Multivariate Analysis

- Multivariate analysis techniques are popular because they enable organizations to create knowledge and thereby improve their decision making.
- **Definition :**
  - In statistical terms,

    Multivariate analysis refers to all statistical techniques that simultaneously analyze multiple measurements on individuals or objects under investigation. Thus any simultaneous analysis of more than two variables can be consider as multivariate analysis

Multivariate Analysis

- Many multivariate techniques are extensions of univariate analysis and bivariate analysis.
- Confusion arises in term multivaraite analysis,
  - Sometimes it is simply mean that the examining relationships between or among more than two variables.
  - Some others mean that it is used for problems in which all the multiple variables are assumed to have multivariate normal distributions.
- Thus the multivariate character lies in the multiple combinations of variables and not only in the number of variables.
- Multivariate analysis will include both multivariable techniques and truly multivariate techniques.

Basic Concepts of Multivariate Analysis

- **The Variate**
  - The building block of multivariate analysis is the variate, a linear combination of variables with empirically determined weights.
  - Variables specified by the researcher.
  - Weights determined by the multivariate techniques.
- **Mathematical Representation of variate**
  - A variate of ‘n’ weighted variables (X1 to Xn) can be stated mathematically as :

$$\text{Variate value} = w_1X_1 + w_2X_2 + w_3X_3 + \cdots + w_nX_n$$

where Xn is the observed variable and Wn is the weight determined by the multivariate technique.

Introduction to Multivariate Analysis

- Multivariate data analysis refers to any statistical techniques used to analyze data that arises from more than one variable.
- This essentially models reality where each situation, product or decision involves more than a single variable. The information age has resulted in masses of data in every field.
- Despite the quantum of data available, the ability to obtain a clear picture of what is going on and make intelligent decision is a challenge.
- When available information is stored in database tables containing rows and columns, multivariate analysis can be used to process the information in a meaningful fashion.

Introduction to Multivariate Analysis

- Multivariate analysis methods typically used for :
  - Consumer and market research
  - Quality control and quality assurance across a range of industries such as food and beverage, pharmaceutics, chemicals, energy, telecommunications etc.
  - Process optimization and process control.
  - Research & development.

Measurement Scales

- The amount of information that can be provided by a variable is its type of measurement scale.
- Specifically variables are classified under two categories :-
  – Nonmetric (qualitative) scale
  – Metric (quantitative) scale
- **Qualitative (categorical) data :-**
  – Qualitative also known as categorical data cannot be measured on a numerical scale (quantified).
  – Example:
    - Gender (Male or Female)
    - Size of T-shirt (S, M, L, XL & XXL)
  – Yet these two variables differ in a sense :
    - Nominal
    - Ordinal

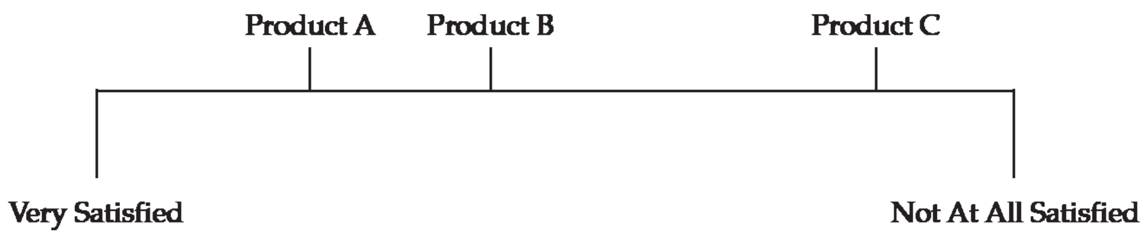
Measurement Scales

- **Nominal (purely categorical) data :**
  - Nominal variables allow for only qualitative classification. That is, they can be measured only in terms of whether the individual items belong to some distinctively different categories, but we cannot quantify or even rank order those categories.
  - Example:-
    - Gender, race, colour, city, marital status etc.
    - Marital Status
      - 1) Never Married
      - 2) Divorced
      - 3) Widowed
      - 4) Married
      - 5) Separated

Measurement Scales

- **Ordinal data :**
  - Ordinal scales are the next “higher” level of measurement precision.
  - Ordinal variables allow us to rank order the items we measure in terms of which has less and which has more of the quality represented by the variable, but still they do not allow us to say how much more.
  - Or else we can define that an ordinal scale rank-order observations. Class rank and horse race results are the examples.
  - There are two salient attributes of an ordinal scale
    - There is an underlying quantitative measure on which the observation differ.
    - The individual ignorance.
  - A typical example of an ordinal variable i.e., the socioeconomic status of families.

Measurement Scales

- For example, different levels of an individual consumer's satisfaction with several new products can be illustrated, first using an ordinal scale. The following scale shows a respondent's view of three products.
- 

- When we measure this variable with an ordinal scale, we "rank order" the products based on satisfaction level.
- We want a measure that reflects that the respondent is more satisfied with Product A and Product B and more satisfied with Product B and Product C.

Measurement Scales

- For example, Employer’s performance
  - 1) Excellent 2) Good 3) Average 4) Poor 5) Very Poor
- **Quantitative (numerical) Data:**
  - Quantitative data can be easily measured on a numerical scale : variables which can be quantified in terms of units are all quantitative.
  - Example :
    - No. Of students per class and height.
    - Yet, these two variables differ in their nature.
      - Discrete
      - Continuous
- **Discrete Data :**
  - Discrete data occur as definite and separate values. It assumes values which are countable so that these are gaps between its successive values.
  - Example :
    - When counting the number of children in a class , we use natural numbers (0,1,2, ... N)

Measurement Scales

- **Continuous Data:**
  - Continuous data occur as the whole set of real numbers or a subset of it. In other words, there are no gaps between successive so that a continuous variable assumes all the values (including all the decimals) between gives boundaries.
  - Example :
    - Temperature , Height, weight and speed
    - Continuous data can be measured on :
      - Interval Scale and
      - Ratio Scale

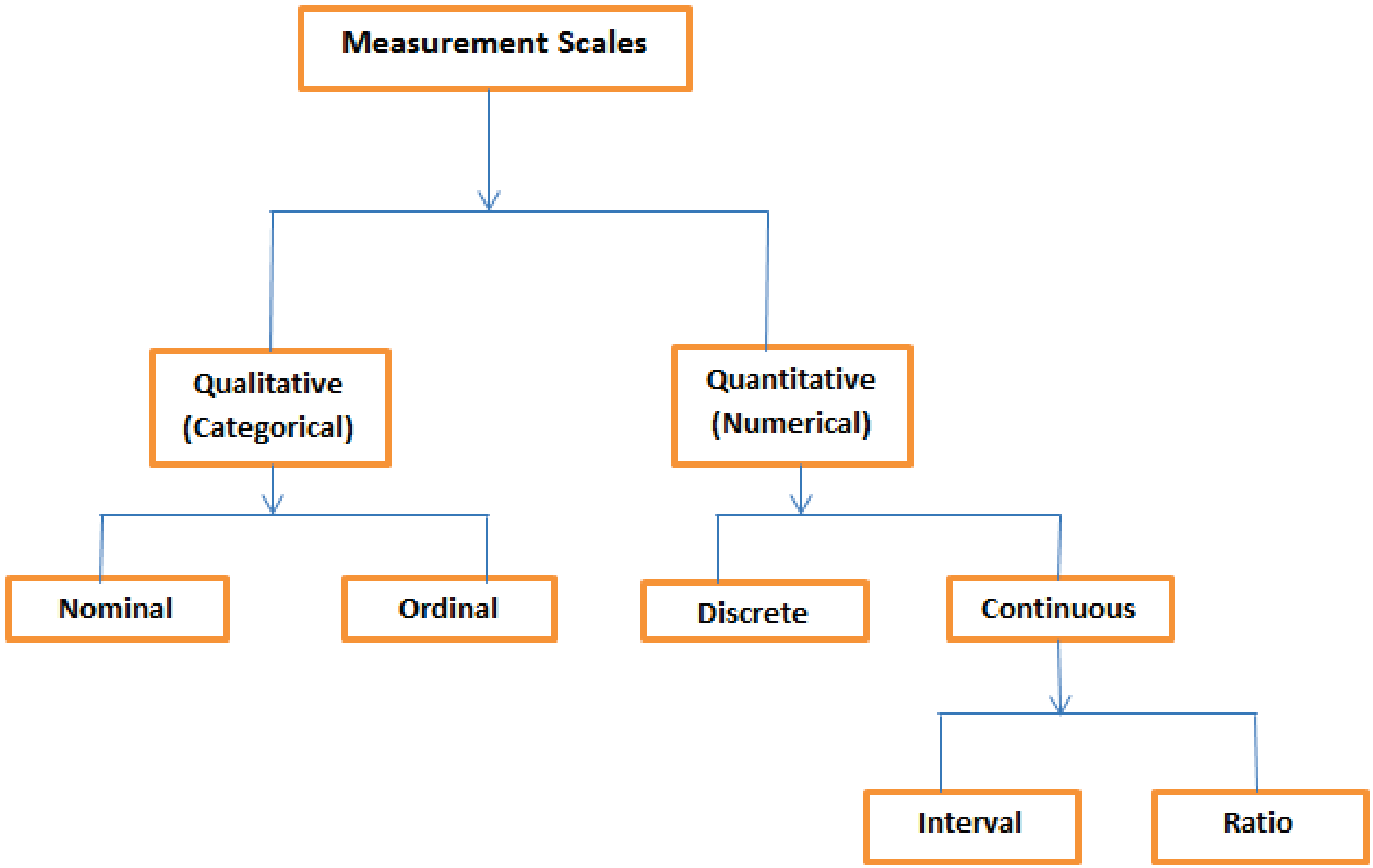
Measurement Scales

- **Interval Scale :**
  - Interval variables allow us not only to rank order the items that are measured, but also to quantify and compare the sizes of difference between them. It don't have absolute zero.
  - These two scales have constant units of measurement, so differences between any two adjacent points on part of the scale are equal.
  - The only real difference between interval and ratio scales is that interval scales use an arbitrary zero point, whereas ratio scales include an absolute zero point.
  - The most familiar interval scales are the Fahrenheit and Celsius temperature scales.

Measurement Scales

- **Ratio Scale :**
  - Ratio variables are very similar to interval variables, in addition to all the properties of interval variables, they feature on identifiable absolute zero point, thus they allow for statements such as x is two times more than y.
  - Typical examples of ratio scales are measures of time and space.
  - Example :
    - Height, since if a person is twice as tall as another, he/she will remain so, irrespective of the units used (centimeter, inches etc.,).

Measurement Scales



Measurement Scales

Provides:	Nominal	Ordinal	Interval	Ratio
The "order" of values is known		✔	✔	✔
"Counts," aka "Frequency of Distribution"	✔	✔	✔	✔
Mode	✔	✔	✔	✔
Median		✔	✔	✔
Mean			✔	✔
Can quantify the difference between each value			✔	✔
Can add or subtract values			✔	✔
Can multiple and divide values				✔
Has "true zero"				✔

Measurement Error

- Measurement Error is the degree to which the observed values are not representative of the true values.
- It arises from many sources, ranging from data entry errors to imprecision of measurement to the inability of respondents to accurately provide information.
- Thus, all variables used in multivariate techniques must be assumed to have some degree of measurement error.
- It adds noise to the observed or measured variable.
- Thus the observed value obtained represents both the “true level” and the “noise”.

Measurement Error

- Our goal is to reduce the degree of measurement error present in any measure.
- By addressing two important characteristics of a measure :
  - Validity
  - Reliability
- **Validity :**
  - Degree to which a measure accurately represents what is its supposed to be.
  - Ensuring validity starts with a thorough understanding of what is to be measured and making the measurement as “correct” and accurate as possible.
  - Accuracy does not measure validity.

Measurement Error

- **Reliability:**
  - Degree to which the observed variable measures the "true" and is "error free"; thus its the opposite of measurement error.
  - More reliable measures will show greater consistency than less reliable measures.
  - Choose the variable with the higher reliability.
- **Employing Multivariate Measurement:**
  - Also called as summated scales, for which several variables are joined in a composite measure to represent a concept.
  - The objective is to avoid the use of only a single variable to represent a concept and instead to use several variables as indicators.
  - The use of multiple indicator enables to precisely specify the desired responses.

Measurement Error

- **Employing Multivariate Measurement:**
  - It does not place a total reliance on a single response, but instead on the average response to a set of related responses.
- **The Impact of measurement error**
  - The impact of measurement error and poor reliability cannot be directly seen because they are embedded in the observed variables.
  - Increase in reliability and validity, which in turn will result in a more accuracy of the variables.
  - Poor results are not always due to measurement error, but the presence of error will make multivariate techniques less powerful.
  - Reducing measurement error, although it takes effort, time and additional resources, it may improve marginal results and strengthen the proven results as well.

Measurement Error

- **Types of statistical Error and Statistical Power :**
  - Interpreting statistical inferences requires to specify the acceptable levels of statistical error that result from using a sample (**known as sampling error**).
  - The most common approach is to specify the level of **Type I error**, also known as **alpha (α)**.
  - Type I error is the probability of rejecting the null hypothesis when it is actually true – generally referred to as a **false positive**.
  - When specifying the level of Type I error, it also determines an associated error, termed **Type II error, or beta (β)**.
  - The Type II error is the probability of not rejecting the null hypothesis when it is actually false.

Measurement Error

- **Types of statistical Error and Statistical Power :**
  - An extension of Type II error is **1 – β**, referred to as the **power** of the statistical inference test.
  - Power is the probability of correctly rejecting the null hypothesis when it should be rejected.
  - Thus, power is the probability that statistical significance will be indicated if it is present.
  - The relationship of the different error probabilities in testing for the difference in two means is shown here

		Real y	
		No Difference	*Difference*
Statistical Decision	H_0: *No Difference*	**$1 - \alpha$**	**β Type II error**
	H_a: *Difference*	**α Type I error**	**$1 - \beta$ Power**

Measurement Error

- Specifying alpha establishes the level of acceptable statistical significances, it is the level of power that gives the probability of success in finding the differences if they actually exist.
- The Type I and Type II errors are inversely related.
- Thus, Type I error becomes more restrictive as the probability of a Type II error increases.
- That is, reducing Type I errors reduces the power of the statistical test.
- Needs a balance between the level of alpha and the resulting power.

Impacts on Statistical Power

- Power is not solely a function of alpha. Power is determined by three factors:
- **Effect size :**
  - The probability of achieving statistical significance is based not only on statistical considerations, but also on the actual size of the effect.
  - Thus, the effect size helps researchers determine whether the observed relationship (difference or correlation) is meaningful.
  - For example, the effect size could be a difference in the means between two groups or the correlation between variables.
  - When examining effect sizes, a larger effect is more likely to be found than a smaller effect and is thus more likely to impact the power of the statistical test.
  - To assess the power of any statistical test, the researcher must first understand the effect being examined.
  - Effect sizes are defined in standardized terms for ease of comparison.
  - Mean differences are stated in terms of standard deviations, thus an effect size of .5 indicates that the mean difference is one-half of a standard deviation.
  - For correlations, the effect size is based on the actual correlation between the variables.

Impacts on Statistical Power

- **Alpha (α) :**
  - As alpha becomes more restrictive, power decreases.
  - Therefore, as the researcher reduces the chance of incorrectly saying an effect is significant when it is not, the probability of correctly finding an effect decreases.
  - Conventional guidelines suggest alpha levels of .05 or .01.
  - Researchers should consider the impact of a particular alpha level on the power before selecting the alpha level.
- **Sample size :**
  - At any given alpha level, increased sample sizes always produce greater power for the statistical test.
  - As sample sizes increase, researchers must decide if the power is too high.
  - By "too high" we mean that by increasing sample size, smaller and smaller effects (e.g., correlations) will be found to be statistically significant, until at very large sample sizes almost any effect is significant.
  - The researcher must always be aware that sample size can affect the statistical test either by making it insensitive (at small sample sizes) or overly sensitive (at very large sample sizes).
  - To achieve such power levels, all three factors—alpha, sample size, and effect size—must be considered simultaneously.

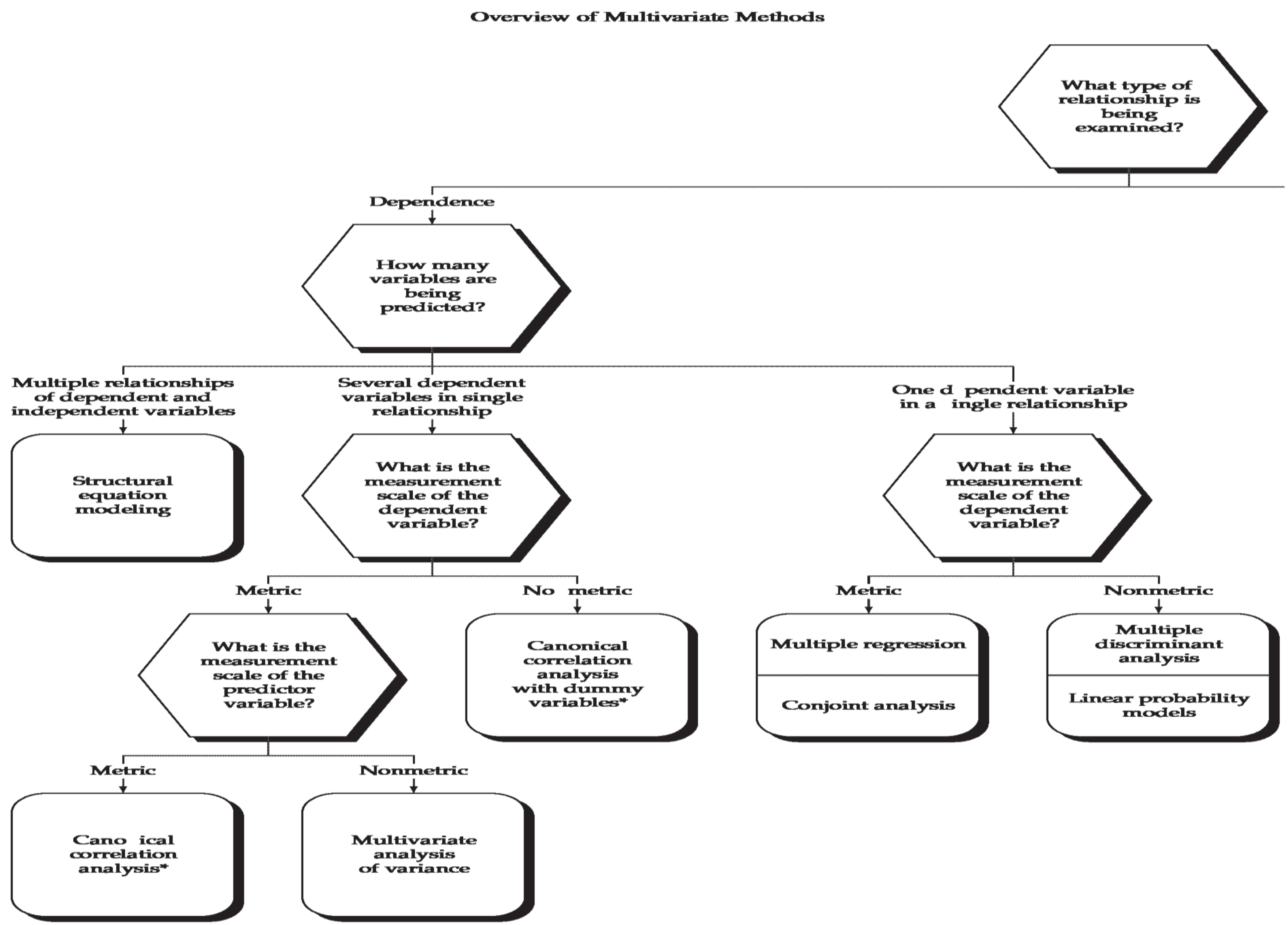
Impacts on Statistical Power

- These interrelationships can be illustrated by a simple example. The example involves testing for the difference between the mean scores of two groups.
- **Using Power with Multivariate Techniques**
- Researchers can use power analysis either in the study design or after data is collected.
- In designing research studies, the sample size and alpha level are selected to achieve the desired power.
- Power also is examined after analysis is completed to determine the actual power achieved so the results can be interpreted.

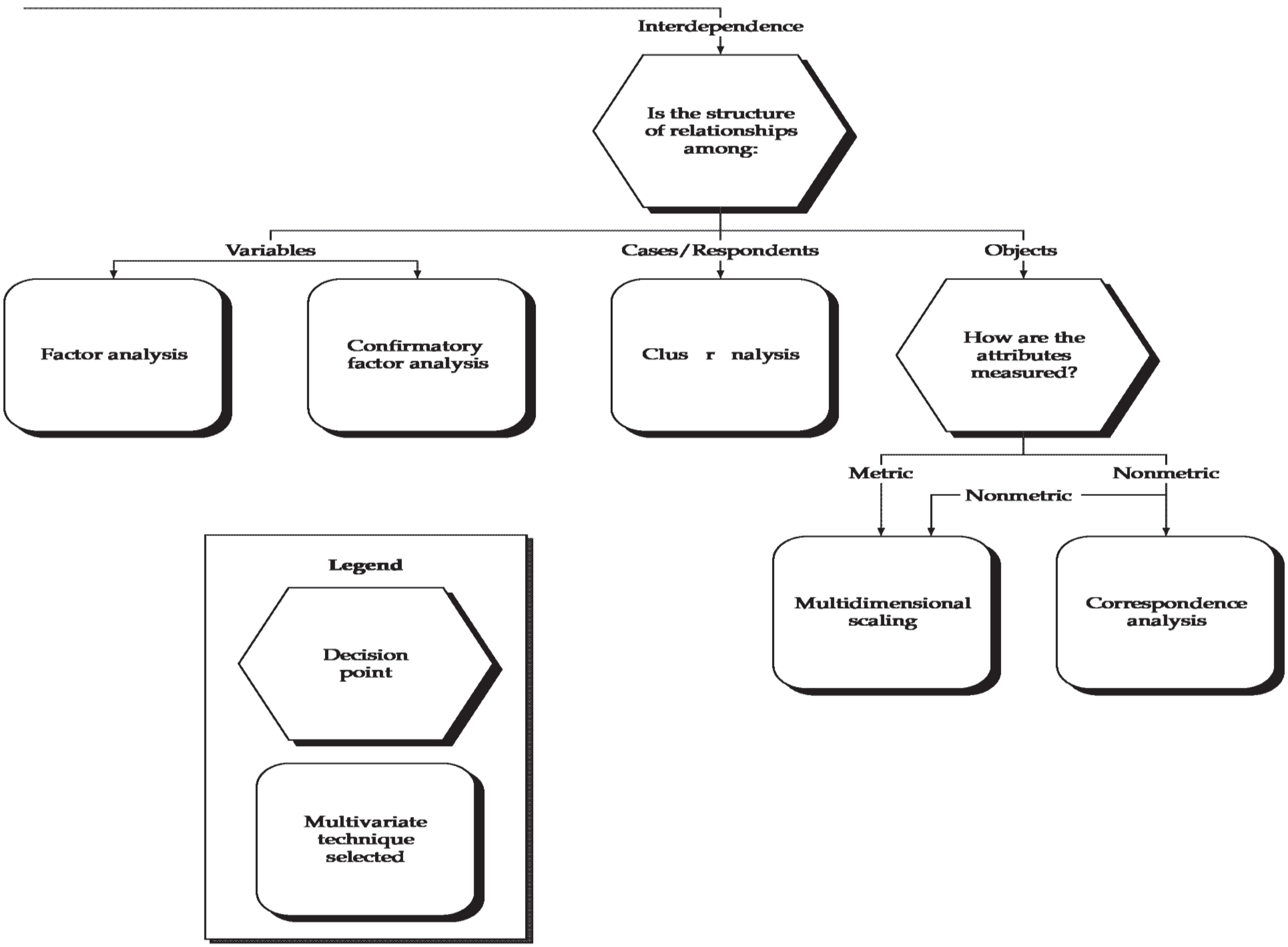
TABLE 1 Power Levels for the Comparison of Two Means: Variations by Sample Size, Significance Level, and Effect Size

	alpha (α) = .05 Effect Size (ES)		alpha (α) = .01 Effect Size (ES)	
Sample Size	**Small (.2)**	**Moderate (.5)**	**Small (.2)**	**Moderate (.5)**
20	.095	.338	.025	.144
40	.143	.598	.045	.349
60	.192	.775	.067	.549
80	.242	.882	.092	.709
100	.290	.940	.120	.823
150	.411	.990	.201	.959
200	.516	.998	.284	.992

Classification of Multivariate Techniques



Classification of Multivariate Techniques



Classification of Multivariate Techniques

- Dependence Technique :
  - Defined as one in which a variable or a set of variables is identified as the dependent variable to be predicted or explained by other variables known as independent variables.
  - Example : Multiple Regression analysis
- Inter-Dependence Technique :
  - Defined as one in which no single variable or a group of variables is defined as being dependent or independent.
  - Example : Factor Analysis

Classification of Multivariate Techniques

- Dependence Techniques :
  - Can be categorized by two characteristics
  - (i) the number of dependent variables
  - (ii) the type of measurement scale employed by the variables
  - Based on dependent variables it can be classified as
    - Single dependent variable
    - Several dependent variables
    - Several dependent/ independent variables
  - Based on type of measurement it can be classified as
    - Metric (quantitative/numerical)
    - Nonmetric (qualitative/categorical)

Classification of Multivariate Techniques

- Dependence Techniques :
  - Single dependent variable and metric
    - Use either **multiple regression analysis or conjoint analysis**
  - Single dependent variable and nonmetric
    - Use either **multiple discriminant analysis and linear probability models**
  - Several dependent variables are metric
    - If independent variables are nonmetric
      - **Use multivariate analysis of variance (MANOVA)**
    - If independent variables are metric
      - **Use canonical correlation**
  - Several dependent variables are nonmetric
    - They can be transformed through dummy variable coding (0-1) and **canonical analysis** can be used.
  - Thus multivariate techniques range from the general method of canonical analysis to specialized technique of structural equation modelling.

TABLE 2 The Relationship Between Multivariate Dependence Methods

Canonical Correlation

$$Y_1 + Y_2 + Y_3 + \cdots + Y_n \quad = \quad X_1 + X_2 + X_3 + \cdots + X_n$$

(metric, nonmetric) (metric, nonmetric)

Multivariate Analysis of Variance

$$Y_1 + Y_2 + Y_3 + \cdots + Y_n \quad = \quad X_1 + X_2 + X_3 + \cdots + X_n$$

(metric) (nonmetric)

Analysis of Variance

$$Y_1 \quad = \quad X_1 + X_2 + X_3 + \cdots + X$$

(metric) (nonmetric)

Multiple Discriminant Analysis

$$Y_1 \quad = \quad X_1 + X_2 + X \quad \cdots + X_n$$

(nonmetric) (m tric)

Multiple Regression   alysis

$$Y_1 \quad = \quad X_1 + X_2 + X_3 + \cdots + X_n$$

(metric) (metric, nonmetric)

Conjoint Analysis

$$Y_1 \quad = \quad X_1 + X_2 + X_3 + \cdots + X_n$$

(nonmetric, metric) (nonmetric)

St uctural Equation Modeling

$$Y_1 \quad = \quad X_{11} + X_{12} + X_{13} + \cdots + X_{1n}$$

$$Y_2 \quad = \quad X_{21} + X_{22} + X_{23} + \cdots + X_{2n}$$

$$Y_m \quad = \quad X_{m1} + X_{m2} + X_{m3} + \cdots + X_{mn}$$

(metric) (metric, nonmetric)

Classification of Multivariate Techniques

- Interdependence Techniques :
  - In this variables cannot be classified as either dependent or independent.
  - All the variables are analysed simultaneously to find an underlying structure to the entire set of variables or subject.
  - If the structure of the variables is to be analysed, then **factor analysis or confirmatory factor analysis** is the appropriate technique.
  - If cases or respondents are to be grouped to represent a structure, the **cluster analysis** is selected.
  - Generally factor analysis and cluster analysis are considered to be metric interdependence techniques.
  - **Non-metric data**
    - Dummy variable coding to use with special forms of factor analysis and cluster analysis.
  - If the inter dependencies of objects measured by non metric data are to be analyzed, **correspondence analysis** is appropriate technique.

Types of Multivariate Techniques

- Principal components and common factor analysis
- Multiple regression and multiple correlation
- Multiple discriminant analysis and logistic regression
- Canonical correlation analysis
- Multivariate analysis of variance and covariance
- Conjoint analysis
- Cluster analysis
- Perceptual mapping (multidimensional scaling)
- Correspondence analysis
- Structural equation modeling and confirmatory factor analysis

Principal Components and Common Factor Analysis

- Factor analysis, including both Principal components and common factor analysis is a statistical approach that can be used to analyze interrelationships among a large number of variables.
- To explain these variables in terms of their common underlying dimensions called **factors**.
- The objective is to find a way of condensing the information contained in a number of original variables into a small set of variates (**factors**) with a minimum loss of information.
- Example :
  - Understand the relationships between customers ratings of fast-food restaurant.
  - Six variables : food taste, food temperature, freshness, waiting time, cleanliness and friendliness of employees
  - By analyzing customer responses,
    - food taste, food temperature and freshness – **food quality (one factor)**
    - waiting time, cleanliness and friendliness of employees – **service quality (one factor)**

Multiple Regression

$$\underset{\text{(metric)}}{Y_1} = \underset{\text{(metric, nonmetric)}}{X_1 + X_2 + X_3 + \cdots + X_n}$$

- It is the appropriate technique, if the problem involves a single metric dependent variable to be related to two or more metric independent variables.
- The objective is to predict the changes in the dependent variable in response to changes in the independent variables.
- Its is achieved through the statistical rule of least squares.
- Example :
  - Monthly expenditures on dining out **(dependent variable)**
  - Family's income, its size and the age of head of household **(independent variables)**
  - Company's sales from its expenditure for advertising, the number of salespeople and the number of product stores.

Multiple Discriminant Analysis (MDA) and Logistic Regression

$$\underset{\text{(nonmetric)}}{Y_1} = \underset{\text{(m tric)}}{X_1 + X_2 + X \cdots + X_n}$$

- MDA is the appropriate multivariate technique if the single dependent variable is dichotomous or multi-chotomous and also nonmetric.
- Multiple regression the independent variables are assumed to be metric.
- It is mainly applicable for the total sample can be divide into groups based on a nonmetric dependent variables characterizing into several known classes.
- The primary objectives are to understand group differences and to predict the likelihood that an entity will belong to a particular class based on several metric independent variables.
- Example :
  - Its is used to distinguishing applications include
  - Heavy product users from light users
  - Males from females
  - National-brand buyers from private-label buyers
  - Good credit risks from poor credit risks.

Multiple Discriminant Analysis and Logistic Regression

- Logistic regression models, are often referred to as logit analysis, are a combination of multiple regression and multiple discriminant analysis,
- It is similar to multiple regression analysis, but a main difference is it uses dependent variable as nonmetric as in discriminant analysis.
- The nonmetric scale of dependent variable requires differences in the estimation method and assumptions about the type of underlying distribution, otherwise its quite similar to multiple regression.
- Logistic models are distinguished from discriminant analysis that they include all types of independent variables (metric and nonmetric) and don not require the assumption of multivariate normality.
- But in many example, with more than two levels of dependent variable, discriminant analysis is the more appropriate technique.
- Example :
  - Financial advisors were trying to selecting emerging firms for start-up investment.
  - Two classes :
    - Successful over a period of five years
    - Unsuccessful after five years
  - For each firm they had financial and managerial data.

Canonical Correlation

$$\underset{\text{(metric, nonmetric)}}{Y_1 + Y_2 + Y_3 + \cdots + Y_n} = \underset{\text{(metric, nonmetric)}}{X_1 + X_2 + X_3 + \cdots + X_n}$$

- Canonical correlation can be viewed as a logical extension of multiple regression analysis.
- Canonical analysis the objective is to correlate simultaneously several metric dependent variables and several metric independent variables.
- Multiple regression involves single dependent variable, whereas canonical correlation involves multiple dependent variables.
- The underlying principle is to develop a linear combination of each set of variables in a manner that maximizes the correlation between the two sets.
- The procedure involves obtaining a set of weights for the dependent and independent variables that provides the maximum simple correlation between the set of independent variables and dependent variables.
- Example :
  - A company conducts a study that collects information on its service quality based on answers to 50 metrically measured questions.
  - Canonical correlation could be used to compare the perceptions of the world-class companies on the 50 questions with the perceptions of the company.

Multivariate Analysis of Variance and Covariance

$$\underset{\text{(metric)}}{Y_1 + Y_2 + Y_3 + \cdots + Y_n} = \underset{\text{(nonmetric)}}{X_1 + X_2 + X_3 + \cdots + X_n}$$

- Multivariate analysis of variance (MANOVA) is a statistical technique that can be used to simultaneously explore the relationship between several categorical independent variables and two or more metric dependent variables.
- It is the extension of univariate analysis of variance (ANOVA).
- Multivariate analysis of covariance (MANCOVA) can be used in conjunction with MANOVA to remove the effect of any uncontrolled metric independent variables (Covariates) on the dependent variables.
- MANOVA is useful for an experimental situation to test hypotheses concerning he variance in group responses on two or more metric dependent variables.
- Example :
  - A company wants to know if a humorous AD will be more effective with its customers than a nonhumorous AD.
  - It could ask its ad agency to develop two ads – one humorous and one nonhumorous and then show group of customers the two ads.
  - MANOVA would be the technique to determine the extent of any statistical differences between the perception of customers who saw the humorous AD versus those who saw the nonhumorous.

Conjoint Analysis

$$\underset{\text{(nonmetric, metric)}}{Y_1} = \underset{\text{(nonmetric)}}{X_1 + X_2 + X_3 + \cdots + X_n}$$

- Conjoint analysis is an emerging technique that bring new sophistication to the evaluation of objects, such as new products, services or ideas.
- It is mainly used in the direct application of new product or service development.
- It is able to assess the importance of attributes as well as the levels of each attribute while consumers evaluate only a few product profiles, which are combinations of product levels.
- Example :
  - A product company has three attributes (price, quality, and color), each at three possible levels (red, yellow and blue).
  - Instead of having to evaluate all 27 (3X3X3) possible combinations, a subset of (9 or more) can be evaluated for their attractiveness of customers.
  - It provides not only the importance of each attribute but also the importance of each level.

