ANOVA (Analysis of Variance)

Before going into the deep of ANOVA, we have to be familiar with F-distribution. It is used in ANOVA.

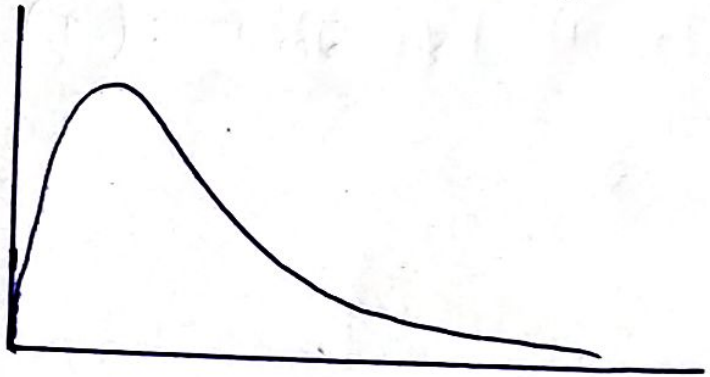
F-distribution

Let S_1 and S_2 are independent random variable with chi-square distribution with degree of freedom d_1 and d_2. then —

$X = \frac{S_1/d_1}{S_2/d_2}$ follows F-distribution.

It is a continuous distribution. having two parameter (degree of freedom).

- It is positively skewed and bounded.



(22) M.

(1) ANOVA Test :-

Basically there are two type of ANOVA test —

① One-way ANOVA test

② two-way ANOVA test

(A) One Way ANOVA Test :-

One way ANOVA test is conducted for comapring the means of three or more independent groups to determine if there are any significant different between them.

(1) The term 'one-way' refers to the fact that there is only one independent variables with multiple groups.

(1) Steps involve in this Test :-

① Create the Null hypothesis as well as Alternative hypothesis.

② Calculate the overall mean or grand mean.

Note that the main aim is to find out F-ratio / F-statistic.

For that the formula is —

$$\text{F ratio / F statistic} = \frac{\text{Mean square between (MSB)}}{\text{Mean square within (MSW)}}.$$

(1) steps for finding the MSB :-

$$\text{MSB} = \frac{\text{sum of square between}}{\text{degree of freedom}}$$

Now what is sum of square between.

$$\text{sum of square between} = \sum (\text{Group Mean} - \text{Grand Mean})^2 \times \text{total No of Number present in group particular}$$

steps for finding the MSW :

Mean square within = $\frac{\text{sum of square within}}{\text{degree of freedom.}}$

what is sum of square within :

Sum of square within = $\left(\text{For each group}\right) \Rightarrow \sum\left(\text{value} - \text{group mean}\right)^2$

After finding the F-statistics, we can find the P-value and comparing with level of significance, we can take decision about our null hypothesis.

Let's discuss all the steps with an example.

Example:

We create three group from a number-population. Now, we check if these three group are same or not.

Number	Category / Group
3	A
6	A
1	B
8	C
8	B
3	A
9	B
6	C
10	C

Answer

Null hypothesis (H_0): Three groups (A, B, C) are same. That means, the means are equal. $\mu_A = \mu_B = \mu_C$

Alternative hypothesis (H_a): One of them is significantly different.

(iii) Now our main target is to find out the F-ratio / F-statistic.

For our calculation purpose we rearrange the data —

Group

A	B	C
3	1	8
6	8	6
3	9	10

Grand Mean $(\bar{\bar{x}}) = \dfrac{(3+6+3)+(1+8+9)+(8+6+10)}{9}$

$= 6$

(iv) **Sum of square between**

Here the formula is —

$\sum$ (group mean − grand mean)^2 × No of point in the group

For Group A $\Rightarrow 3 \times (4-6)^2$ $(\bar{X}_A = 4)$

For group B $\Rightarrow 3(6-6)^2$ $(\bar{X}_B = 6)$

For group C $\Rightarrow 3(8-6)^2$ $(\bar{X}_C = 8)$

So, sum of square between $(SSB) = (12+12) = 24$.

and corresponding degree of freedom $= (3-1) = 2$

(as we have 3 group)

(v) **Sum of square with in**.

For group A : $\sum (x_i - \bar{X}_A)^2$ $(x_i \in A)$

$= (3-4)^2 + (6-4)^2 + (3-4)^2$

$= 1+4+1 = 6$

similarly for group B : $\sum (x_i - \bar{X}_B)^2 \quad (x_i \in B)$

$= (1-6)^2 + (8-6)^2 + (9-6)^2$

$= 25 + 4 + 9$

$= 38$

For group C : $(8-8)^2 + (6-8)^2 + (10-8)^2$

$= 4 + 4 = 8$

so, sum of square within (ssW) = $(6 + 38 + 8)$

$= 52$

and corrosponding the degree of freedom

$= (3-1) + (3-1) + (3-1)$

$= 6$

> Another way for finding the degree of freedom with respect to ssB and ssW :
>
> For ssB, df = (total group -1)
>
> For ssW, df = (total datapoint $-$ total group)

Now, F-ratio = $\frac{24/2}{52/6} = \frac{12 \times 6}{52} = 1.385$

① Now we have to check the condition

so, corrosponding the level of significance (0.05) and $df_1 = 2$ and $df_2 = 6$, we got 5.1433.

So,

f-ratio < 5.1433.

so, we cannot reject our Null hypothesis.

② Another way (p-value) :

We got P-value = 0.32 with respect to $df_1 = 2$ $df_2 = 6$ and F-statistic = 1.385.

So,

P-value > 0.05 (α)

So, We cannot reject the null hypothesis.

So, these group are same. That mean $\mu_A = \mu_B = \mu_C$ (Proved)

---

(11) **What if our Null hypothesis be Wrong**

If our Null-hypothesis be wrong, then we perform post-hoc test to find out which specific group or ~~group~~ pair of groups have significantly different means.

The main purpose of doing the post-hoc test is to control family-wise error rate (FWER) and adjust the significance level for multiple comparision to avoid inflated Type-I error.

(12) **Family-wise error rate**:

If We perform t-test pairwise with significance level 0.05, then for 3 pair of test significance level would be $(0.05 \times 3) = 0.15$. This is called Family-wise error rate.

To decrease this we do two thing —

① **Bonferroni Correction**

This method adjusts the significance level (α) by dividing it by the number of comparision we made.

suppose we have 3 pair. So our α value would be $(0.05/3)$.

But it may have lower statistical power with respect to large number of pair.

⑪ Tuskey's HSD (Honest significant Difference) Test :-

This test controls the FWER and is used when the sample size are equal and the variance are assumed to be equal accross the group.
