F-Test (Regression Analysis)

M T W T F S S | Page No.: | Date: | YOUVA

- Tells if relation b/w input & output is linear or not.

① State the null & alternative hypothesis'

For 3 input cols $\quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

H0: $\beta_1 = \beta_2 = \beta_3 = 0$

H1: atleast one of regression coef is not zero.

For 1 input col

$$y = \beta_0 + \beta_1 x.$$

H0: $\beta_1 = 0$

H1: $\beta_1 \neq 0$

② Calculate F-statistic

$$\text{F-stat} = \frac{\text{MSR (Mean squared regression)}}{\text{MSE (Mean squared error)}}$$

Where

$$\text{MSR} = \frac{\text{ESS}}{\text{df\_model}} = \frac{\text{ESS}}{K \text{ (no of input col)}}$$

$$\text{MSE} = \frac{\text{RSS}}{\text{df\_residuals}} = \frac{\text{RSS}}{n-K-1}$$

↳ input no of rows.

$$\text{F-Stat} = \frac{\text{ESS}}{\text{df\_model}} = \frac{\text{TSS} - \text{RSS}}{K}$$

$$\frac{\text{RSS}}{\text{df\_residual}} \qquad \frac{(y_i - \hat{y}_i)^2}{(n-K-1)}$$

It is call f-stat, because this will follow F distribution as you know F-dis = $\left[\frac{\frac{chi^2}{df1}}{\frac{chi}{}}\right]$

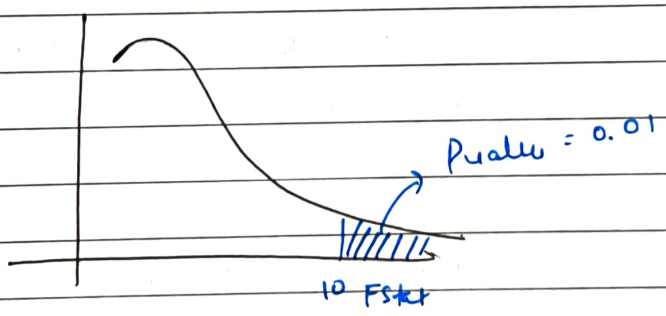
& chi² distribution is (Normal distribution)²

$(y_i - \hat{y})^2$ will follow normal distribution,
Hence $(y_i - \hat{y})^2$ will follow chi² distribution

Same goes for ESS = TSS - RSS is ~~normal~~ chi² distribution

$$F\text{-stat} = \frac{\frac{ESS}{df\_model}}{\frac{RSS}{df\_residual}} = \frac{\frac{ESS}{K}}{\frac{RSS}{n-K-1}}$$

(3) Calculate P value.



Since Pvalue is less than 0.05 (significance) we will reject the null hypothesis.
i.e mean $\beta_1 = 0$ is wrong.

M T W T F S S

Page No | YOUVA

Date

INTUTION

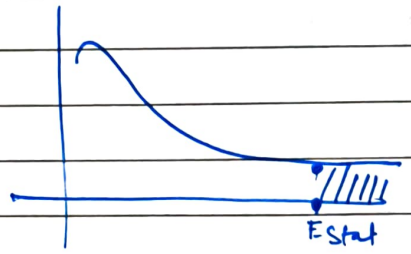
$$F\text{-stat} = \frac{\frac{ESS}{K}}{\frac{RSS}{n-k-1}}$$

ESS → Explained variance by regression per degree of freedom

RSS → Unexplained variance by regression per degree of freedom

(1) if F-stat is very large, that means ESS is very large. (Regression explained large var)

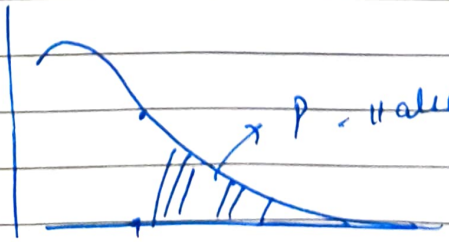
& if Fstat large that means, In F-dis it will lie at → & P-value is very low.



Conclusion: Since ESS is more, which means It is working ie means there is a linear relationship

(2) if F-stat is very small, ie means unexplained variance (RSS) is very large, which means our regression line wasn't able to explain much variance.

∴ F-Stat will be small



Hence we can't reject null hypo

It means, there is no linear relationship

R SQUARED (R2 score)

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS}$$

$$= 1 - \frac{RSS}{TSS}$$