Cluster Analysis

- Cluster analysis is an analytical technique for developing meaningful subgroups of individuals or objects.
- Specifically, the objective is to classify a sample of entities (individuals or objects) into a small number of mutually exclusive groups based on the similarities among the entities.
- In cluster analysis, unlike discriminant analysis, the groups are not predefined. Instead, the technique is used to identify the groups.
- Cluster analysis usually involves at least three steps.
  - The first is the measurement of some form of similarity or association among the entities to determine how many groups really exist in the sample.
  - The second step is the actual clustering process, whereby entities are partitioned into groups (clusters).
  - The final step is to profile the persons or variables to determine their composition.
- Many times this profiling may be accomplished by applying discriminant analysis to the groups identified by the cluster technique.
- Example :
  - assume a restaurant owner wants to know whether customers are patronizing the restaurant for different reasons.
  - Data could be collected on perceptions of pricing, food quality, and so forth.
  - Two clusters can be formed based on
    - highly motivated by low prices versus
    - those who are much less motivated to come to the restaurant based on price considerations.

Perceptual Mapping

- In perceptual mapping (also known as multidimensional scaling), the objective is to transform consumer judgments of similarity or preference (e.g., preference for stores or brands) into distances represented in multidimensional space.
- If objects A and B are judged by respondents as being the most similar compared with all other possible pairs of objects, perceptual mapping techniques will position objects A and B in such a way that the distance between them in multidimensional space is smaller than the distance between any other pairs of objects.
- The resulting perceptual maps show the relative positioning of all objects, but additional analyses are needed to describe or assess which attributes predict the position of each object.
- Example :
  – assume an owner of a Burger King franchise wants to know whether the strongest competitor is McDonald's or Wendy's.
  – A sample of customers is given a survey and asked to rate the pairs of restaurants from most similar to least similar.
  – The results show that the Burger King is most similar to Wendy's, so the owners know that the strongest competitor is the Wendy's restaurant because it is thought to be the most similar.
  – Follow-up analysis can identify what attributes influence perceptions of similarity or dissimilarity.

Correspondence Analysis

- Correspondence analysis is a recently developed interdependence technique that facilitates the perceptual mapping of objects (e.g., products, persons) on a set of nonmetric attributes.
- Researchers are constantly faced with the need to “quantify the qualitative data” found in nominal variables.
- Correspondence analysis differs from the interdependence techniques discussed earlier in its ability to accommodate both nonmetric data and nonlinear relationships.
- In its most basic form, correspondence analysis employs a contingency table, which is the cross-tabulation of two categorical variables.
- It then transforms the nonmetric data to a metric level and performs dimensional reduction (similar to factor analysis) and perceptual mapping.
- Correspondence analysis provides a multivariate representation of interdependence for nonmetric data that is not possible with other methods.
- Example :
  - respondents’ brand preferences can be cross-tabulated on demographic variables (e.g., gender, income categories, occupation)
  - by indicating how many people preferring each brand fall into each category of the demographic variables.
  - Brands perceived as similar are located close to one another.

Structural Equation Modeling

$$\begin{array}{ccc} Y_1 & = & X_{11} + X_{12} + X_{13} + \cdots + X_{1n} \\ Y_2 & = & X_{21} + X_{22} + X_{23} + \cdots + X_{2n} \\ Y_m & = & X_{m1} + X_{m2} + X_{m3} + \cdots + X_{mn} \\ \text{(metric)} & & \text{(metric, nonmetric)} \end{array}$$

- Structural equation modeling (SEM) is a technique that allows separate relationships for each of a set of dependent variables.
- In its simplest sense, structural equation modeling provides the appropriate and most efficient estimation technique for a series of separate multiple regression equations estimated simultaneously.
- It is characterized by two basic components:
  - (1) the structural model and
  - (2) the measurement model.
- The structural model is the path model, which relates independent to dependent variables.
- In such situations, theory, prior experience, or other guidelines enable the researcher to distinguish which independent variables predict each dependent variable.
- Models discussed previously that accommodate multiple dependent variables—multivariate analysis of variance and canonical correlation—are not applicable in this situation because they allow only a single relationship between dependent and independent variables.
- The measurement model enables the researcher to use several variables (indicators) for a single independent or dependent variable.

Structural Equation Modeling

- Example :
  - A study by management consultants identified several factors that affect worker satisfaction: supervisor support, work environment, and job performance.
  - In addition to this relationship, they noted a separate relationship wherein supervisor support and work environment were unique predictors of job performance.
  - Hence, they had two separate, but interrelated relationships.
- SEM provides a means of not only assessing each of the relationships simultaneously rather than in separate analyses, but also incorporating the multi-item scales in the analysis to account for measurement error associated with each of the scales.

A Structured Approach to Multivariate Model Building

- Numerous multivariate techniques are available and the large set of issues involved in their application.
- Successful completion of a multivariate analysis involves the selection of a correct method.
- Issues ranging from problem definition to a critical diagnosis of the results must be addressed.
- To aid the researcher or user in applying multivariate methods, a six-step approach to multivariate analysis is presented.
- The intent is not to provide a set of procedures to follow but to provide a series of guidelines that emphasize a model-building approach.
- This six-step model-building process provides a framework for developing, interpreting, and validating any multivariate analysis.

A Structured Approach to Multivariate Model Building

- **Stage 1 : Define the Research Problem, Objectives, and Multivariate technique to be used**
  - The starting point for any multivariate analysis is to define the research problem and analysis objectives in conceptual terms before specifying ant variables or measures.
  - The role of conceptual model development cannot be overstated.
  - First view the problem in conceptual terms by defining the concepts and identifying the fundamental relationships to be investigated.
  - It should not be complex and detailed. It can be a just simple representation of the relationships to be studied.
  - For a dependence problem specify both the dependent and independent concepts. For an interdependent application, the dimensions of structure or similarity should be specified.
  - This minimizes the chance that relevant concepts will be omitted in the effort to develop measures and to define the specifics of the research design.

A Structured Approach to Multivariate Model Building

- **Stage 1 : Define the Research Problem, Objectives, and Multivariate technique to be used**
  - With the objective and conceptual model specified, the researcher has only to choose the appropriate multivariate technique based on the measurement characteristics of the dependent and independent variables.
  - Variables for each concept are specified prior to the study in its design, but may be respecified or transformed after the data have been collected.
- **Stage 2 : Develop the Analysis Plan**
  - With the concept model established and multivariate technique selected, attention turns to the implementation issues.
  - The issues include general considerations such as minimum or desired sample sizes and a required type of variables (metric versus nonmetric) and estimation methods.

A Structured Approach to Multivariate Model Building

- **Stage 3 : Evaluate the Assumptions underlying the Multivariate Technique**
  - Once the data collected, the first task is not to estimate the multivariate model but to evaluate its underlying assumptions, both statistical and conceptual, that substantially affect their ability to represent multivariate relationships.
  - For the techniques based on statistical inference, the assumptions of multivariate normality, linearity, independence of the error terms, and equality if variances must all be met.
  - Before any model estimation is attempted, must ensure that both statistical and conceptual assumptions are met.

A Structured Approach to Multivariate Model Building

- **Stage 4 : Estimate the Multivariate Model and Assess Overall Model Fit**
  - With the assumptions satisfied, the analysis proceeds to the actual estimation of the multivariate model and an assessment of overall model fit.
  - In the estimation process, choose options to meet specific characteristics of the data (e.g. Use of Covariances in MANOVA) or maximize the fit to the data (e.g. rotation of factors or discriminant functions).
  - After the model is estimated, the overall model fit is evaluated to achieve,
    - Acceptable levels of statistical criteria (level of significance)
    - Identifies the proposed relationships
    - Achieves practical significance
  - Model will be respecified to achieve better levels of overall fit.
  - Determine whether the results affected by single or small set observations will indicate that the results may be unstable or not generalizable.
  - III-fitting observations are identified as outliers, influential observations or other disparate results.

A Structured Approach to Multivariate Model Building

- **Stage 5 : Interpret the Variates**
  - With an acceptable level of model fit, interpreting the variates reveals the nature of the multivariate relationship.
  - The interpretation effects for individual variables is made by examining the estimated coefficeints (weights) for each variable in the variate.
  - For multiple variates, represent underlying dimensions of comparison or association.
  - The interpretation may lead to, respecification of the variables and model formulation. Model is re-estiamated and then interpreted again.
  - The objective is to identify the empirical evidence of multivariate relationships in the sample data that can be generalized to the total population.

A Structured Approach to Multivariate Model Building

- **Stage 6 : Validate the Multivariate Model**
  - Before accepting the results, must subject them to one final set of diagnostic analyses that assess the degree of generalizability of the results by the available validation methods.
  - The attempts to validate the model are directed toward demonstrating the generalizability of the results to the total population.
  - These diagnostic analyses add little to the interpretation of the results but can be viewed as "insurance" that the results are the most descriptive of the data, yet generalizable to the population.
- **A decision Flowchart**
  - For each multivariate technique, the six-step approach to multivariate model building will be given in a decision flowchart partitioned into two sections.
  - The first section **(Stages 1 through 3)** : deals with the issues addressed while preparing for actual model estimation (i.e., research objectives, research design considerations and testing for assumptions).
  - The second section of the decision flowchart **(stages 4 through 6)** deals with the issues pertaining to model estimation, interpretation and validation).

Applications of Multivariate Techniques

- The published applications of multivariate methods have increased tremendously in recent years.
- In order to give some indication of the usefulness of multivariate techniques, following are the short descriptions of the results of studies from several disciplines.
- These descriptions are organized according to the categories of objectives of multivariate techniques.

Applications of Multivariate Techniques

- A taxonomy of multivariate statistical analyses shows that most techniques fall into one of the following categories:

1. Data reduction or structural simplification.
2. Sorting and grouping.
3. Investigation of the dependence among variables.
4. Prediction.
5. Hypothesis construction and testing.

Applications of Multivariate Techniques

- **Data Reduction or Simplification :**
  - Using data on several variables related to cancer patient responses to radio therapy, a simple measure of patient response to radiotherapy was constructed.
  - Track records from many nations were used to develop an index of performance for both male and female athletes.
  - Multispectral image data collected by a high-altitude scanner were reduced to a form that could be viewed as images (pictures) of a shoreline in two dimensions.
  - Data on several variables relating to yield and protein content were used to create an index to select parents of subsequent generations of improved bean plants.

Applications of Multivariate Techniques

- **Sorting and Grouping :**
  - Data on several variables related to computer use were employed to create clusters of categories of computer jobs that allow a better determination of existing (or planned) computer utilization.
  - Measurements of several psychological variables were used to develop a screening procedure that discriminates alcoholics from non-alcoholics.
  - Data related to responses to visual stimuli were used to develop a rule for separating people suffering from a multiple-sclerosis-caused visual pathology from those not suffering from the disease.
  - The U.S Internal Revenue Service uses data collected from tax returns to sort taxpayers into two groups: those that will be audited and those that will not.

Applications of Multivariate Techniques

- **Investigation of the dependence among variables :**
  - Data on several variables were used to identify factors that were responsible for client success in hiring external consultants.
  - Measurements of variables related to innovation, on the one hand, and variables related to the business environment and business organization , on the other hand, were used to discover why some firms are product innovators and some firms are not.
  - Data on variables representing the outcomes of the 10 decathlon events in the Olympics were used to determine the physical factors responsible for the success in the decathlon.
  - The associations between measures of risk-taking propensity and measures of socioeconomic characteristics for top-level business executives were used to assess the relation between risk-taking behavior and performance.

Applications of Multivariate Techniques

- **Prediction :**
  - The associations between test scores and several high school performance variables and several college performance variables were used to develop predictors of success in college.
  - Data on several variables related to the size distributions of sediments were used to develop rules for predicting different depositional environments.
  - Measurements on several accounting and financial variables were used to develop a method for identifying potentially insolvent property-liability insurers.
  - Data on several variables for chickweed plants were used to develop a method for predicting the species of a new plant.

Applications of Multivariate Techniques

- **Hypotheses Testing :**
  - Several pollution-related variables were measured to determine whether levels for a large metropolitan area were roughly constant throughout the week, or whether there was a noticeable difference between weekdays and weekends.
  - Experimental data on several variables were used to see whether the nature of the instructions makes any difference in perceived risks, as quantified by test scores.
  - Data on many variables were used to investigate the differences in structure of American occupations to determine the support for one of two competing sociological theories.
  - Data on several variables were used to determine whether different types of firms in newly industrialized countries exhibited different patterns of innovation.

  The preceding descriptions offer glimpses into the use of multivariate methods in widely diverse fields.

UNIT - II

Factor Analysis

Prepared BY: A. Shobana Devi.
SRM, IT Dept.

Factor Analysis

- **WHAT IS FACTOR ANALYSIS?**
- Factor analysis is an interdependence technique whose primary purpose is to define the underlying structure among the variables in the analysis.
- Variables play a key role in any multivariate analysis.
- we must have a set of variables upon which to form relationships (e.g., What variables best predict sales or success/ failure?).
- Variables are the building blocks of relationships.
- As we employ multivariate techniques, by their very nature, the number of variables increases.
- Univariate techniques are limited to a single variable, but multivariate techniques can have tens, hundreds, or even thousands of variables.

Factor Analysis

- As we add more and more variables, more and more overlap (i.e., correlation) is likely among the variables.
- As the variables become correlated, we needs ways in which to manage these variables—
  - grouping highly correlated variables together, labeling or naming the groups, and perhaps even creating a new composite measure that can represent each group of variables.
- Factor analysis is introduced as first multivariate technique because it can play a unique role in the application of other multivariate techniques.
- **What is a Factor?**
  - A factor is a linear combination of variables. It is a construct that is not directly observed but that needs to be inferred from the input variables.
- **Definition :**
  - Factor analysis provides the tools for analyzing the structure of the interrelationships (correlations) among a large number of variables (e.g., test scores, test items, questionnaire responses) by defining sets of variables that are highly interrelated, known as factors. These groups of variables (factors), which are by definition highly inter correlated, are assumed to represent dimensions within the data.

Factor Analysis

- Factor analysis presents several ways of representing these groups of variables for use in other multivariate techniques.
- **Factor Analysis is commonly used in :**
  - Data Reduction
  - Scale Development
  - The evaluation of psychometric quality of measure and
  - The Assessment of the dimensionality of a set of variables.
- **Two Types of Factor Analysis :**
  - Exploratory Factor Analysis
  - Confirmatory Factor Analysis
- **Exploratory Factor analysis :**
  - It is useful in searching for structure among a set of variables or as a data reduction method. In this perspective, factor analytic techniques "take what the data give you" and do not set any a priori constraints on the estimation of components or the number of components to be extracted.

Factor Analysis

- **Confirmatory Factor analysis :**
  - When the researcher has preconceived thoughts on the actual structure of the data, based on theoretical support or prior research.
  - For example, the researcher may wish to test hypotheses involving issues such as which variables should be grouped together on a factor or the precise number of factors.
  - This is to assess the degree to which the data meet the expected structure.
  - In this chapter, we view factor analytic techniques principally from an exploratory or non-confirmatory viewpoint.

Factor Analysis

- **A HYPOTHETICAL EXAMPLE OF FACTOR ANALYSIS :**
  - Assume that through qualitative research a retail firm identified 80 different characteristics of retail stores and their service that consumers mentioned as affecting their patronage choice among stores.
  - To identify these broader dimensions, the retailer could commission a survey asking for consumer evaluations on each of the 80 specific items.
  - Factor analysis would then be used to identify the broader underlying evaluative dimensions.
  - Specific items that correlate highly are assumed to be a member of that broader dimension.
  - These dimensions become composites of specific variables, which in turn allow the dimensions to be interpreted and described.
  - In our example, the factor analysis might identify such dimensions as product assortment, product quality, prices, store personnel, service, and store atmosphere as the broader evaluative dimensions used by the respondents.
  - Each of these dimensions contains specific items that are a facet of the broader evaluative dimension. From these findings, the retailer may then use the dimensions (factors) to define broad areas for planning and action.

Factor Analysis

Exploratory Factor Analysis

PART 1: ORIGINAL CORRELATION MATRIX

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
V_1 Price Level	1.000								
V_2 Store Personnel	.427	1.000							
V_3 Return Policy	.302	.771	1.000						
V_4 Product Availability	.470	.497	.427	1.000					
V_5 Product Quality	.765	.406	.307	.427	1.000				
V_6 Assortment Depth	.281	.445	.423	.713	.325	1.000			
V_7 Assortment Width	.345	.490	.471	.719	.378	.724	1.000		
V_8 In-store Service	.242	.719	.733	.428	.240	.311	.435	1.000	
V_9 Store Atmosphere	.372	.737	.774	.479	.326	.429	.466	.710	.000

PART 2: CORRELATION MATRIX OF VARIABLES AFTER GROUPING ACCORDING TO FACTOR ANALYSIS

	V_3	V_8	V_9	V_2	V_6	V_7	V_4	V_1	V_5
V_3 Return Policy	1.000								
V_8 In-store Service	.773	1.000							
V_9 Store Atmosphere	.771	.710	1.000						
V_2 Store Personnel	.771	.719	.737	1.000					
V_6 Assortment Depth	.423	.311	.429	.445	1.000				
V_7 Assortment Width	.471	.435	.466	.490	.724	1.000			
V_4 Product Availability	.427	.428	.479	.497	.713	.719	1.000		
V_1 Price Level	.302	.242	.372	.427	. 81	.354	.470	1.000	
V_5 Product Quality	.307	.240	.326	.406	325	.378	.427	.765	1.000

Factor Analysis

- **A ILLUSTRATIVE EXAMPLE OF FACTOR ANALYSIS :**
  - An illustrative example of a simple application of factor analysis is shown in Figure 1, which represents the correlation matrix for nine store image elements.
  - Included in this set are measures of the product offering, store personnel, price levels, and in-store service and experiences.
  - The question a researcher may wish to address is:
    - Are all of these elements separate in their evaluative properties or do they group into some more general areas of evaluation?
    - For example, do all of the product elements group together?
    - Where does price level fit, or is it separate?
    - How do the in-store features (e.g., store personnel, service, and atmosphere) relate to one another?

Factor Analysis

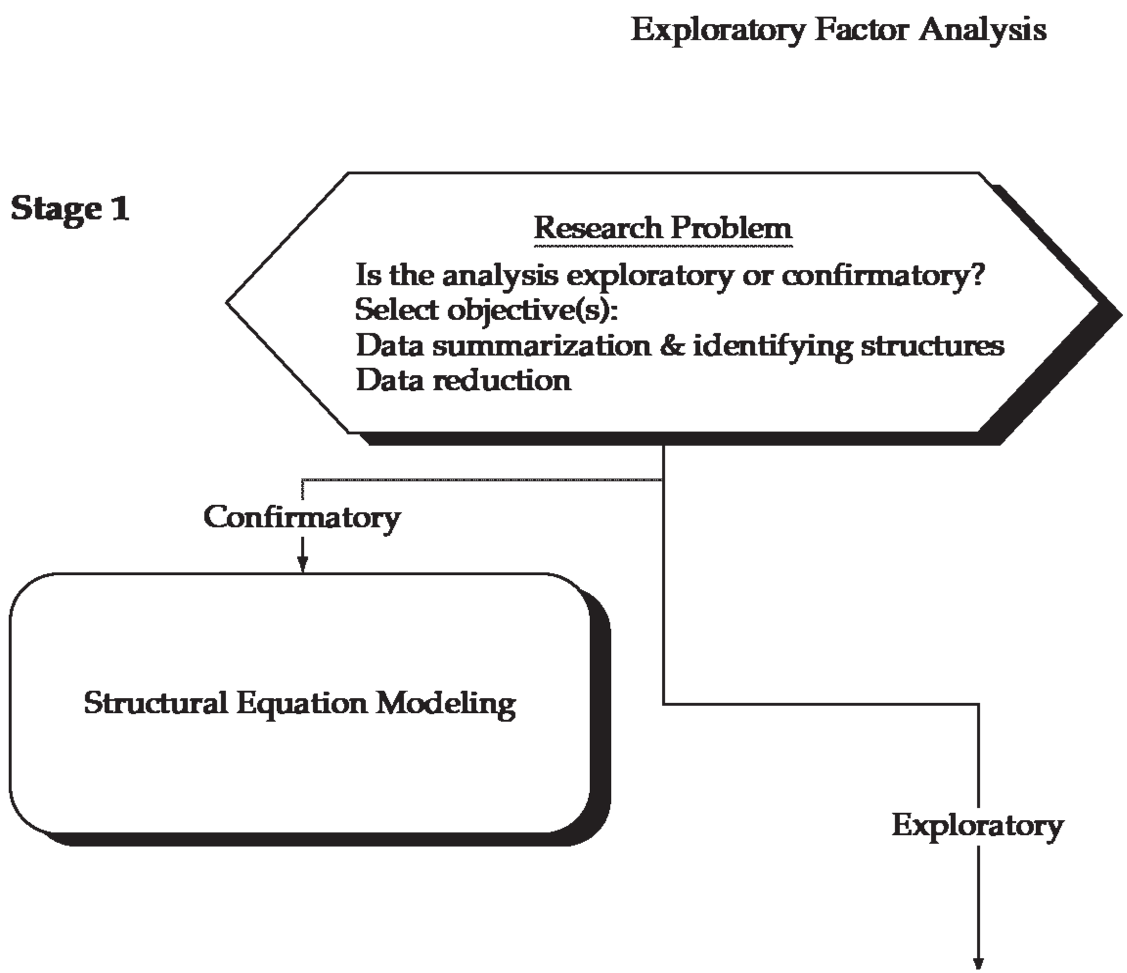
- **AN ILLUSTRATIVE EXAMPLE OF FACTOR ANALYSIS :**
  - Visual inspection of the original correlation matrix (Figure 1, part 1) does not easily reveal any specific pattern.
  - Among scattered high correlations, variable groupings are not apparent. The application of factor analysis results in the grouping of variables as reflected in part 2 of Figure 1.
  - Here some interesting patterns emerge.
    - First, four variables all relating to the in-store experience of shoppers are grouped together.
    - Then, three variables describing the product assortment and availability are grouped together.
    - Finally, product quality and price levels are grouped.
    - Each group represents a set of highly interrelated variables that may reflect a more general evaluative dimension.
    - In this case, we might label the three variable groupings by the labels in-store experience, product offerings, and value.
  - This simple example of factor analysis demonstrates its basic objective of grouping highly inter correlated variables into distinct sets (factors).
  - In many situations, these factors can provide a wealth of information about the interrelationships of the variables.

Factor Analysis Decision Process

- Discussion of factor analysis on the six-stage model-building paradigm.
- The following figure shows the first three stages of the structured approach to multivariate model building,
- Next figure details the final three stages, plus an additional stage (stage 7) beyond the estimation, interpretation, and validation of the factor models, which aids in selecting surrogate variables, computing factor scores, or creating summated scales for use in other multivariate techniques.

Factor Analysis Decision Process

- **Stage 1 : Objective of Factor Analysis**



Factor Analysis Decision Process

- **Stage 1 : Objective of Factor Analysis**
  - The starting point in factor analysis, as with other statistical techniques, is the research problem.
  - The general purpose of factor analytic techniques is to find a way to condense (summarize) the information contained in a number of original variables into a smaller set of new, composite dimensions or variates (factors) with a minimum loss of information.
  - That is, to search for and define the fundamental constructs or dimensions assumed to underlie the original variables.
  - In meeting its objectives, factor analysis is keyed to four issues:
    - specifying the unit of analysis,
    - achieving data summarization and/or data reduction,
    - variable selection, and
    - using factor analysis results with other multivariate techniques.

Factor Analysis Decision Process

- **Specifying the Unit of Analysis :**
  - Factor analysis is actually a more general model in that it can identify the structure of relationships among either variables or respondents by examining either the correlations between the variables or the correlations between the respondents.
  - If the objective of the research were to summarize the characteristics, factor analysis would be applied to a **correlation matrix** of the variables. This most common type of factor analysis, referred to as **R factor analysis**, analyzes a set of variables to identify the dimensions that are latent (not easily observed).
  - Factor analysis also may be applied to a correlation matrix of the individual respondents based on their characteristics. Referred to as **Q factor analysis**, this method combines or condenses large numbers of people into distinctly different groups within a larger population.
  - The Q factor analysis approach is not utilized frequently because of computational difficulties. Instead, most researchers utilize some type of cluster analysis to group individual respondents.

Factor Analysis Decision Process

- **Specifying the Unit of Analysis :**
  - Thus, the researcher must first select the unit of analysis for factor analysis:
    - variables or respondents.
  - Even though we will focus primarily on structuring variables, the option of employing factor analysis among respondents as an alternative to cluster analysis is also available.
- **Achieving Data Summarization Versus Data Reduction**
  - Factor analysis provides the researcher with two distinct, but interrelated, outcomes:
    - data summarization and data reduction.
  - In summarizing the data, factor analysis derives underlying dimensions that, when interpreted and understood, describe the data in a much smaller number of concepts than the original individual variables.
  - Data reduction extends this process by deriving an empirical value (factor score) for each dimension (factor) and then substituting this value for the original values.

Factor Analysis Decision Process

- **Achieving Data Summarization Versus Data Reduction**
  - **DATA SUMMARIZATION**
  - The fundamental concept involved in data summarization is the definition of structure.
  - Through structure, the researcher can view the set of variables at various levels of generalization, ranging from the most detailed level (individual variables themselves) to the more generalized level, where individual variables are grouped and then viewed not for what they represent individually, but for what they represent collectively in expressing a concept.
  - For example, variables at the individual level might be:
    - "I shop for specials," "I usually look for the lowest possible prices," "I shop for bargains," "National brands are worth more than store brands." Collectively, these variables might be used to identify consumers who are "price conscious" or "bargain hunters."
  - Factor analysis, as an interdependence technique, differs from the dependence techniques (i.e., multiple regression, discriminant analysis, multivariate analysis of variance, or conjoint analysis)
    - where one or more variables are explicitly considered the criterion or dependent variables and all others are the predictor or independent variables.

Factor Analysis Decision Process

- **Achieving Data Summarization Versus Data Reduction**
  - **DATA SUMMARIZATION**
  - In Factor analysis, all variables are simultaneously considered with no distinction as to dependent or independent variables.
  - Factor analysis still employs the concept of the variate, the linear composite of variables, but in factor analysis, the variates (factors) are formed to maximize their explanation of the entire variable set, not to predict a dependent variable(s).
  - The goal of data summarization is achieved by defining a small number of factors that adequately represent the original set of variables.
  - Each factor (variate) as a dependent variable that is a function of the entire set of observed variables.
  - It gives the differences in purpose between dependence (prediction) and interdependence (identification of structure) techniques.
  - Structure is defined by the interrelatedness among variables allowing for the specification of a smaller number of dimensions (factors) representing the original set of variables.

Factor Analysis Decision Process

- **Achieving Data Summarization Versus Data Reduction**
  - **DATA REDUCTION**
  - Factor analysis can also be used to achieve data reduction by
    - (1) identifying representative variables from a much larger set of variables for use in subsequent multivariate analyses, or
    - (2) creating an entirely new set of variables, much smaller in number, to partially or completely replace the original set of variables.
  - In both instances, the purpose is to retain the nature and character of the original variables, but reduce their number to simplify the subsequent multivariate analysis. Both conceptual and empirical issues support the creation of composite measures.
  - Factor analysis provides the empirical basis for assessing the structure of variables and the potential for creating these composite measures or selecting a subset of representative variables for further analysis.
  - Data summarization makes the identification of the underlying dimensions or factors ends in themselves.
  - Thus, estimates of the factors and the contributions of each variable to the factors (termed loadings) are all that is required for the analysis.
  - Data reduction relies on the factor loadings as well, but uses them as the basis for either identifying variables for subsequent analysis with other techniques or making estimates of the factors themselves (factor scores or summated scales), which then replace the original variables in subsequent analyses.

Factor Analysis Decision Process

- **Variable Selection**
  - Whether factor analysis is used for data reduction and/or summarization, the researcher should always consider the conceptual underpinnings of the variables and use judgment as to the appropriateness of the variables for factor analysis.
  - In both uses of factor analysis, the researcher implicitly specifies the potential dimensions that can be identified through the character and nature of the variables submitted to factor analysis.
  - For example, in assessing the dimensions of store image, if no questions on store personnel were included, factor analysis would not be able to identify this dimension.
  - The researcher also must remember that factor analysis will always produce factors. Thus, factor analysis is always a potential candidate for the “garbage in, garbage out” phenomenon.
  - The quality and meaning of the derived factors reflect the conceptual underpinnings of the variables included in the analysis.
  - Obviously, the use of factor analysis as a data summarization technique is based on having a conceptual basis for any variables analyzed.
  - But even if used solely for data reduction, factor analysis is most efficient when conceptually defined dimensions can be represented by the derived factors.

Factor Analysis Decision Process

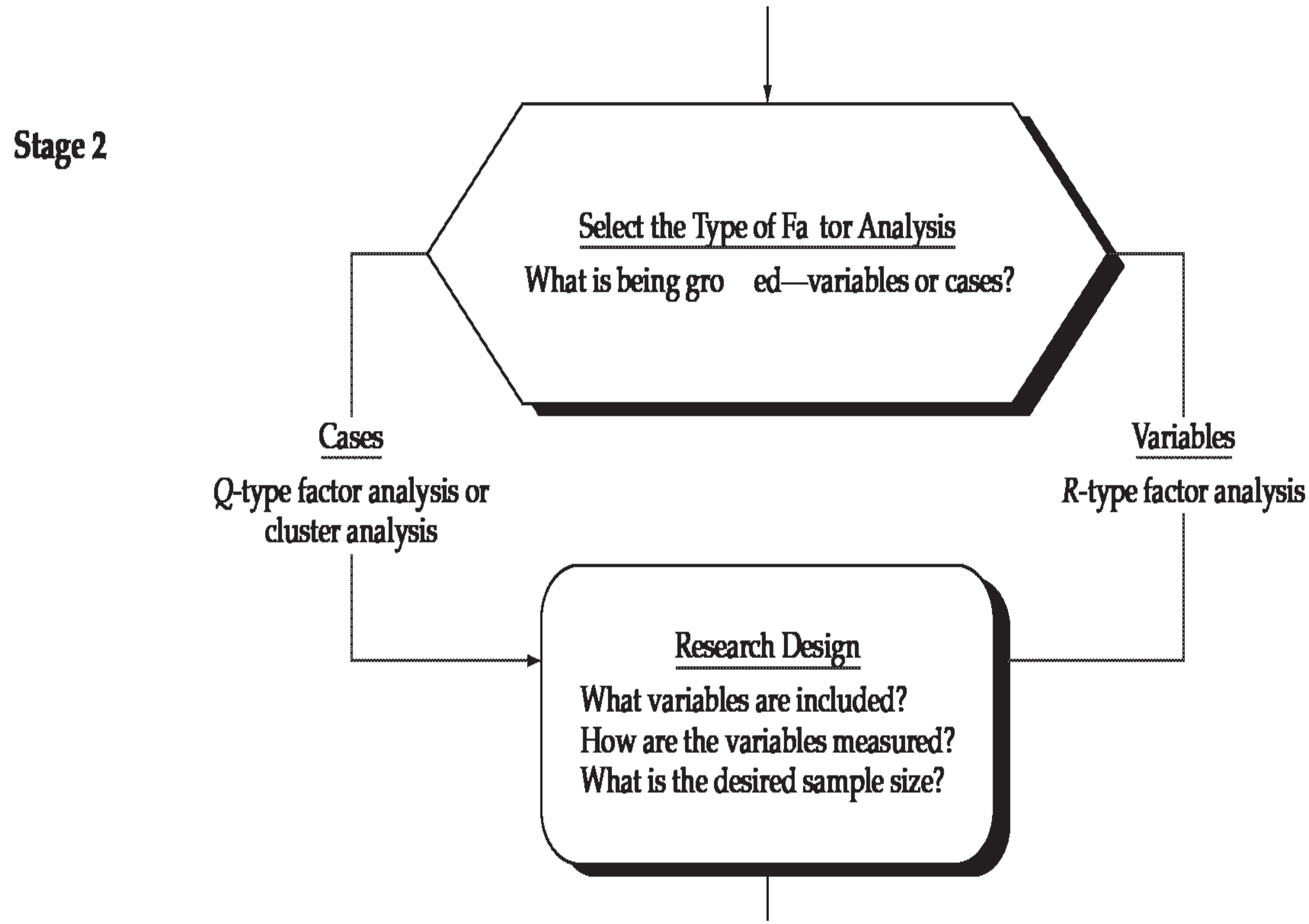
- **Using Factor Analysis with Other Multivariate Techniques**
  - Factor analysis, by providing insight into the interrelationships among variables and the underlying structure of the data, is an excellent starting point for many other multivariate techniques.
  - Variables determined to be highly correlated and members of the same factor would be expected to have similar profiles of differences across groups in multivariate analysis of variance or in discriminant analysis.
  - Highly correlated variables, such as those within a single factor, affect the stepwise procedures of multiple regression and discriminant analysis that sequentially enter variables based on their incremental predictive power over variables already in the model.
  - As one variable from a factor is entered, it becomes less likely that additional variables from that same factor would also be included due to their high correlations with variable(s) already in the model, meaning they have little incremental predictive power.
  - It does not mean that the other variables of the factor are less important or have less impact, but instead their effect is already represented by the included variable from the factor.

Factor Analysis Decision Process

- **Using Factor Analysis with Other Multivariate Techniques**
  - Factor analysis provides the basis for creating a new set of variables that incorporate the character and nature of the original variables in a much smaller number of new variables, whether using representative variables, factor scores, or summated scales.
  - In this manner, problems associated with large numbers of variables or high intercorrelations among variables can be substantially reduced by substitution of the new variables.
  - The researcher can benefit from both the empirical estimation of relationships and the insight into the conceptual foundation and interpretation of the results.

