Multivariate Analysis

Bijay Lal Pradhan

Multivariate Data Analysis

Multivariate analysis comprises a set of techniques dedicated to the analysis of data sets with more than one variable simultaneously.

Accordingly, the term correlation, simple and multiple regression analysis are also Multivariate data analysis which we have covered already.

Concepts

- The Variate
- Measurement Scales
  - Nonmetric
  - Metric
- Multivariate Measurement
- Measurement Error
- Types of Techniques

Variate

The variate is a linear combination of variables with empirically determined weights.
Weights are determined to best achieve the objective of the specific multivariate technique.

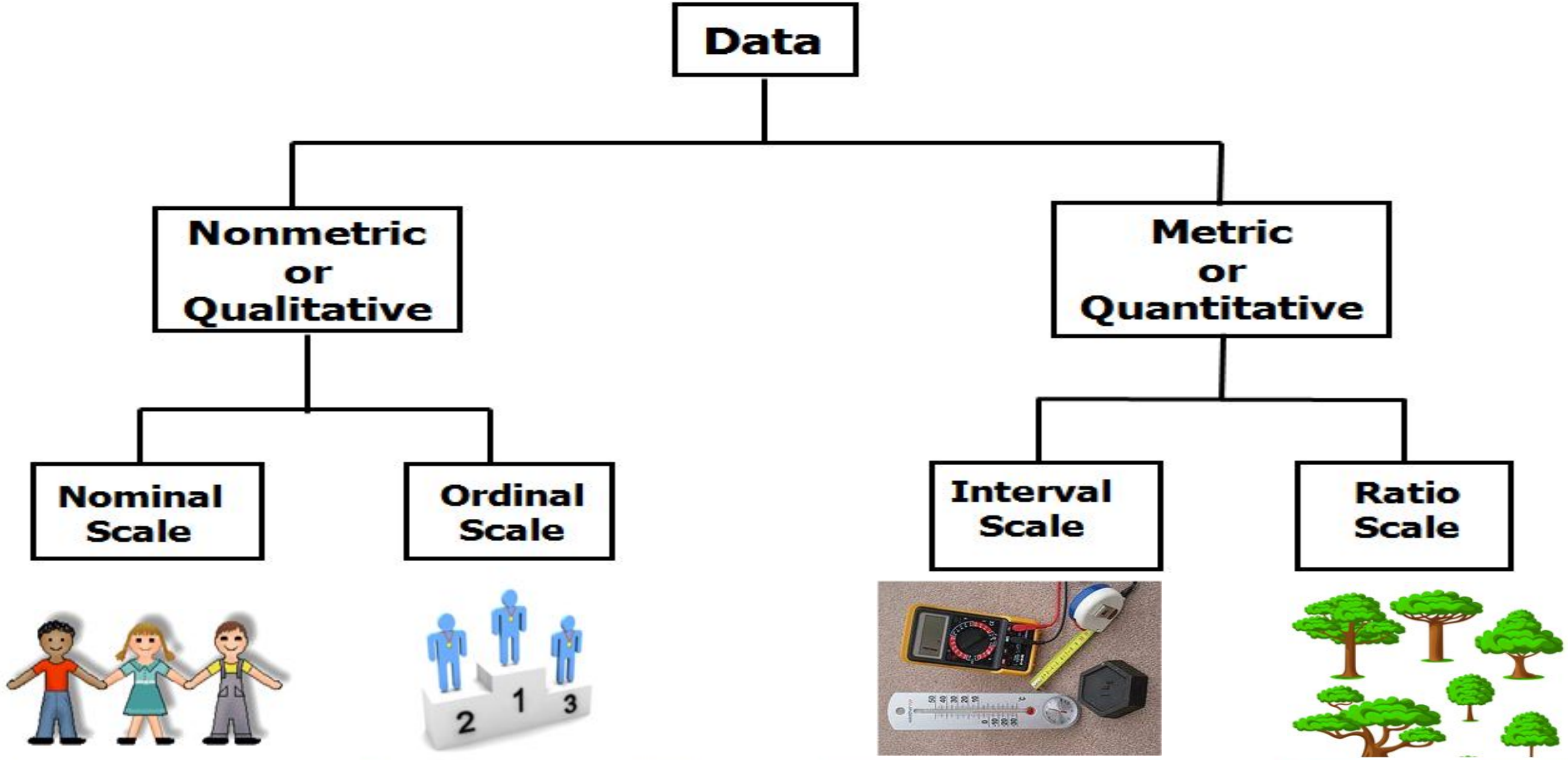
Variate equation: $(Y') = W_1 X_1 + W_2 X_2 + \ldots + W_n X_n$

