UNIT IV DATA PREPARATION AND ANALYSIS

–Applications of Bivariate and Multivariate statistical techniques

I . UNIVARIATE ANALYSIS

It focus on the level coverage and distribution (variance)) of a single variable

II. BIVARIATE ANALYSIS

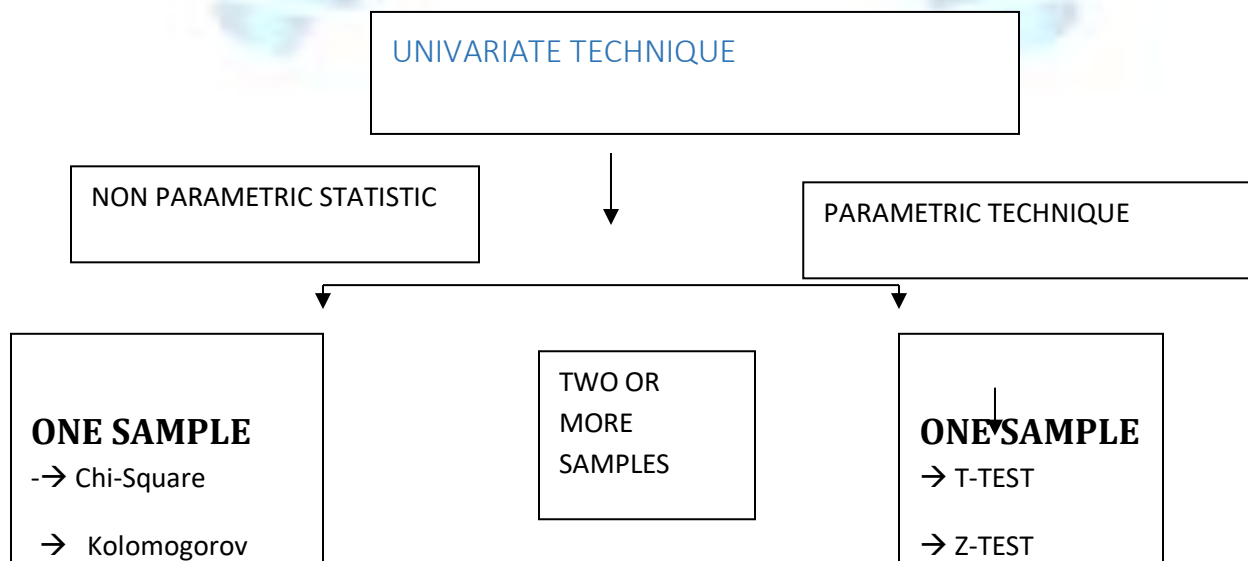It focus on the degree of relationship (correlation or covariance) between two variables.
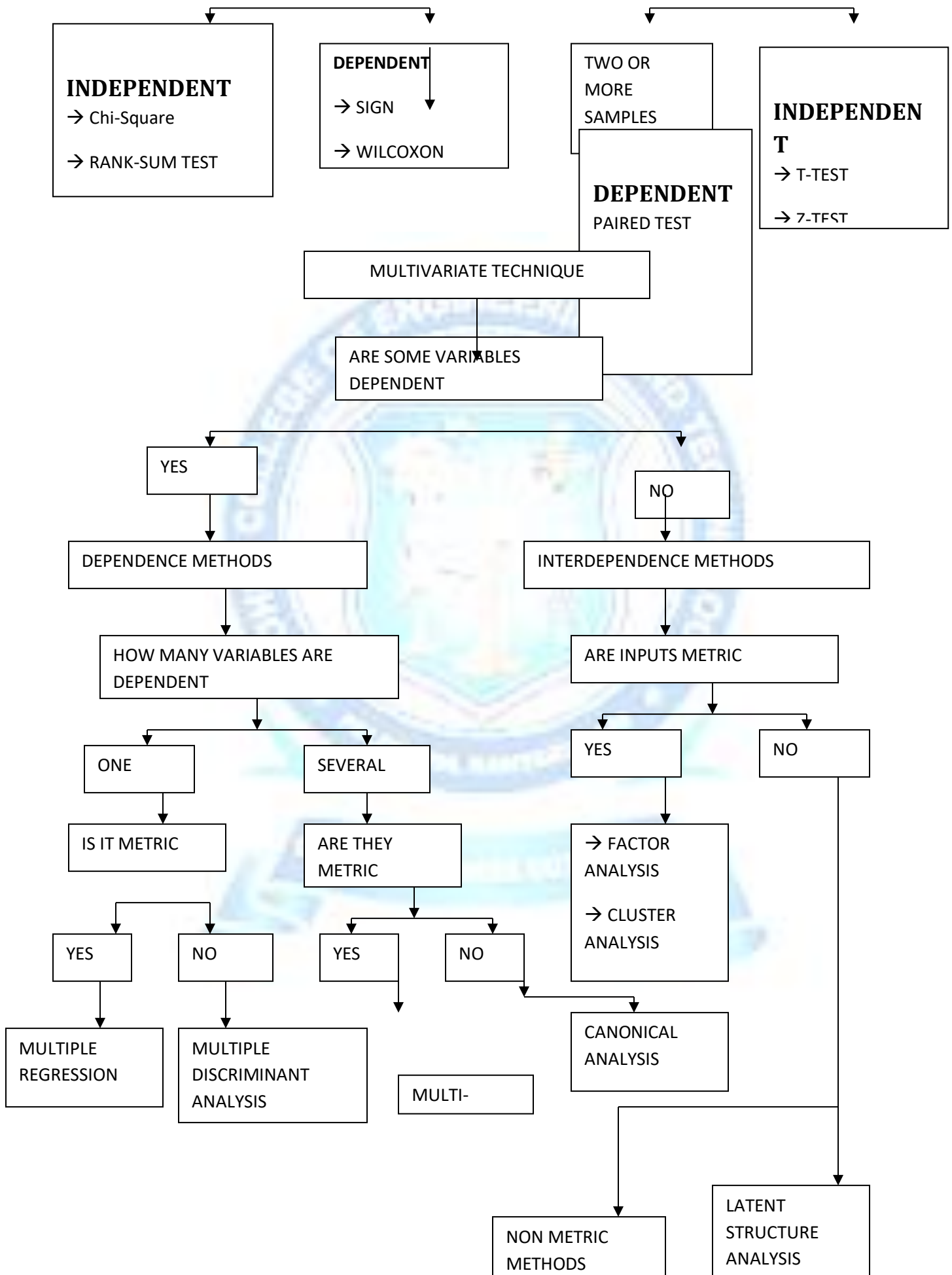
III. MULTIVARIATE ANALYSIS

