1) ANOVA

2) Chi Square

3) EDA

ANOVA ⟹ Analysis of Variances

This is a test to compare multiple items

ex: C_1 C_2 C_3

H_0: $\mu_1 = \mu_2 = \mu_3$ ⇒ Null hypothesis (all are equal)

H_1: $\mu_1 \neq \mu_2 \neq \mu_3$

$\mu_1 \neq \mu_2 = \mu_3$

$\mu_1 = \mu_2 \neq \mu_3$

$\mu_3 = \mu_2 \neq \mu_1$

⇒ Alternate hypothesis

(Atleast one of them is different)

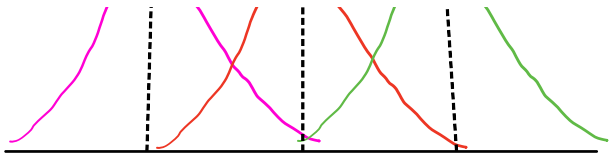
→ Collect data from all

→ Understand the distribution of each value.

→ In graphical distribution, if they are close to each other i.e distance between them

- if its closer than they are correlated.
- if they are far from each other they are not correlated.

'close' isn't clearly defined



'Close' ⇒ Inter group distance is negligible compared to Intra-group distance.

⇓

We average this to get better comparision

Inter group distance → distance between group

Intra group distance → distance within group

Inter group distance

ex

$[x_1 \dots x_n]$ C_1

$[y_1 \dots y_n]$ C_2

$[z_1 \dots z_n]$ C_3

$\mu_1 = \frac{\sum x_i}{n}$

$\mu_2 = \frac{\sum y_i}{n}$

$\mu_3 = \frac{\sum z_i}{n}$

grand mean $= \frac{\mu_1 + \mu_2 + \mu_3}{3} = \mu$

grand μ $\mu = \frac{\mu_1 + \mu_2 + \mu_3}{3}$

$d_1 = (\mu - \mu_1)^2$

$d_2 = (\mu - \mu_2)^2$

$d_3 = (\mu - \mu_3)^2$

⇒ 3 variables, then degrees of freedom is 2

assign weightage

$n_1 d_1$

$n_2 d_2$

$n_3 d_3$

Average Inter group = $\frac{n_1 d_1 + n_2 d_2 + n_3 d_3}{2}$

Intra - group distance

$$\frac{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2 + (y_1 - y_2)^2 + (y_2 + \dots y_2)^2 + (y_n + \dots y_n)^2 + (z_1 - y_3)^2 + \dots (z_n + y_3)^2}{3(n-1)}$$

Average intra group distance $\rightarrow \frac{\Sigma (x_i - y_1)^2 + \Sigma (y_i - y_2)^2 + \Sigma (z_i - y_3)^2}{3(n-1)}$

$\left(\frac{\text{Distance Between Groups}}{\text{Distance Within Groups}} \right)$ $\Rightarrow$ If this ratio is low, H_0

$\Rightarrow$ If this ratio is high, H_1

Mean : Atkins : 4.5

GM : 5.33

South Beach : 7

grand mean : 5.6111

Statistical Average

Intergroup

$$= \frac{12(4.5 - 5.6111)^2 + 12(5.33 - 5.6111)^2 + 12(7 - 5.6111)^2}{2}$$

Intra - group

Statistical Distribution

$$\text{F-ratio} = \frac{(\text{SS between Groups} \mid \text{DF})}{(\text{SS within Groups} \mid \text{DF})}$$

Conclusion : If calculated F ratio is greater than F value then we reject null hypothesis H_0

Chi - Square

EDA → Exploratory Data Analysis

Extremely important ⇒ Cleaning i.e pre-processing the data