Each respondent has a variate value (Y').
The Y' <u>value</u> is a <u>linear combination</u> of the entire set of variables. It is the dependent variable.

<u>Example Independent Variables</u>:

X1 = income   X2 = education
X3 = family size   X4 = ??

Types of data and measurement scale



Types of Multivariate Techniques

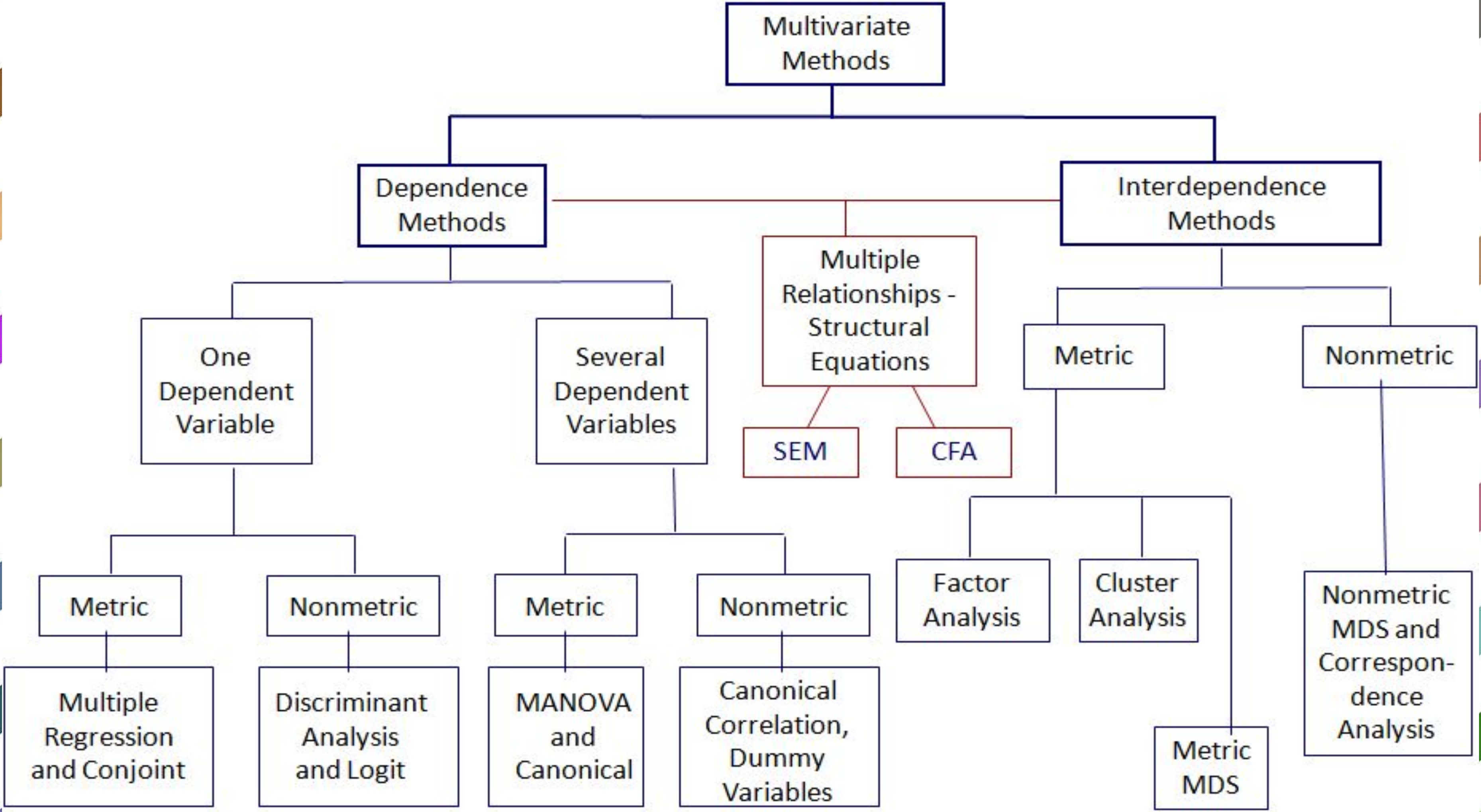
Dependence techniques: a variable or set of variables is identified as the dependent variable to be predicted or explained by other variables known as independent variables.