It is the analysis of the simultaneous relationship among three or more variables. In a multivariate analysis, the focus shift from paired relationships to the more complex simultaneous relationships among three or more variables.

The multivariate technique can be broadly classifed as dependence technique as dependence technique or interdependence techniques.

Dependence techniques are appropriate when one or more variables can be identified as dependent variable and the remaining as independent variables.

In the interdependence techniques the variables are not classified as dependant or independent.

UNIVARIATE TECHNIQUE

NON PARAMETRIC STATISTIC		PARAMETRIC TECHNIQUE

ONE SAMPLE	TWO OR MORE SAMPLES	ONE SAMPLE
-→ Chi-Square		→ T-TEST
→ Kolomogorov		→ Z-TEST

INDEPENDENT

→ Chi-Square

→ RANK-SUM TEST

DEPENDENT

→ SIGN

→ WILCOXON

TWO OR MORE SAMPLES

INDEPENDENT

→ T-TEST

→ Z-TEST

DEPENDENT

PAIRED TEST

MULTIVARIATE TECHNIQUE

ARE SOME VARIABLES DEPENDENT

YES

NO

DEPENDENCE METHODS

INTERDEPENDENCE METHODS

HOW MANY VARIABLES ARE DEPENDENT

ARE INPUTS METRIC

ONE

SEVERAL

YES

NO

IS IT METRIC

ARE THEY METRIC

→ FACTOR ANALYSIS

→ CLUSTER ANALYSIS

YES

NO

YES

NO

MULTIPLE REGRESSION

MULTIPLE DISCRIMINANT ANALYSIS

MULTI-

CANONICAL ANALYSIS

NON METRIC METHODS

LATENT STRUCTURE ANALYSIS

Variable In Multivariate Analysis

1. **Exploratory Variable**

   Explanatory variables are called external or exogenous variables. If x a considered t be the cause of y, then x is the explanatory variable (or) causal or independent variables.

2. **Criterion Variable**

   These are called endogenous variables. In the above stated case y is called as criterion (or) dependent resultant variable.

3. **Observable & Latent variable**

   When explanatory variable are directly observable variable, they are termed observable variable. However some unobservable variable may influence criterion variable in which case they are called latent variables.

4. **Discrete variable & Continuous Variable**

   Discrete variables are those that lake only the integer value when measured. Continuous variables are those that when measured, can assume any real volume.

5. **Dummy Variable**

   This term is used in a technical sense and is useful in algebraic manipulations in the context of multivariate analysis

Application:

Used to predict the dependent variable , given knowledge of independent variable.

To understand the relationship between the dependent variable and independent variable.

- **Inputs :**

  Variable value for dependent and the independent variable.

- **Output:**

  It will output the regression coefficients and their associated beta coefficient and t-values which can be used to evaluate the strength of the relationship between the respective independent variable and the dependent variable.

Statistical Test:

The hypothesis that a regression parameter obtained from the sample evidence is zero or not is based on the t-value.

Limitation

- ❖ The knowledge of a regression coeffient and it's t-value can suggest the extent of association or influence that an independent variable has on the dependent variable.
- ❖ The regression coeffient will reflect the impact of the omitted variable on the dependent variable.
- ❖ The model is based on collected data that represent certain environmental conditions.
- ❖ The model is limited by the methodology associated with the data collection including the sample size and measures used.