Factor Analysis Decision Process

- **Stage 2 : Designing A Factor Analysis**



Factor Analysis Decision Process

- **Stage 2 : Designing A Factor Analysis**
  - The design of a factor analysis involves three basic decisions:
  - (1) calculation of the input data (a correlation matrix) to meet the specified objectives of grouping variables or respondents;
  - (2) design of the study in terms of number of variables, measurement properties of variables, and the types of allowable variables; and
  - (3) the sample size necessary, both in absolute terms and as a function of the number of variables in the analysis.
  - **Correlations Among Variables or Respondents**
  - The first decision in the design of a factor analysis focuses on calculating the input data for the analysis.
  - Two forms of factor analysis:
    - R-type versus Q-type factor analysis.
  - Both types of factor analysis utilize a correlation matrix as the basic data input.
  - With R-type factor analysis, the researcher would use a traditional correlation matrix (correlations among variables) as input.
  - In this Q-type factor analysis, the results would be a factor matrix that would identify similar individuals.

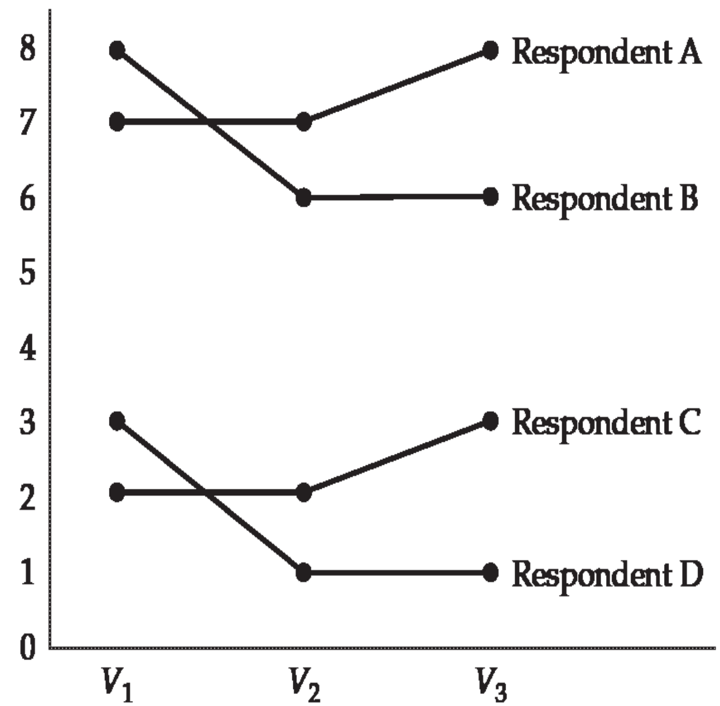
Factor Analysis Decision Process

- **Stage 2 : Designing A Factor Analysis**
  - From the results of a Q factor analysis, we could identify groups or clusters of individuals that demonstrate a similar pattern on the variables included in the analysis.
  - How does Q-type factor analysis differ from cluster analysis?
  - The answer is that Q-type factor analysis is based on the intercorrelations between the respondents, whereas cluster analysis forms groupings based on a distance-based similarity measure between the respondents' scores on the variables being analyzed.
  - To illustrate this difference, consider Figure 3, which contains the scores of four respondents over three different variables.
  - A Q-type factor analysis of these four respondents would yield two groups with similar covariance structures, consisting of respondents A and C versus B and D.
  - In contrast, the clustering approach would be sensitive to the actual distances among the respondents' scores and would lead to a grouping of the closest pairs. Thus, with a cluster analysis approach, respondents A and B would be placed in one group and C and D in the other group.

Factor Analysis Decision Process

- **Stage 2 : Designing A Factor Analysis**
  - Focuses on R-type factor analysis, the grouping of variables rather than respondents.

	Variables		
Respondent	V_1	V_2	V_3
A	7	7	8
B	8	6	6
C	2	2	3
D	3	1	1



Factor Analysis Decision Process

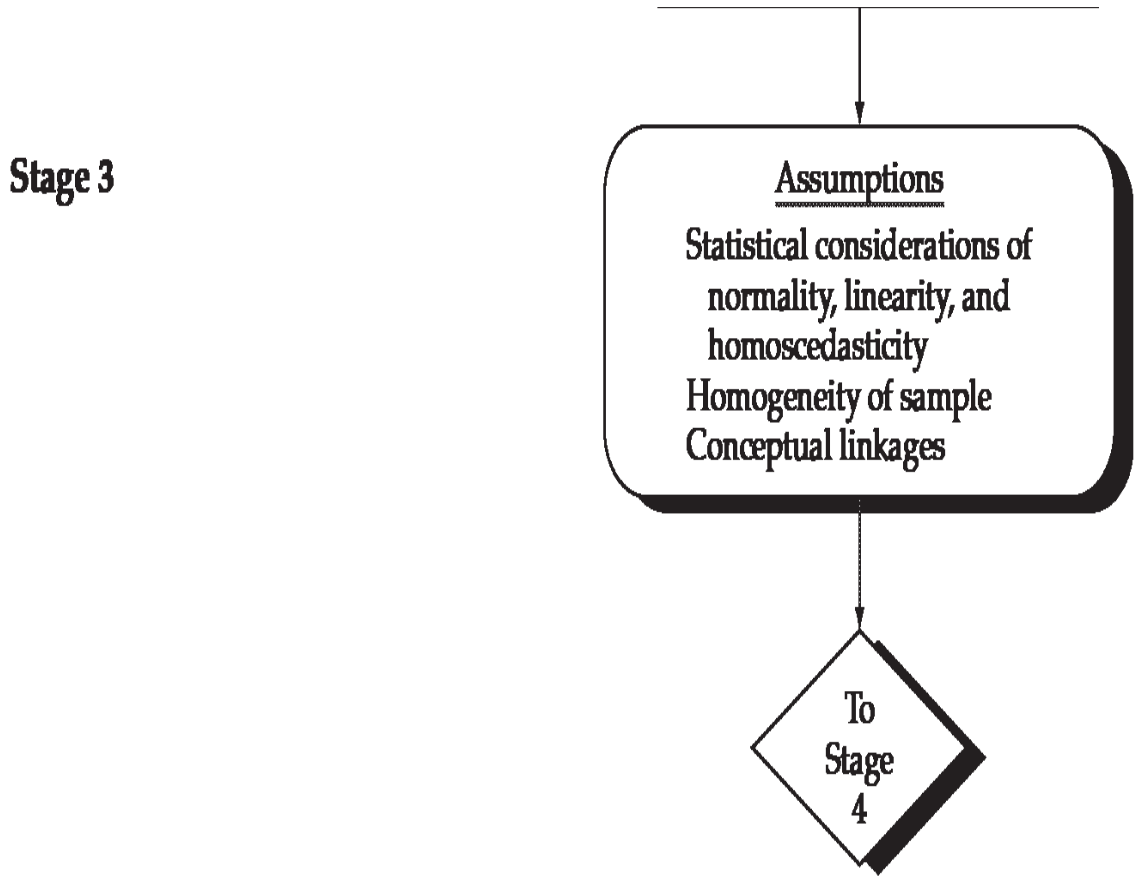
- **Stage 2 : Designing A Factor Analysis**
  - **Variable Selection and Measurement Issues :**
  - Two specific questions must be answered at this point:
    - (1) What type of variables can be used in factor analysis?
    - (2) How many variables should be included?
  - In terms of the types of variables included, the primary requirement is that a correlation value can be calculated among all variables.
  - Metric variables are easily measured by several types of correlations.
  - Nonmetric variables, however, are more problematic because they cannot use the same types of correlation measures used by metric variables.
  - Although some specialized methods calculate correlations among nonmetric variables, the most prudent approach is to avoid nonmetric variables.
  - If a nonmetric variable must be included, one approach is to define dummy variables (coded 0–1) to represent categories of nonmetric variables.
  - If all the variables are dummy variables, then specialized forms of factor analysis, such as Boolean factor analysis, are more appropriate.
  - The researcher should also attempt to minimize the number of variables included but still maintain a reasonable number of variables per factor.

Factor Analysis Decision Process

- **Stage 2 : Designing A Factor Analysis**
  - **Variable Selection and Measurement Issues :**
  - If a study is being designed to assess a proposed structure, the researcher should be sure to include several variables (five or more) that may represent each proposed factor.
  - Finally, when designing a study to be factor analyzed, the researcher should, if possible, identify several key variables (sometimes referred to as key indicants or marker variables) that closely reflect the hypothesized underlying factors.
  - **Sample Size :**
  - Regarding the sample size question, the researcher generally would not factor analyze a sample of fewer than 50 observations, and preferably the sample size should be 100 or larger.
  - As a general rule, the minimum is to have at least five times as many observations as the number of variables to be analyzed, and the more acceptable sample size would have a 10:1 ratio.
  - The researcher should always try to obtain the highest cases-per-variable ratio to minimize the chances of overfitting the data (i.e., deriving factors that are sample-specific with little generalizability).

Factor Analysis Decision Process

- **Stage 3 : Assumptions in factor analysis**



Factor Analysis Decision Process

- **Stage 3 : Assumptions in factor analysis**
  - The critical assumptions underlying factor analysis are more conceptual than statistical.
  - The researcher is always concerned with meeting the statistical requirement for any multivariate technique, but in factor analysis the overriding concerns center as much on the character and composition of the variables included in the analysis as on their statistical qualities.
  - **Conceptual Issues :**
  - The conceptual assumptions underlying factor analysis relate to the set of variables selected and the sample chosen.
  - **Basic Assumption :**
    - Some underlying structure do exist between the variables in the data set, They are related in some way
  - **Homogeneity :**
    - The sample of the data contains ingredients that are similar in nature.
    - For example, mixing dependent and independent variables in a single factor analysis and then using the derived factors to support dependence relationships is inappropriate.
  - The researcher must also ensure that the sample is homogeneous with respect to the underlying factor structure.
  - It is inappropriate to apply factor analysis to a sample of males and females for a set of items known to differ because of gender.

Factor Analysis Decision Process

- **Stage 3 : Assumptions in factor analysis**
  - When the two subsamples (males and females) are combined, the resulting correlations and factor structure will be a poor representation of the unique structure of each group.
  - Thus, whenever differing groups are expected in the sample, separate factor analyses should be performed, and the results should be compared to identify differences not reflected in the results of the combined sample.
  - **Statistical Issues :**
  - From a statistical standpoint, departures from normality, homoscedasticity, and linearity apply only to the extent that they diminish the observed correlations.
  - Only normality is necessary if a statistical test is applied to the significance of the factors, but these tests are rarely used.
  - In fact, some degree of multicollinearity is desirable, because the objective is to identify interrelated sets of variables.
  - **Multicollinearity :**
    - Means that the variables are so related that they can be explained by each other.
  - The next step is to ensure that the variables are sufficiently intercorrelated to produce representative factors.
  - As we will see, we can assess this degree of interrelatedness from both overall and individual variable perspectives.
  - The following are several empirical measures to aid in diagnosing the factorability of the correlation matrix.

Factor Analysis Decision Process

- **Stage 3 : Assumptions in factor analysis**
  - **Overall Measures of Intercorrelation :**
  - In addition to the statistical bases for the correlations of the data matrix, also ensure that the data matrix has sufficient correlations to justify the application of factor analysis.
  - If it is found that all of the correlations are low, or that all of the correlations are equal (denoting that no structure exists to group variables), then it is difficult for the application of factor analysis.
  - **Methods :**
  - **1. Visual Inspection :**
  - By inspecting the correlation matrix , we see the correlations between the variables.
  - If there is no substantial number of correlations greater than .30, then factor analysis is probably inappropriate.
  - The correlations among variables can also be analyzed by computing the partial correlations among variables.
  - A **partial correlation** is the correlation that is unexplained when the effects of other variables are taken into account.
  - If “true” factors exist in the data, the partial correlation should be small, because the variable can be explained by the variables loading on the factors.
  - If the partial correlations are high, indicating no underlying factors, then factor analysis is inappropriate.
  - Partial correlation above 0.7 means that it is too high (anti-image correlations) and not suited to factor analysis.

Factor Analysis Decision Process

- **Stage 3 : Assumptions in factor analysis**
  - **2. Bartlett's test of sphericity :**
  - A statistical test for the presence of correlations among the variables.
  - It provides the statistical significance that the correlation matrix has significant correlations among at least some of the variables.
  - Tests the hypothesis that there are correlations between the variables.(H0=No Correlations)
  - This value should be low. (0.05 < is considered as significant)
  - That increasing the sample size causes the Bartlett test to become more sensitive in detecting correlations among the variables.
  - **3. MSA – Measure of Sampling Adequacy :**
  - This method explains how well each variable is explained or predicted by other variables.
  - This index ranges from 0 to 1, reaching 1 when each variable is perfectly predicted without error by the other variables.
  - The measure can be interpreted with the following guidelines:
    - .80 or above, meritorious;
    - .70 or above, middling;
    - .60 or above, mediocre;
    - .50 or above, miserable; and
    - below .50, unacceptable.

Factor Analysis Decision Process

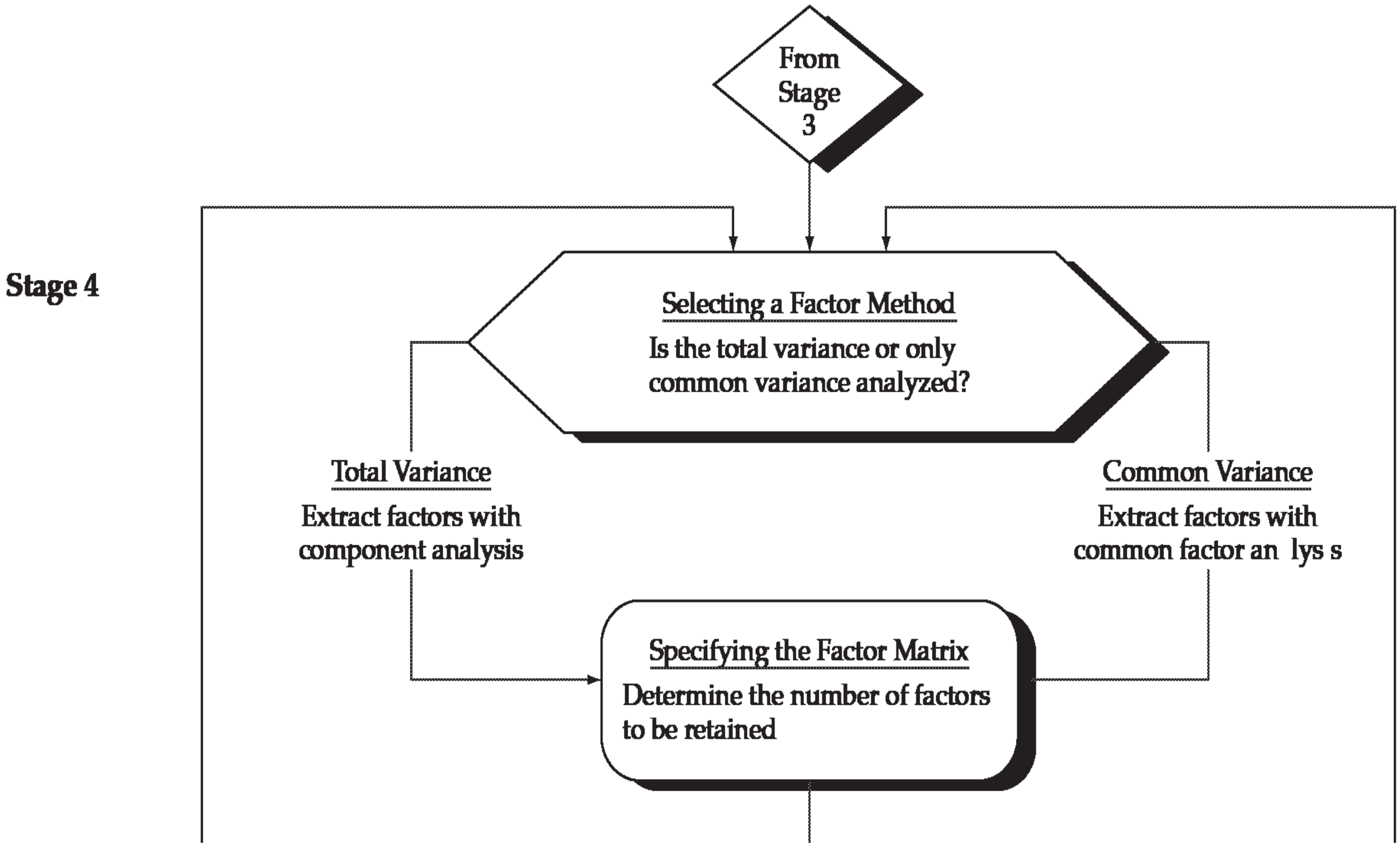
- **Stage 3 : Assumptions in factor analysis**
  - The MSA increases as
    - (1) the sample size increases,
    - (2) the average correlations increase,
    - (3) the number of variables increases, or
    - (4) the number of factors decreases.
  - The researcher should always have an overall MSA value of above .50 before proceeding with the factor analysis.
  - **Variable-specific measures of intercorrelation :**
  - The MSA guidelines can be extended to individual variables.
  - We should examine the MSA values for each variable and exclude those falling in the unacceptable range.
  - In deleting variables, we should first delete the variable with the lowest MSA and then recalculate the factor analysis.
  - Continue this process of deleting the variable with the lowest MSA value under .50 until all variables have an acceptable MSA value.
  - Once the individual variables achieve an acceptable level, then the overall MSA can be evaluated and a decision made on continuance of the factor analysis.

Factor Analysis Decision Process

- **Stage 3 : Assumptions in factor analysis**
  - **Assumptions to be fulfilled for Running Factor Analysis**
  - 1. No outliers in the data set
  - 2. Normality of the data set
  - 3. Adequate Sample size
  - 4. **Multi-Collinearity** and **singularity** among the variables does not exist.
  - 5. **Homoscedasticity** does not exist between the variables because factor analysis is a linear function of measured variables.
  - 6. Variables should be linear in nature.
  - 7. Data should be metric in nature i.e. on interval and ratio scale.

Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**



Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**
  - Once the variables are specified and the correlation matrix is prepared, the researcher is ready to apply factor analysis to identify the underlying structure of relationships.
  - In doing so, decisions must be made concerning
  - (1) the method of extracting the factors (common factor analysis versus components analysis) and
  - (2) the number of factors selected to represent the underlying structure in the data.
  - **Selecting the Factor Extraction Method**
  - The researcher can choose from two similar, yet unique, methods for defining (extracting) the factors to represent the structure of the variables in the analysis.
  - This decision on the method to use must combine the objectives of the factor analysis with knowledge about some basic characteristics of the relationships between variables.

Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**
  - **Partitioning the variance of a variable :**
  - In order to select between the two methods of factor extraction, we must first have some understanding of the variance for a variable and how it is divided or partitioned.
  - First, remember that variance is a value (i.e., the square of the standard deviation) that represents the total amount of dispersion of values for a single variable about its mean.
  - When a variable is correlated with another variable, means it shares variance with the other variable, and the amount of sharing between just two variables is simply the squared correlation.
  - For example, if two variables have a correlation of .50, each variable shares 25 percent (.502) of its variance with the other variable.
  - In factor analysis, we group variables by their correlations, such that variables in a group (factor) have high correlations with each other.
  - Thus, for the purposes of factor analysis, it is important to understand how much of a variable's variance is shared with other variables in that factor versus what cannot be shared (e.g., unexplained).

Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**
  - The total variance of any variable can be divided (partitioned) into three types of variance:
  - **1. Common variance :**
  - It is defined as that variance in a variable that is shared with all other variables in the analysis.
  - This variance is accounted for (shared) based on a variable's correlations with all other variables in the analysis.
  - A variable's communality is the estimate of its shared, or common, variance among the variables as represented by the derived factors.
  - **2. Specific variance :** (also known as unique variance)
  - It is that variance associated with only a specific variable.
  - This variance cannot be explained by the correlations to the other variables but is still associated uniquely with a single variable.
  - **3. Error variance :**
  - It is also variance that cannot be explained by correlations with other variables, but it is due to unreliability in the data-gathering process, measurement error, or a random component in the measured phenomenon.

Factor Analysis Decision Process

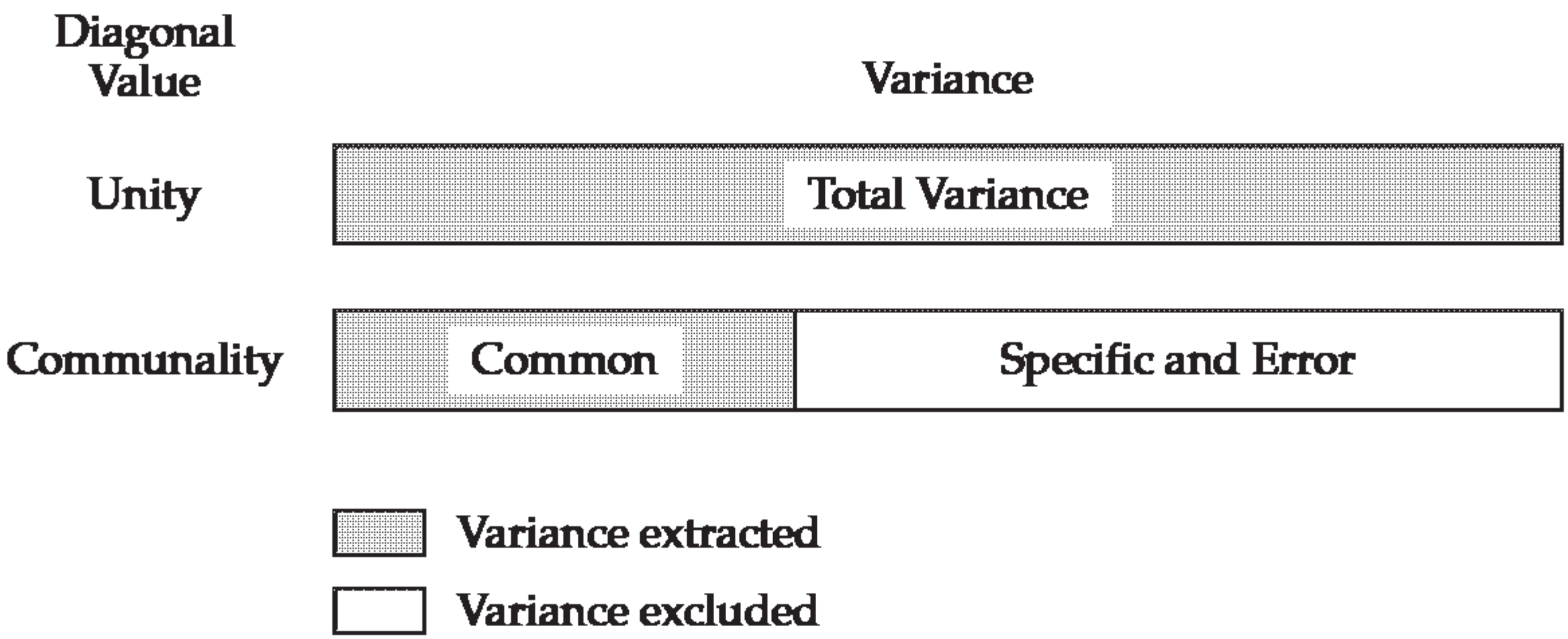
- **Stage 4 : Deriving factors and assessing overall fit**
  - Thus, the total variance of any variable is composed of its common, unique, and error variances.
  - As a variable is more highly correlated with one or more variables, the common variance (communality) increases.
  - However, if unreliable measures or other sources of extraneous error variance are introduced, then the amount of possible common variance and the ability to relate the variable to any other variable are reduced.
  - **Common factor analysis versus component analysis**
  - With a basic understanding of how variance can be partitioned, the researcher is ready to address the differences between the two methods, known as common factor analysis and component analysis.
  - The selection of one method over the other is based on two criteria:
  - (1) the objectives of the factor analysis and
  - (2) the amount of prior knowledge about the variance in the variables.
  - Component analysis is used when the objective is to summarize most of the original information (variance) in a minimum number of factors for prediction purposes.
  - In contrast, common factor analysis is used primarily to identify underlying factors or dimensions that reflect what the variables share in common.
  - The most direct comparison between the two methods is by their use of the explained versus unexplained variance:

Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**
  - **Component analysis** :
  - It is also known as principal components analysis, considers the total variance and derives factors that contain small proportions of unique variance and, in some instances, error variance.
  - However, the first few factors do not contain enough unique or error variance to distort the overall factor structure.
  - Specifically, with component analysis, unities (values of 1.0) are inserted in the diagonal of the correlation matrix, so that the full variance is brought into the factor matrix.
  - The following figure portrays the use of the total variance in component analysis and the differences when compared to common factor analysis.
  - **Common factor analysis** :
  - In contrast, it considers only the common or shared variance, assuming that both the unique and error variance are not of interest in defining the structure of the variables.
  - To employ only common variance in the estimation of the factors, communalities (instead of unities) are inserted in the diagonal.
  - Thus, factors resulting from common factor analysis are based only on the common variance.
  - As shown in the following Figure, common factor analysis excludes a portion of the variance included in a component analysis.

Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**



– First, the common factor and component analysis models are both widely used.
– As a practical matter, the components model is the typical default method of most statistical programs when performing factor analysis.
– Beyond the program defaults, distinct instances indicate which of the two methods is most appropriate:

Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**
  - **Component factor analysis is most appropriate when:**
    - Data reduction is a primary concern, focusing on the minimum number of factors needed to account for the maximum portion of the total variance represented in the original set of variables, and
    - Prior knowledge suggests that specific and error variance represent a relatively small proportion of the total variance.
  - **Common factor analysis is most appropriate when:**
    - The primary objective is to identify the latent dimensions or constructs represented in the original variables, and
    - The researcher has little knowledge about the amount of specific and error variance and therefore wishes to eliminate this variance.
  - Common factor analysis, with its more restrictive assumptions and use of only the latent dimensions (shared variance), is often viewed as more theoretically based.
  - Although theoretically sound, however, common factor analysis has several problems.
  - First, common factor analysis suffers from factor indeterminacy, which means that for any individual respondent, several different factor scores can be calculated from a single factor model result.
  - No single unique solution is found, as in component analysis, but in most instances the differences are not substantial.
  - The second issue involves the calculation of the estimated communalities used to represent the shared variance. Sometimes the communalities are not estimable or may be invalid (e.g., values greater than 1 or less than 0), requiring the deletion of the variable from the analysis.

Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**
  - **Criteria for the Number of Factors to Extract :**
  - Both factor analysis methods are interested in the best linear combination of variables—best in the sense that the particular combination of original variables accounts for more of the variance in the data as a whole than any other linear combination of variables.
  - Therefore, the first factor may be viewed as the single best summary of linear relationships exhibited in the data.
  - The second factor is defined as the second-best linear combination of the variables, subject to the constraint that it is orthogonal to the first factor.
  - To be orthogonal to the first factor, the second factor must be derived from the variance remaining after the first factor has been extracted.
  - Thus, the second factor may be defined as the linear combination of variables that accounts for the most variance that is still unexplained after the effect of the first factor has been removed from the data.
  - The process continues extracting factors accounting for smaller and smaller amounts of variance until all of the variance is explained.
  - For example, the components method actually extracts n factors, where n is the number of variables in the analysis. Thus, if 30 variables are in the analysis, 30 factors are extracted.

Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**
  - An exact quantitative basis for deciding the number of factors to extract has not been developed.
  - However, the following stopping criteria for the number of factors to extract are currently being utilized.
  - **LATENT ROOT CRITERION :**
  - The most commonly used technique is the latent root criterion.
  - This technique is simple to apply to either components analysis or common factor analysis.
  - The rationale for the latent root criterion is that any individual factor should account for the variance of at least a single variable if it is to be retained for interpretation.
  - With component analysis each variable contributes a value of 1 to the total eigenvalue.
  - Thus, only the factors having latent roots or eigenvalues greater than 1 are considered significant; all factors with latent roots less than 1 are considered insignificant and are disregarded.
  - Using the eigenvalue for establishing a cutoff is most reliable when the number of variables is between 20 and 50.
  - If the number of variables is less than 20, the tendency is for this method to extract a conservative number of factors (too few);
  - whereas if more than 50 variables are involved, it is not uncommon for too many factors to be extracted.

Factor Analysis Decision Process

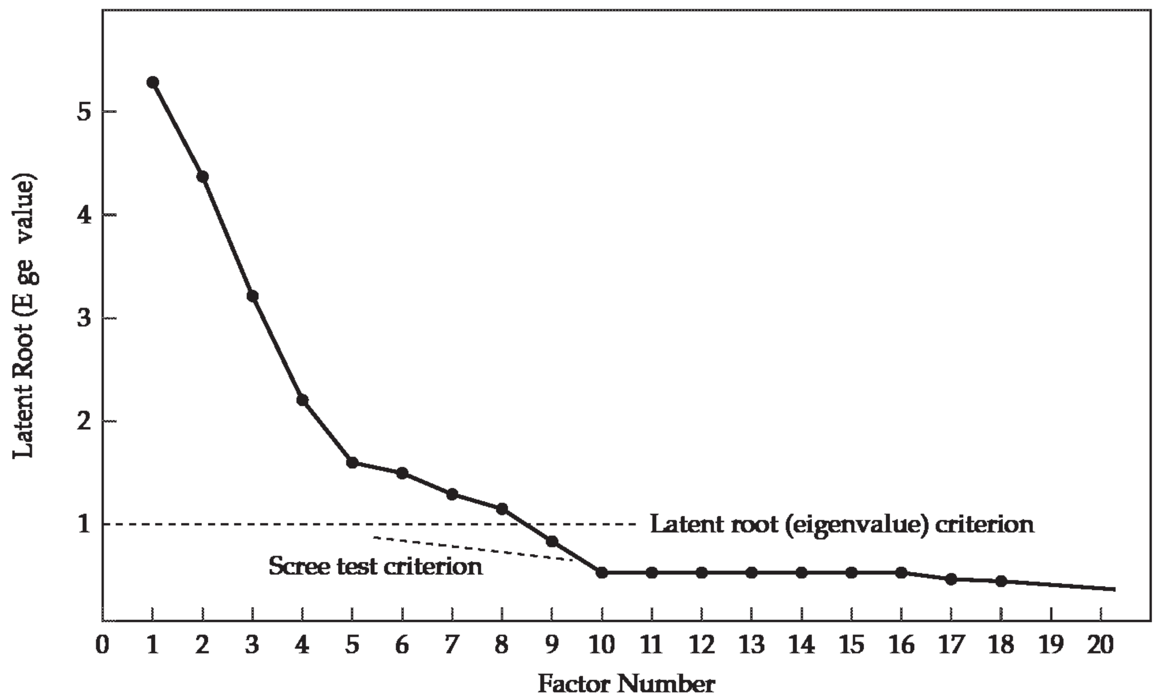
- **Stage 4 : Deriving factors and assessing overall fit**
  - **A PRIORI CRITERION :**
  - The a priori criterion is a simple yet reasonable criterion under certain circumstances.
  - When applying it, we already knows how many factors to extract before undertaking the factor analysis.
  - The researcher simply instructs the computer to stop the analysis when the desired number of factors has been extracted.
  - This approach is useful when testing a theory or hypothesis about the number of factors to be extracted.
  - It also can be justified in attempting to replicate another researcher’s work and extract the same number of factors that was previously found.
  - **PERCENTAGE OF VARIANCE CRITERION :**
  - The percentage of variance criterion is an approach based on achieving a specified cumulative percentage of total variance extracted by successive factors.
  - The purpose is to ensure practical significance for the derived factors by ensuring that they explain at least a specified amount of variance.
  - No absolute threshold has been adopted for all applications.
  - However, in the natural sciences the factoring procedure usually should not be stopped until the extracted factors account for at least 95 percent of the variance or until the last factor accounts for only a small portion (less than 5%).
  - In contrast, in the social sciences, where information is often less precise, it is not uncommon to consider a solution that accounts for 60 percent of the total variance (and in some instances even less) as satisfactory.

Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**
  - **SCREE TEST CRITERION :**
  - In the component analysis factor model the later factors extracted contain both common and unique variance.
  - Although all factors contain at least some unique variance, the proportion of unique variance is substantially higher in later factors.
  - The scree test is used to identify the optimum number of factors that can be extracted before the amount of unique variance begins to dominate the common variance structure.
  - The scree test is derived by plotting the latent roots against the number of factors in their order of extraction, and the shape of the resulting curve is used to evaluate the cutoff point.
  - The following Figure plots the first 18 factors extracted in a study.
  - Starting with the first factor, the plot slopes steeply downward initially and then slowly becomes an approximately horizontal line.
  - The point at which the curve first begins to straighten out **(“elbow”)** is considered to indicate the maximum number of factors to extract.

Factor Analysis Decision Process

- **Stage 4 : Deriving factors and assessing overall fit**



  - In the present case, the first 10 factors would qualify.
  - Beyond 10, too large a proportion of unique variance would be included; thus these factors would not be acceptable.
  - Note that in using the latent root criterion only 8 factors would have been considered. In contrast, using the scree test provides us with 2 more factors.
  - As a general rule, the scree test results in at least one and sometimes two or three more factors being considered for inclusion than does the latent root criterion.

THANK YOU

Prepared BY: A. Shobana Devi.
SRM, IT Dept.