- Multiple Regression
- Multiple Discriminant Analysis
- Logit/Logistic Regression
- Multivariate Analysis of Variance (MANOVA) and Covariance
- Conjoint Analysis
- Canonical Correlation
- Structural Equations Modeling (SEM)

Types of Multivariate Techniques

Interdependence techniques: involve the simultaneous analysis of all variables in the set, without distinction between dependent variables and independent variables.

- Principal Components and Factor Analysis
- Cluster Analysis
- Multidimensional Scaling (perceptual mapping)
- Correspondence Analysis


Multivariate Methods
Dependence Methods
Interdependence Methods
Multiple Relationships - Structural Equations
SEM
CFA
One Dependent Variable
Several Dependent Variables
Metric
Nonmetric
Metric
Nonmetric
Metric
Nonmetric
Multiple Regression and Conjoint
Discriminant Analysis and Logit
MANOVA and Canonical
Canonical Correlation, Dummy Variables
Factor Analysis
Cluster Analysis
Metric MDS
Nonmetric MDS and Correspon-dence Analysis

Difficulties in Attitudinal Scale

- Which aspects of a situation or issue should be included when seeking to measure an attitude?
- What procedure should be adopted for combining the different aspects to obtain an overall picture?
- How can one ensure that a scale really is measuring what it is supposed to measure?

Likert Scale

Self- report technique for attitude measurement in which the subjects are asked their degree of agreement and disagreement with each of statements.

The summing score is the total attitude score, is the Likert scale.

Most people can be trusted.

Strongly agree5_ Agree4_ Undecided3_ Disagree2_ Strongly disagree1

Any body select 1 to 5

Take other similar items- to find out general attitude of indication

Add the total score- the top 25% - the most favourable attitude and bottom 25%- the least favourable attitude toward the topic being studied;

Assumption – Each statement on the scale has equal attitudinal value, importance

The Lecturer	Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
1. Knows subject well					
2. Is unenthusiastic about teaching					
3. Shows concern for students					
4. Makes unreasonable demands					
5. Has poor communication skills					
6. Knows how to teach					
7. Can explain difficult concepts in simple terms					
8. Is hard to approach					
9. Is liked by some students and not by all					
10. It is difficult to get along with					

The Lecturer	Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
1. Knows subject well	Q				Ã
2. Is unenthusiastic about teaching		Ã			Q
3. Shows concern for students			Q		Ã
4. Makes unreasonable demands		Ã			Q
5. Has poor communication skills		Ã			Q
6. Knows how to teach		Q		Ã	
7. Can explain difficult concepts in simple terms	Q	Ã			
8. Is hard to approach			QÃ		
9. Is liked by some students and not by all				Q Ã	
10. It is difficult to get along with			Ã		Q

Calculations

Statement 1 2 3 4 5 6 7 8 9 10

Respondent Q: 5+5+3+5+5+4+5+3+2+5= 42

Respondent Ã 1+2+1+2+2+2+4+3+2+3 =22

The analysis shows that overall respondent Q has a more positive attitude towards the lecturer than respondent Ã .

