Non Parametric Test

Parametric Test

- Normal Distribution.
- More powerful.
- Complete information about population.
- Compare mean and std.deviation.
- Quantiative Data.
- E.g T-Test,Z-Test,F-Test,ANOVA.

Non Parametric Test

- Skewed Distribution
- Qualitative Data
- Nominal and ordinal Data
- Less Powerful
- Incomplete information about Population
- Compare % and proportion
- E.g chi Square Test,Krushal walis Test,Mann whitney U test

Chi Square

- Given by Karl Pearson in 1900
- Compare expected frequency with observed frequency
- Observation are independent and random
- Skewed distribution(Non-negative value)
- Hypothesis Testing
- If observed value is more than critical value,Null Hypothesis is rejected

Chi Square Test

- Goodness of fit: to measure consistency of sample with population.
- Df==(r-1)(c-1)
- Df is no of Independent value that can be assigned to statistical distribution
- Also called test of independence of data
- Frequency must be atleast 50
- Used as parametric test to test population variance on the basis of small size sample variance.

Mann whitney U Test

- Data is not normal distributed
- Measure median of 2 smaple of 1 population
- Non parametric variant of T-test
- 5-20 units in one sample
- Data is ranked from highest to lowest
- Ordinal data
- Also called mann whitney Wilcoxon or mann whitney Wilcoxon

Rank sum test.

Mann whitney U Test

- In parametric we comare independent sample means with the assumption that population is normally distributed
- Mann-whitney test allow us to compare two population where underlying distribution are not normal but they have similar shapes
- Therefore we use the ranks of observations(location)

Mann and Whitney U test Formula

- where:
  - n_1: number of samples in sample 1
  - n_2: number of samples in sample 2
  - R_1: Rank sum of sample 1
  - R_2: Rank sum of sample 2

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$
$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

- Now, our test statistic (U) will be smaller of U_1 and U_2.
- Now, we look to the critical values in the table with respect to n_1 and n_2 (take it U_0).
  - if $U <= U_0$: we reject the null hypothesis.
  - else, we do not reject the null hypothesis.

Krushal wallis H test

- Data is not normally distributed.
- Alternate to anova
- Determine statistically significant difference between more than 2 sample
- One way anova on rank
- Used rank instead of actual data
- Extension of mann whitney U test

Krushal wallis H test

- Comparing more than two independent samples using a parametric test would lead us to anova procedure.
- We compare mean with assumption that population is normally distributed.
- The kruskal wallis test allow us to compare two or more population that are not normally distributed
- We use sum of ranks for observations.

F-Test

- F test is named in the honour of statistician R.A.Fisher
- It is used to find out whether two independent estimates of population variance differ significantly.
- To carry out the test of significance we calculate the ratio of F
- F= variance of first data set/variance of second data set
- Variance is square of std deviation
- variance of first data set > variance of second data set

F –Test

- In simplified way F –Test can be expressed as
- F= Larger estimate of variance / smaller estimate of variance

$$\text{F\_value} = \frac{variance1}{variance2} \qquad \sigma^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

$$\text{i.e. F\_value} = \frac{\sigma_1^2}{\sigma_2^2}$$

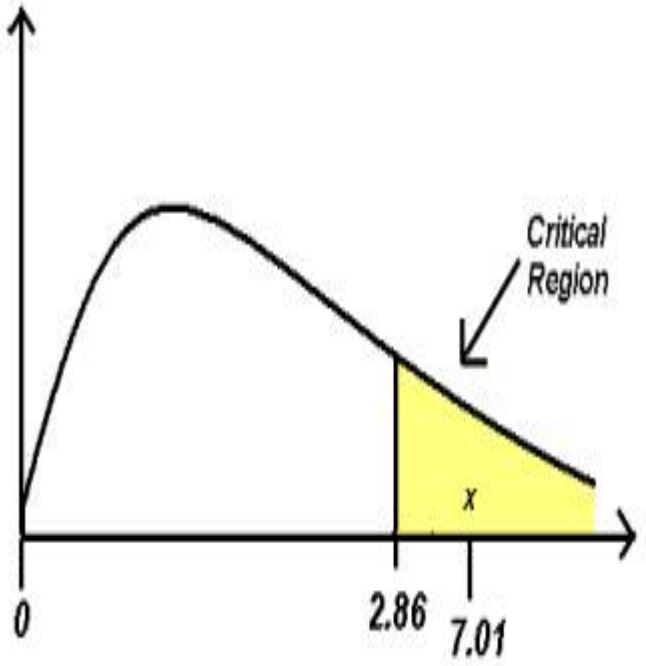
- Degree of freedom (V)
- V1(numerator) = n1-1 (larger variance)
- V2(denominator) = n2-1 (smaller variance)
- Calculated F – value will be compared with table F –value for v1 and v2 D.F at 1% or 5% level of significance

- If calculated F-Value < table F- value
- H0 is accepted (no significant difference between two variance)
- Otherwise H0 rejected (significant difference between two variable)

- F test is based on ratio of two variance , it is also known variance ratio test.

Properties of F distribution

- F distribution is skewed towards right with the range o to infinity having rough median value as 1
- Shape of f distribution id dependent on d.f of numerator and denominator
- F distribution curve is never symmetrical, but if d.f will be increased the it will be more similar to symmetrical shape
- Degree of skewness decrease with increase in degree of freedom



F Test Statistic Formula Assumptions

- F Test equation involves several assumptions. In order to use the F - test Formula, the population should be distributed normally. The samples considered for the test should be independent events. In addition to these, it is also important to consider the following points.
- Calculation of right-tailed tests is easier. To force the test into a right-tailed test, the larger variance is pushed in the numerator.
- In the case of two-tailed tests, alpha is divided by two prior to the determination of critical value.
- Variances are the squares of the standard deviations.

- If the obtained degree of freedom is not listed in the F table, it is always better to use a larger critical value to decrease the probability of type 1 errors.