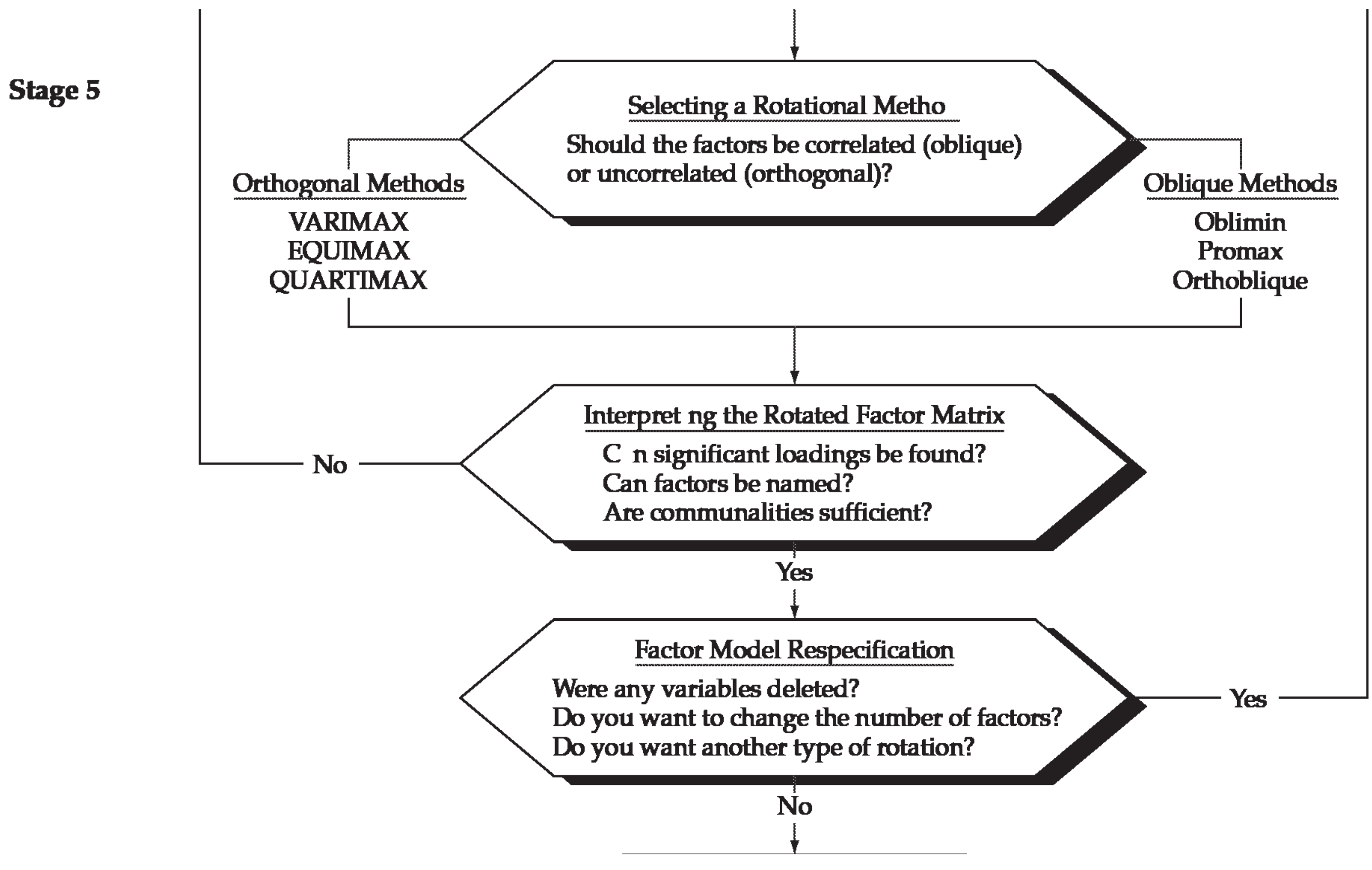
UNIT - II

Factor Analysis

Prepared BY: A. Shobana Devi.
SRM, IT Dept.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**



Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - To assist in the process of interpreting a factor structure and selecting a final factor solution, three fundamental processes are described.
  - Within each process, several substantive issues (factor rotation, factor-loading significance, and factor interpretation) are encountered.
  - **The Three Processes of Factor Interpretation**
  - Factor interpretation is circular in nature.
  - The researcher first evaluates the initial results, then makes a number of judgments in viewing and refining these results, with the distinct possibility that the analysis is respecified, requiring a return to the evaluative step.
  - **1. ESTIMATE THE FACTOR MATRIX**
  - First, the initial unrotated factor matrix is computed, containing the factor loadings for each variable on each factor.
  - Factor loadings are the correlation of each variable and the factor.
  - Loadings indicate the degree of correspondence between the variable and the factor, with higher loadings making the variable representative of the factor.
  - Factor loadings are the means of interpreting the role each variable plays in defining each factor.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **2. FACTOR ROTATION**
  - Unrotated factor solutions achieve the objective of data reduction, but we must ask whether the unrotated factor solution will provide information that offers the most adequate interpretation of the variables under examination.
  - In most instances the answer to this question is no, because factor rotation should simplify the factor structure.
  - Therefore, we should next employs a rotational method to achieve simpler and theoretically more meaningful factor solutions.
  - In most cases rotation of the factors improves the interpretation by reducing some of the ambiguities that often accompany initial unrotated factor solutions.
  - **3. FACTOR INTERPRETATION AND RESPECIFICATION**
  - As a final process, we evaluates the (rotated) factor loadings for each variable in order to determine that variable's role and contribution in determining the factor structure.
  - In the course of this evaluative process, the need may arise to respecify the factor model owing to

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - (1) the deletion of a variable(s) from the analysis,
  - (2) the desire to employ a different rotational method for interpretation,
  - (3) the need to extract a different number of factors, or
  - (4) the desire to change from one extraction method to another.
  - Respecification of a factor model involves returning to the extraction stage (stage 4), extracting factors, and then beginning the process of interpretation once again.
  - **Rotation of Factors**
  - The most important tool in interpreting factors is factor rotation.
  - The term rotation means exactly what it implies. Specifically, the reference axes of the factors are turned about the origin until some other position has been reached.
  - As indicated earlier, unrotated factor solutions extract factors in the order of their variance extracted.
  - The first factor tends to be a general factor with almost every variable loading significantly, and it accounts for the largest amount of variance.
  - The second and subsequent factors are then based on the residual amount of variance.
  - Each accounts for successively smaller portions of variance.
  - The ultimate effect of rotating the factor matrix is to redistribute the variance from earlier factors to later ones to achieve a simpler, theoretically more meaningful factor pattern.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - The simplest case of rotation is an **orthogonal factor rotation**, in which the axes are maintained at 90 degrees.
  - It is also possible to rotate the axes and not retain the 90-degree angle between the reference axes.
  - When not constrained to being orthogonal, the rotational procedure is called an **oblique factor rotation**.
  - Orthogonal and oblique factor rotations are demonstrated in Figures 7 and 8, respectively.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**

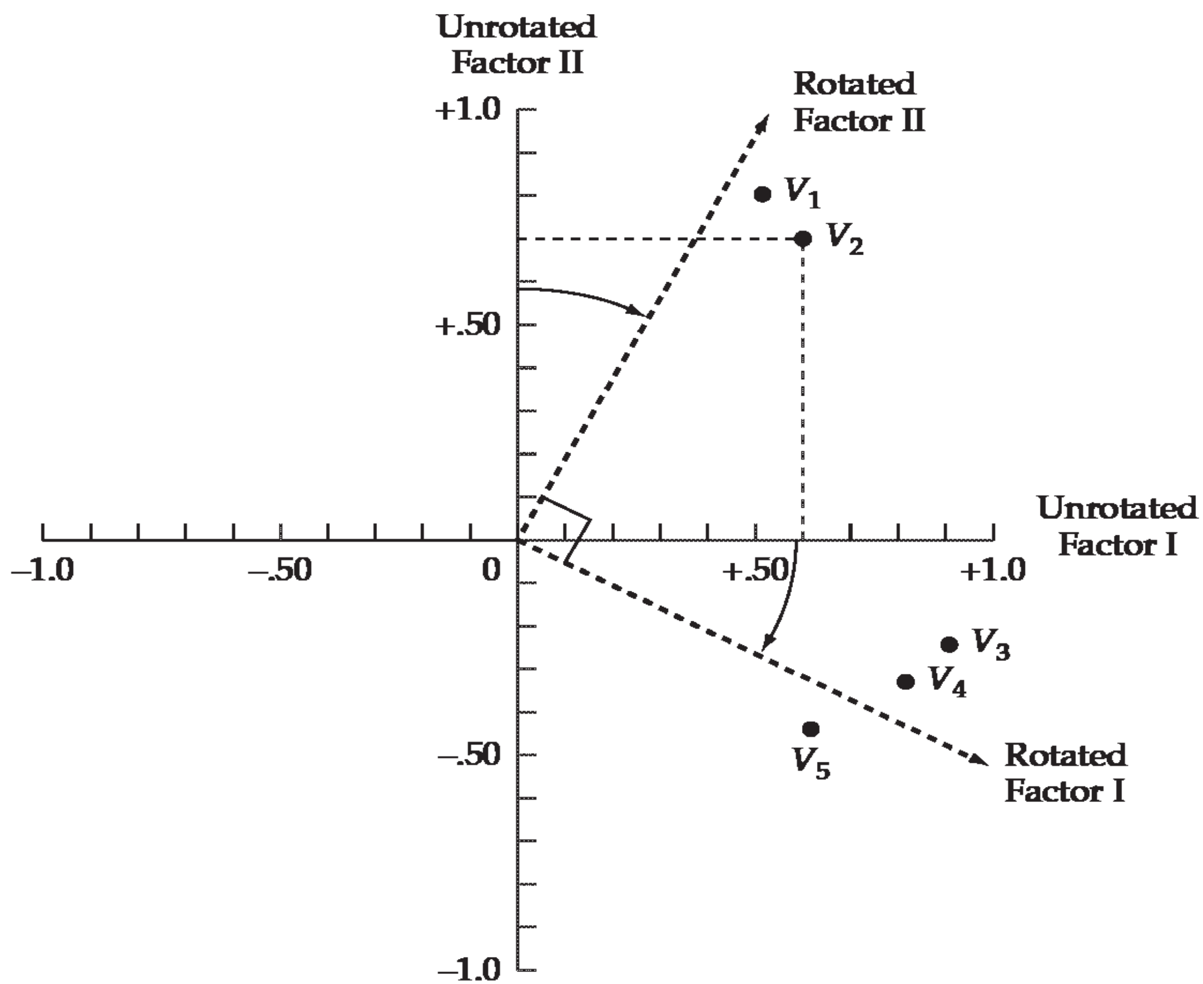


FIGURE 7 Orthogonal Factor Rotation

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - In first figure, in which five variables are depicted in a two-dimensional factor diagram, illustrates factor rotation.
  - The vertical axis represents unrotated factor II, and the horizontal axis represents unrotated factor I.
  - The axes are labeled with 0 at the origin and extend outward to +1.0 or -1.0.
  - The numbers on the axes represent the factor loadings.
  - The five variables are labeled V1, V2, V3, V4, and V5.
  - The factor loading for variable 2 (V2) on unrotated factor II is determined by drawing a dashed line horizontally from the data point to the vertical axis for factor II.
  - Similarly, a vertical line is drawn from variable 2 to the horizontal axis of unrotated factor I to determine the loading of variable 2 on factor I.
  - A similar procedure followed for the remaining variables determines the factor loadings for the unrotated and rotated solutions, as displayed in Table 1 for comparison purposes.
  - On the unrotated first factor, all the variables load fairly high.
  - On the unrotated second factor, variables 1 and 2 are very high in the positive direction.
  - Variable 5 is moderately high in the negative direction, and variables 3 and 4 have considerably lower loadings in the negative direction.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**

TABLE 1 Comparison Between Rotated and Unrotated Factor Loadings

	Unrotated Factor Loadings		*Rotated Factor Loadings*	
Variables	**I**	**II**	**I**	**II**
V_1	.50	.80	.03	.94
V_2	.60	.70	.16	.90
V_3	.90	−.25	.95	.24
V_4	.80	−.30	.84	.15
V_5	.60	−.50	.76	−.13

  - From visual inspection of Figure 7, two clusters of variables are obvious. Variables 1 and 2 go together, as do variables 3, 4, and 5. However, such patterning of variables is not so obvious from the unrotated factor loading
  - By rotating the original axes clockwise, as indicated in Figure 7, we obtain a completely different factor-loading pattern.
  - Note that in rotating the factors, the axes are maintained at 90 degrees.
  - After rotating the factor axes, variables 3, 4, and 5 load high on factor I, and variables 1 and 2 load high on factor II.
  - Thus, the clustering or patterning of these variables into two groups is more obvious after the rotation than before, even though the relative position or configuration of the variables remains unchanged.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**

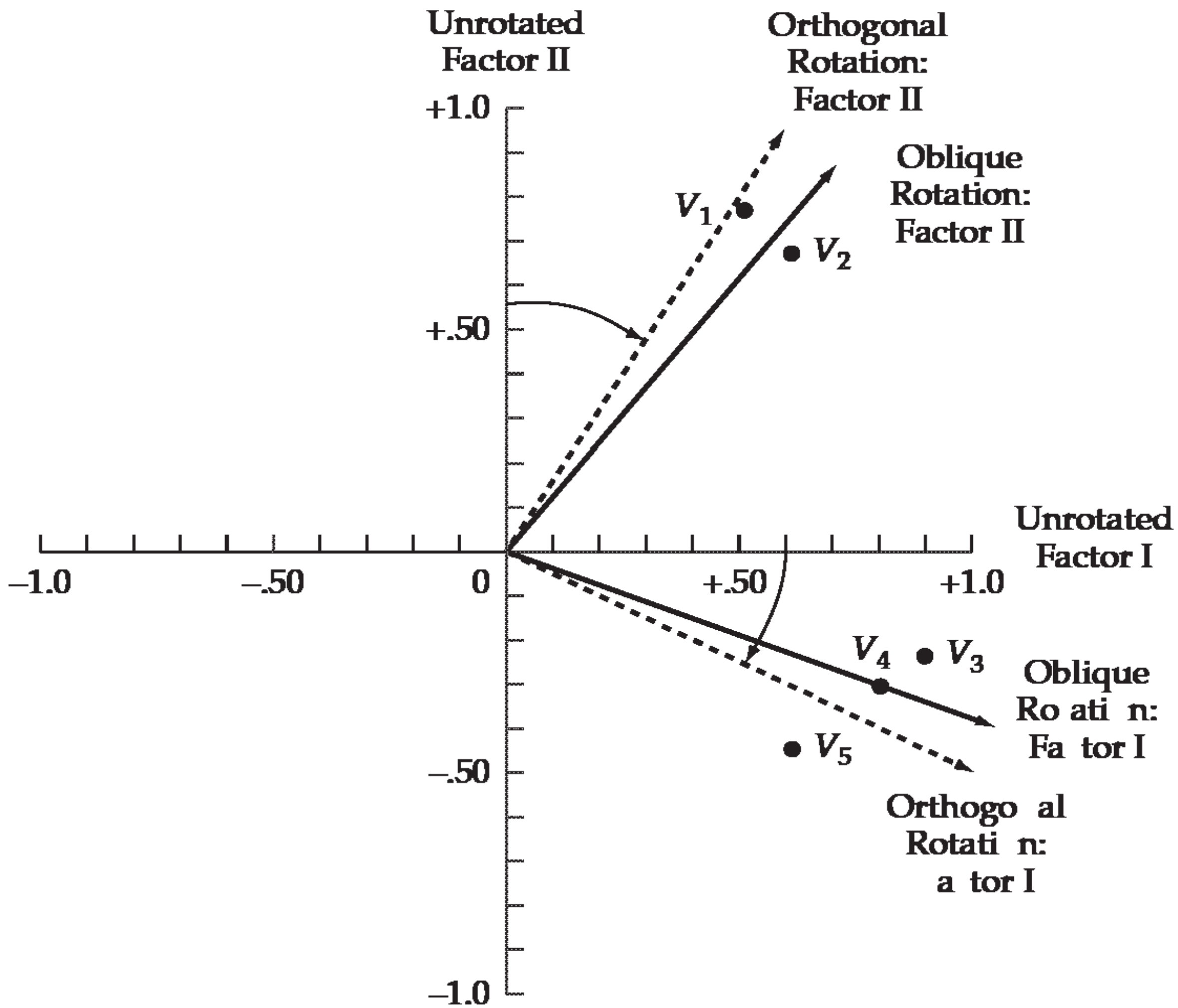


FIGURE 8 Oblique Factor Rotation

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - The same general principles of orthogonal rotations pertain to oblique rotations.
  - The oblique rotational method is more flexible, however, because the factor axes need not be orthogonal.
  - It is also more realistic because the theoretically important underlying dimensions are not assumed to be uncorrelated with each other.
  - In Figure 8 the two rotational methods are compared. Note that the oblique factor rotation represents the clustering of variables more accurately.
  - This accuracy is a result of the fact that each rotated factor axis is now closer to the respective group of variables.
  - Also, the oblique solution provides information about the extent to which the factors are actually correlated with each other.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **ORTHOGONAL ROTATION METHODS**
  - In practice, the objective of all methods of rotation is to simplify the rows and columns of the factor matrix to facilitate interpretation.
  - In a factor matrix, columns represent factors, with each row corresponding to a variable's loading across the factors.
  - By simplifying the rows, we mean making as many values in each row as close to zero as possible (i.e., maximizing a variable's loading on a single factor).
  - By simplifying the columns, we mean making as many values in each column as close to zero as possible (i.e., making the number of high loadings as few as possible).
  - Three major orthogonal approaches have been developed:
  - 1. QUARTIMAX
  - 2. VARIMAX
  - 3. EQUIMAX

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **ORTHOGONAL ROTATION METHODS**
  - 1. The ultimate goal of a QUARTIMAX rotation is to simplify the rows of a factor matrix; that is, QUARTIMAX focuses on rotating the initial factor so that a variable loads high on one factor and as low as possible on all other factors.
  - In these rotations, many variables can load high or near high on the same factor because the technique centers on simplifying the rows.
  - The QUARTIMAX method has not proved especially successful in producing simpler structures.
  - Its difficulty is that it tends to produce a general factor as the first factor on which most, if not all, of the variables have high loadings.
  - Regardless of one's concept of a simpler structure, inevitably it involves dealing with clusters of variables; a method that tends to create a large general factor (i.e., QUARTIMAX) is not in line with the goals of rotation.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **ORTHOGONAL ROTATION METHODS**
  - 2. In contrast to QUARTIMAX, the VARIMAX criterion centers on simplifying the columns of the factor matrix.
  - With the VARIMAX rotational approach, the maximum possible simplification is reached if there are only 1s and 0s in a column.
  - That is, the VARIMAX method maximizes the sum of variances of required loadings of the factor matrix.
  - With the VARIMAX rotational approach, some high loadings (i.e., close to –1 or +1) are likely, as are some loadings near 0 in each column of the matrix.
  - The logic is that interpretation is easiest when the variable-factor correlations are (1) close to either +1 or –1, thus indicating a clear positive or negative association between the variable and the factor; or
  - (2) close to 0, indicating a clear lack of association.
  - This structure is fundamentally simple.
  - Although the QUARTIMAX solution is analytically simpler than the VARIMAX solution, VARIMAX seems to give a clearer separation of the factors.
  - The VARIMAX method has proved successful as an analytic approach to obtaining an orthogonal rotation of factors.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **ORTHOGONAL ROTATION METHODS**
  - 3. The EQUIMAX approach is a compromise between the QUARTIMAX and VARIMAX approaches.
  - Rather than concentrating either on simplification of the rows or on simplification of the columns, it tries to accomplish some of each.
  - EQUIMAX has not gained widespread acceptance and is used infrequently.
  - **OBLIQUE ROTATION METHODS**
  - Oblique rotations are similar to orthogonal rotations, except that oblique rotations allow correlated factors instead of maintaining independence between the rotated factors.
  - Where several choices are available among orthogonal approaches, however, most statistical packages typically provide only limited choices for oblique rotations.
  - For example, SPSS provides OBLIMIN; SAS has PROMAX and ORTHOBLIQUE; and BMDP provides DQUART, DOBLIMIN, and ORTHOBLIQUE.
  - The objectives of simplification are comparable to the orthogonal methods, with the added feature of correlated factors.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **SELECTING AMONG ROTATIONAL METHODS**
  - No specific rules have been developed to guide us in selecting a particular orthogonal or oblique rotational technique.
  - In most instances, we simply utilizes the rotational technique provided by the computer program.
  - Most programs have the default rotation of VARIMAX, but all the major rotational methods are widely available.
  - However, no compelling analytical reason suggests favoring one rotational method over another.
  - The choice of an orthogonal or oblique rotation should be made on the basis of the particular needs of a given research problem.
  - **Judging the Significance of Factor Loadings**
  - In interpreting factors, a decision must be made regarding the factor loadings worth consideration and attention.
  - The following discussion details issues regarding practical and statistical significance, as well as the number of variables, that affect the interpretation of factor loadings.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **ENSURING PRACTICAL SIGNIFICANCE**
  - Making a preliminary examination of the factor matrix in terms of the factor loadings.
  - Because a factor loading is the correlation of the variable and the factor, the squared loading is the amount of the variable's total variance accounted for by the factor.
  - Thus, a .30 loading translates to approximately 10 percent explanation, and a .50 loading denotes that 25 percent of the variance is accounted for by the factor.
  - The loading must exceed .70 for the factor to account for 50 percent of the variance of a variable.
  - Thus, the larger the absolute size of the factor loading, the more important the loading in interpreting the factor matrix.
  - Using practical significance as the criteria, we can assess the loadings as follows:
  - • Factor loadings in the range of ±.30 to ±.40 are considered to meet the minimal level for interpretation of structure.
  - • Loadings ±.50 or greater are considered practically significant.
  - • Loadings exceeding 1.70 are considered indicative of well-defined structure and are the goal of any factor analysis.
  - These guidelines are applicable when the sample size is 100 or larger and where the emphasis is on practical, not statistical, significance.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **ASSESSING STATISTICAL SIGNIFICANCE**
  - As previously noted, a factor loading represents the correlation between an original variable and its factor.
  - In determining a significance level for the interpretation of loadings, an approach similar to determining the statistical significance of correlation coefficients could be used.
  - Thus, factor loadings should be evaluated at considerably stricter levels.
  - The researcher can employ the concept of statistical power to specify factor loadings considered significant for differing sample sizes.
  - With the stated objective of obtaining a power level of 80 percent, the use of a .05 significance level, and the proposed inflation of the standard errors of factor loadings,
  - Table 2 contains the sample sizes necessary for each factor loading value to be considered significant.
  - For example, in a sample of 100 respondents, factor loadings of .55 and above are significant.
  - However, in a sample of 50, a factor loading of .75 is required for significance.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - In comparison with the prior rule of thumb, which denoted all loadings of .30 as having practical significance, this approach would consider loadings of .30 significant only for sample sizes of 350 or greater.
  - Thus, these guidelines should be used as a starting point in factor-loading interpretation, with lower loadings considered significant and added to the interpretation based on other considerations.
  - The next section details the interpretation process and the role that other considerations can play.

TABLE 2 Guidelines for Identifying Significant Factor Loadings Based on Sample Size

Factor Loading	Sample Size Needed for Significance[a]
.30	350
.35	250
.40	200
.45	150
.50	120
.55	100
.60	85
.65	70
.70	60
.75	50

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **ADJUSTMENTS BASED ON THE NUMBER OF VARIABLES**
  - A disadvantage of both of the prior approaches is that the number of variables being analyzed and the specific factor being examined are not considered.
  - It has been shown that as we moves from the first factor to later factors, the acceptable level for a loading to be judged significant should increase.
  - The fact that unique variance and error variance begin to appear in later factors means that some upward adjustment in the level of significance should be included.
  - The number of variables being analyzed is also important in deciding which loadings are significant.
  - As the number of variables being analyzed increases, the acceptable level for considering a loading significantly decreases.
  - Adjustment for the number of variables is increasingly important as one moves from the first factor extracted to later factors.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **Interpreting a Factor Matrix**
  - The task of interpreting a factor-loading matrix to identify the structure among the variables can at first seem overwhelming.
  - Even a fairly simple analysis of 15 variables on four factors necessitates evaluating and interpreting 60 factor loadings.
  - Thus, interpreting the complex interrelationships represented in a factor matrix requires a combination of applying objective criteria with managerial judgment.
  - By following the five-step procedure outlined next, the process can be simplified considerably.
  - **STEP 1: EXAMINE THE FACTOR MATRIX OF LOADINGS**
  - The factor-loading matrix contains the factor loading of each variable on each factor.
  - They may be either rotated or unrotated loadings, but as discussed earlier, rotated loadings are usually used in factor interpretation unless data reduction is the sole objective.
  - Typically, the factors are arranged as columns; thus, each column of numbers represents the loadings of a single factor.
  - If an oblique rotation has been used, two matrices of factor loadings are provided.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - The first is the factor pattern matrix, which has loadings that represent the unique contribution of each variable to the factor.
  - The second is the factor structure matrix, which has simple correlations between variables and factors, but these loadings contain both the unique variance between variables and factors and the correlation among factors.
  - As the correlation among factors becomes greater, it becomes more difficult to distinguish which variables load uniquely on each factor in the factor structure matrix.
  - Thus, most researchers report the results of the factor pattern matrix.
  - **STEP 2: IDENTIFY THE SIGNIFICANT LOADING(S) FOR EACH VARIABLE**
  - The interpretation should start with the first variable on the first factor and move horizontally from left to right, looking for the highest loading for that variable on any factor.
  - When the highest loading (largest absolute factor loading) is identified, it should be underlined if significant as determined by the criteria discussed earlier.
  - Attention then focuses on the second variable and, again moving from left to right horizontally, looking for the highest loading for that variable on any factor and underlining it.
  - This procedure should continue for each variable until all variables have been reviewed for their highest loading on a factor.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - Most factor solutions, however, do not result in a simple structure solution (a single high loading for each variable on only one factor).
  - Thus, we will, after underlining the highest loading for a variable, continue to evaluate the factor matrix by underlining all significant loadings for a variable on all the factors.
  - The process of interpretation would be greatly simplified if each variable had only one significant variable.
  - In practice, however, we may find that one or more variables each has moderate-size loadings on several factors, all of which are significant, and the job of interpreting the factors is much more difficult.
  - When a variable is found to have more than one significant loading (0.50) on one or more than one factor, it is termed a cross-loading.
  - Ultimately, the objective is to minimize the number of significant loadings on each row of the factor matrix (i.e., make each variable associate with only one factor).
  - We may find that different rotation methods eliminate any cross-loadings and thus define a simple structure.
  - If a variable persists in having cross-loadings, it becomes a candidate for deletion.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **STEP 3: ASSESS THE COMMUNALITIES OF THE VARIABLES**
  - Once all the significant loadings have been identified, the researcher should look for any variables that are not adequately accounted for by the factor solution.
  - One simple approach is to identify any variable(s) lacking at least one significant loading.
  - Another approach is to examine each variable's communality, representing the amount of variance accounted for by the factor solution for each variable.
  - We should view the communalities to assess whether the variables meet acceptable levels of explanation.
  - For example, we may specify that at least one-half of the variance of each variable must be taken into account.
  - Using this guideline, we would identify all variables with communalities less than .50 as not having sufficient explanation.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - **STEP 4: RESPECIFY THE FACTOR MODEL IF NEEDED**
  - Once all the significant loadings have been identified and the communalities examined, we may find any one of several problems:
  - (a) a variable has no significant loadings;
  - (b) even with a significant loading, a variable's communality is deemed too low; or
  - (c) a variable has a cross-loading.
  - In this situation, we can take any combination of the following remedies, listed from least to most extreme:
  - • Ignore those problematic variables and interpret the solution as it is, which is appropriate if the objective is solely data reduction, but the variables in question are poorly represented in the factor solution.
  - • Evaluate each of those variables for possible deletion, depending on the variable's overall contribution to the research as well as its communality index.
  - If the variable is of minor importance to the study's objective or has an unacceptable communality value, it may be eliminated and then the factor model respecified by deriving a new factor solution with those variables eliminated.
  - • Employ an alternative rotation method, particularly an oblique method if only orthogonal methods had been used.
  - • Decrease/increase the number of factors retained to see whether a smaller/larger factor structure will represent those problematic variables.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - • Modify the type of factor model used (component versus common factor) to assess whether varying the type of variance considered affects the factor structure.
  - **STEP 5: LABEL THE FACTORS**
  - When an acceptable factor solution has been obtained in which all variables have a significant loading on a factor, we attempts to assign some meaning to the pattern of factor loadings.
  - Variables with higher loadings are considered more important and have greater influence on the name or label selected to represent a factor.
  - Thus, we will examine all the significant variables for a particular factor and, placing greater emphasis on those variables with higher loadings, will attempt to assign a name or label to a factor that accurately reflects the variables loading on that factor.
  - The signs are interpreted just as with any other correlation coefficients.
  - On each factor, like signs mean the variables are positively related, and opposite signs mean the variables are negatively related.
  - In orthogonal solutions the factors are independent of one another. Therefore, the signs for factor loading relate only to the factor on which they appear, not to other factors in the solution.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**
  - This label is not derived or assigned by the factor analysis computer program; rather, the label is intuitively developed by us based on its appropriateness for representing the underlying dimensions of a particular factor.
  - This procedure is followed for each extracted factor.
  - The final result will be a name or label that represents each of the derived factors as accurately as possible.
  - **AN EXAMPLE OF FACTOR INTERPRETATION**
  - To serve as an illustration of factor interpretation, nine measures were obtained in a pilot test based on a sample of 202 respondents.
  - After estimation of the initial results, further analysis indicated a three-factor solution was appropriate.
  - Thus, we now has the task of interpreting the factor loadings of the nine variables.
  - The following Table contains a series of factor-loading matrices.
  - The first to be considered is the unrotated factor matrix (part a).
  - We will examine the unrotated and rotated factor-loading matrices through the five-step process described earlier.

Factor Analysis Decision Process

- **Stage 5 : Interpreting The Factors**

TABLE 3 Interpretation of a Hypothetical Factor-Loading Mat x

(a) Unrotated Factor-Loading Matrix

	Factor 1	Factor 2	Factor 3
V_1	.611	.250	−.204
V_2	.614	−.446	.264
V_3	.295	−.447	.107
V_4	.561	−.176	−.550
V_5	.589	−.467	314
V_6	.630	−.102	−.285
V_7	.498	.611	.160
V_8	.310	.300	.649
V_9	.492	.597	−.094

(b) VARIMAX Rotated Factor-Loading Matrix

	Factor 1	Factor 2	Factor 3	Communality
V_1	.462	.099	.505	.477
V_2	.101	.778	.173	.644
V_3	−.134	.517	.114	.299
V_4	−.005	.184	.784	.648
V_5	.087	.801	.119	.664
V_6	.180	.302	.605	.489
V_7	.795	−.032	.120	.647
V_8	.623	.293	−.366	.608
V_9	.694	−.147	.323	.608

(c) Simplified Rotated Factor-Loading M t ix

	Component 1	Component 2	Component 3
V_7	795		
V_9	.694		
V_8	.623		
V_5		.801	
V_2		.778	
V_3		.517	
V_4			.784
V_6			.605
V_1	.462		.505

[1] *Loadings less than .40 are not shown and variables are sorted by highest loading.*

(d) Rotated Factor-Loading Matrix with V_1 Deleted[2]

	Factor 1	Factor 2	Factor 3
V_2	.807		
V_5	.803		
V_3	.524		
V_7		.802	
V_9		.686	
V_8		.655	
V_4			.851
V_6			.717

[2] *V_1 deleted from the analysis, loadings less than .40 are not shown, and variables are sorted by highest loading.*

Factor Analysis Decision Process

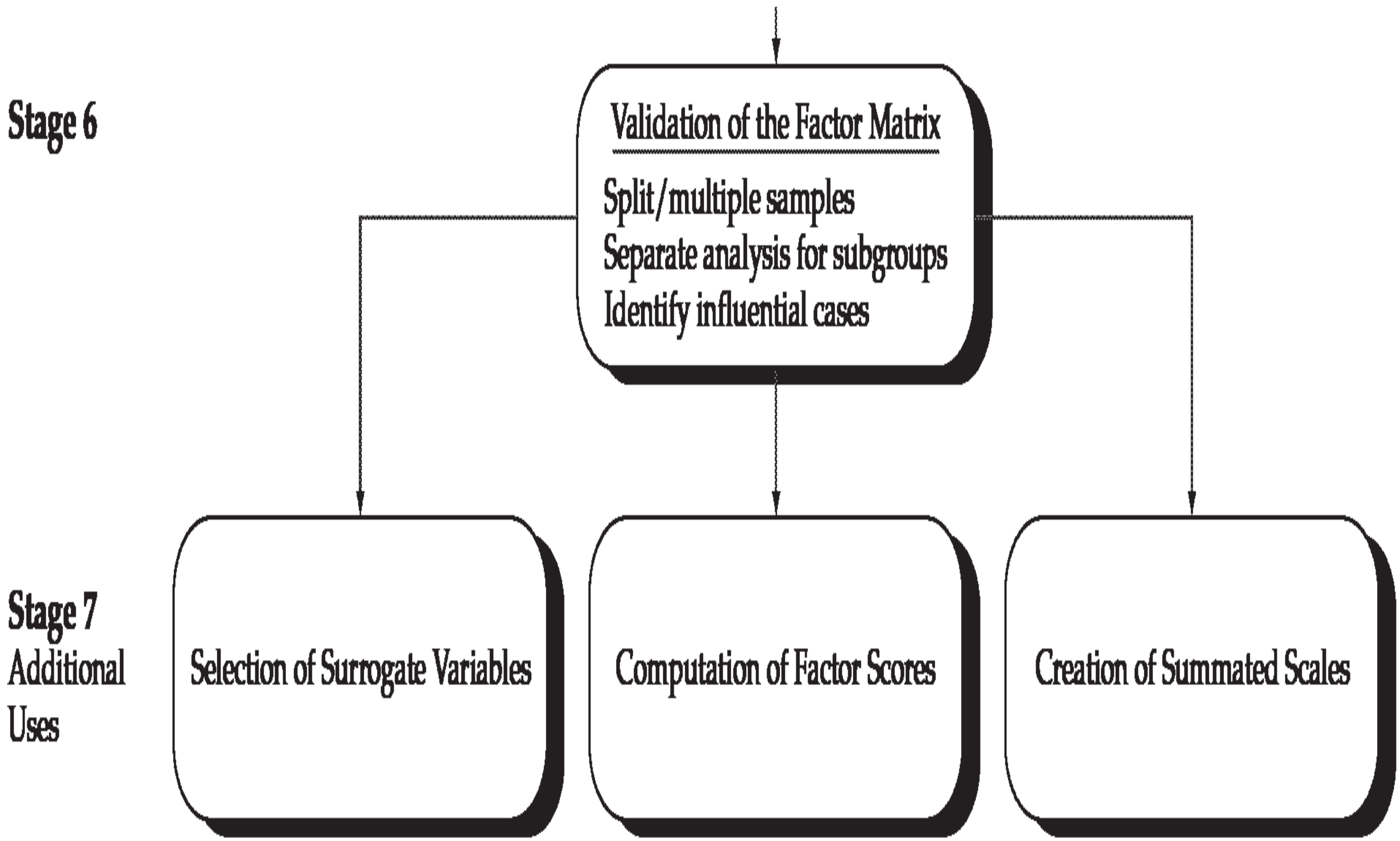
- **Stage 5 : Interpreting The Factors**
  - **Steps 1 and 2: Examine the Factor-Loading Matrix and Identify Significant Loadings.**
  - Given the sample size of 202, factor loadings of .40 and higher will be considered significant for interpretative purposes.
  - Using this threshold for the factor loadings, we can see that the unrotated matrix does little to identify any form of simple structure.
  - Five of the nine variables have cross-loadings, and for many of the other variables the significant loadings are fairly low.
  - In this situation, rotation may improve our understanding of the relationship among variables.
  - As shown in Table 3b, the VARIMAX rotation improves the structure considerably in two noticeable ways.
  - First, the loadings are improved for almost every variable, with the loadings more closely aligned to the objective of having a high loading on only a single factor.
  - Second, now only one variable (V1) has a cross-loading.
  - **Step 3: Assess Communalities.**
  - Only V3 has a communality that is low (.299).
  - For our purposes V3 will be retained, but a researcher may consider deletion of such variables in other research contexts.
  - It illustrates the instance in which a variable has a significant loading, but may still be poorly accounted for by the factor solution.

