Analysis of Variance (ANOVA):-

* ANOVA is the strong statistical technique that is used to show [redacted] or check whether there is difference between 3 or more [redacted] unrelated or independent groups.

* Basically it is the generalisation of independent t-test

* It uses F-test to determine the hypothesis.

*It has 2 types they are one-way ANOVA and two-way ANOVA

* Example of Hypothesis in ANOVA

H_0: $H_1 = H_2 = H_3 = \cdots H_k$ [redacted]

H_1: Atleast [redacted] one group differ from other

F-test

* F-test in one-way ANOVA is used to assess whether the expected values of a quantitative variable within several Pre-defined grps differ from each other.

$$F = \frac{\text{Variance between groups}}{\text{Variance within groups}}$$

(OR)

$$F = \frac{\text{Sum of Squars Between groups (SSB)}}{c-1}$$

$$\frac{}{\text{Sum of Squares within groups (SSW)}}$$

$$n-c$$

* where n is ~~no of sample~~ sample size,
and c is no of groups

* $c-1$, $n-c$ are respective degrees of freedom of numerator and denominator.

* Basic formula

$$SST = SSW + SSB$$

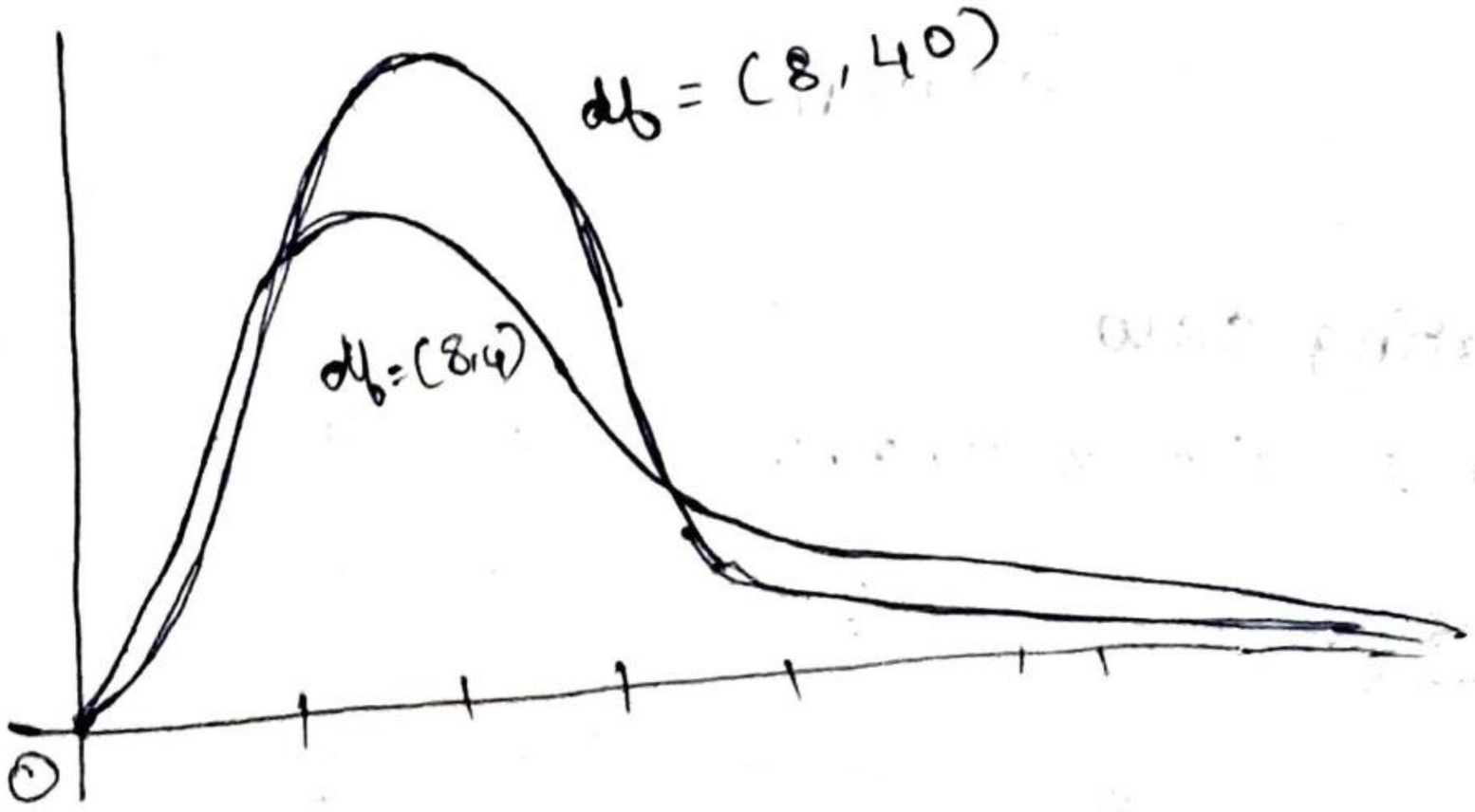
where SST is Total sum of Squares.

