Statistics

Collection of methods for collecting,displaying,analyzing and drawing conclusion from data.

language of statistics

- ***Average*** income in Pakistan
- **Highest** score in cricket match
- Fastest (**Maximum**) bowler
- 40% (**percentage**) Teachers in pakistan are females
- Weather forecast
- Stock market

Data Types

1- Cross sectional and Time series

2- Univariate and Multi variate

Variable types

1- Binomial and Multinomial

2- Categorical (Nominal)

- ordinal

3- Ratio data

4- Interval variables/data

Measurment of Central tendencies

MEAN ,MEDIAN AND MODE

TESTS AND THEIR TYPES

PARAMETRIC TESTS

- More reliable results
- first we have to meet the assumptions

Column A	Column B
2	25
5	38
16	52

Not equal !

NON PARAMETRIC TESTS

- less reliable
- Callculate rank of the data
- No need to make assumptions

Column A	Column B
1	1
2	2
3	3

Equal? on the base of ranking

--- *Before starting the data analysis*

Step 1 Normality Test

If the bell curve on graph is normal means that data is normal . Tests that are used most commonly are

- Shapiro-Wilk Test (more specific and reliable)
- Kolmogorov-Smirnov Test (General and less reliablr)

Step 2 Homogeneity Test

The variance of the variable in data are equal.

Tests that are used are

- Levene Test

Step 3 Purpose

know the purpose of your reserach question.

Two types Of purposes

- comparison (difference)

e.g male vs Female

control vs disease group

- Relationship (connection)

  e.g Can food predict weight of group of indivisual

  Do Fertilizer application increase crop growth

Step 4 Data TYpe

Knowing type of data we are working with .

TWO TYPES OF DATA:

1- Catagorical

- Qualitative
- no numerical meaning
- represented in text
- e.g>charater,factor,yes or no

2-Contineous

- Quantitave
- numerical
- mostly represented in numbers
- e.g>numerical variable,amount,no,plant height,age,fertilizer amount

Step 5 Statistical Tests

choosing statistical test from three main families .these are

- chi-squared

> purpose is comparison

> data is catagorical only

> can be use with any number of level

> must remember purpose and data type

- t-Test /Anova

> purpose is comparison

> data iscategorical and cntineous

- Correlation

> purpose is frelationship

> data is contineous only

> types include

```
    - pearson's correlation >shows how closly two values are
    - Regression>shows specific methametical equationthat  decide relationship .
```

Important things

- if data is following gaussian distribution ,that means data is normally distributed

- data normalization methods are
    - Log transformation
    - Min-max scalling
    - Z-score normalization
    - unit vector transformation
    - mean normalization
    - Box-cox normalization

Anova

3+level or groups are involved

1-1 way anova

```
even on of the group is significant ,you will get significant results,but does not
tell you which one .
```

2-Two way anova

```
Two factor are envoloved
```

3-Repeated measures of anova

```
Three +paired groups are envolved
```

Anvoca

Analysis of co variance compare the means of three +independent groups which cannot be tested by anova because the variables are affected by co-variance .

Manova

Multi - variate analysis of variance

Mancova

Multi variate analysis of CO - variance

OTHER TESTS

- Reliability tests
- Validity tests
- Sample size computation
- Inter rater reliability tests