Factor Analysis Decision Process

- ## Stage 5 : Interpreting The Factors
  - **Step 4: Respecify the Factor Model if Needed.**
  - If we set a threshold value of .40 for loading significance and rearrange the variables according to loadings, the pattern shown in Table 3c emerges.
  - Variables V7, V9, and V8 all load highly on factor 1; factor 2 is characterized by variables V5, V2, and V3; and factor 3 has two distinctive characteristics (V4 and V6).
  - Only V1 is problematic, with significant loadings on both factors 1 and 3.
  - Given that at least two variables are given on both of these factors, V1 is deleted from the analysis and the loadings recalculated.
  - **Step 5: Label the Factors.**
  - As shown in Table 3d, the factor structure for the remaining eight variables is now very well defined, representing three distinct groups of variables that the researcher may now utilize in further research.
  - As the preceding example shows, the process of factor interpretation involves both objective and subjective judgments.

Factor Analysis Decision Process

- **STAGE 6: VALIDATION OF FACTOR ANALYSIS**



Factor Analysis Decision Process

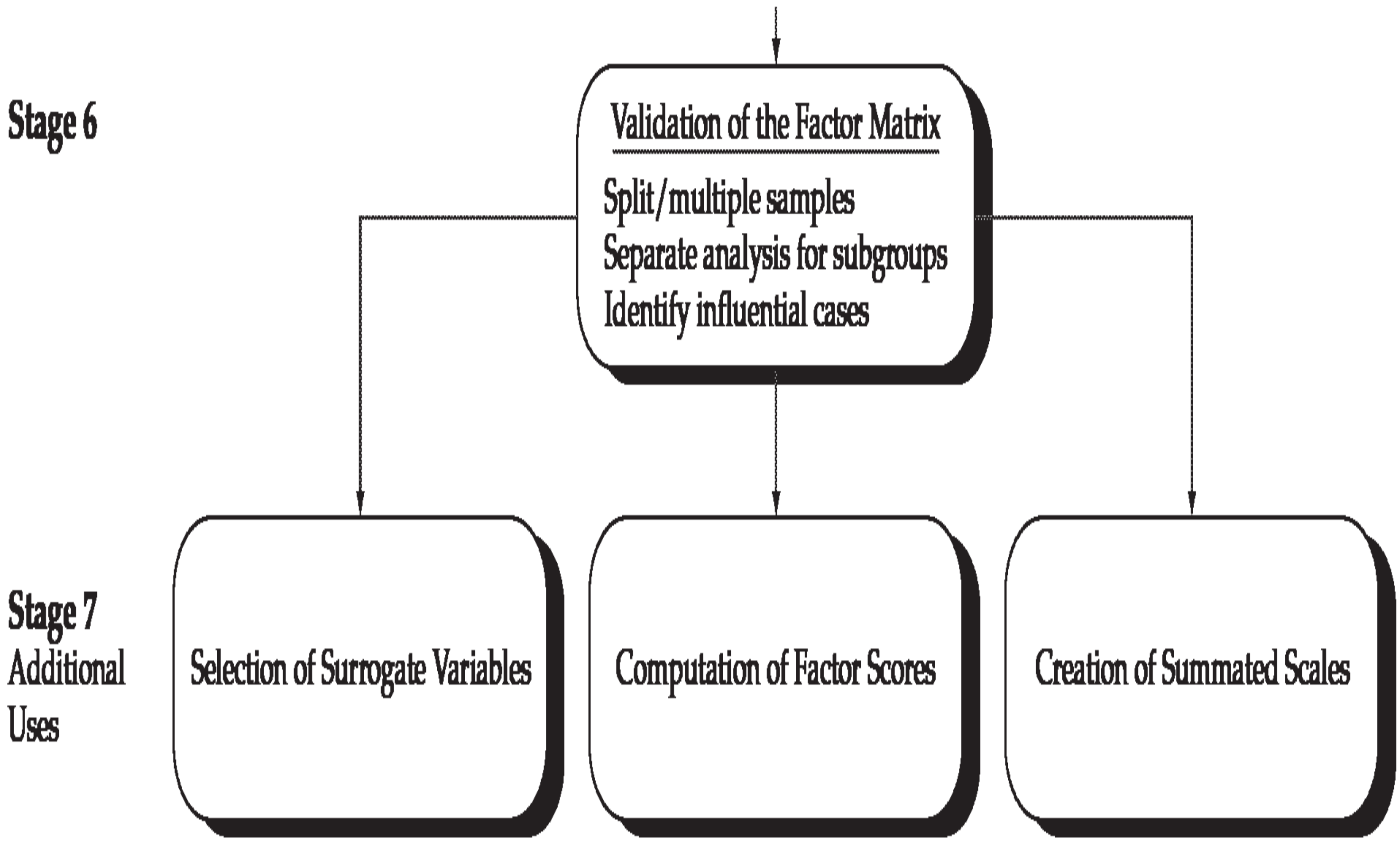
- **STAGE 6: VALIDATION OF FACTOR ANALYSIS**
  - The sixth stage involves assessing the degree of generalizability of the results to the population and the potential influence of individual cases or respondents on the overall results.
  - The issue of generalizability is critical for each of the multivariate methods, but it is especially relevant for the interdependence methods because they describe a data structure that should be representative of the population as well.
  - **Use of a Confirmatory Perspective**
  - The most direct method of validating the results is to move to a confirmatory perspective and assess the replicability of the results, either with a split sample in the original data set or with a separate sample.
  - The comparison of two or more factor model results has always been problematic. However, several options exist for making an objective comparison.
  - The emergence of confirmatory factor analysis (CFA) through structural equation modeling has provided one option, but it is generally more complicated and requires additional software packages, such as LISREL or EQS [4, 21]. These methods have had sporadic use, owing in part to
  - (1) their perceived lack of sophistication, and
  - (2) the unavailability of software or analytical programs to automate the comparisons.
  - Thus, when CFA is not appropriate, these methods provide some objective basis for comparison.

Factor Analysis Decision Process

- **STAGE 6: VALIDATION OF FACTOR ANALYSIS**
  - **Assessing Factor Structure Stability**
  - Another aspect of generalizability is the stability of the factor model results. Factor stability is primarily dependent on the sample size and on the number of cases per variable.
  - We always encouraged to obtain the largest sample possible and develop parsimonious models to increase the cases-to-variables ratio.
  - If sample size permits, we may wish to randomly split the sample into two subsets and estimate factor models for each subset.
  - Comparison of the two resulting factor matrices will provide an assessment of the robustness of the solution across the sample.
  - **Detecting Influential Observations**
  - In addition to generalizability, another issue of importance to the validation of factor analysis is the detection of influential observations.
  - We encouraged to estimate the model with and without observations identified as outliers to assess their impact on the results.
  - If omission of the outliers is justified, the results should have greater generalizability.
  - Also, several measures of influence that reflect one observation's position relative to all others (e.g., covariance ratio) are applicable to factor analysis as well.
  - Finally, the complexity of methods proposed for identifying influential observations specific to factor analysis limits the application of these methods.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**



Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - Depending upon the objectives for applying factor analysis, we may stop with factor interpretation or further engage in one of the methods for data reduction.
  - If the objective is simply to identify logical combinations of variables and better understand the interrelationships among variables, then factor interpretation will suffice.
  - It provides an empirical basis for judging the structure of the variables and the impact of this structure when interpreting the results from other multivariate techniques.
  - If the objective, however, is to identify appropriate variables for subsequent application to other statistical techniques, then some form of data reduction will be employed.
  - The two options include the following:
    - Selecting the variable with the highest factor loading as a surrogate representative for a particular factor dimension
    - Replacing the original set of variables with an entirely new, smaller set of variables created either from summated scales or factor scores

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - Either option will provide new variables for use,
  - for example, as the independent variables in a regression or discriminant analysis,
  - as dependent variables in multivariate analysis of variance,
  - or even as the clustering variables in cluster analysis.
  - **Selecting Surrogate Variables for Subsequent Analysis**
  - If our objective is simply to identify appropriate variables for subsequent application with other statistical techniques, then the option of examining the factor matrix and selecting the variable with the highest factor loading on each factor to act as a surrogate variable that is representative of that factor.
  - This approach is simple and direct only when one variable has a factor loading that is substantially higher than all other factor loadings.
  - In many instances, however, the selection process is more difficult because two or more variables have loadings that are significant and fairly close to each other, yet only one is chosen as representative of a particular dimension.
  - This decision should be based on the priori knowledge of theory that may suggest that one variable more than the others would logically be representative of the dimension.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - The approach of selecting a single surrogate variable as representative of the factor—although simple and maintaining the original variable—has several potential disadvantages.
  - It does not address the issue of measurement error encountered when using single measures.
  - It also runs the risk of potentially misleading results by selecting only a single variable to represent a perhaps more complex result.
  - For example, assume that variables representing price competitiveness, product quality, and value were all found to load highly on a single factor. The selection of any one of these separate variables would create different interpretations.
  - **Creating Summated Scales**
  - Summated scale - which is formed by combining several individual variables into a single composite measure.
  - In simple terms, all of the variables loading highly on a factor are combined, and the total—or more commonly the average score of the variables—is used as a replacement variable.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - A summated scale provides two specific benefits.
  - 1. First, it provides a means of overcoming to some extent the measurement error inherent in all measured variables.
  - Measurement error is the degree to which the observed values are not representative of the actual values due to any number of reasons, ranging from actual errors (e.g., data entry errors) to the inability of individuals to accurately provide information.
  - The impact of measurement error is to partially mask any relationships and make the estimation of multivariate models more difficult.
  - The summated scale reduces measurement error by using multiple indicators (variables) to reduce the reliance on a single response.
  - 2. A second benefit of the summated scale is its ability to represent the multiple aspects of a concept in a single measure.
  - Many times we employ more variables in our multivariate models in an attempt to represent the many facets of a concept that we know is quite complex.
  - It may complicate the interpretation of the results because of the redundancy in the items associated with the concept.
  - The summated scale, when properly constructed, does combine the multiple indicators into a single measure representing what is held in common across the set of measures.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - Four issues basic to the construction of any summated scale:
  - conceptual definition
  - dimensionality
  - reliability and
  - validity
  - **CONCEPTUAL DEFINITION**
  - The starting point for creating any summated scale is its conceptual definition.
  - The conceptual definition specifies the theoretical basis for the summated scale by defining the concept being represented in terms applicable to the research context.
  - In academic research, theoretical definitions are based on prior research that defines the character and nature of a concept.
  - In a managerial setting, specific concepts may be defined that relate to proposed objectives, such as image, value, or satisfaction.
  - In either instance, creating a summated scale is always guided by the conceptual definition specifying the type and character of the items that are candidates for inclusion in the scale.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - **Content validity** is the assessment of the correspondence of the variables to be included in a summated scale and its conceptual definition.
  - This form of validity, also known as face validity, subjectively assesses the correspondence between the individual items and the concept through ratings by expert judges, pretests with multiple subpopulations, or other means.
  - The objective is to ensure that the selection of scale items also include theoretical and practical considerations.
  - **DIMENSIONALITY**
  - An underlying assumption and essential requirement for creating a summated scale is that the items are unidimensional, meaning that they are strongly associated with each other and represent a single concept.
  - Factor analysis plays a pivotal role in making an empirical assessment of the dimensionality of a set of items by determining the number of factors and the loadings of each variable on the factor(s).
  - The test of unidimensionality is that each summated scale should consist of items loading highly on a single factor.
  - If a summated scale is proposed to have multiple dimensions, each dimension should be reflected by a separate factor.
  - We can assess unidimensionality with either exploratory factor analysis, or confirmatory factor analysis.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - **RELIABILITY**
  - Reliability is an assessment of the degree of consistency between multiple measurements of a variable.
  - One form of reliability is test–retest, by which consistency is measured between the responses for an individual at two points in time.
  - The objective is to ensure that responses are not too varied across time periods so that a measurement taken at any point in time is reliable.
  - A second and more commonly used measure of reliability is internal consistency, which applies to the consistency among the variables in a summated scale.
  - The rationale for internal consistency is that the individual items or indicators of the scale should all be measuring the same construct and thus be highly intercorrelated.
  - Because no single item is a perfect measure of a concept, we must rely on a series of diagnostic measures to assess internal consistency.
  - • The first measures we consider relate to each separate item, including the item-to-total correlation (the correlation of the item to the summated scale score) and the inter-item correlation (the correlation among items).
  - • The second type of diagnostic measure is the reliability coefficient, which assesses the consistency of the entire scale, with Cronbach's alpha being the most widely used measure.
  - The generally agreed upon lower limit for Cronbach's alpha is .70 , although it may decrease to .60 in exploratory research.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - • Also available are reliability measures derived from confirmatory factor analysis. Included in these measures are the composite reliability and the average variance extracted.
  - Any summated scale should be analyzed for reliability to ensure its appropriateness before proceeding to an assessment of its validity.
  - **VALIDITY**
  - Having ensured that a scale
  - (1) conforms to its conceptual definition,
  - (2) is unidimensional, and
  - (3) meets the necessary levels of reliability,
  - We must make one final assessment: scale validity.
  - **Validity** is the extent to which a scale or set of measures accurately represents the concept of interest.
  - We already described one form of validity—content or face validity—in the conceptual definitions.
  - Other forms of validity are measured empirically by the correlation between theoretically defined sets of variables.
  - The three most widely accepted forms of validity **are convergent, discriminant, and nomological validity**

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - • **Convergent validity** assesses the degree to which two measures of the same concept are correlated.
  - Here we may look for alternative measures of a concept and then correlate them with the summated scale. High correlations here indicate that the scale is measuring its intended concept.
  - • **Discriminant validity** is the degree to which two conceptually similar concepts are distinct.
  - The empirical test is again the correlation among measures, but this time the summated scale is correlated with a similar, but conceptually distinct, measure.
  - Now the correlation should be low, demonstrating that the summated scale is sufficiently different from the other similar concept.
  - • Finally, **nomological validity** refers to the degree that the summated scale makes accurate predictions of other concepts in a theoretically based model.
  - We must identify theoretically supported relationships from prior research or accepted principles and then assess whether the scale has corresponding relationships.
  - In summary, **convergent validity** confirms that the scale is correlated with other known measures of the concept;
  - **discriminant validity** ensures that the scale is sufficiently different from other similar concepts to be distinct; and
  - **nomological validity** determines whether the scale demonstrates the relationships shown to exist based on theory or prior research.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - **CALCULATING SUMMATED SCALES**
  - Calculating a summated scale is a straightforward process whereby the items comprising the summated scale (i.e., the items with high loadings from the factor analysis) are summed or averaged.
  - Whenever variables have both positive and negative loadings within the same factor, either the variables with the positive or the negative loadings must have their data values reversed.
  - Typically, the variables with the negative loadings are reverse scored so that the correlations, and the loadings, are now all positive within the factor.
  - Reverse scoring is the process by which the data values for a variable are reversed so that its correlations with other variables are reversed (i.e., go from negative to positive).
  - For example, on our scale of 0 to 10, we would reverse score a variable by subtracting the original value from 10 (i.e., reverse score = 10 – original value).
  - In this way, original scores of 10 and 0 now have the reversed scores of 0 and 10. All distributional characteristics are retained; only the distribution is reversed.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - **CALCULATING SUMMATED SCALES**
  - The purpose of reverse scoring is to prevent a canceling out of variables with positive and negative loadings. Let us use as an example of two variables with a negative correlation.
  - We are interested in combining V1 and V2, with V1 having a positive loading and V2 a negative loading.
  - If 10 is the top score on V1, the top score on V2 would be 0.
  - Now assume two cases.
  - In case 1, V1 has a value of 10 and V2 has a value of 0 (the best case).
  - In the second case, V1 has a value of 0 and V2 has a value of 10 (the worst case).
  - If V2 is not reverse scored, then the scale score calculated by adding the two variables for both cases 1 and 2 is 10, showing no difference, whereas we know that case 1 is the best and case 2 is the worst.
  - If we reverse score V2, however, the situation changes.
  - Now case 1 has values of 10 and 10 on V1 and V2, respectively, and case 2 has values of 0 and 0.
  - The summated scale scores are now 20 for case 1 and 0 for case 2, which distinguishes them as the best and worst situations.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - **Computing Factor Scores**
  - The third option for creating a smaller set of variables to replace the original set is the computation of factor scores.
  - Factor scores are also composite measures of each factor computed for each subject.
  - Conceptually the factor score represents the degree to which each individual scores high on the group of items with high loadings on a factor.
  - Thus, higher values on the variables with high loadings on a factor will result in a higher factor score.
  - The one key characteristic that differentiates a factor score from a summated scale is that the factor score is computed based on the factor loadings of all variables on the factor, whereas the summated scale is calculated by combining only selected variables.
  - Most statistical programs can easily compute factor scores for each respondent. By selecting the factor score option, these scores are saved for use in subsequent analyses.
  - The one disadvantage of factor scores is that they are not easily replicated across studies because they are based on the factor matrix, which is derived separately in each study.
  - Replication of the same factor matrix across studies requires substantial computational programming.

Factor Analysis Decision Process

- **STAGE 7: ADDITIONAL USES OF FACTOR ANALYSIS RESULTS**
  - **Selecting Among the Three Methods**
  - To select among the three data reduction options, the researcher must make a series of decisions, weighing the advantages and disadvantages of each approach with the research objectives.
  - The decision rule, therefore, would be as follows:
  - • If data are used only in the original sample or orthogonality must be maintained, factor scores are suitable.
  - • If generalizability or transferability is desired, then summated scales or surrogate variables are more appropriate. If the summated scale is a well-constructed, valid, and reliable instrument, then it is probably the best alternative.
  - • If the summated scale is untested and exploratory, with little or no evidence of reliability or validity, surrogate variables should be considered if additional analysis is not possible to improve the summated scale.

UNIT - III

Cluster Analysis

Cluster Analysis

- **Cluster analysis** is a group of multivariate techniques whose primary purpose is to group objects based on the characteristics they possess.
- It has been referred to as Q analysis, typology construction, classification analysis, and numerical taxonomy.
- This variety of names is due to the usage of clustering methods in such diverse disciplines as psychology, biology, sociology, economics, engineering, and business.
- Although the names differ across disciplines, the methods all have a common dimension: classification according to relationships among the objects being clustered.
- This common dimension represents the essence of all clustering approaches—the classification of data as suggested by natural groupings of the data themselves.
- Cluster analysis is comparable to factor analysis in its objective of assessing structure.
- **Cluster analysis differs from factor analysis**, however, in that cluster analysis groups objects, whereas factor analysis is primarily concerned with grouping variables.
- Additionally, factor analysis makes the groupings based on patterns of variation (correlation) in the data whereas cluster analysis makes groupings on the basis of distance (proximity).

Cluster Analysis

- **Cluster Analysis as a Multivariate Technique**
- Cluster analysis classifies objects (e.g., respondents, products, or other entities), on a set of user selected characteristics.
- The resulting clusters should exhibit high internal (within-cluster) homogeneity and high external (between-cluster) heterogeneity.
- Thus, if the classification is successful, the objects within clusters will be close together when plotted geometrically, and different clusters will be far apart.
- The variate in cluster analysis is determined quite differently from other multivariate techniques.
- Cluster analysis is the only multivariate technique that does not estimate the variate empirically but instead uses the variate as specified by the researcher.
- **Conceptual Development with Cluster Analysis**
- Cluster analysis has been used in every research setting imaginable.
- Ranging from the derivation of taxonomies in biology for grouping all living organisms, to psychological classifications based on personality and other personal traits, to segmentation analyses of markets, cluster analysis applications have focused largely on grouping individuals.
- However, cluster analysis can classify objects other than individual people, including the market structure, analyses of the similarities and differences among new products, and performance evaluations of firms to identify groupings based on the firms' strategies or strategic orientations.

Cluster Analysis

- The more common roles cluster analysis can play in conceptual development include the following:
- **Data reduction:**
  - A researcher may be faced with a large number of observations that are meaningless unless classified into manageable groups.
  - Cluster analysis can perform this data reduction procedure objectively by reducing the information from an entire population or sample to information about specific groups.
  - For example, if we can understand the attitudes of a population by identifying the major groups within the population, then we have reduced the data for the entire population into profiles of a number of groups.
- **Hypothesis generation:**
  - Cluster analysis is also useful when a researcher wishes to develop hypotheses concerning the nature of the data or to examine previously stated hypotheses.
  - For example, a researcher may believe that attitudes toward the consumption of diet versus regular soft drinks could be used to separate soft-drink consumers into logical segments or groups.
  - Cluster analysis can classify soft-drink consumers by their attitudes about diet versus regular soft drinks, and the resulting clusters, if any, can be profiled for demographic similarities and differences.

Cluster Analysis

- **Necessity of Conceptual Support in Cluster Analysis :**
  - Even if cluster analysis is being used in conceptual development as just mentioned, some conceptual rationale is essential.
  - The following are the most common criticisms that must be addressed by conceptual rather than empirical support:
  - **Cluster analysis is descriptive, atheoretical, and noninferential.**
  - Cluster analysis has no statistical basis upon which to draw inferences from a sample to a population, and many contend that it is only an exploratory technique.
  - Nothing guarantees unique solutions, because the cluster membership for any number of solutions is dependent upon many elements of the procedure, and many different solutions can be obtained by varying one or more elements.
  - **Cluster analysis will always create clusters, regardless of the actual existence of any structure in the data.**
  - When using cluster analysis, the researcher is making an assumption of some structure among the objects.
  - The researcher should always remember that just because clusters can be found does not validate their existence.
  - Only with strong conceptual support and then validation are the clusters potentially meaningful and relevant.

Cluster Analysis

- **Necessity of Conceptual Support in Cluster Analysis :**
  - **The cluster solution is not generalizable because it is totally dependent upon the variables used as the basis for the similarity measure.**
  - This criticism can be made against any statistical technique, but cluster analysis is generally considered more dependent on the measures used to characterize the objects than other multivariate techniques.
  - With the cluster variate completely specified by the researcher, the addition of spurious variables or the deletion of relevant variables can have a substantial impact on the resulting solution.
  - As a result, the researcher must be especially cognizant of the variables used in the analysis, ensuring that they have strong conceptual support.
  - Thus, in any use of cluster analysis the researcher must take particular care in ensuring that strong conceptual support predates the application of the technique.

Cluster Analysis

- **HOW DOES CLUSTER ANALYSIS WORK?**
  - Cluster analysis performs a task innate to all individuals—pattern recognition and grouping.
  - The human ability to process even slight differences in innumerable characteristics is a cognitive process inherent in human beings that is not easily matched with all of our technological advances.
  - Take for example the task of analyzing and grouping human faces. Even from birth, individuals can quickly identify slight differences in facial expressions and group different faces in homogeneous groups while considering hundreds of facial characteristics.
  - Yet we still struggle with facial recognition programs to accomplish the same task. The process of identifying natural groupings is one that can become quite complex rather quickly.
  - To demonstrate how cluster analysis operates, we examine a simple example that illustrates some of the key issues: measuring similarity, forming clusters, and deciding on the number of clusters that best represent structure.

Cluster Analysis

- **HOW DOES CLUSTER ANALYSIS WORK?**
  - **A Simple Example**
  - The nature of cluster analysis and the basic decisions on the part of the researcher will be illustrated by a simple example involving identification of customer segments in a retail setting.
  - Suppose a marketing researcher wishes to determine market segments in a community based on patterns of loyalty to brands and stores.
  - A small sample of seven respondents is selected as a pilot test of how cluster analysis is applied.
  - Two measures of loyalty—V1 (store loyalty) and V2 (brand loyalty)— were measured for each respondent on a 0–10 scale.
  - The values for each of the seven respondents are shown in Figure 1, along with a scatter diagram depicting each observation on the two variables.
  - The primary objective of cluster analysis is to define the structure of the data by placing the most similar observations into groups.

Cluster Analysis

- **HOW DOES CLUSTER ANALYSIS WORK?**
  - To accomplish this task, we must address three basic questions:
  - **1. How do we measure similarity?**
  - We require a method of simultaneously comparing observations on the two clustering variables (V1 and V2).
  - Several methods are possible, including the correlation between objects or perhaps a measure of their proximity in two-dimensional space such that the distance between observations indicates similarity.
  - **2. How do we form clusters?**
  - No matter how similarity is measured, the procedure must group those observations that are most similar into a cluster, thereby determining the cluster group membership of each observation for each set of clusters formed.
  - **3. How many groups do we form?**
  - The final task is to select one set of clusters as the final solution.
  - In doing so, the researcher faces a trade-off: fewer clusters and less homogeneity within clusters versus a larger number of clusters and more within-group homogeneity.
  - Yet as the number of clusters decreases, the heterogeneity within the clusters necessarily increases.
  - Thus, a balance must be made between defining the most basic structure (fewer clusters) that still achieves an acceptable level of heterogeneity between the clusters.

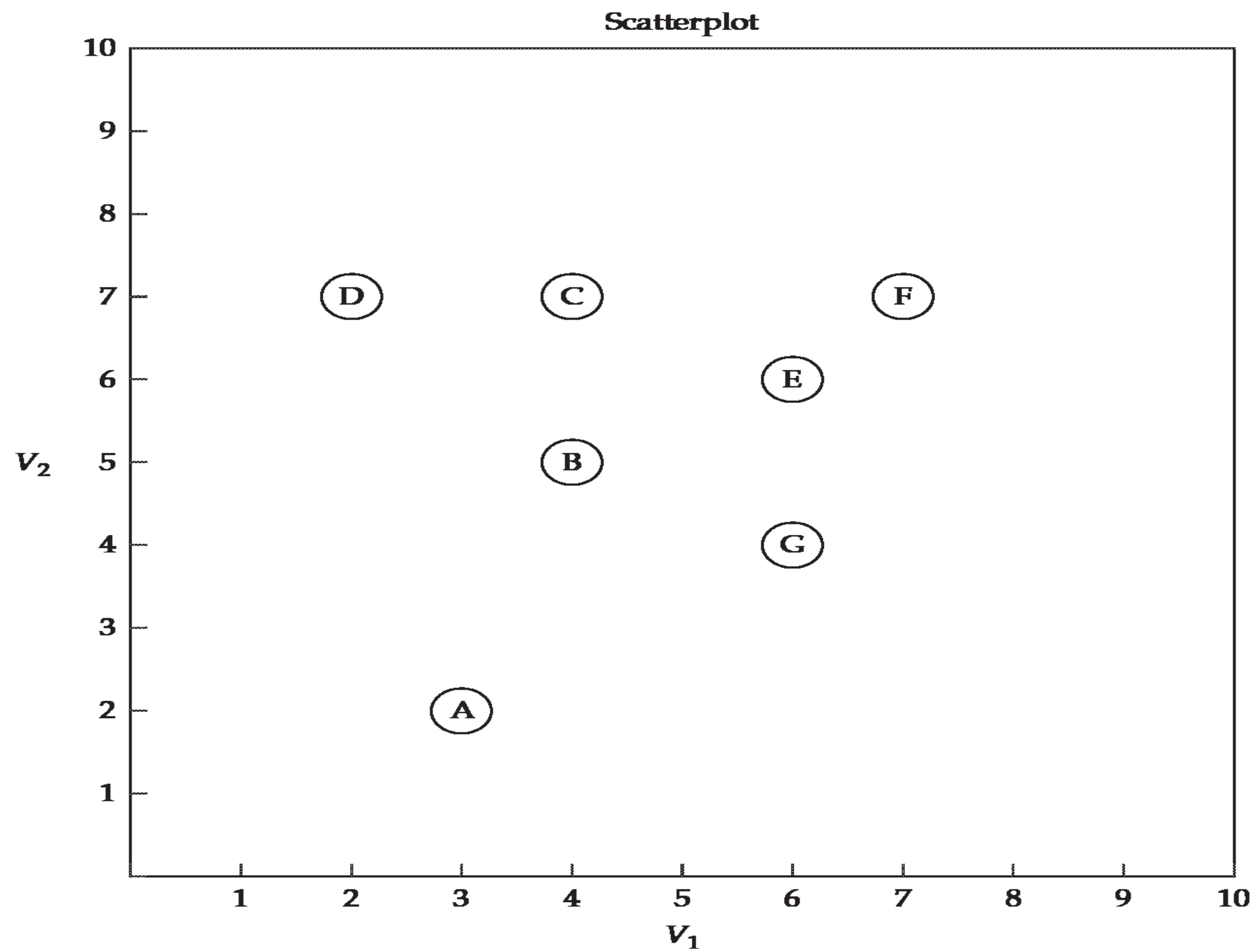
Cluster Analysis

Cluster Analysis

Data Values

Clustering Variable	Respondents						
	A	B	C	D	E	F	G
V_1	3	4	4	2	6	7	6
V_2	2	5	7	7	6	7	4

Cluster Analysis



Cluster Analysis

- **MEASURING SIMILARITY :**
  - The first task is developing some measure of similarity between each object to be used in the clustering process.
  - Similarity represents the degree of correspondence among objects across all of the characteristics used in the analysis.
  - In a way, similarity measures are more descriptively dissimilarity measures in that smaller numbers represent greater similarity and larger numbers represent less similarity.
  - Similarity must be determined between each of the seven observations (respondents A–G) to enable each observation to be compared to each other.
  - In this example, similarity will be measured according to the Euclidean (straight-line) distance between each pair of observations (see Table 1) based on the two characteristics (V1 and V2).
  - In this two-dimensional case (where each characteristic forms one axis of the graph) we can view distance as the proximity of each point to the others.
  - In using distance as the measure of proximity, we must remember that smaller distances indicate greater similarity, such that observations E and F are the most similar (1.414), and A and F are the most dissimilar (6.403).

Cluster Analysis

- **MEASURING SIMILARITY :**

TABLE 1 Proximity Matrix of Euclidean Distances Between Observations

	Observation						
Observation	**A**	**B**	**C**	**D**	**E**	**F**	**G**
A	—						
B	3.162	—					
C	5.099	2.000	—				
D	5.099	2.828	2.000	—			
E	5.000	2.236	2.236	4.123	—		
F	6.403	3.606	3.000	5.000	1.414	—	
G	3.606	2.236	3.606	5.000	2.000	3.162	—

Cluster Analysis

- **FORMING CLUSTERS :**
  - With similarity measures calculated, we now move to forming clusters based on the similarity measure of each observation.
  - Typically we form a number of cluster solutions (a two-cluster solution, a three-cluster solution, etc.).
  - Once clusters are formed, we then select the final cluster solution from the set of possible solutions.
  - First we will discuss how clusters are formed and then examine the process for selecting a final cluster solution.
  - Having calculated the similarity measure, we must develop a procedure for forming clusters.
  - We use this simple rule:
  - Identify the two most similar (closest) observations not already in the same cluster and combine them.
  - We apply this rule repeatedly to generate a number of cluster solutions, starting with each observation as its own “cluster” and then combining two clusters at a time until all observations are in a single cluster.
  - This process is termed a hierarchical procedure because it moves in a stepwise fashion to form an entire range of cluster solutions. It is also an agglomerative method because clusters are formed by combining existing clusters.