Strong Negative	Mild Negative	Mild Positive	Strong Positive

10 20 30 40 50

Factor Analysis

Factor Analysis (FA) is a powerful multivariate statistical technique used to identify the underlying structure (latent variables or factors) among a large set of observed variables. It is primarily used in scale development, data reduction, and understanding the dimensionality of constructs in social and behavioral sciences.

FA assumes that observed variables are influenced by fewer unobserved (latent) factors and that these factors account for the patterns of correlations among the variables. Factor analysis groups attributes that are alike.

Factor Analysis

This technique can be used to examine interrelationships among many variables (called items) and to explain these variables in terms of their common underlying and unobservable dimensions (called “factors”).

Researchers use factor analysis to reduce the information contained in several original variables into a smaller, more manageable set of variables while losing as little information as possible.

Objective of Factor Analysis

- To identify the relationships among measured variables.
- To reduce a large number of items into a smaller set of factors.
- To validate measurement instruments (e.g., survey scales).
- To develop and refine theoretical constructs.

Types of Factor Analysis

a. Exploratory Factor Analysis (EFA):

- Used when the structure of the data is not known in advance.
- Helps to explore the number and nature of latent factors.
- No prior specification of the factor structure.

b. Confirmatory Factor Analysis (CFA):

- Used to test hypotheses about the factor structure.
- Involves specifying a model based on theory & evaluating its fit to the data.
- Typically conducted within the Structural Equation Modeling (SEM).

Key Concepts and Terminology

Latent Variable (Factor): An unobservable construct that explains the observed variables.

Factor Loadings: The correlation coefficients between observed variables and the factor.

Communality (h^2): The proportion of variance in an observed variable explained by the factors.

Eigenvalues: Represent the amount of variance accounted by each factor.

Factor Rotation: A method to improve interpretability of factors (e.g., Varimax for orthogonal, Promax for oblique).

Factor Scores: Estimated values of the latent variables for each case.

Assumption of Factor Analysis

- Linearity among variables.
- Sufficient sample size (often $N \geq 100$ or $N \geq 5–10$ cases per variable).
- Multivariate normality (especially important for CFA).
- Adequate correlation between variables.

Steps in Conducting Exploratory Factor Analysis

Assess Suitability of Data:

Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (should be > 0.6).

Bartlett's Test of Sphericity (should be significant).

Extraction of Factors:

Common methods: Principal Component Analysis (PCA), Principal Axis Factoring (PAF), Maximum Likelihood.

Choose number of factors using:

Eigenvalues > 1 rule

Scree plot

Steps in Conducting Exploratory Factor Analysis

Rotation of Factors:

Orthogonal (e.g., Varimax) or oblique (e.g., Oblimin, Promax) based on theory.

Interpretation:

Examine factor loadings (commonly > 0.40 or 0.50).

Label factors based on loaded items.

Reliability and Validity:

Assess internal consistency (e.g., Cronbach's alpha).

Evaluate validity.

Example

Suppose, for example, we were interested in determining whether people who said they were anxious were also more likely to report being depressed. We made up six questions for accessing anxiety and depression:

V1	I get tense easily	V4	I often feel depressed
V2	I am often anxious	V5	I am usually happy
V3	I am generally relaxed	V6	Life is generally dull

Data on Likert Scale

Each question is answered on a five-point Likert scale ranging from 'Strongly agree' (coded 1) through Neither agree nor disagree' (coded 3) to 'Strongly disagree' (coded 5). Six observation taken

