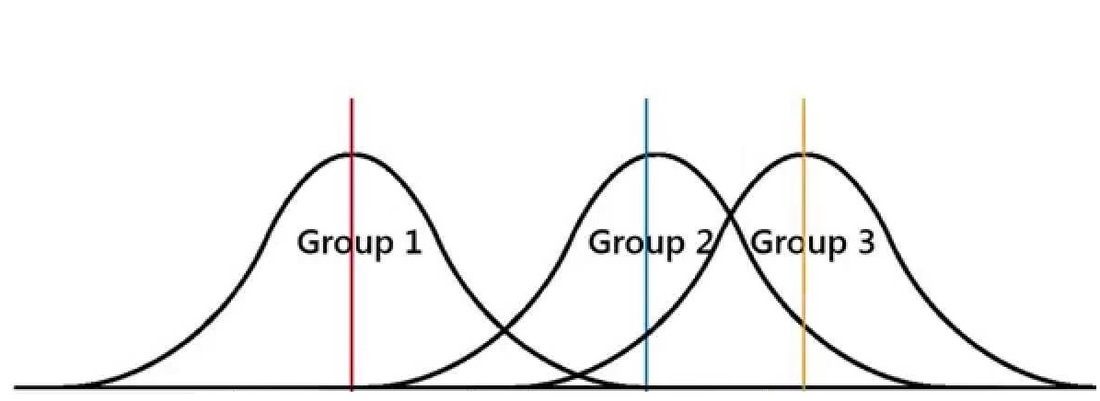
ANOVA

ANOVA is a type of regression where independent variables are nominal variables. Nominal variable is one that have two or more levels, but there is no intrinsic ordering for the levels. ANOVA stands for Analysis of Variance. It is used to compare more than two means. As the name suggest, it estimate an variance and based on the variance, it allow us to make a conclusion about the comparison of means. One of the most important algebraic equation in ANOVA is

TSS=SST+SSE (1)

The total sum of square can be broken down in to sum of square due to treatments and sum of square due to error. Here, the treatment represents the levels of nominal variable. If the nominal variable takes 4 level, then the number of treatment is 4.

MST=SST/D.O.F SST (2)

MSE=SSE/D.O.F SSE (3)

D.O.F = Degree of freedom

MST = Mean Sum of Square due to treatment

MSE = Mean Sum of Square due to error

MSE is an estimate of σ2. MST is also an estimate of σ2 if the treatment means are equal. It is an obvious that if there is a significant differences in the treatment means, MST will be larger compared to MSE. We can tell these thing more formally by conducting a F-test between MST and MSE.

F=MST/MSE (4)

If F-statistic is very large, then we can conclude that at least one of the treatment mean is different from other treatment means.

What is an Anova test used for?

The one-way analysis of variance (**ANOVA**) is **used to** determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups

What is the difference between Anova and t test?

Summary: The **t**-**test** is used when determining whether two averages or means are the same or **different**. The **ANOVA** is preferred when comparing three or more averages or means. A **t**-**test** has more odds of committing an error the more means are used, which is why **ANOVA** is used when comparing two or more means

What is difference between F test and t test?

**t**-**test** is used to **test** if two sample have the same mean. The assumptions are that they are samples from normal distribution. **f**-**test** is used to **test** if two sample have the same variance

What is F value in Anova?

An **F statistic** is a **value** you get when you run an **ANOVA** test or a regression analysis to find out if the means between two populations are significantly different

Example problem:

| Let's say we have three groups - A, B and C which have the below samples picked.A | | | | | Group B | Group C |
| --- | --- | --- | --- | --- | --- | --- |
| A | B | C |
| 37 | 62 | 50 |
| 60 | 27 | 63 |
| 52 | 69 | 58 |
| 43 | 64 | 54 |
| 40 | 43 | 49 |
| 52 | 54 | 52 |
| 55 | 44 | 53 |
| 39 | 31 | 43 |
| 39 | 49 | 65 |
| 23 | 57 | 43 |
|  |  |  |
| 440 | 500 | 530 |
|  |  |  |

|  |
| --- |
| SOLUTION:   1. Hypothesis:     H0: µ1= µ2= µ3  H1: µ1≠ µ2≠ µ3   1. Level of significance α = 0.05 2. Critical value DF = (2,27) 3. Test statistic: |
|  |
|  |
|  |
|  |
| T= ∑X1+∑X2+∑X3 = 1470  N= 10+10+10 = 30  CORRECTION FACTOR = T2/N = 72030  SST=22358+ 29242+ 31395-C.F = 3720  SSC = (440)2/10 + (500)210 + (530)2/10 = 420    INFERENCE:  Table value of F for ѵ1 = 2, ѵ2 = 27 at 5% level = 3.35. since the calculated value of F is less than table value H0 is fail to reject. Hence there is significant difference in the average among the 3 groups. |