Cluster Analysis

- **FORMING CLUSTERS :**

TABLE 2 Agglomerative Hierarchical Clustering Process

	AGGLOMERATION PROCESS		CLUSTER SOLUTION		
Step	*Minimum Distance Between Unclustered Observations*[a]	*Observation Pair*	*Cluster Membership*	*Number of Clusters*	*Overall Similarity Measure (Average Within-Cluster Distance)*
	Initial Solution		**(A) (B) (C) (D) (E) (F) (G)**	**7**	**0**
1	1.414	E-F	(A) (B) (C) (D) (E-F) (G)	6	1.414
2	2.000	E-G	(A) (B) (C) (D) (E-F-G)	5	2.192
3	2.000	C-D	(A) (B) (C-D) (E-F-G)	4	2.144
4	2.000	B-C	(A) (B-C-D) (E-F-G)	3	2.234
5	2.236	B-E	(A) (B-C-D-E-F-G)	2	2.896
6	3.162	A-B	(A-B-C-D-E-F-G)	1	3.420

[a]Euclidean distance between observations

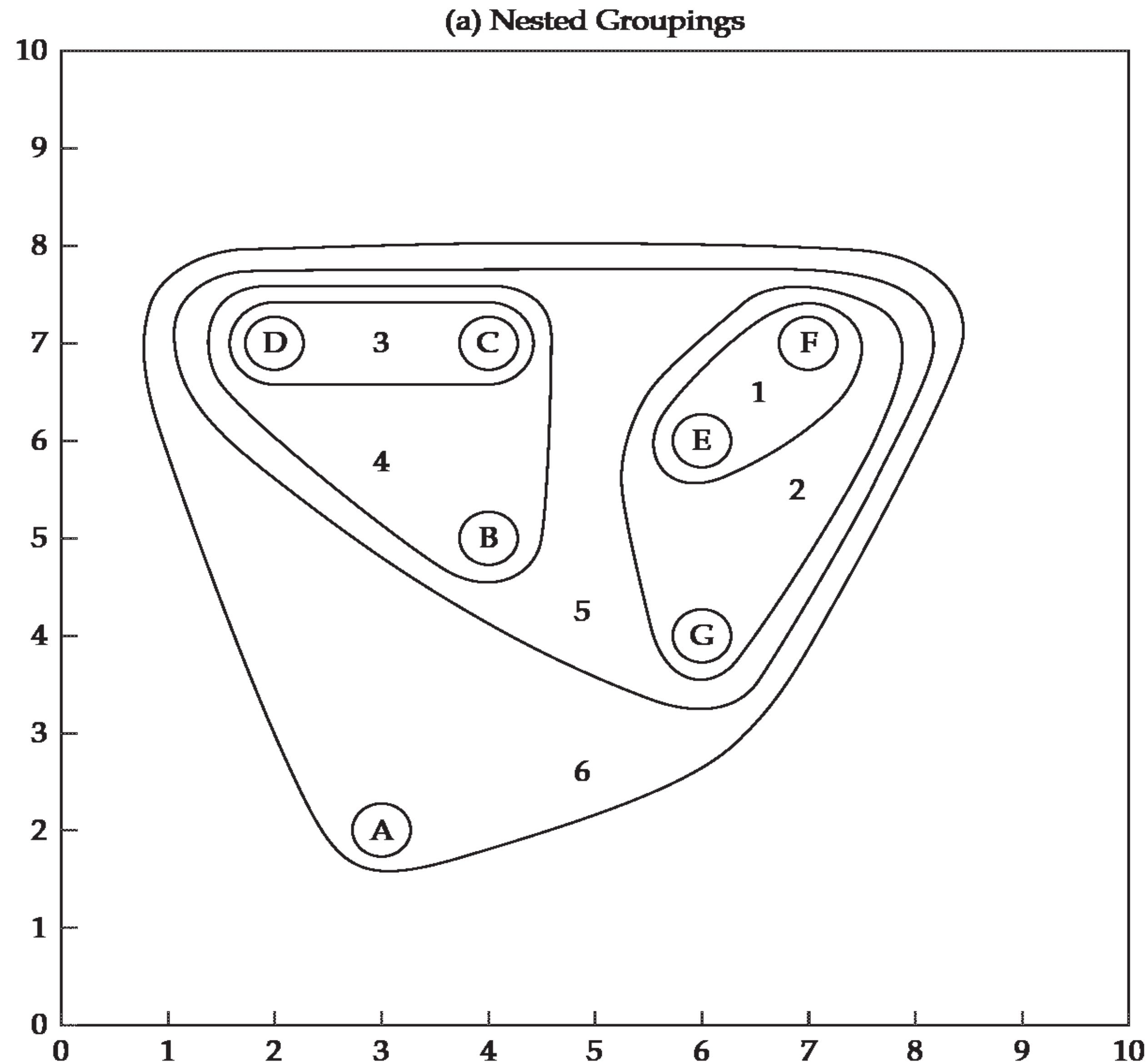
Cluster Analysis

- **FORMING CLUSTERS :**
  - The six-step clustering process is described here:
  - **Step 1:** Identify the two closest observations (E and F) and combine them into a cluster, moving from seven to six clusters.
  - **Step 2:** Find the next closest pairs of observations. In this case, three pairs have the same distance of 2.000 (E-G, C-D, and B-C). For our purposes, choose the observations E-G. G is a single-member cluster, but E was combined in the prior step with F. So, the cluster formed at this stage now has three members: G, E, and F.
  - **Step 3:** Combine the single-member clusters of C and D so that we now have four clusters.
  - **Step 4:** Combine B with the two-member cluster C-D that was formed in step 3. At this point, we now have three clusters: cluster 1 (A), cluster 2 (B, C, and D), and cluster 3 (E, F, and G).
  - **Step 5:** Combine the two three-member clusters into a single six-member cluster. The next smallest distance is 2.236 for three pairs of observations (E-B, B-G, and C-E). We use only one of these distances, however, as each observation pair contains a member from each of the two existing clusters (B, C, and D versus E, F, and G).
  - **Step 6**: Combine observation A with the remaining cluster (six observations) into a single cluster at a distance of 3.162. You will note that distances smaller or equal to 3.162 are not used because they are between members of the same cluster.

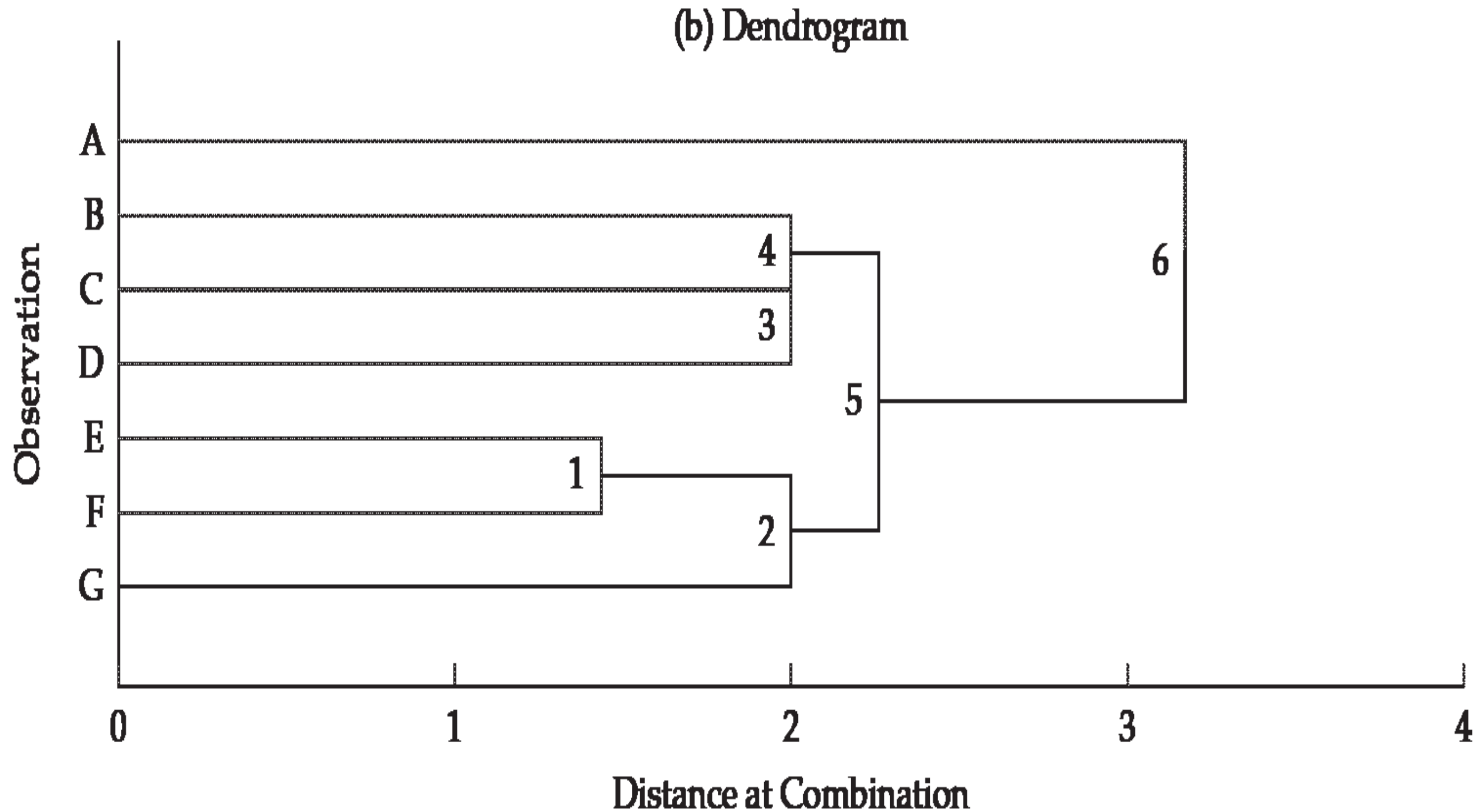
Cluster Analysis

- **FORMING CLUSTERS :**
  - The hierarchical clustering process can be portrayed graphically in several ways. Figure 2 illustrates two such methods.
  - First, because the process is hierarchical, the clustering process can be shown as a series of nested groupings (see Figure 2a).
  - This process, however, can represent the proximity of the observations for only two or three clustering variables in the scatterplot or three dimensional graph.
  - A more common approach is a **dendrogram**, which represents the clustering process in a treelike graph.
  - The horizontal axis represents the agglomeration coefficient, in this instance the distance used in joining clusters.
  - This approach is particularly useful in identifying outliers, such as observation A.
  - It also depicts the relative size of varying clusters, although it becomes unwieldy when the number of observations increases.

Cluster Analysis



Cluster Analysis



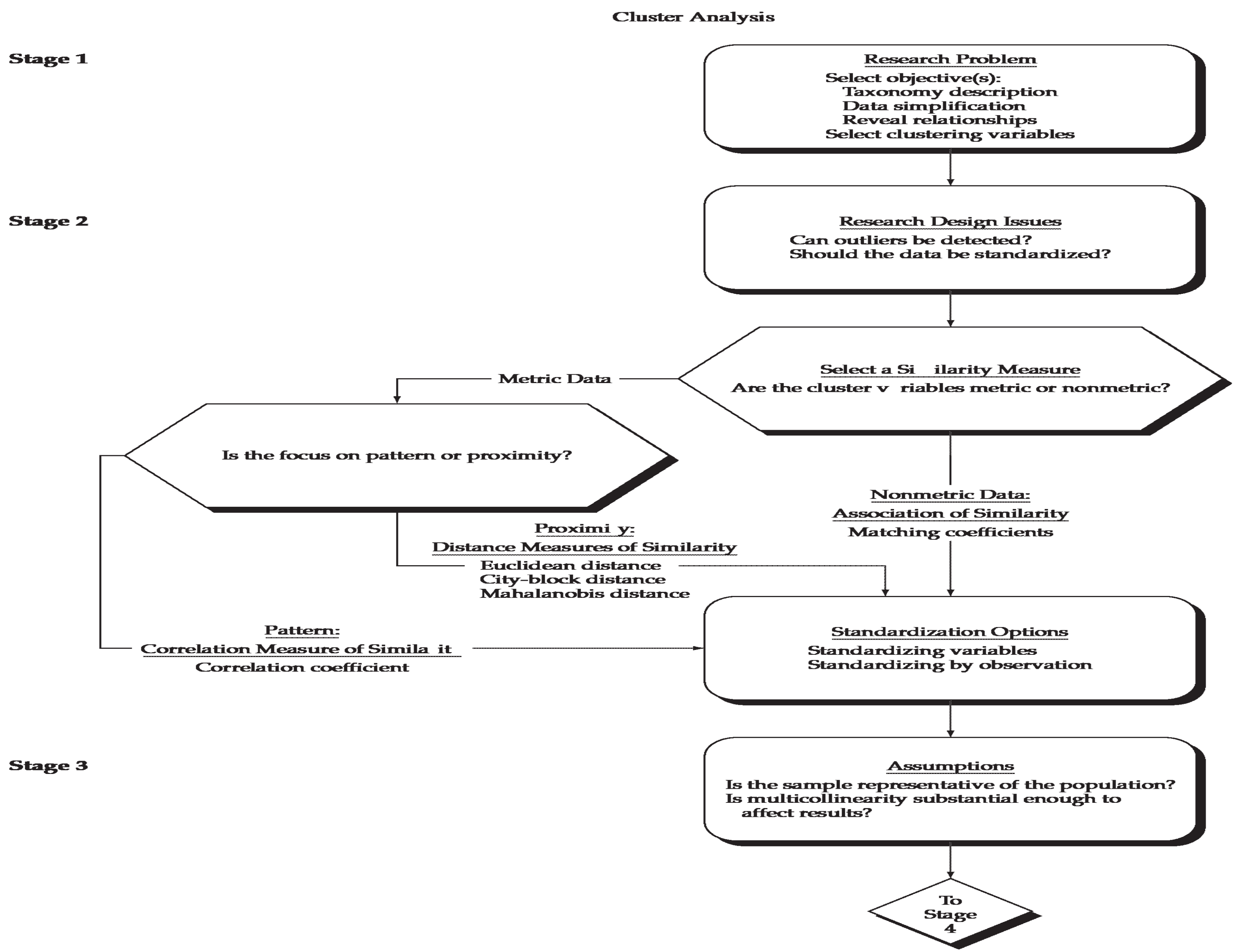
Cluster Analysis

- **DETERMINING THE NUMBER OF CLUSTERS IN THE FINAL SOLUTION :**
  - A hierarchical method results in a number of cluster solutions—in this case starting with a seven-cluster solution and ending in a one-cluster solution.
  - If all observations are treated as their own unique cluster, no data reduction has taken place and no true segments have been found.
  - The goal is identifying segments by combining observations, but at the same time introducing only small amounts of heterogeneity.
  - **Measuring Heterogeneity :**
  - Any measure of heterogeneity of a cluster solution should represent the overall diversity among observations in all clusters.
  - In the initial solution of an agglomerative approach where all observations are in separate clusters, no heterogeneity exists.
  - As observations are combined to form clusters, heterogeneity increases.
  - The measure of heterogeneity thus should start with a value of zero and increase to show the level of heterogeneity as clusters are combined.

Cluster Analysis

- **DETERMINING THE NUMBER OF CLUSTERS IN THE FINAL SOLUTION :**
  - **Measuring Heterogeneity :**
  - In this example, we use a simple measure of heterogeneity:
  - the average of all distances between observations within clusters (see Table 2).
  - As already described, the measure should increase as clusters are combined:
  - • In the initial solution with seven clusters, our overall similarity measure is 0—no observation is paired with another.
  - • **Six clusters:** The overall similarity is the distance between the two observations (1.414) joined in step 1.
  - • **Five clusters:** Step 2 forms a three-member cluster (E, F, and G), so that the overall similarity measure is the mean of the distances between E and F (1.414), E and G (2.000), and F and G (3.162), for an average of 2.192.
  - • **Four clusters:** In the next step a new two-member cluster is formed with a distance of 2.000, which causes the overall average to fall slightly to 2.144.
  - • **Three, two, and one clusters:** The final three steps form new clusters in this manner until a single-cluster solution is formed (step 6), in which the average of all distances in the distance matrix is 3.420.

Cluster Analysis Decision Process



Cluster Analysis Decision Process

- Cluster analysis, like the other multivariate techniques discussed earlier, can be viewed from a six-stage model-building approach.
- Starting with research objectives that can be either exploratory or confirmatory, the design of a cluster analysis deals with the following:
  - • Partitioning the data set to form clusters and selecting a cluster solution
  - • Interpreting the clusters to understand the characteristics of each cluster and develop a name or label that appropriately defines its nature
  - • Validating the results of the final cluster solution (i.e., determining its stability and generalizability), along with describing the characteristics of each cluster to explain how they may differ on relevant dimensions such as demographics
- **Stage 1: Objectives of Cluster Analysis**
  - The primary goal of cluster analysis is to partition a set of objects into two or more groups based on the similarity of the objects for a set of specified characteristics (cluster variate).
  - In fulfilling this basic objective, the researcher must address two key issues:
  - the research questions being addressed in this analysis and
  - the variables used to characterize objects in the clustering process.

Cluster Analysis Decision Process

- **RESEARCH QUESTIONS IN CLUSTER ANALYSIS**
- In forming homogeneous groups, cluster analysis may address any combination of three basic research questions:
- **1. Taxonomy description.**
  - The most traditional use of cluster analysis has been for exploratory purposes and the formation of a taxonomy—an empirically based classification of objects.
  - Cluster analysis can also generate hypotheses related to the structure of the objects.
  - Finally, although viewed principally as an exploratory technique, cluster analysis can be used for confirmatory purposes.
  - In such cases, a proposed typology (theoretically based classification) can be compared to that derived from the cluster analysis.
- **2. Data simplification.**
  - By defining structure among the observations, cluster analysis also develops a simplified perspective by grouping observations for further analysis.
  - Whereas factor analysis attempts to provide dimensions or structure to variables, cluster analysis performs the same task for observations.
  - Thus, instead of viewing all of the observations as unique, they can be viewed as members of clusters and profiled by their general characteristics

Cluster Analysis Decision Process

- **3. Relationship identification.**
  - With the clusters defined and the underlying structure of the data represented in the clusters, the researcher has a means of revealing relationships among the observations that typically is not possible with the individual observations.
  - Whether analyses such as discriminant analysis are used to empirically identify relationships, or the groups are examined by more qualitative methods, the simplified structure from cluster analysis often identifies relationships or similarities and differences not previously revealed.
- **SELECTION OF CLUSTERING VARIABLES**
  - The objectives of cluster analysis cannot be separated from the selection of variables used to characterize the objects being clustered.
  - Whether the objective is exploratory or confirmatory, the researcher effectively constrains the possible results by the variables selected for use.
  - The derived clusters reflect the inherent structure of the data and are defined only by the variables.
  - Thus, selecting the variables to be included in the cluster variate must be done with regard to theoretical and conceptual as well as practical considerations.

Cluster Analysis Decision Process

- **Conceptual Considerations**
  - Any application of cluster analysis must have some rationale upon which variables are selected.
  - Whether the rationale is based on an explicit theory, past research, or supposition, the researcher must realize the importance of including only those variables that
  - (1) characterize the objects being clustered
  - (2) relate specifically to the objectives of the cluster analysis.
  - The cluster analysis technique has no means of differentiating relevant from irrelevant variables and derives the most consistent, yet distinct, groups of objects across all variables.
  - Thus, one should never include variables indiscriminately.
  - Instead, carefully choose the variables with the research objective as the criterion for selection.
- **Practical Considerations**
  - Cluster analysis can be affected dramatically by the inclusion of only one or two inappropriate or undifferentiated variables.
  - We always encouraged to examine the results and to eliminate the variables that are not distinctive (i.e., that do not differ significantly) across the derived clusters.
  - This procedure enables the cluster techniques to maximally define clusters based only on those variables exhibiting differences across the objects.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - With the objectives defined and variables selected, we must address four questions before starting the partitioning process:
  - 1. Is the sample size adequate?
  - 2. Can outliers be detected and, if so, should they be deleted?
  - 3. How should object similarity be measured?
  - 4. Should the data be standardized?
  - Many different approaches can be used to answer these questions. However, none of them has been evaluated sufficiently to provide a definitive answer to any of these questions, and unfortunately, many of the approaches provide different results for the same data set.
  - Thus, cluster analysis, along with factor analysis, is as much an art as a science.
  - **SAMPLE SIZE**
  - The issue of sample size in cluster analysis does not relate to any statistical inference issues (i.e., statistical power).
  - Instead the sample size must be large enough to provide sufficient representation of small groups within the population and represent the underlying structure.
  - Small groups will naturally appear as small numbers of observations, particularly when the sample size is small.
  - For example, when a sample contains only 100 or fewer observations, groups that actually make up 10 percent of the population may be represented by only one or two observations due to the sampling process.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - In such instances the distinction between outliers and representatives of a small group is much harder to make.
  - Larger samples increase the chance that small groups will be represented by enough cases to make their presence more easily identified.
  - As a result, the researcher should ensure the sample size is sufficiently large to adequately represent all of the relevant groups of the population.
  - Obviously, if the analysis objectives require identification of small groups within the population, the researcher should strive for larger sample sizes.
  - If we are interested only in larger groups (e.g., major segments for promotional campaigns), however, then the distinction between an outlier and a representative of a small group is less important and they can both be handled in a similar manner.

Cluster Analysis Decision Process

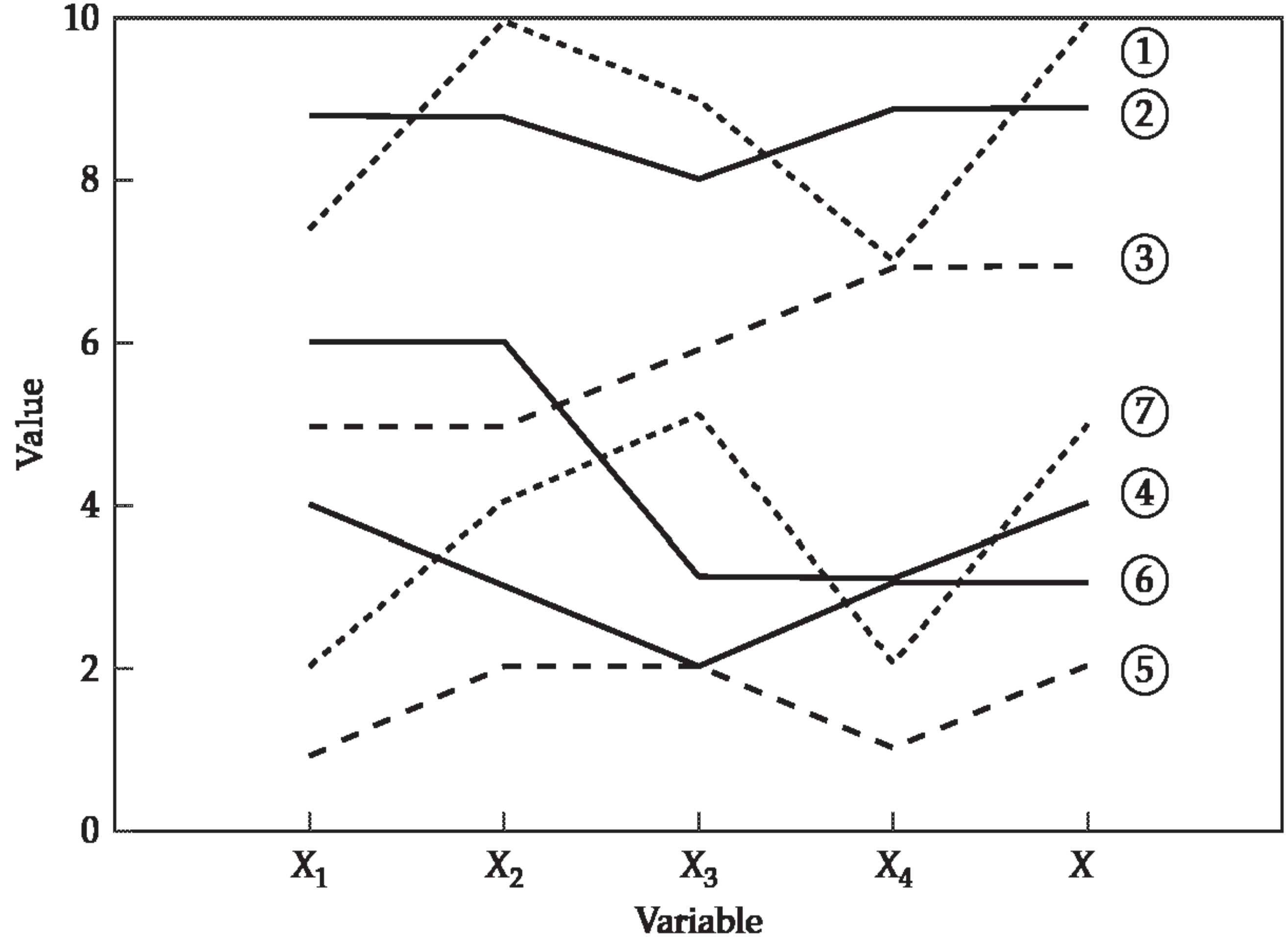
- **Stage 2: Research Design in Cluster Analysis**
  - **DETECTING OUTLIERS**
  - In its search for structure, cluster analysis is sensitive to the inclusion of irrelevant variables. But cluster analysis is also sensitive to outliers (objects different from all others).
  - Outliers can represent either:
  - Truly abnormal observations that are not representative of the general population.
    - In this case they distort the actual structure & make the derived clusters unrepresentative of the population structure → DELETE IT.
  - Representative observations of small or insignificant groups of objects within the population
    - In the second case, the outlier is removed so that the resulting clusters more accurately represent the relevant segments in the populations. → DELETE IT.
  - An undersampling of actual group(s) in the population that causes poor representation of the group(s) in the sample
    - However, in the third case the outliers should be included in the cluster solutions, even if they are underrepresented in the sample, because they represent valid and relevant groups. → KEEP IT.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **DETECTING OUTLIERS**
  - For this reason, a preliminary screening for outliers is always necessary.
  - **Graphical Approaches**
  - One of the simplest ways to screen data for outliers is to prepare a graphic profile diagram, listing the variables along the horizontal axis and the variable values along the vertical axis.
  - Each point on the graph represents the value of the corresponding variable, and the points are connected to facilitate visual interpretation.
  - Profiles for all objects are then plotted on the graph, a line for each object.
  - Outliers are those respondents that have very different profiles from the more typical respondents.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **PROFILE DIAGRAM :**



Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **DETECTING OUTLIERS**
  - **Empirical Approaches**
  - Although quite simple, the graphical procedures become difficult with a large number of objects and even more difficult as the number of variables increases.
  - Moreover, detecting outliers must extend beyond a univariate approach, because outliers also may be defined in a multivariate sense as having unique profiles across an entire set of variables that distinguish them from all of the other observations.
  - As a result, an empirical measure is needed to facilitate comparisons across objects.
  - Another approach is to identify outliers through the measures of similarity.
  - The most obvious examples of outliers are single observations that are the most dissimilar to the other observations.
  - Before the analysis, the similarities of all observations can be compared to the overall group centroid (typical respondent).
  - Isolated observations showing great dissimilarity can be dropped. Clustering patterns can also be observed once the cluster program has been run.
  - Thus, emphasis should be placed on identifying outliers before the analysis begins. Since, they can affect the result of how clusters are formed.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **MEASURING SIMILARITY**
  - The concept of similarity is fundamental to cluster analysis.
  - Interobject similarity is an empirical measure of correspondence, or resemblance, between objects to be clustered.
  - Comparing the two interdependence techniques (factor analysis and cluster analysis) will demonstrate how similarity works to define structure in both instances.
  - In factor analysis :
    - the correlation matrix between all pairs of variables was used to group variables into factors.
    - The correlation coefficient represented the similarity of each variable to another variable when viewed across all observations.
    - Thus, factor analysis grouped together variables that had high correlations among themselves.
  - A comparable process occurs in cluster analysis :
    - Here, the similarity measure is calculated for all pairs of objects, with similarity based on the profile of each observation across the characteristics specified by the researcher.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
    - In this way, any object can be compared to any other object through the similarity measure, just as we used correlations between variables in factor analysis.
    - The cluster analysis procedure then proceeds to group similar objects together into clusters.
  - Interobject similarity can be measured in a variety of ways, but three methods dominate the applications of cluster analysis:
    - correlational measures, distance measures, and association measures.
  - Both the correlational and distance measures require metric data, whereas the association measures are for nonmetric data.
  - **Correlational Measures**
  - The interobject measure of similarity that probably comes to mind first is the correlation coefficient between a pair of objects measured on several variables.
  - In effect, instead of correlating two sets of variables, we invert the data matrix so that the columns represent the objects and the rows represent the variables.
  - Thus, the correlation coefficient between the two columns of numbers is the correlation (or similarity) between the profiles of the two objects.
  - High correlations indicate similarity (the correspondence of patterns across the characteristics) and low correlations denote a lack of it.
  - This procedure is also followed in the application of Q-type factor analysis.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**

TABLE 3 Calculating Correlational and Distance Measures of Similarity

Original Data

Case	X_1	X_2	X_3	X_4	X_5
1	7	10	9	7	10
2	9	9	8	9	9
3	5	5	6	7	7
4	6	6	3	3	4
5	1	2	2	1	2
6	4	3	2	3	3
7	2	4	5	2	5

Similarity Measure: Correlation

	Case						
Case	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	1.00						
2	−.147	1.00					
3	.000	.000	1.00				
4	.087	.516	−.824	1.00			
5	.963	−.408	.000	— 060	1.00		
6	−.466	.791	−.354	699	−.645	1.00	
7	.891	−.516	.165	— 239	.963	−.699	1.00

Similarity Measure: Euclidean Distance

	Case						
Case	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	*nc*						
2	3.32	*nc*					
3	6.86	6.63	*nc*				
4	10.25	10.20	6.00	*nc*			
5	15.78	16.19	10.10	7.07	*nc*		
6	13 1	13.00	7.28	3.87	3.87	*nc*	
7	11 27	12.16	6.32	5.10	4.90	4.36	*nc*

nc = distances ot calculated.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - A correlational measure of similarity does not look at the observed mean value, or magnitude, but instead at the patterns of movement seen as one traces the data for each case over the variables measured; in other words, the similarity in the profiles for each case.
  - In Table 3, which contains the correlations among these seven observations, we can see two distinct groups.
  - First, cases 1, 5, and 7 all have similar patterns and corresponding high positive correlations.
  - Likewise, cases 2, 4, and 6 also have high positive correlations among themselves but low or negative correlations with the other observations.
  - Case 3 has low or negative correlations with all other cases, thereby perhaps forming a group by itself.
  - Correlations represent patterns across the variables rather than the magnitudes, which are comparable to a Q-type factor analysis.
  - Correlational measures are rarely used because emphasis in most applications of cluster analysis is on the magnitudes of the objects, not the patterns of values.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **Distance Measures**
  - Even though correlational measures have an intuitive appeal and are used in many other multivariate techniques, they are not the most commonly used measure of similarity in cluster analysis.
  - Instead, the most commonly used measures of similarity are distance measures.
  - These distance measures represent similarity as the proximity of observations to one another across the variables in the cluster variate.
  - Distance measures are actually a measure of dissimilarity, with larger values denoting lesser similarity.
  - Distance is converted into a similarity measure by using an inverse relationship.
  - A simple illustration of using distance measures was shown in our hypothetical example, in which clusters of observations were defined based on the proximity of observations to one another when each observation's scores on two variables were plotted graphically.
  - Even though proximity may seem to be a simple concept, several distance measures are available, each with specific characteristics.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **Distance Measures**
  - **Euclidean distance**
  - Euclidean distance is the most commonly recognized measure of distance, many times referred to as straight-line distance.
  - Suppose that two points in two dimensions have coordinates (X1, Y1) and (X2, Y2), respectively. The Euclidean distance between the points is the length of the hypotenuse of a right triangle, as calculated by the formula under the figure.
  - This concept is easily generalized to more than two variables.

$$D(i,j) = \sqrt{A^2 + B^2} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2}.$$

  - **Squared (or absolute) Euclidean distance**
  - Squared (or absolute) Euclidean distance is the sum of the squared differences without taking the square root.
  - The squared Euclidean distance has the advantage of not having to take the square root, which speeds computations markedly.
  - It is the recommended distance measure for the centroid and Ward's methods of clustering.

$$D_2(i,j) = A^2 + B^2 = (X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2.$$

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **Distance Measures**
  - **City-block (Manhattan) distance**
  - City-block (Manhattan) distance is not based on Euclidean distance.
  - Instead, it uses the sum of the absolute differences of the variables (i.e., the two sides of a right triangle rather than the hypotenuse).
  - This procedure is the simplest to calculate, but may lead to invalid clusters if the clustering variables are highly correlated.

$$D_3(i,j) = |A| + |B| = |X_{1i} - X_{1j}| + |X_{2i} - X_{2j}|.$$

  - **Mahalanobis distance**
  - Mahalanobis distance (D2) is a generalized distance measure that accounts for the correlations among variables in a way that weights each variable equally.
  - It also relies on standardized variables and will be discussed in more detail in a following section.
  - Although desirable in many situations, it is not available as a proximity measure in either SAS or SPSS.
  - Other distance measures are available in many clustering programs.
  - We are encouraged to explore alternative cluster solutions obtained when using different distance measures in an effort to best represent the underlying data patterns.
  - Although these distance measures are said to represent similarity, in a very real sense they better represent dissimilarity, because higher values typically mean relatively less similarity.
  - Greater distance means observations are less similar.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **Comparison to Correlational Measures**
  - Distance measures focus on the magnitude of the values and portray as similar the objects that are close together, even if they have different patterns across the variables.
  - In contrast, correlation measures focus on the patterns across the variables and do not consider the magnitude of the differences between objects.
  - Let us look at our seven observations to see how these approaches differ.
  - Table 3 contains the values for the seven observations on the five variables (X1 to X5), along with both distance and correlation measures of similarity.
  - Cluster solutions using either similarity measure seem to indicate three clusters, but the membership in each cluster is quite different.
  - With smaller distances representing greater similarity, we see that cases 1 and 2 form one group (distance of 3.32), and cases 4, 5, 6, and 7 (distances ranging from 3.87 to 7.07) make up another group.
  - The distinctiveness of these two groups from each other is shown in that the smallest distance between the two clusters is 10.20.
  - These two clusters represent observations with higher versus lower values.
  - A third group, consisting of only case 3, differs from the other two groups because it has values that are both low and high.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **Comparison to Correlational Measures**
  - Using the correlation as the measure of similarity, three clusters also emerge. First, cases 1, 5, and 7 are all highly correlated (.891 to .963), as are cases 2, 4, and 6 (.516 to .791).
  - Moreover, the correlations between clusters are generally close to zero or even negative.
  - Finally, case 3 is again distinct from the other two clusters and forms a single-member cluster.
  - **Which Distance Measure Is Best?**
  - In attempting to select a particular distance measure, the researcher should remember the following caveats:
  - Different distance measures or a change in the scales of the variables may lead to different cluster solutions.
  - Thus, it is advisable to use several measures and compare the results with theoretical or known patterns.
  - When the variables are correlated (either positively or negatively), the Mahalanobis distance measure is likely to be the most appropriate because it adjusts for correlations and weights all variables equally.
  - Alternatively, we may wish to avoid using highly redundant variables as input to cluster analysis.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **Association Measures**
  - Association measures of similarity are used to compare objects whose characteristics are measured only in nonmetric terms (nominal or ordinal measurement).
  - As an example, respondents could answer yes or no on a number of statements.
  - An association measure could assess the degree of agreement or matching between each pair of respondents.
  - The simplest form of association measure would be the percentage of times agreement occurred (both respondents said yes to a question or both said no) across the set of questions.
  - Extensions of this simple matching coefficient have been developed to accommodate multicategory nominal variables and even ordinal measures.
  - **Selecting a Similarity Measure**
  - Although three different forms of similarity measures are available, the most frequently used and preferred form is the distance measure for several reasons.
  - First, the distance measure best represents the concept of proximity, which is fundamental to cluster analysis. Correlational measures, although having widespread application in other techniques, represent patterns rather than proximity.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **Selecting a Similarity Measure**
  - Second, cluster analysis is typically associated with characteristics measured by metric variables.
  - In some applications, nonmetric characteristics dominate, but most often the characteristics are represented by metric measures making distance again the preferred measure.
  - **STANDARDIZING THE DATA**
  - Clustering variables that are not all of the same scale should be standardized whenever necessary to avoid instances where a variable's influence on the cluster solution is greater than it should be.
  - We will now examine several approaches to standardization:
  - **Standardizing the Variables**
  - The most common form of standardization is the conversion of each variable to standard scores (also known as Z scores) by subtracting the mean and dividing by the standard deviation for each variable.
  - This option can be found in all computer programs and many times is even directly included in the cluster analysis procedure.
  - The process converts each raw data score into a standardized value with a mean of 0 and a standard deviation of 1, and in turn, eliminates the bias introduced by the differences in the scales of the several attributes or variables used in the analysis.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **Standardizing the Variables**
  - There are two primary benefits from standardization.
  - First, it is much easier to compare between variables because they are on the same scale (a mean of 0 and standard deviation of 1).
  - Positive values are above the mean, and negative values are below.
  - The magnitude represents the number of standard deviations the original value is from the mean.
  - Second, no difference occurs in the standardized values when only the scale changes.
  - For example, when we standardize a measure of time duration, the standardized values are the same whether measured in minutes or seconds.
  - Thus, using standardized variables truly eliminates the effects due to scale differences not only across variables, but for the same variable as well.
  - The need for standardization is minimized when all of the variables are measured on the same response scale (e.g., a series of attitudinal questions), but becomes quite important whenever variables using quite different measurement scales are included in the cluster variate.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **Using a Standardized Distance Measure**
  - A measure of Euclidean distance that directly incorporates a standardization procedure is the Mahalanobis distance (D2).
  - The Mahalanobis approach not only performs a standardization process on the data by scaling in terms of the standard deviations but it also sums the pooled within-group variance–covariance, which adjusts for correlations among the variables.
  - Highly correlated sets of variables in cluster analysis can implicitly overweight one set of variables in the clustering procedures.
  - In short, the Mahalanobis generalized distance procedure computes a distance measure between objects comparable to R2 in regression analysis.
  - Although many situations are appropriate for use of the Mahalanobis distance, not all programs include it as a measure of similarity.
  - In such cases, we usually selects the squared Euclidean distance.

