Statistics for Data Science

Test and their Types

- Parametric Test
- Non-Parametric Test

Parametric Test

1. More relaible results
2. First we have to meet the assumptions Example: Not Equal Data(The Comparison of girls and boys *Age*)

Non-Parametric Test

1. Less Relaible results
2. Calculate the rank of Data
3. No need to meet the Assumptions Example: Based on Ranking(the age of boy 1,2..)

Step-1

- **Normality Test**
    - Test to be Used:
        1. Shapiro-Wilk Test Specific(Relaible)
        2. Kolmogorov-Smirnov Test General(Less Relaible)

Step-2

- **Homogenity Test**
    - The Variance of the Variable
    - Test to be Used:
        1. Levene's Test

Step-3

- **Purpose**
    - Know thw Purpose of your Resesarch Question
    - There are two types of Purposes
        1. Comparison (Difference)
        2. Relationship (Connection)

1. **Comparison**
    - Atleast two groups
        - Example:
            - Male vs Female

- Control Group vs Treatment Group
- Grouping Individual by Color Preferences

1. **Relatioship**
   - Find a Connection
     - Examples:
       - Can food predict weight of a group of individuals
       - Do fertilizer application increase crop growth
   - We seek following here
   - Connection
   - Correlation
   - Causation
   - Prediction

Step-4

- **Data Type**
  - Know the Type Of Data you are working with
  - Two types of Data
    1. Categorical
    2. Continuous

1- **Categorical**

- Qualitative
- Non-Numerical Meaning
- Represented in Texts Example:
  1. Character
  2. Factors

Example:

- Yes or No answer
  - Do you like Mangoes
  - Are you learning with Baba Aammar

2- **Continuous**

- Quantitative
- Numerical
- Mostly Represented in number
- Example:
  - Numerical Variable(int and float)

Examples:

- Amount
- Number

- Age
- Plants Height
- Number of Plants Colonies
- Chlorophyl Content
- Fertilizer Ammount

Step-5

- **Statistical Test**
  - Choose a Statistical Test for 3 main families
    1. Chi-Squared
    2. t-Test/ANOVA
    3. Correlation

- **Chi-squared**(It is a Non-Parametric Test)
  - Purpose: Comparison
  - Data: Categorical Only
    - Types
    1. Test of Homogeneity
    2. Test of Independance
    - When to use?
    1. Nothing affect this
    2. Can be used with number of level or groups

- **t-Test/ANOVA**
  - Purpose: Comparison
  - Data: Categorical and Continuous
    - Types
  1. One-Sample t-Test
    - For one sample group with a known mean
  2. Two-Sample t-Test
    - Unpaired t-Test
      - Two different group
    - Paired t-Test
      - Same group twice
  3. ANOVA
    - Analysis of Variance
    - 3+Level of groups are involved
    - One way ANOVA
    - Even one group is significant you wil get significant results,but doesn't tell which one
    - Two way ANOVA
    - Repeated measures of ANOVA
    - 3+paired groups, scale up of paired t-Test

- **Correlation**
  - Purpose: Relationship
  - Data: Continuous Only
    - Types
  1. Pearson's Correlation
     - One Independent & One Dependent Variabele
  2. Regression
     - One Independent & One Dependent Variabele
     - **Correlation:** Tell us how closely two variable are connected " Is food a predictor of weight gain"
     - **Regression:** Tell us specific mathematical equation that describe the relationship e.g Missing values can be predicted like this.

Difference betweeen ANOVA, ANCOVA, MANOVA & MANCOVA

The core component is ANOVA. ANOVA test three or more groups for mean difference of continuous response varaible(dependent variable)

- **ANCOVA**
  - C stands for Covariance
  - It has single continuous variable like ANOVA
  - ANCOVA compare a response variable by both a factor and a continuous independent variable
  - The continuous independent variable used in ANCOVA is called **Covariate**

- **MANOVA**
  - MANOVA is an ANOVA with two or more continuous variable
  - M stands for Multivariate
  - MANOVA has both one way and two way types

- **MANCOVA**
  - Both MANOVA and MANCOVA has two or more response varaible