SN	V1	V2	V3	V4	V5	V6
1	5	3	2	3	4	2
2	2	1	4	3	2	4
3	4	3	2	4	1	4
4	3	5	1	2	3	2
5	2	1	5	4	2	4
6	3	2	4	3	4	1

Correlation matrix

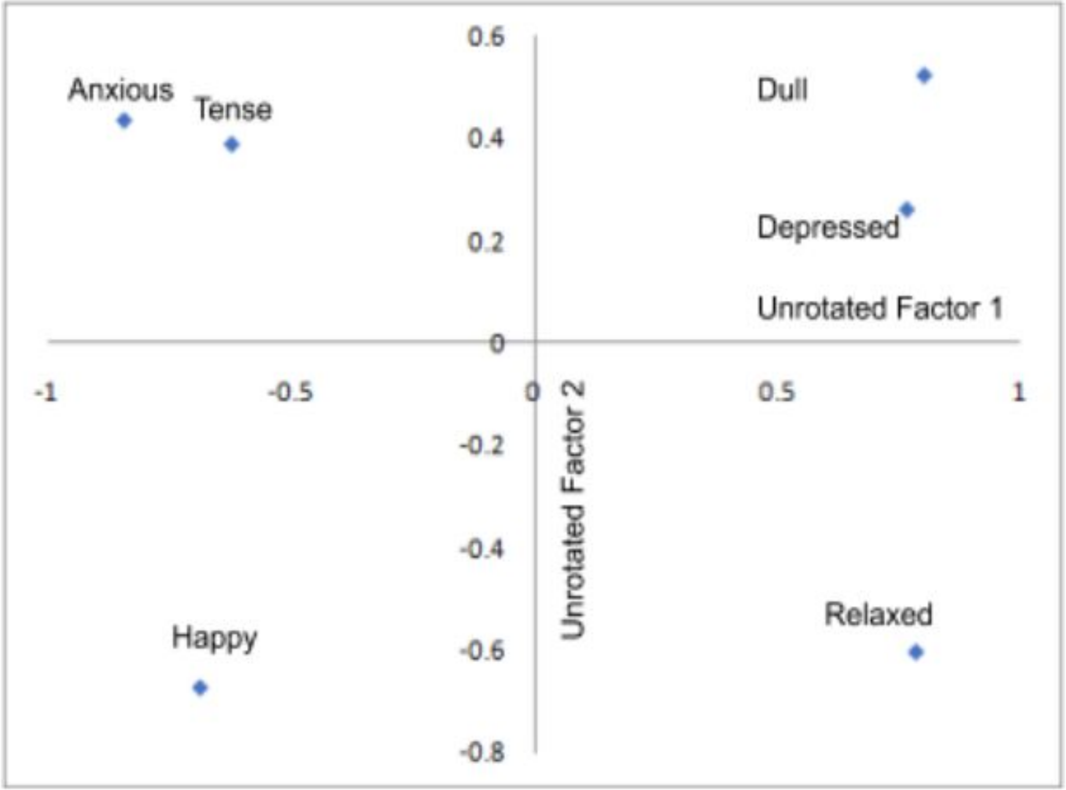
	Tense	Anxious	Relaxed	Depressed	Happy	Dull
Tense	1	.508	-.663	-.038	.330	-.365
Anxious	.508	1	-.936	-.613	.218	-.446
Relaxed	-.663	-.936	1	.514	-.107	.291
Depressed	-.038	-.613	.514	1	-.585	.633
Happy	.330	.218	-.107	-.585	1	-.911
Dull	-.365	-.446	.291	.633	-.911	1

Principal Components

	1	2	3	4	5	6
Tense	-.620	.390	0.668	-0.129	-0.021	0.00
Anxious	-.841	.437	-0.234	0.198	0.089	0.00
Relaxed	.787	-.605	0.103	0.054	0.034	0.00
Depressed	.768	.262	0.544	0.214	0.038	0.00
Happy	-.685	-.675	0.240	-0.082	0.107	0.00
Dull	.804	.525	-0.146	-0.217	0.099	0.00
Eigen Value	3.414	1.509	.887	.158	.032	0.00
% Eigen value	56.904	25.149	14.784	2.625	.537	0.00
Cumulative	56.904	82.054	96.838	99.463	100.000	100.000