Cluster Analysis Decision Process

- **Stage 2: Research Design in Cluster Analysis**
  - **Standardizing by Observation**
  - Suppose we collected a number of ratings on a 10-point scale of the importance for several attributes used in purchase decisions for a product.
  - We could apply cluster analysis and obtain clusters, but one distinct possibility is that what we would get are clusters of people who said everything was important, some who said everything had little importance, and perhaps some clusters in between.
  - What we are seeing are patterns of responses specific to an individual.
  - These patterns may reflect a specific way of responding to a set of questions, such as yes or no.
  - If we want to identify groups according to their response style and even control for these patterns, then the typical standardization through calculating Z scores is not appropriate.
  - In this instance, standardizing by respondent would standardize each question not to the sample's average but instead to that respondent's average score.
  - This within-case or row-centering standardization can be quite effective in removing response style effects and is especially suited to many forms of attitudinal data.

Cluster Analysis Decision Process

- **Stage 3: Assumptions in Cluster Analysis**
  - Cluster analysis is not a statistical inference technique in which parameters from a sample are assessed as representing a population.
  - Instead, cluster analysis is a method for quantifying the structural characteristics of a set of observations.
  - As such, it has strong mathematical properties but not statistical foundations.
  - The requirements of normality, linearity, and homoscedasticity that were so important in other techniques really have little bearing on cluster analysis.
  - We must focus, however, on two other critical issues: representativeness of the sample and multicollinearity among variables in the cluster variate.
  - **REPRESENTATIVENESS OF THE SAMPLE**
  - Rarely does we have a census of the population to use in the cluster analysis.
  - Usually, a sample of cases is obtained and the clusters are derived in the hope that they represent the structure of the population.
  - We must therefore be confident that the obtained sample is truly representative of the population.
  - As mentioned earlier, outliers may really be only an undersampling of divergent groups that, when discarded, introduce bias in the estimation of structure.
  - We must realize that cluster analysis is only as good as the representativeness of the sample.
  - Therefore, all efforts should be made to ensure that the sample is representative and the results are generalizable to the population of interest.

Cluster Analysis Decision Process

- **Stage 3: Assumptions in Cluster Analysis**
  - **MULTICOLLINEARITY**
  - Multicollinearity is a state of very high intercorrelations or inter-associations among the independent variables. It is therefore a type of disturbance in the data, and if present in the data the statistical inferences made about the data may not be reliable.
  - **Reasons why multicollinearity occurs :**
  - It is caused by an inaccurate use of dummy variables.
  - It is caused by the inclusion of a variable which is computed from other variables in the data set.
  - Multicollinearity can also result from the repetition of the same kind of variable.
  - Generally occurs when the variables are highly correlated to each other.

Cluster Analysis Decision Process

- **Stage 3: Assumptions in Cluster Analysis**
  - **IMPACT OF MULTICOLLINEARITY**
  - Multicollinearity was an issue in other multivariate techniques because of the difficulty in discerning the true impact of multicollinear variables.
  - In cluster analysis the effect is different because multicollinearity is actually a form of implicit weighting.
  - Multicollinearity acts as a weighting process not apparent to the observer but affecting the analysis nonetheless.
  - For this reason, the researcher is encouraged to examine the variables used in cluster analysis for substantial multicollinearity and, if found, either reduce the variables to equal numbers in each set or use a distance measure that takes multicollinearity into account.
  - Another possible solution involves factoring the variables prior to clustering and either selecting one cluster variable from each factor or using the resulting factor scores as cluster variables.
  - Recall that principal components or varimax rotated factors are uncorrelated. In this way, the research can take a proactive approach to dealing with multicollinearity.

UNIT - III

Cluster Analysis

Cluster Analysis Decision Process

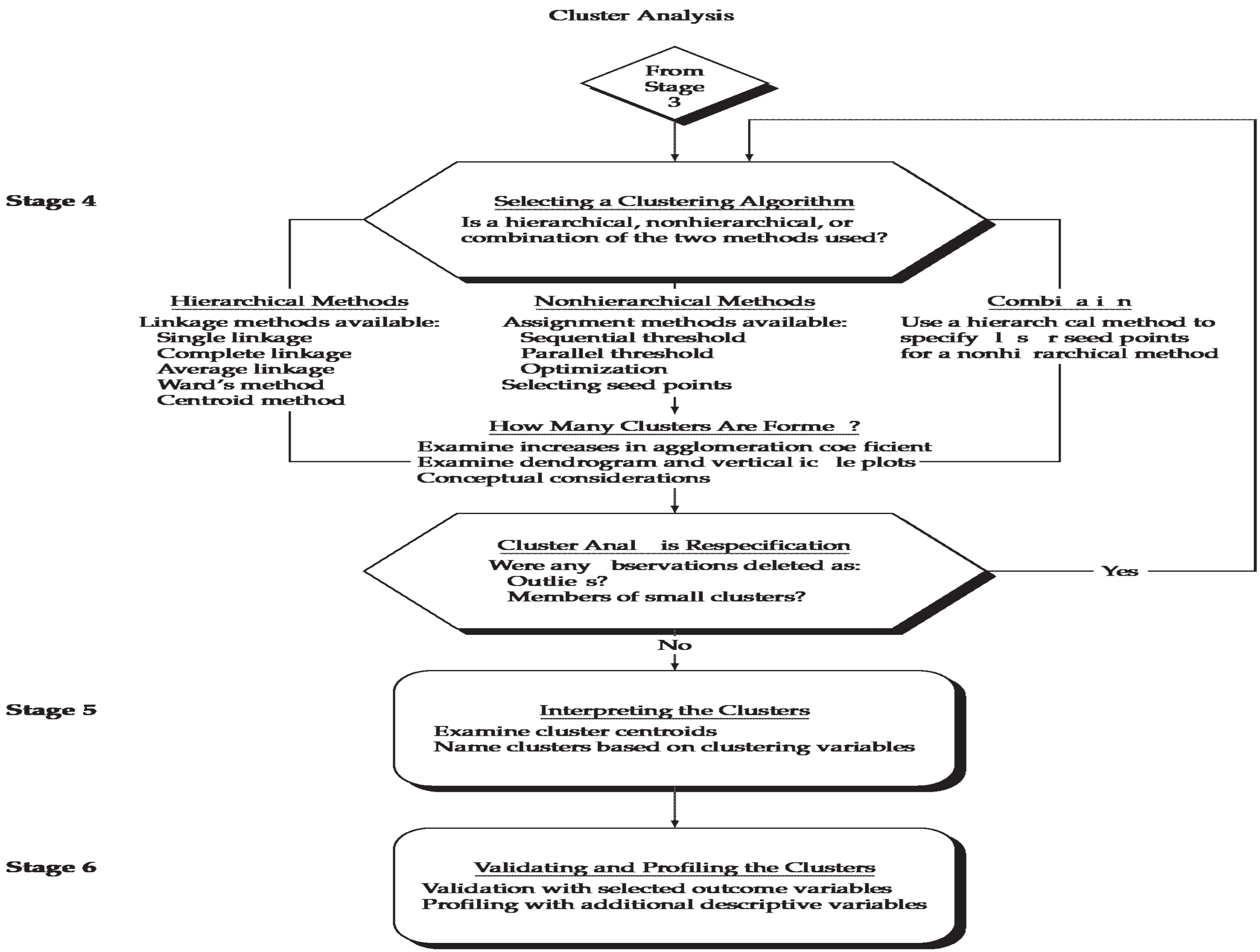


FIGURE 6 Stages 4–6 of the Cluster Analysis Decision Diagram

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - With the clustering variables selected and the similarity matrix calculated, the partitioning process begins.
  - • Select the partitioning procedure used for forming clusters.
  - • Make the decision on the number of clusters to be formed.
  - Both decisions have substantial implications not only on the results that will be obtained but also on the interpretation that can be derived from the results .
  - First, we examine the available partitioning procedures and then discuss the options available for deciding on a cluster solution by defining the number of clusters and membership for each observation.
  - Partitioning procedures work on a simple principle. They seek to maximize the distance between groups while minimizing the differences of in-group members.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**

Cluster Analysis

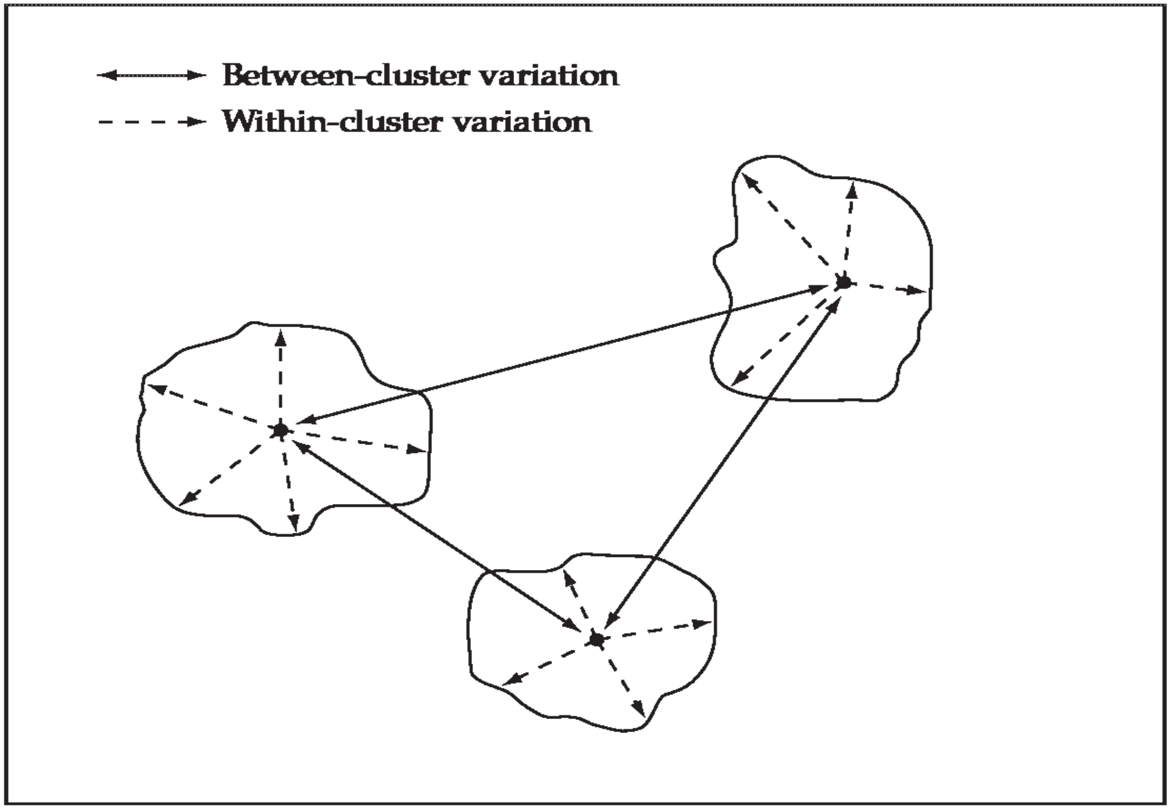


FIGURE 7 Cluster Diagram Showing Be w en- and Within-Cluster Variation

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - Hierarchical procedures involve a series of n – 1 clustering decisions (where n equals the number of observations) that combine observations into a hierarchy or a treelike structure.
  - The two basic types of hierarchical clustering procedures are agglomerative and divisive.
  - In the **agglomerative methods**, each object or observation starts out as its own cluster and is successively joined, the two most similar clusters at a time until only a single cluster remains.
  - In **divisive methods** all observations start in a single cluster and are successively divided (first into two clusters, then three, and so forth) until each is a single-member cluster.
  - In Figure 8, agglomerative methods move from left to right, and divisive methods move from right to left.
  - we focus here on the agglomerative methods.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**

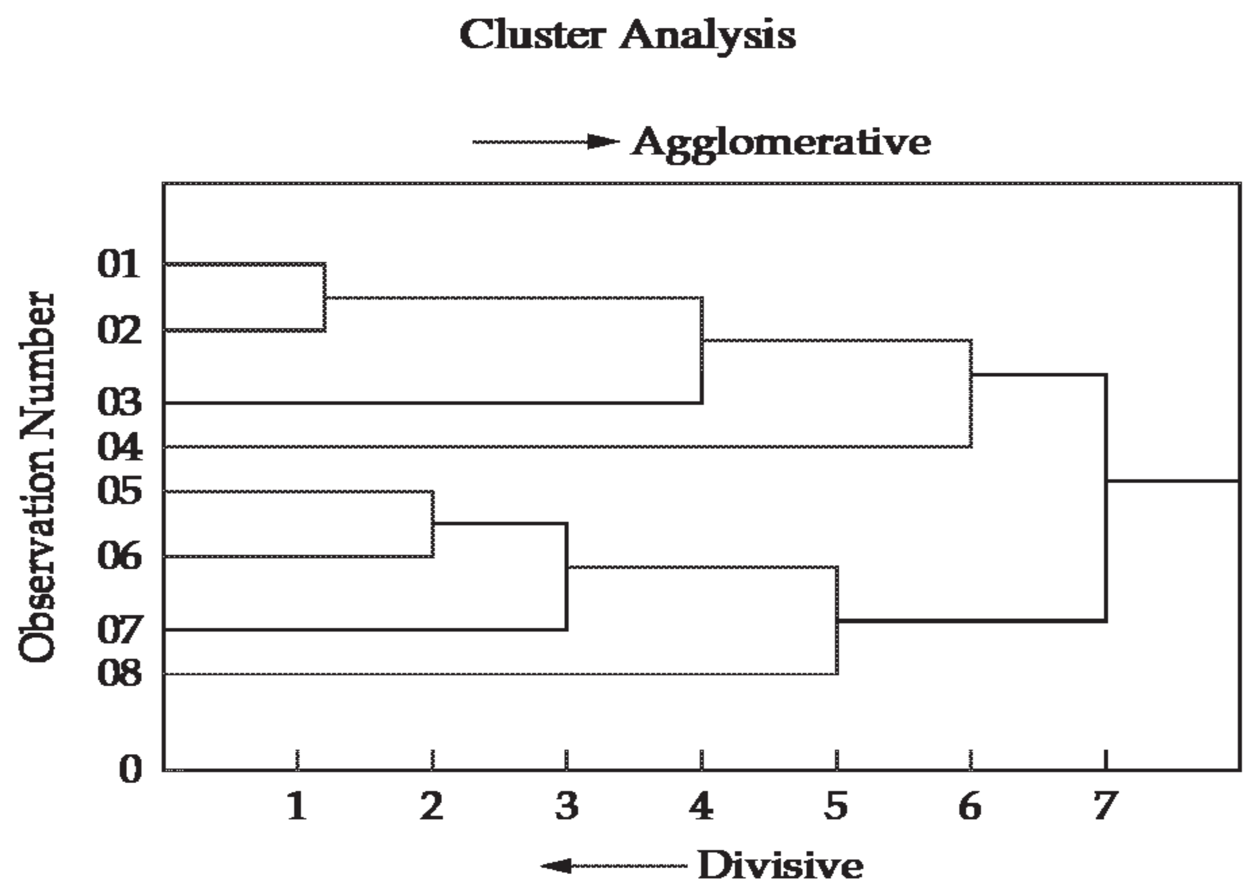


FIGURE 8 Dendrogram Illustrating Hierarchical Clustering

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - To understand how a hierarchical procedure works, we will examine the most common form—the agglomerative method—which follows a simple, repetitive process:
  - 1. Start with all observations as their own cluster (i.e., each observation forms a single-member cluster), so that the number of clusters equals the number of observations.
  - 2. Using the similarity measure, combine the two most similar clusters into a new cluster (now containing two observations), thus reducing the number of clusters by one.
  - 3. Repeat the clustering process again, using the similarity measure to combine the two most similar clusters into a new cluster.
  - 4. Continue this process, at each step combining the two most similar clusters into a new cluster. Repeat the process a total of n – 1 times until all observations are contained in a single cluster.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - Assume that we had 100 observations.
  - We would initially start with 100 separate clusters, each containing a single observation.
  - At the first step, the two most similar clusters would be combined, leaving us with 99 clusters.
  - At the next step, we combine the next two most similar clusters, so that we then have 98 clusters.
  - This process continues until the last step where the final two remaining clusters are combined into a single cluster.
  - An important characteristic of hierarchical procedures is that the results at an earlier stage are always nested within the results at a later stage, creating a similarity to a tree.
  - For example, an agglomerative six-cluster solution is obtained by joining two of the clusters found at the seven cluster stage.
  - Because clusters are formed only by joining existing clusters, any member of a cluster can trace its membership in an unbroken path to its beginning as a single observation.
  - This process is shown in Figure 8; the representation is referred to as a **dendrogram or tree graph,** which can be useful, but becomes unwieldy with large applications.
  - The dendrogram is widely available in most clustering software.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - **Clustering Algorithms.**
  - The clustering algorithm in a hierarchical procedure defines how similarity is defined between multiple-member clusters in the clustering process.
  - So how do we measure similarity between clusters when one or both clusters have multiple members?
  - Do we select one member to act as a typical member and measure similarity between these members of each cluster, or do we create some composite member to represent the cluster, or even combine the similarities between all members of each cluster?
  - We could employ any of these approaches, or even devise other ways to measure similarity between multiple-member clusters.
  - Among numerous approaches, the five most popular agglomerative algorithms are (1) Single-linkage
  - (2) Complete-linkage
  - (3) Average linkage
  - (4) Centroid method and
  - (5) Ward's method.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - **Single-Linkage**
  - The single-linkage method (also called the nearest-neighbor method) defines the similarity between clusters as the shortest distance from any object in one cluster to any object in the other.
  - This rule was applied in the example at the beginning of this chapter and enables us to use the original distance matrix between observations without calculating new distance measures.
  - Just find all the distances between observations in the two clusters and select the smallest as the measure of cluster similarity.
  - This method is probably the most versatile agglomerative algorithm, because it can define a wide range of clustering patterns (e.g., it can represent clusters that are concentric circles, like rings of a bull's-eye).

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - An example of this arrangement is shown in Figure 9. Three clusters (A, B, and C) are to be joined.
  - The single-linkage algorithm, focusing on only the closest points in each cluster, would link clusters A and B because of their short distance at the extreme ends of the clusters.
  - Joining clusters A and B creates a cluster that encircles cluster C. Yet in striving for within cluster homogeneity, it would be much better to join cluster C with either A or B.
  - This figure shows the principal disadvantage of the single-linkage algorithm.

Cluster Analysis

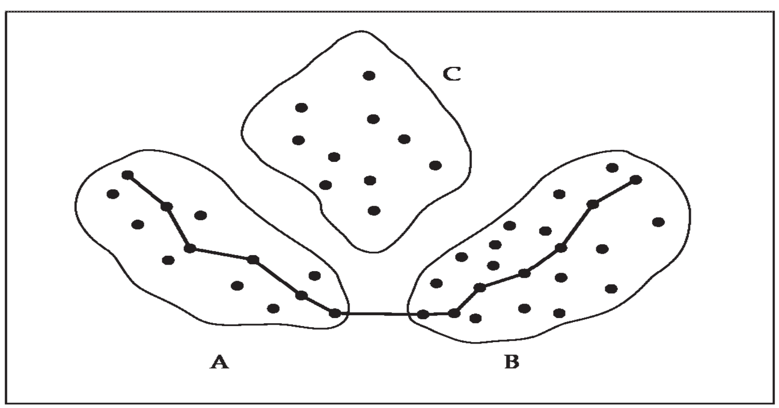


FIGURE 9 Example of Single Linkage Joining Diss mi ar Clusters A and B

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - **• Complete-Linkage**
  - The complete-linkage method (also known as the farthest-neighbor or diameter method) is comparable to the single-linkage algorithm, except that cluster similarity is based on maximum distance between observations in each cluster.
  - Similarity between clusters is the smallest (minimum diameter) sphere that can enclose all observations in both clusters.
  - This method is called complete-linkage because all objects in a cluster are linked to each other at some maximum distance.
  - Thus, within-group similarity equals group diameter. This technique eliminates the chaining problem identified with single-linkage and has been found to generate the most compact clustering solutions.
  - Even though it represents only one aspect of the data (i.e., the farthest distance between members), many researchers find it the most appropriate for a wide range of clustering applications.
  - Figure 10 compares the shortest (single-linkage) and longest (complete-linkage) distances in representing similarity between clusters.
  - Both measures reflect only one aspect of the data.
  - The use of the single-linkage reflects only a closest single pair of objects, and the complete-linkage also reflects a single pair, this time the two most extreme.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  – **HIERARCHICAL CLUSTER PROCEDURES**

Cluster Analysis

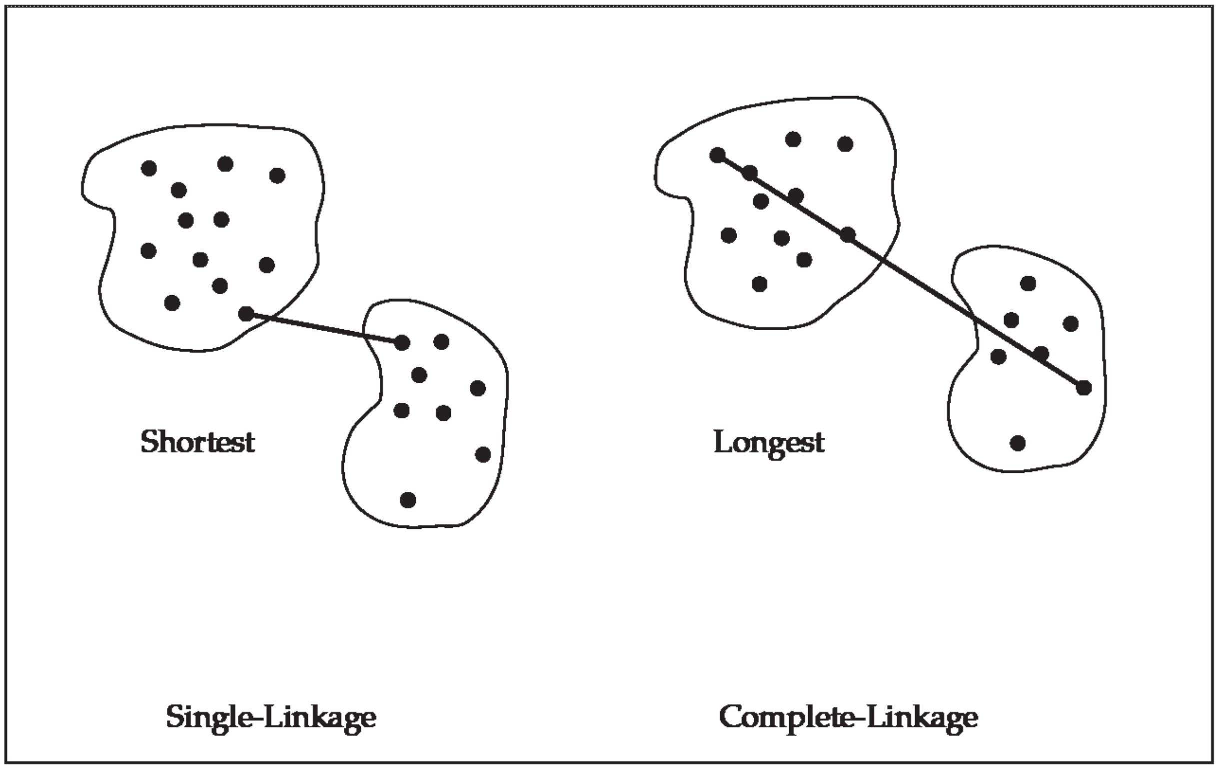


FIGURE 10 Comparison of Distance Measures for Single-Linkage and Complete-Linkage

