Numerical - Categorical Relationship

⇒ Tests → T- test
→ Z- test

⇒ Constraint on Num - Categories

⇒ num- categories = 2

Questions What test would you use for following Variable pairs

① Product vs Gender

Chi - square

② Income vs Gender

T- test / z - test

③ Product vs Income?

(> 2 Categories) (Numerical)

⇒ $^{n}C_{2}$ T- Tests (n ⇒ num_categories)

⇒ One Way ANOVA

One-Way ANOVA Test

- * One ⇒ A single categorical Variable (Ex: Product)
- * Way ⇒ Dependent Variable (Ex. Income)
- * ANOVA ⇒ Analysis of Variance
- * Uses F-Distribution to generate F-statistic / F-ratio
- * This test is suitable when dealing with

  Numerical Variable

  vs

  Categorical Variable (num_categories > 2)

- * Aerofit : Is there a significant Difference between Income of Different Product Buyers

H_0 : Means across all groups are Same

H_a : Atleast mean of One Group is Different

In Anova

① Analys Variance of Data using F Distribution

② Two types <
- Variance Between the Groups
- Variance within the groups

$$\boxed{\text{F - ratio}}$$

$$\boxed{\frac{\text{Variance Between the Groups}}{\text{Variance Within the Groups}}}$$

F - ratio

F statistic

$$\text{F - ratio} \propto \frac{1}{P}$$

* Interpretation

① F - score is 1 (if Both within and Between Variances are close to each other

② Population is probably Common for all Group/Samples

② F-score >> 1

⇒ Samples might be from Population with Different means

Assumptions of ANOVA

① Normality of Data (Residual)

② Homogeneous Variances (Residual)
(Variances across groups should be approximately Equal)

③ Independent Observations

Questions

① Do we have Alternative when Data is Not Gaussian?

② How do we check if Data is Gaussian?

Kruskal-Wallis Test

Ho : All groups have same population median

Ha : There is a significant difference in medians of Atleast two group

Advantages

1. No assumption of Normality

2. Robust to Outliers

Limitation

1. Less powerful compared to ANOVA when assumption are met.

② How do we check if Data is Gaussion?

Ans ① Plot Historgram

② Plot QQ-plot

③ Shapiro Wilkins Test

H_0 ⇒ Data is Gaussian

H_a ⇒ Data is Not Gaussian

α ⇒ 0.05

test_stat, pvalue ⇒ Shapiro(data)

Levene - Test

Used for checking if variances in 2 samples are equal or Not

H_0 : Variances are Equal i.e. $\sigma_1^2 = \sigma_2^2$

H_a : Variance are not Equal

=> Returns a p-value and a Test-statistic

=> P-Value can be compare with α to reject or fail to Reject: $H_0 (\sigma_1^2 = \sigma_2^2)$