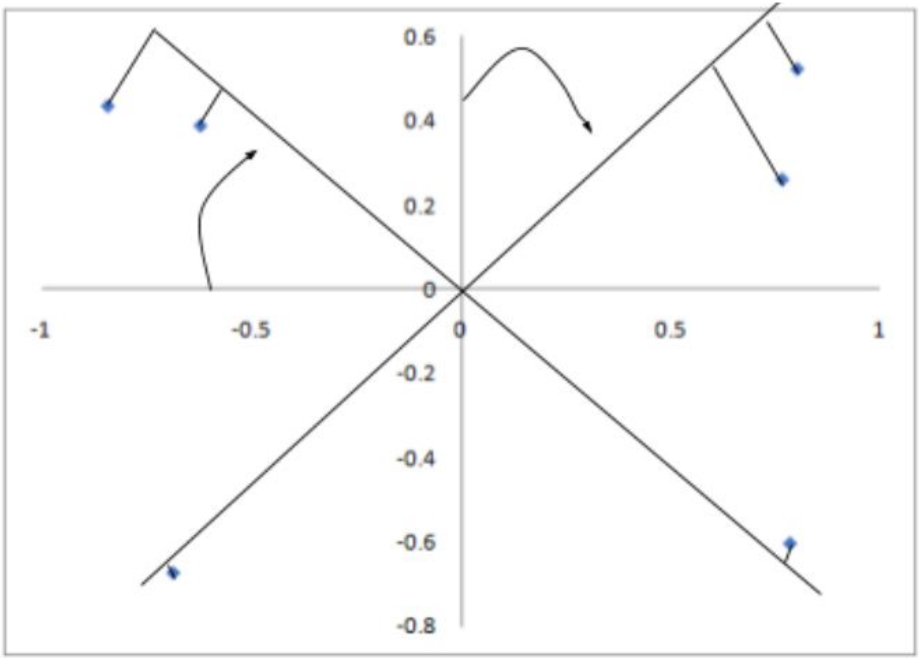
Correlation plot with two unrotated factors

Axes are called reference axes. In figure, the horizontal axis represents the first factor and vertical axis the second factor. The scale on the axes indicates the factor loading and varies from -1.0 to +1.0 The item on anxiousness (V2) for example has a load of -0.841 on the first factor and 0.437 on the second.



