

Bivariate Analysis Concepts

1.Covariance:

- Measures how two variables vary together.
- A positive covariance means the variables tend to increase or decrease together.
- A negative covariance means one variable tends to increase when the other decreases.
- Covariance values range from $-\infty$ to $+\infty$.

2.Corelation:

- Measure the linear relationship between two variables.
- Standardized version of covariance. values range from -1 to +1.

3.MultiColinearity:

- Multicollinearity is a situation in multiple regression analysis where two or more independent (predictor) variables are highly correlated with each other affecting model interpretation and reliability.
- Structural multicollinearity- X , X2 and Data multicollinearity-data itself is multicollinear.
- Ways to Address Multicollinearity:
 - Remove or combine highly correlated predictors.
 - Use dimensionality reduction techniques like Principal Component Analysis (PCA).
 - Regularization methods such as Ridge regression (L2 regularization).

4.Variance Inflation Factor:

- VIF is a simple way to measure how much multicollinearity is inflating the variance (or uncertainty) of a regression coefficient estimate.
- VIF=1 no collinearity
- VIF=1 to 5, moderate
- VIF >5 or 10 , problematic collinearity.

5.Homoscedasticity:

- Means "same variance" of errors across all levels of the independent variables.
- Homoscedasticity supports valid inference from models in regression.

6.Heteroscedasticity:

- Means "different variance" or uneven scatter of residuals across levels of the independent variables.
- Heteroscedasticity can lead to incorrect conclusions and less reliable models.

7.T-test:

- The T-test is a technique for testing if two groups differ in their average values(mean).

- It compares the means of two samples and tells you whether any observed differences are likely to be real or just due to random chance.
- The larger the T value, the bigger the difference between groups. P-value up to 0.05 is acceptable.

Types:

- **Independent samples T-test(unpaired)**: Compares the means between two separate groups with same condition. (e.g., treatment vs. control).
- **Paired samples T-test**: Compares means from the same group at different condition (e.g., before and after treatment).
- **One-sample T-test**: Compares the mean of a single group to a known or hypothesized population mean.

8.Hypothesis testing:

- It involves testing an assumption (called a hypothesis) about a population parameter using data from a sample.
- **Null Hypothesis** : The default assumption, usually stating no effect or no difference.[p-value <0.05 then reject null hypothesis]
- **Alternative Hypothesis** : The statement you want to test for, usually a difference or effect.

9.Analysis of variance:

- Analysis of Variance (ANOVA) is a method used to compare the means of three or more groups to determine if there are any significant differences between them.
- F-ratio=ratio of(between-group variance to/within-group variance). A higher F indicates the group means are more different compared to the variability inside each group.
- p-value <0.05 then reject null hypothesis

Types of ANOVA:

- One-way ANOVA: Examines the effect of one independent variable with multiple groups on a dependent variable.
- Two-way ANOVA: Assesses the effects of two independent variables and their interaction on a dependent variable.