F- distribution:-

* F-distribution also called as Variance ratio distribution as it usually defines the ratio of the variances of two normally distributed variances.

* F- distribution is skewed distribution because, the F- value can never be negative. Its either 0 or positive value.

* The F- distribution depends on the degree of freedom of numerator and denominator.



One way ANOVA

* One way ANOVA is a technique that is used to compare or test three or more groups whether it is significantly differ from or not.

* Here we have one dependent Var affecting dependent variable in

① $H_0 : M_1 = M_2 = M_3$

H_1 : Atleast one group different from other

Sample 1 = [2, 3, 7, 2, 6]

Sample 2 = [10, 8, 7, 5, 10]

Sample 3 = [10, 13, 14, 13, 15]

$$F = \frac{SSB / C-1}{SSW / n-C}$$

Calculating SSW -

Sample 1 (2, 3, 7, 2, 6) $\bar{x} = 4$

$\sum (x_i - \bar{x})^2 = 22$

Sample 2 (10, 8, 7, 5, 10) $\bar{x} = 8$

$\sum (x_i - \bar{x})^2 = 18$

Sample 3 (10, 13, 14, 13, 15) $\bar{x} = 13$

$\sum (x_i - \bar{x}) = 14$

$$SSW = 22+18+14 = 54$$

$$\boxed{SSW = 54}$$

Since $SST = SSW + SSB$

calculating SST :-

combine all the samples and make it as one group $\bar{x} = 8.3$

$$\Sigma (x_1 - \bar{x}) = 257.3$$

$$SSB = SST - SSW$$

$$= 257.3 - 54$$

$$\boxed{SSB = 203.3}$$

s_1 s_2 s_3

$5+5+5 = 15$

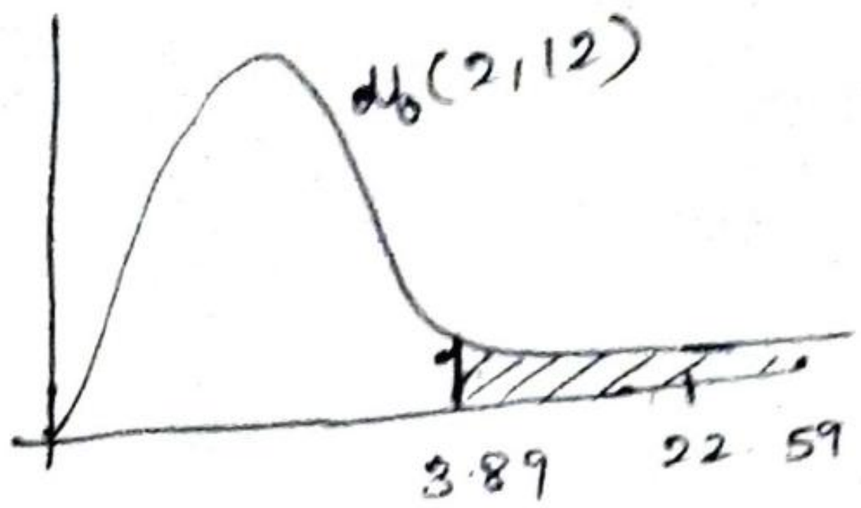
$$F = \frac{203.3/2}{54/15-3}$$

$$F = \frac{203.3/2}{54/12} = \frac{101.66}{4.5}$$

$$\boxed{F = 22.59}$$

for $F(2,12)$ the F-value is 3.89

means



∴ Reject the null hypothesis

∴ There is statistical difference between the groups.

* we can also compare a numerical feature and categorical feature with more categories (2 way anova)

Two way ANOVA:- (Factorial analysis)

* Two way ANOVA is used to estimate how the mean of quantitative Variable or numerical Variable changes according to the levels of two categorical variables.

Basically it tells how 2 categorical ~~variables~~ feature have impact on numerical feature.

* Two way ANOVA with interaction tests 3 null hypotheses at the same time

1) There is no difference in group means at any level of the first independent Variable.

ii) There is no difference in group means at any level of the second independent Variable.

iii) The effect of one independent Variable does not depend on the effect of the other independent Variable (no. effect)