Factor Rotation

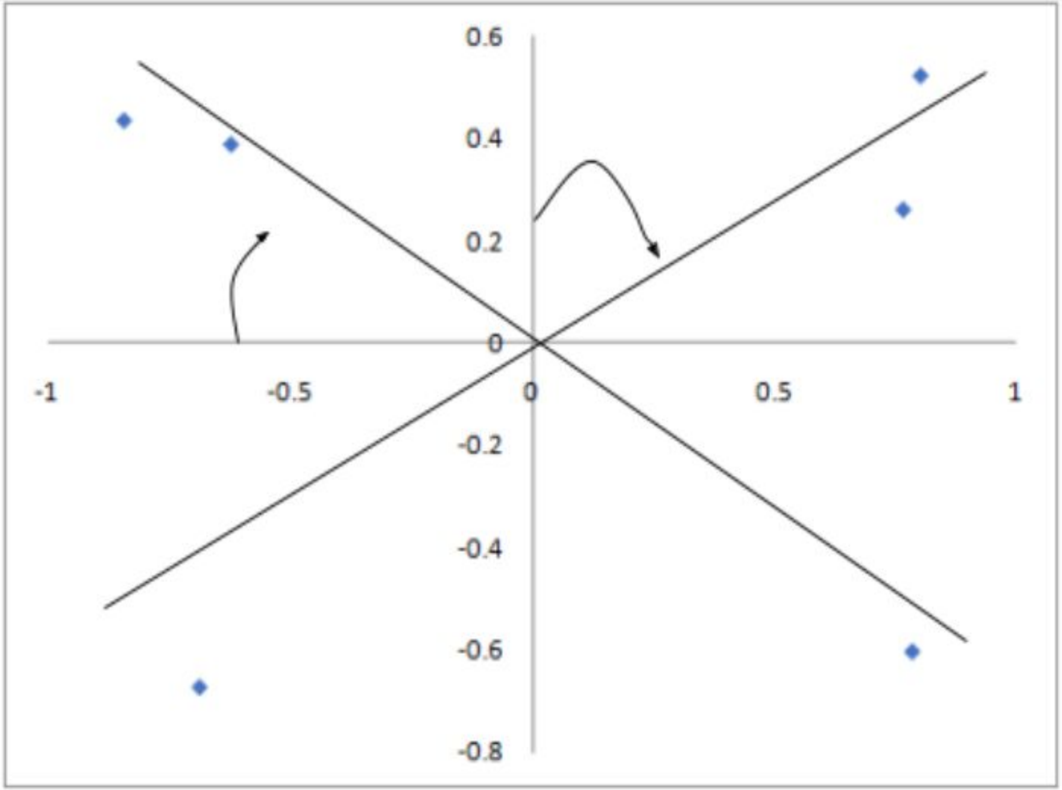
It may be apparent that the two axes do not run as close as they could to the points representing the variables. If we rotate the axes around their origin, then these two axes could be made to pass nearer to these points as shown in the figure.



Axes may be rotated in one of two ways; at a **right angle** to each other, as is the case in figure above. This is known as **orthogonal rotation**. The factors are independent with one another.

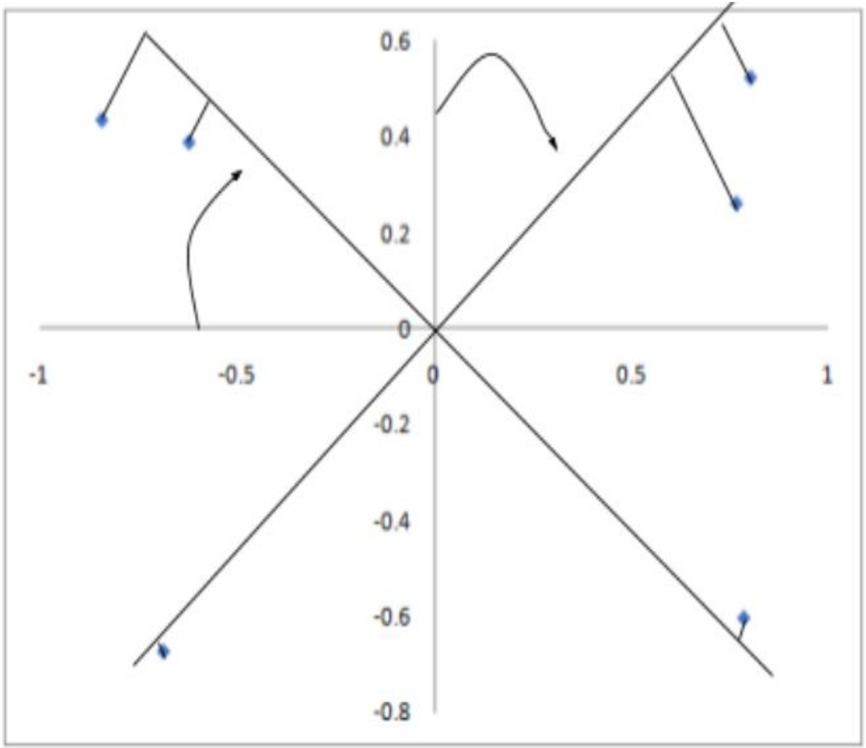
Factor Rotation

Second, the factors may be allowed to be related and to vary from being at right angles to one another, as illustrated in Figure. This is known as **oblique rotation**. The advantage of this method is that the factors may more accurately reflect what occurs in real life.



