value of the calculated F-statistic is more than the F-critical value (for a specific α /significance level), then we reject the null hypothesis and can say that the treatment had a significant effect.

F-Statistic



F-Statistic

- Unlike the z and t-distributions, the F-distribution does not have any negative values because between and within-group variability are always positive due to squaring each deviation.



F-Statistic

- There is only one critical region, in the right tail (shown as the blue shaded region above).
- If the F-statistic lands in the critical region, we can conclude that the means are significantly different and we reject the null hypothesis.
- Again, we have to find the critical value to determine the cut-off for the critical region. We'll use the F-table for this purpose.
- We need to look at different F-values for each alpha/significance level because the F-critical value is a function of two things: df_{within} and df_{between}

One Way ANOVA

- A recent study claims that using music in a class enhances the concentration and consequently helps students absorb more information. As a teacher, your first reaction would be skepticism.
- What if it affected the results of the students in a negative way? Or what kind of music would be a good choice for this?
- Considering all this, it would be immensely helpful to have some proof that it actually works.

One Way ANOVA

- To figure this out, we decided to implement it on a smaller group of randomly selected students from three different classes.
- The idea is similar to conducting a survey. We take three different groups of ten randomly selected students (all of the same age) from three different classrooms.
- Each classroom was provided with a different environment for students to study.
- Classroom A had constant music being played in the background, classroom B had variable music being played and classroom C was a regular class with no music playing. After one month, we conducted a test for all the three groups and collected their test scores.

One Way ANOVA

	Test scores of students (out of 10)									
Class A (constant sound)	7	9	5	8	6	8	6	10	7	4
Class B (variable sound)	4	3	6	2	7	5	5	4	1	3
Class C (no sound)	6	1	3	5	3	4	6	5	7	3

	Test scores of students (out of 10)										Mean
Class A (constant sound)	7	9	5	8	6	8	6	10	7	4	7
Class B (variable sound)	4	3	6	2	7	5	5	4	1	3	4
Class C (no sound)	6	1	3	5	3	4	6	5	7	3	4.3
Grand mean ->										5.1	

One Way ANOVA

- Looking at the above table, we might assume that the mean score of students from Group A is definitely greater than the other two groups, so the treatment must be helpful.
- Maybe it's true, but there is also a slight chance that we happened to select the best students from class A, which resulted in better test scores (remember, the selection was done at random).

One Way ANOVA

- This leads to a few questions, like:
 - How do we decide that these three groups performed differently because of the different situations and not merely by chance?
 - In a statistical sense, how different are these three samples from each other?
 - What is the probability of group A students performing so differently than the other two groups?

One Way ANOVA

- To answer all these questions, first we will calculate the F-statistic which can be expressed as the ratio of Between Group variability and Within Group Variability.

Let's complete the ANOVA test for our example with $\alpha = 0.05$.

$$SS_{\text{between}} = 54.6$$

$$SS_{\text{within}} = 90.1$$

$$df_{\text{between}} = 2$$

$$df_{\text{within}} = 27$$

$$MS_{\text{between}} = 27.3$$

$$MS_{\text{within}} = 3.33$$

$$F = 8.18$$

Limitations

- A one-way ANOVA tells us that at least two groups are different from each other. But it won't tell us which groups are different.
- If our test returns a significant f-statistic, we may need to run a post-hoc test to tell us exactly which groups have a difference in means.

Thank you

This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License



@mitu_skillologies



/mITuSkillologies



@mitu_group



/company/mitu-
skillologies



MITUSkillologies

Web Resources

<https://mitu.co.in>
<http://tusharkute.com>

contact@mitu.co.in
tushar@tusharkute.com