Orthogonal Rotation

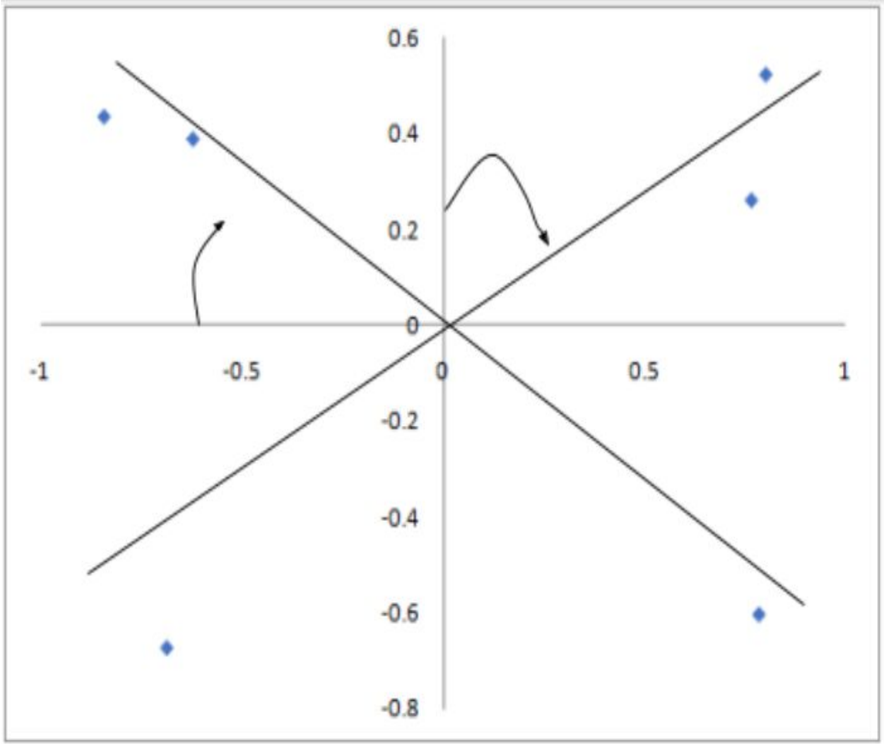
Factor	1	2
Tense	.717	-.147
Anxious	.909	-.267
Relaxed	-.987	.108
Depressed	-.372	.721
Happy	.027	-.961
Dull	-.216	.935



In SPSS it is called varimax

Oblique Rotation

	1	2
Tense	-.732	-.275
Anxious	-.943	-.428
Relaxed	.990	.286
Depressed	.499	.776
Happy	-.204	-.950
Dull	.386	.959



In SPSS it is called direct oblimin

Scale construction and naming

The results of the factor analysis are used to determine which items should be combined to form the scale for measuring a particular construct. Items loading highly on the relevant factor and not on the other factors should be used to form the scale. Generally higher scores on the scale indicate greater quantities of the variable being measured. The numerical codes for negative statement the responses should have to be reversed to reflect the factor. For instance, the numerical codes for the anxiety items V1 and V2 need to be reversed so that strong agreement with these items is re-coded as 5. The items will be added to form a factor and the proper naming of the factor should be done.

Confirmatory Factor Analysis (CFA)

CFA is used to test whether a hypothesized factor model fits the observed data. It is an essential part of Structural Equation Modeling (SEM).

- **Steps in CFA:**
  - Specify the model (e.g., how observed variables load onto latent variables).
  - Estimate the model (using ML, GLS, or other estimators).
  - Assess model fit:
    - **Chi-square (χ^2)**
    - **CFI (Comparative Fit Index) > 0.90**
    - **TLI (Tucker-Lewis Index) > 0.90**
    - **RMSEA < 0.08**
- CFA allows the testing of: Construct validity
- SPSS only does not have capacity to operate CFA.