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - **• Average Linkage**
  - The average linkage procedure differs from the single-linkage or complete-linkage procedures in that the similarity of any two clusters is the average similarity of all individuals in one cluster with all individuals in another.
  - This algorithm does not depend on extreme values (closest or farthest pairs) as do single-linkage or complete-linkage.
  - Instead, similarity is based on all members of the clusters rather than on a single pair of extreme members and are thus less affected by outliers.
  - Average linkage approaches, as a type of compromise between single- and complete-linkage methods, tend to generate clusters with small within-cluster variation.
  - They also tend toward the production of clusters with approximately equal within-group variance.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - **• Centroid Method**
  - In the centroid method the similarity between two clusters is the distance between the cluster centroids.
  - Cluster centroids are the mean values of the observations on the variables in the cluster variate.
  - In this method, every time individuals are grouped, a new centroid is computed. Cluster centroids migrate as cluster mergers take place.
  - In other words, a cluster centroid changes every time a new individual or group of individuals is added to an existing cluster.
  - These methods are the most popular in the physical and life sciences (e.g., biology) but may produce messy and often confusing results.
  - The confusion occurs because of reversals, that is, instances when the distance between the centroids of one pair may be less than the distance between the centroids of another pair merged at an earlier combination.
  - The advantage of this method, like the average linkage method, is that it is less affected by outliers than are other hierarchical methods.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - **• Ward's Method**
  - The Ward's method differs from the previous methods in that the similarity between two clusters is not a single measure of similarity, but rather the sum of squares within the clusters summed over all variables.
  - It is quite similar to the simple heterogeneity measure used in the example at the beginning of the chapter to assist in determining the number of clusters.
  - In the Ward's procedure, the selection of which two clusters to combine is based on which combination of clusters minimizes the within-cluster sum of squares across the complete set of disjoint or separate clusters.
  - At each step, the two clusters combined are those that minimize the increase in the total sum of squares across all variables in all clusters.
  - This procedure tends to combine clusters with a small number of observations, because the sum of squares is directly related to the number of observations involved.
  - The use of a sum of squares measure makes this method easily distorted by outliers.
  - Moreover, the Ward's method also tends to produce clusters with approximately the same number of observations.
  - However, the use of this method also makes it more difficult to identify clusters representing small proportions of the sample.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HIERARCHICAL CLUSTER PROCEDURES**
  - **• Overview**
  - Hierarchical clustering procedures are a combination of a repetitive clustering process combined with a clustering algorithm to define the similarity between clusters with multiple members.
  - The process of creating clusters generates a treelike diagram that represents the combinations/divisions of clusters to form the complete range of cluster solutions.
  - Hierarchical procedures generate a complete set of cluster solutions, ranging from all single-member clusters to the one-cluster solution where all observations are in a single cluster.
  - In doing so, the hierarchical procedure provides an excellent framework with which to compare any set of cluster solutions and help in judging how many clusters should be retained.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **NON-HIERARCHICAL CLUSTERING PROCEDURES**
  - In contrast to hierarchical methods, nonhierarchical procedures do not involve the treelike construction process. Instead, they assign objects into clusters once the number of clusters is specified.
  - For example, a six-cluster solution is not just a combination of two clusters from the seven-cluster solution, but is based only on finding the best six-cluster solution.
  - The nonhierarchical cluster software programs usually proceed through two steps:
  - **1. Specify cluster seeds:**
  - The first task is to identify starting points, known as cluster seeds, for each cluster. A cluster seed may be pre-specified by the researcher or observations may be selected, usually in a random process.
  - **2. Assignment:**
  - With the cluster seeds defined, the next step is to assign each observation to one of the cluster seeds based on similarity.
  - Many approaches are available for making this assignment (see later discussion in this section), but the basic objective is to assign each observation to the most similar cluster seed.
  - In some approaches, observations may be reassigned to clusters that are more similar than their original cluster assignment.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **NON-HIERARCHICAL CLUSTERING PROCEDURES**
  - We discuss several different approaches for selecting cluster seeds and assigning objects in the next sections.
  - **Selecting Seed Points**
  - Even though the nonhierarchical clustering algorithms discussed in the next section differ in the manner in which they assign observations to the seed points, they all face the same problem:
  - How do we select the cluster seeds?
  - The different approaches can be classified into two basic categories:
  - **1. Researcher specified.**
  - In this approach, the researcher provides the seed points based on external data.
  - The two most common sources of the seed points are prior research or data from another multivariate analysis.
  - Many times the researcher has knowledge of the cluster profiles being researched.
  - For example, prior research may have defined segment profiles and the task of the cluster analysis is to assign individuals to the most appropriate segment cluster.
  - It is also possible that other multivariate techniques may be used to generate the seed points.
  - One common example is the use of a hierarchical clustering algorithm to establish the number of clusters and then generate seed points from these results.
  - The common element is that the researcher, while knowing the number of clusters to be formed, also has information about the basic character of these clusters.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **NON-HIERARCHICAL CLUSTERING PROCEDURES**
  - **2. Sample generated.**
  - The second approach is to generate the cluster seeds from the observations of the sample, either in some systematic fashion or simply through random selection.
  - For example, in the FASTCLUS program in SAS, the first seed is the first observation in the data set with no missing values.
  - The second seed is the next complete observation (no missing data) that is separated from the first seed by a specified minimum distance. The default option is a zero minimum distance.
  - After all seeds are selected, the program assigns each observation to the cluster with the nearest seed.
  - The K-means program in SPSS can select the necessary seed points randomly from among the observations. In any of these approaches, the researcher relies on the selection process to choose seed points that reflect natural clusters as starting points for the clustering algorithms.
  - A possible limitation is that replication of the results is difficult if the observations are reordered.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **NON-HIERARCHICAL CLUSTERING PROCEDURES**
  - In either approach, the researcher must be aware of the impact of the cluster seed selection process on the final results.
  - All of the clustering algorithms, even those of an optimizing nature (see the following discussion), will generate different cluster solutions depending on the initial cluster seeds.
  - The differences among cluster solutions will hopefully be minimal using different seed points, but they underscore the importance of cluster seed selection and its impact on the final cluster solution.
  - **Nonhierarchical Clustering Algorithms.**
  - Several clustering algorithms have been proposed. The most frequently cited are sequential, parallel and optimization.
  - The sequential threshold method starts by selecting one cluster seed and includes all objects within a pre-specified distance.
  - A second cluster seed is then selected, and all objects within the pre-specified distance of that seed are included.
  - A third seed is then selected, and the process continues as before.
  - The primary disadvantage of this approach is that when an observation is assigned to a cluster, it cannot be reassigned to another cluster, even if that cluster seed is more similar.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **NON-HIERARCHICAL CLUSTERING PROCEDURES**
  - The parallel threshold method considers all cluster seeds simultaneously and assigns observations within the threshold distance to the nearest seed.
  - The third method, referred to as the optimizing procedure, is similar to the other two nonhierarchical procedures except that it allows for reassignment of observations to a seed other than the one with which it was originally associated.
  - All of these belong to a group of clustering algorithms known as **K-means**.
  - **K-means algorithms** work by portioning the data into a user-specified number of clusters and then iteratively reassigning observations to clusters until some numerical criterion is met.
  - The criterion specifies a goal related to minimizing the distance of observations from one another in a cluster and maximizing the distance between clusters.
  - K-means is so commonly used that the term is used by some to refer to nonhierarchical cluster analysis in general.
  - For example, in SPSS, the nonhierarchical clustering routine is referred to as K-means.
  - An optimizing procedure allows for reassignment of observations based on the goal of creating the most distinct clusters.
  - If, in the course of assigning observations, an observation becomes closer to another cluster that is not the cluster to which it is currently assigned, then an optimizing procedure switches the observation to the more similar (closer) cluster.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **SHOULD HIERARCHICAL OR NONHIERARCHICAL METHODS BE USED?**
  - We can examine the strengths and weaknesses of each method to determine which is most appropriate for a unique research setting.
  - **Pros and Cons of Hierarchical Methods.**
  - Hierarchical clustering techniques have long been the more popular clustering method, with Ward's method and average linkage probably being the best available.
  - Besides the fact that hierarchical procedures were the first clustering methods developed, they do offer several advantages that lead to their widespread usage:
  - **1. Simplicity:**
  - Hierarchical techniques, with their development of the treelike structures depicting the clustering process, do afford the researcher with a simple, yet comprehensive, portrayal of the entire range of clustering solutions.
  - In doing so, the researcher can evaluate any of the possible clustering solutions from one analysis.
  - **2. Measures of similarity:**
  - The widespread use of the hierarchical methods led to an extensive development of similarity measures for almost any type of clustering variables.
  - Hierarchical techniques can be applied to almost any type of research question.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **SHOULD HIERARCHICAL OR NONHIERARCHICAL METHODS BE USED?**
  - **3. Speed:**
  - Hierarchical methods have the advantage of generating an entire set of clustering solutions (from all separate clusters to one cluster) in an expedient manner.
  - This ability enables the researcher to examine a wide range of alternative clustering solutions, varying measures of similarities and linkage methods, in an efficient manner.
  - Even though hierarchical techniques have been widely used, they do have several distinct **disadvantages** that affect any of their cluster solutions:
  - 1. Hierarchical methods can be misleading because undesirable early combinations may persist throughout the analysis and lead to artificial results. Of specific concern is the substantial impact of outliers on hierarchical methods, particularly with the complete-linkage method.
  - 2. To reduce the impact of outliers, the researcher may wish to cluster analyze the data several times, each time deleting problem observations or outliers.
  - The deletion of cases, however, even those not found to be outliers, can many times distort the solution.
  - Thus, the researcher must employ extreme care in the deletion of observations for any reason.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **SHOULD HIERARCHICAL OR NONHIERARCHICAL METHODS BE USED?**
  - 3. Although computations of the clustering process are relatively fast, hierarchical methods are not amenable to analyzing large samples or even large numbers of variables.
  - As sample size increases, the data storage requirements increase dramatically.
  - For example, a sample of 400 cases requires storage of approximately 80,000 similarities, which increases to almost 125,000 for a sample of 500.
  - Even with the computing power of today's personal computers, these data requirements can limit the application in many instances.
  - The researcher may take a random sample of the original observations to reduce sample size but must now question the representativeness of the sample taken from the original sample.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **SHOULD HIERARCHICAL OR NONHIERARCHICAL METHODS BE USED?**
  - **Emergence of Nonhierarchical Methods.**
  - Nonhierarchical methods have gained increased acceptability and usage, but any application depends on the ability of the researcher to select the seed points according to some practical, objective, or theoretical basis.
  - In these instances, nonhierarchical methods offer several advantages over hierarchical techniques.
  - 1. The results are less susceptible to outliers in the data, the distance measure used, and the inclusion of irrelevant or inappropriate variables.
  - 2. Nonhierarchical methods can analyze extremely large data sets because they do not require the calculation of similarity matrices among all observations, but instead just the similarity of each observation to the cluster centroids.
  - Even the optimizing algorithms that allow for reassignment of observations between clusters can be readily applied to all sizes of data sets.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **SHOULD HIERARCHICAL OR NONHIERARCHICAL METHODS BE USED?**
  - **Emergence of Nonhierarchical Methods.**
  - Although nonhierarchical methods do have several distinct advantages, several shortcomings can markedly affect their use in many types of applications.
  - 1. The benefits of any nonhierarchical method are realized only with the use of nonrandom (i.e., specified) seed points. Thus, the use of nonhierarchical techniques with random seed points is generally considered inferior to hierarchical techniques.
  - 2. Even a nonrandom starting solution does not guarantee an optimal clustering of observations.
  - In fact, in many instances the researcher will get a different final solution for each set of specified seed points.
  - Only by analysis and validation can the researcher select what is considered the best representation of structure, realizing that many alternatives may be as acceptable.
  - 3. Nonhierarchical methods are also not as efficient when examining a large number of potential cluster solutions.
  - Each cluster solution is a separate analysis, in contrast to the hierarchical techniques that generate all possible cluster solutions in a single analysis.
  - Thus, nonhierarchical techniques are not as well suited to exploring a wide range of solutions based on varying elements such as similarity measures, observations included, and potential seed points.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **SHOULD HIERARCHICAL OR NONHIERARCHICAL METHODS BE USED?**
  - **A Combination of Both Methods.**
  - Many researchers recommend a combination approach using both methods. In this way, the advantages of one approach can compensate for the weaknesses or the other. This is accomplished in two steps:
  - 1. First, a hierarchical technique is used to generate a complete set of cluster solutions, establish the applicable cluster solutions, and establish the appropriate number of clusters.
  - 2. After outliers are eliminated, the remaining observations can then be clustered by a nonhierarchical method. In this way, the advantages of the hierarchical methods are complemented by the ability of the nonhierarchical methods to refine the results by allowing the switching of cluster membership.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **HOW MANY CLUSTERS SHOULD BE FORMED?**
  - Stopping rule that suggests two or more cluster solutions which can be compared before making the final decision.
  - Unfortunately, no standard objective selection procedure exists. Because no internal statistical criterion is used for inference, such as the statistical significance tests of other multivariate methods, researchers have developed many criteria for approaching the problem.
  - The principal issues facing any of these stopping rules include the following:
  - • These ad hoc procedures must be computed by the researcher and often involve fairly complex.
  - • Many of these criteria are specific to a particular software program and are not easily calculated if not provided by the program.
  - • A natural increase in heterogeneity comes from the reduction in number of clusters. Thus, the researcher must look at the trend in the values of these stopping rules across cluster solutions to identify marked increases. If not, in most instances the two-cluster solution would always be chosen because the value of any stopping rule is normally highest when going from two to one cluster.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **Measures of Heterogeneity Change.**
  - One class of stopping rules examines some measure of heterogeneity between clusters at each successive step, with the cluster solution defined when the heterogeneity measure exceeds a specified value or when the successive values between steps makes a sudden jump.
  - Heterogeneity refers to how different the observations in a cluster are from each other (i.e., heterogeneity refers to a lack of similarity among group members).
  - A simple example was used at the beginning of the chapter, which looked for large increases in the average within-cluster distance.
  - When a large increase occurs, the researcher selects the prior cluster solution on the logic that its combination caused a substantial increase in heterogeneity.
  - This type of stopping rule has been shown to provide fairly accurate decisions in empirical studies, but it is not uncommon for a number of cluster solutions to be identified by these large increases in heterogeneity.
  - It is then the researcher’s task to select a final cluster solution from these selected cluster solutions. Various stopping rules follow this general approach.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **Percentage Changes in Heterogeneity**
  - Probably the simplest and most widespread rule is a simple percentage change in some measure of heterogeneity.
  - A typical example is using the agglomeration coefficient in SPSS, which measures heterogeneity as the distance at which clusters are formed (if a distance measure of similarity is used) or the within-cluster sum of squares if the Ward's method is used.
  - With this measure, the percentage increase in the agglomeration coefficient can be calculated for each cluster solution.
  - Then the researcher selects cluster solutions as a potential final solution when the percentage increase is markedly larger than occurring at other steps.
  - **Measures of Variance Change**
  - The root mean square standard deviation (RMSSTD) is the square root of the variance of the new cluster formed by joining the two clusters.
  - The variance for the newly formed cluster is calculated as the variance across all clustering variables.
  - Large increases in the RMSSTD suggest the joining of two quite dissimilar clusters, indicating the previous cluster solution (in which the two clusters were separate) was a candidate for selection as the final cluster solution.

Cluster Analysis Decision Process

- **Stage 4: Deriving Clusters and Assessing Overall Fit**
  - **Statistical Measures of Heterogeneity Change**
  - A series of test statistics attempts to portray the degree of heterogeneity for each new cluster solution (i.e., joining of two clusters).
  - One of the most widely used is a pseudo F statistic, which compares the goodness-of-fit of k clusters to k – 1 clusters.
  - Highly significant values indicate that the k – 1 cluster solution is more appropriate than the k cluster solution.
  - The researcher should not consider any significant value, but instead look to those values markedly more significant than for other solutions.
  - **Direct Measures of Heterogeneity.**
  - A second general class of stopping rules attempts to directly measure heterogeneity of each cluster solution.
  - The most common measure in this class is the cubic clustering criterion (CCC) contained in SAS, a measure of the deviation of the clusters from an expected distribution of points formed by a multivariate uniform distribution.
  - Here the researcher selects the cluster solution with the largest value of CCC (i.e., the cluster solution where CCC peaks).
  - Despite its inclusion in SAS and its advantage of selecting a single-cluster solution, it has been shown to many times generate too many clusters as the final solution and is based on the assumption that the variables are uncorrelated.
  - However, it is a widely used measure and is generally as efficient as any other stopping rule.

Cluster Analysis Decision Process

- **Stage 5: Interpretation of the Clusters**
  - The interpretation stage involves examining each cluster in terms of the cluster variate to name or assign a label accurately describing the nature of the clusters.
  - To clarify this process, let us refer to the example of diet versus regular soft drinks.
  - Let us assume that an attitude scale was developed that consisted of statements regarding consumption of soft drinks, such as “diet soft drinks taste harsher,” “regular soft drinks have a fuller taste,” “diet drinks are healthier,” and so forth.
  - Further, let us assume that demographic and softdrink consumption data were also collected.
  - When starting the interpretation process, one measure frequently used is the cluster’s centroid.
  - If the clustering procedure was performed on the raw data, it would be a logical description.
  - If the data were standardized or if the cluster analysis was performed using factor analysis (component factors), the researcher would have to go back to the raw scores for the original variables and develop profile diagrams using these data.

Cluster Analysis Decision Process

- **Stage 5: Interpretation of the Clusters**
  - The profiling and interpretation of the clusters, however, achieve more than just description and are essential elements in selecting between cluster solutions when the stopping rules indicate more than one appropriate cluster solution.
  - • They provide a means for assessing the correspondence of the derived clusters to those proposed by prior theory or practical experience. If used in a confirmatory mode, the cluster analysis profiles provide a direct means of assessing the correspondence.
  - • The cluster profiles also provide a route for making assessments of practical significance. The researcher may require that substantial differences exist on a set of clustering variables and the cluster solution be expanded until such differences arise.
  - In assessing either correspondence or practical significance, the researcher compares the derived clusters to a preconceived typology.
  - This more subjective judgment by the researcher combines with the empirical judgment of the stopping rules to determine the final cluster solution to represent the data structure of the sample.

Cluster Analysis Decision Process

- **Stage 6: Validation and Profiling of the Clusters**
  - **VALIDATING THE CLUSTER SOLUTION**
  - Validation includes attempts by the researcher to assure that the cluster solution is representative of the general population, and thus is generalizable to other objects and is stable over time.
  - **Cross-Validation.**
  - The most direct approach in this regard is to cluster analyze separate samples, comparing the cluster solutions and assessing the correspondence of the results.
  - This approach, however, is often impractical because of time or cost constraints or the unavailability of objects (particularly consumers) for multiple cluster analyses.
  - In these instances, a common approach is to split the sample into two groups.
  - Each cluster is analyzed separately, and the results are then compared.
  - Cross-tabulation also can be used, because the members of any specific cluster in one solution should stay together in a cluster in another solution.
  - Therefore, the cross-tabulation should display obvious patterns of matching cluster membership.
  - Other approaches include
  - (1) a modified form of split sampling whereby cluster centers obtained from one cluster solution are employed to define clusters from the other observations and the results are compared, and
  - (2) a direct form of cross-validation.

Cluster Analysis Decision Process

- **Stage 6: Validation and Profiling of the Clusters**
  - **VALIDATING THE CLUSTER SOLUTION**
  - For any of these methods, stability of the cluster results can be assessed by the number of cases assigned to the same cluster across cluster solutions.
  - Generally, a very stable solution would be produced with less than 10 percent of observations being assigned to a different cluster.
  - A stable solution would result with between 10 and 20 percent assigned to a different group, and a somewhat stable solution when between 20 and 25 percent of the observations are to a different cluster than the initial one.
  - **Establishing Criterion Validity.**
  - The researcher may also attempt to establish some form of criterion or predictive validity.
  - To do so, the researcher selects variable(s) not used to form the clusters but known to vary across the clusters.
  - In our example, we may know from past research that attitudes toward diet soft drinks vary by age.
  - Thus, we can statistically test for the differences in age between those clusters that are favorable to diet soft drinks and those that are not.
  - The variable(s) used to assess predictive validity should have strong theoretical or practical support because they become the benchmark for selecting among the cluster solutions.

Cluster Analysis Decision Process

- **Stage 6: Validation and Profiling of the Clusters**
  - **PROFILING THE CLUSTER SOLUTION**
  - The profiling stage involves describing the characteristics of each cluster to explain how it may differ on relevant dimensions.
  - This process typically involves the use of discriminant analysis. The procedure begins after the clusters are identified.
  - The researcher utilizes data not previously included in the cluster procedure to profile the characteristics of each cluster.
  - These data typically are demographic characteristics, psychographic profiles, consumption patterns, and so forth.
  - Although no theoretical rationale may exist for their difference across the clusters, such as required for predictive validity assessment, they should at least have practical importance.
  - Using discriminant analysis, the researcher compares average score profiles for the clusters.
  - The categorical dependent variable is the previously identified clusters, and the independent variables are the demographics, psychographics, and so on.

Cluster Analysis Decision Process

- **Stage 6: Validation and Profiling of the Clusters**
  - **PROFILING THE CLUSTER SOLUTION**
  - From this analysis, assuming statistical significance, the researcher could conclude, for example, that the "health- and calorie-conscious" cluster from our previous diet soft drink example consists of better-educated, higher-income professionals who are moderate consumers of soft drinks.
  - In short, the profile analysis focuses on describing not what directly determines the clusters but rather on the characteristics of the clusters after they are identified.
  - Moreover, the emphasis is on the characteristics that differ significantly across the clusters and those that could predict membership in a particular cluster.
  - Profiling often is an important practical step in clustering procedures, because identifying characteristics like demographics enables segments to be identified or located with easily obtained information.

UNIT - IV

Discriminant Analysis

Discriminant Analysis

- Discriminant analysis is the appropriate statistical techniques when the dependent variable is a categorical (nominal or nonmetric) variable and the independent variables are metric variables.
- In many cases, the dependent variable consists of two groups or classifications, for example, male versus female or high versus low.
- In other instances, more than two groups are involved, such as low, medium, and high classifications.
- Discriminant analysis is capable of handling either two groups or multiple (three or more) groups.
- When two classifications are involved, the technique is referred to as two-group discriminant analysis.
- When three or more classifications are identified, the technique is referred to as multiple discriminant analysis (MDA).
- Logistic regression is limited in its basic form to two groups, although other formulations can handle more groups.

Discriminant Analysis

- **Discriminant Analysis**
- Discriminant analysis involves deriving a variate.
- The discriminant variate is the linear combination of the two (or more) independent variables that will discriminate best between the objects (persons, firms, etc.) in the groups defined a priori.
- Discrimination is achieved by calculating the variate's weights for each independent variable to maximize the differences between the groups (i.e., the between-group variance relative to the within-group variance).
- The variate for a discriminant analysis, also known as the discriminant function, is derived from an equation much like that seen in multiple regression.
- It takes the following form:

$$Z_{jk} = a + W_1 X_{1k} + W_2 X_{2k} + \dots + W_n X_{nk}$$

where

Z_{jk} = discriminant Z score of discriminant function j for object k
a = intercept
W_i = discriminant weight for independent variable i
X_{ik} = independent variable i for object k

Discriminant Analysis

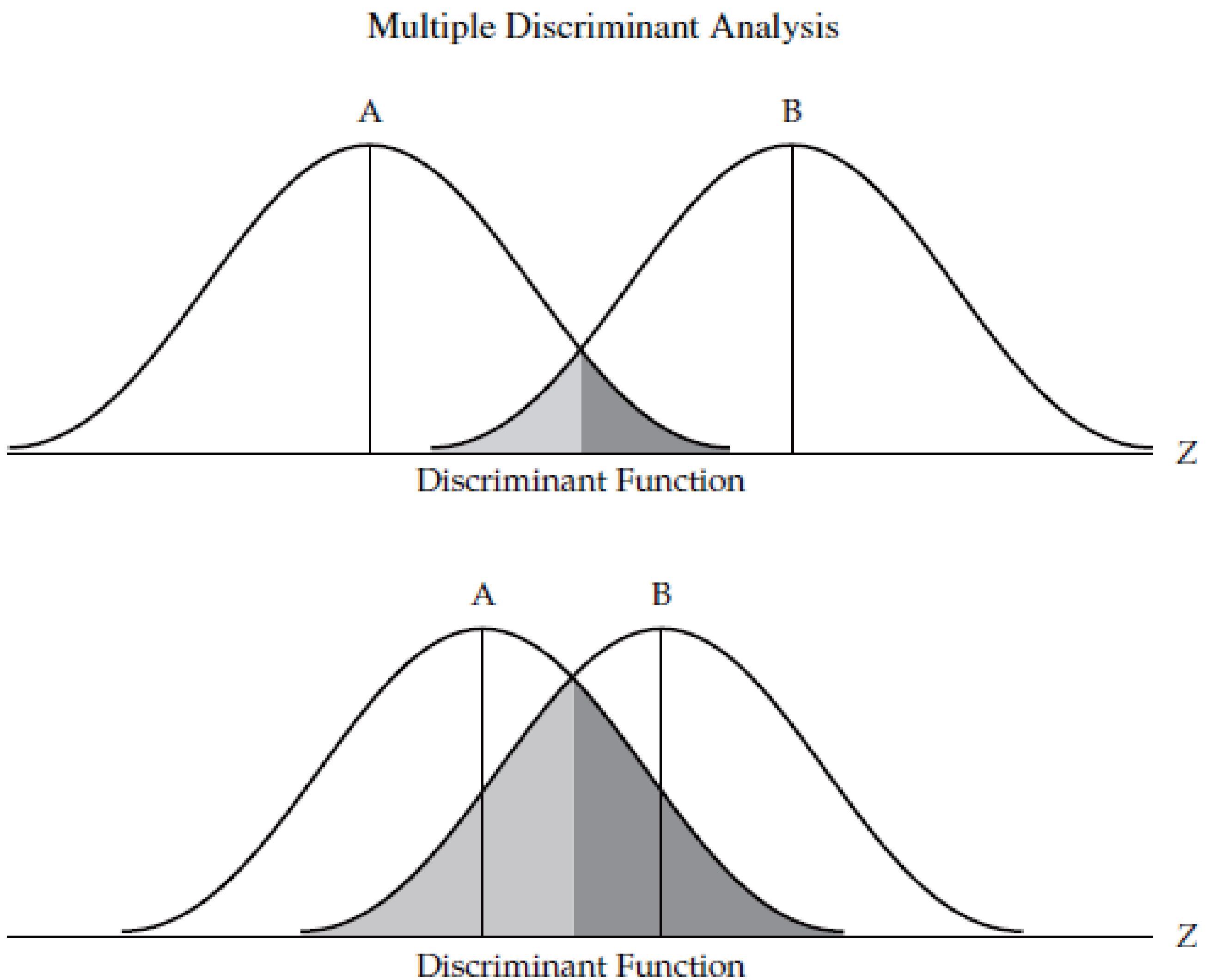
- **Discriminant Analysis**
- As with the variate in regression or any other multivariate technique we see the discriminant score for each object in the analysis (person, firm, etc.) being a summation of the values obtained by multiplying each independent variable by its discriminant weight.
- What is unique about discriminant analysis is that more than one discriminant function may be present, resulting in each object possibly having more than one discriminant score.
- Discriminant analysis is the appropriate statistical technique for testing the hypothesis that the group means of a set of independent variables for two or more groups are equal.
- By averaging the discriminant scores for all the individuals within a particular group, we arrive at the group mean.
- This group mean is referred to as a centroid.
- When the analysis involves two groups, there are two centroids; with three groups, there are three centroids; and so forth.

Discriminant Analysis

- **Discriminant Analysis**
- The centroids indicate the most typical location of any member from a particular group, and a comparison of the group centroids shows how far apart the groups are in terms of that discriminant function.
- The test for the statistical significance of the discriminant function is a generalized measure of the distance between the group centroids.
- It is computed by comparing the distributions of the discriminant scores for the groups.
- Multiple discriminant analysis is unique in one characteristic among the dependence relationships.
- If the dependent variable consists of more than two groups, discriminant analysis will calculate more than one discriminant function.
- As a matter of fact, it will calculate NG – 1 functions, where NG is the number of groups.
- Each discriminant function will calculate a separate discriminant Z score.
- In the case of a three-group dependent variable, each object (respondent, firm, etc.) will have a separate score for discriminant functions one and two, which enables the objects to be plotted in two dimensions, with each dimension representing a discriminant function.
- Thus, discriminant analysis is not limited to a single variate, as is multiple regression, but creates multiple variates representing dimensions of discrimination among the groups.

Discriminant Analysis

- **Discriminant Analysis**



Discriminant Analysis

- **HYPOTHETICAL EXAMPLE OF DISCRIMINANT ANALYSIS**
- Discriminant analysis is applicable to any research question with the objective of understanding group membership, whether the groups comprise individuals (e.g., customers versus noncustomers), firms (e.g., profitable versus unprofitable), products (e.g., successful versus unsuccessful), or any other object that can be evaluated on a series of independent variables.
- **Two-Group Discriminant Analysis: Purchasers Versus Non-purchasers**
- Suppose KitchenAid wants to determine whether one of its new products—a new and improved food mixer—will be commercially successful. In carrying out the investigation, KitchenAid is primarily interested in identifying those consumers who would purchase the new product versus those who would not.
- To assist in identifying potential purchasers, KitchenAid devised rating scales on three characteristics—durability, performance, and style—to be used by consumers in evaluating the new product.
- Rather than relying on each scale as a separate measure, KitchenAid hopes that a weighted combination of all three would better predict purchase likelihood of consumers.

Discriminant Analysis

- **HYPOTHETICAL EXAMPLE OF DISCRIMINANT ANALYSIS**
- **IDENTIFYING DISCRIMINATING VARIABLES**
- To identify variables that may be useful in discriminating between groups (i.e., purchasers versus nonpurchasers), emphasis is placed on group differences rather than measures of correlation used in multiple regression.

TABLE 1 KitchenAid Survey Results for the Evaluation of a New Consumer Product

	*Evaluation of New Product**		
Groups Based on Purchase Intention	**X_1 Durability**	**X_2 Performance**	**X_3 Style**
Group 1: Would purchase			
Subject 1	8	9	6
Subject 2	6	7	5
Subject 3	10	6	3
Subject 4	9	4	4
Subject 5	4	8	2
Group mean	7.4	6.8	4.0
Group 2: Would not purchase			
Subject 6	5	4	7
Subject 7	3	7	2
Subject 8	4	5	5
Subject 9	2	4	3
Subject 10	2	2	2
Group mean	3.2	4.4	3.8
Difference between group means	4.2	2.4	0.2

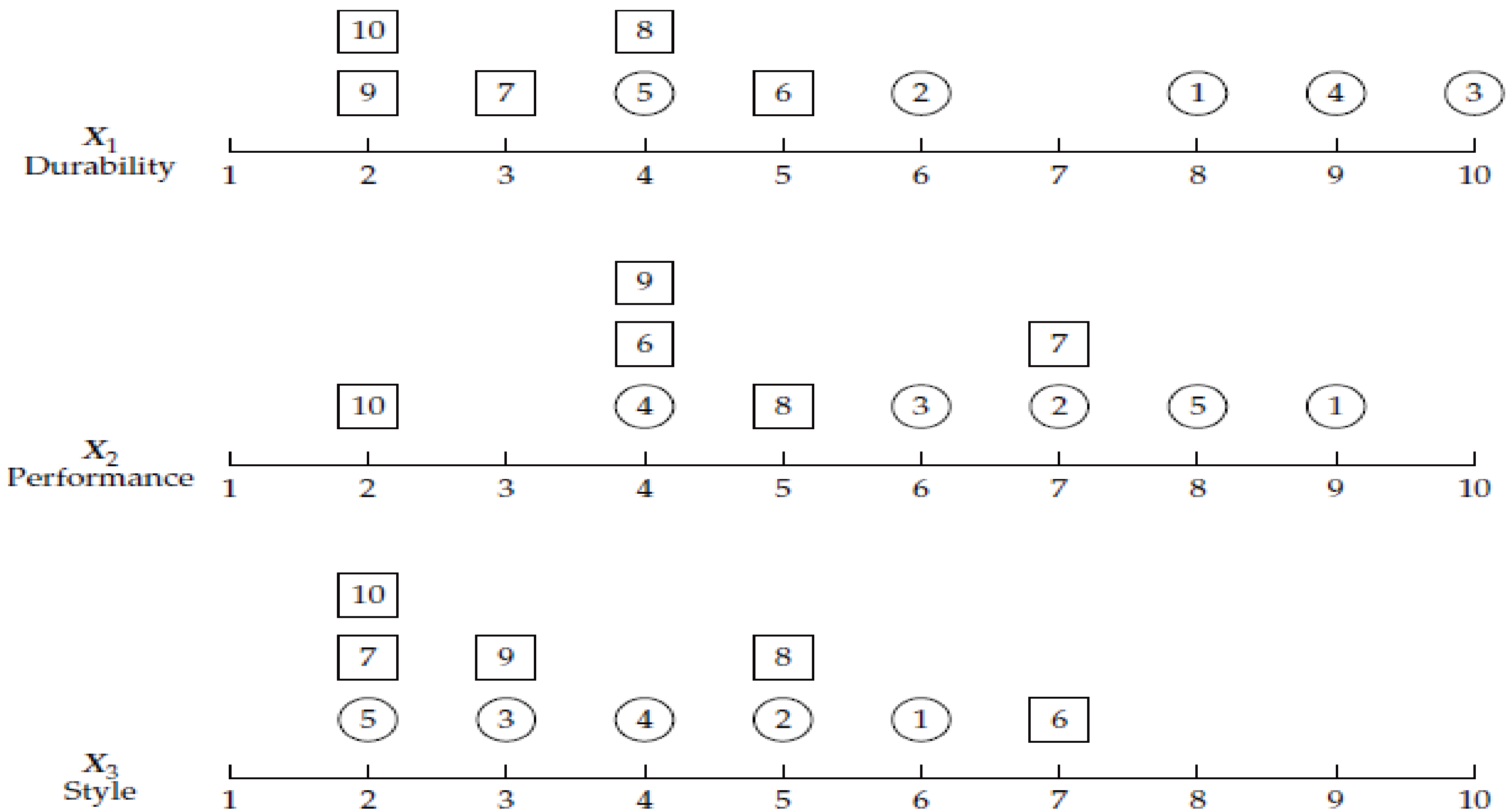
*Evaluations are made on a 10-point scale (1 = very poor to 10 = excellent).

Discriminant Analysis

- **HYPOTHETICAL EXAMPLE OF DISCRIMINANT ANALYSIS**
- **IDENTIFYING DISCRIMINATING VARIABLES**
- Because we have only 10 respondents in two groups and three independent variables, we can also look at the data graphically to determine what discriminant analysis is trying to accomplish.
- Figure 2 shows the 10 respondents on each of the three variables. The “would purchase” group is represented by circles and the “would not purchase” group by the squares. Respondent identification numbers are inside the shapes.
- • *X*1 (Durability) had a substantial difference in mean scores, enabling us to almost perfectly discriminate between the groups using only this variable. If we established the value of 5.5 as our cutoff point to discriminate between the two groups, then we would misclassify only respondent 5, one of the “would purchase” group members. This variable illustrates the discriminatory power in having a large difference in the means for the two groups and a lack of overlap between the distributions of the two groups.
- • *X*2 (Performance) provides a less clear-cut distinction between the two groups. However, this variable does provide high discrimination for respondent 5, who was misclassified if we used only *X*1. In addition, the respondents who would be misclassified using *X*2 are well separated on *X*1. Thus, *X*1 and *X*2 might be used quite effectively in combination to predict group membership.

Discriminant Analysis

- **HYPOTHETICAL EXAMPLE OF DISCRIMINANT ANALYSIS**
- **IDENTIFYING DISCRIMINATING VARIABLES**
- • *X*3 (Style) shows little differentiation between the groups. Thus, by forming a variate of only *X*1 and *X*2, and omitting *X*3, a discriminant function may be formed that maximizes the separation of the groups on the discriminant score.



Discriminant Analysis

- **HYPOTHETICAL EXAMPLE OF DISCRIMINANT ANALYSIS**
- **CALCULATING A DISCRIMINANT FUNCTION**
- With the three potential discriminating variables identified, attention shifts toward investigation of the possibility of using the discriminating variables in combination to improve upon the discriminating power of any individual variable.
- To this end, a variate can be formed with two or more discriminating variables to act together in discriminating between the groups.
- Table 2 contains the results for three different formulations of a discriminant function, each representing different combinations of the three independent variables.

Discriminant Analysis

- **HYPOTHETICAL EXAMPLE OF DISCRIMINANT ANALYSIS**
- **CALCULATING A DISCRIMINANT FUNCTION**

TABLE 2 Creating Discriminant Functions to Predict Purchasers Versus Nonpurchasers

Group	Calculated Discriminant Z Scores: Function 1: $Z = X_1$	Function 2: $Z = X_1 + X_2$	Function 3: $Z = -4.53 + .476X_1 + .359X_2$
Group 1: Would purchase			
Subject 1	8	17	2.51
Subject 2	6	13	.84
Subject 3	10	16	2.38
Subject 4	9	13	1.19
Subject 5	4	12	.25
Group 2: Would not purchase			
Subject 6	5	9	−.71
Subject 7	3	10	−.59
Subject 8	4	9	−.83
Subject 9	2	6	−2.14
Subject 10	2	4	−2.86
Cutting score	5.5	11	0.0

Classification Accuracy:

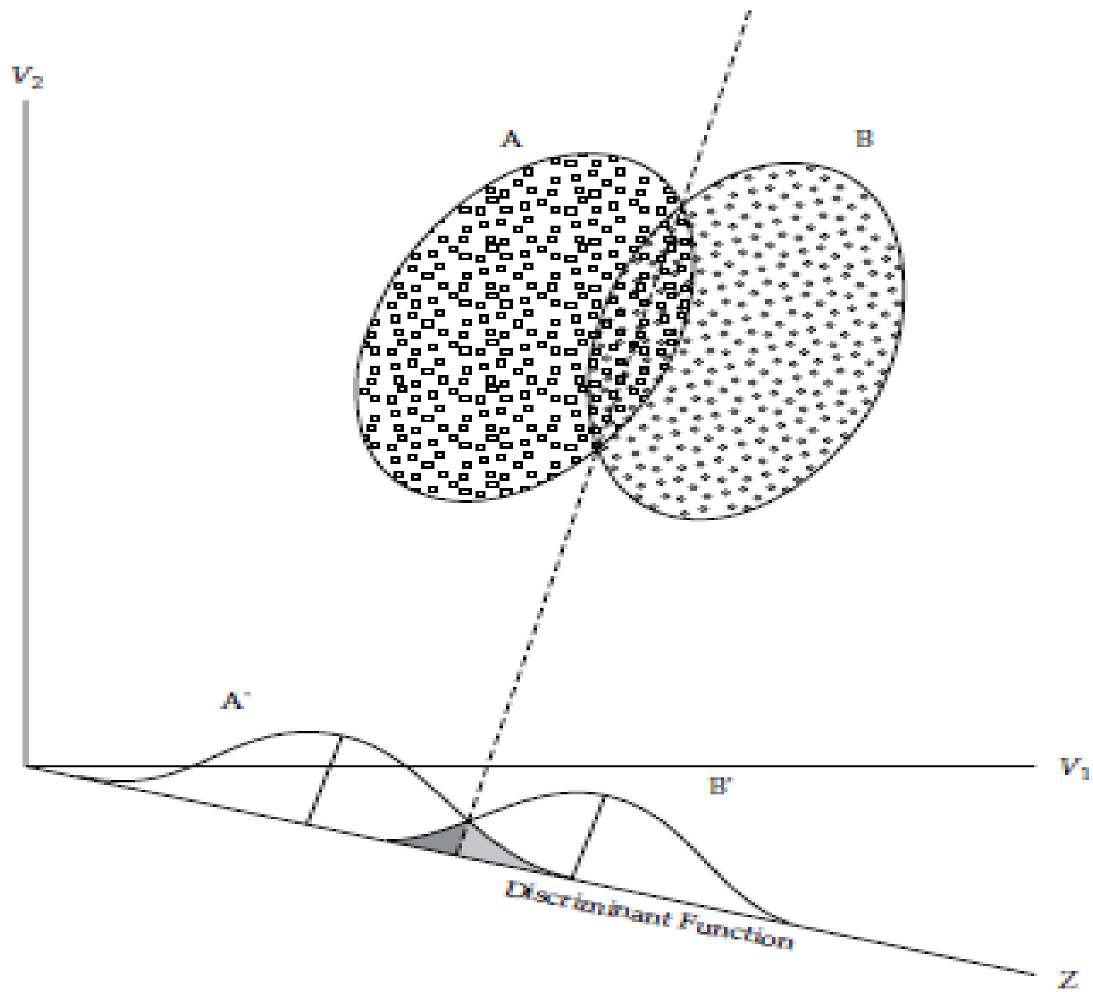
Actual Group	Predicted Group: 1	2	Predicted Group: 1	2	Predicted Group: 1	2
1: Would purchase	4	1	5	0	5	0
2: Would not purchase	0	5	0	5	0	5

Discriminant Analysis

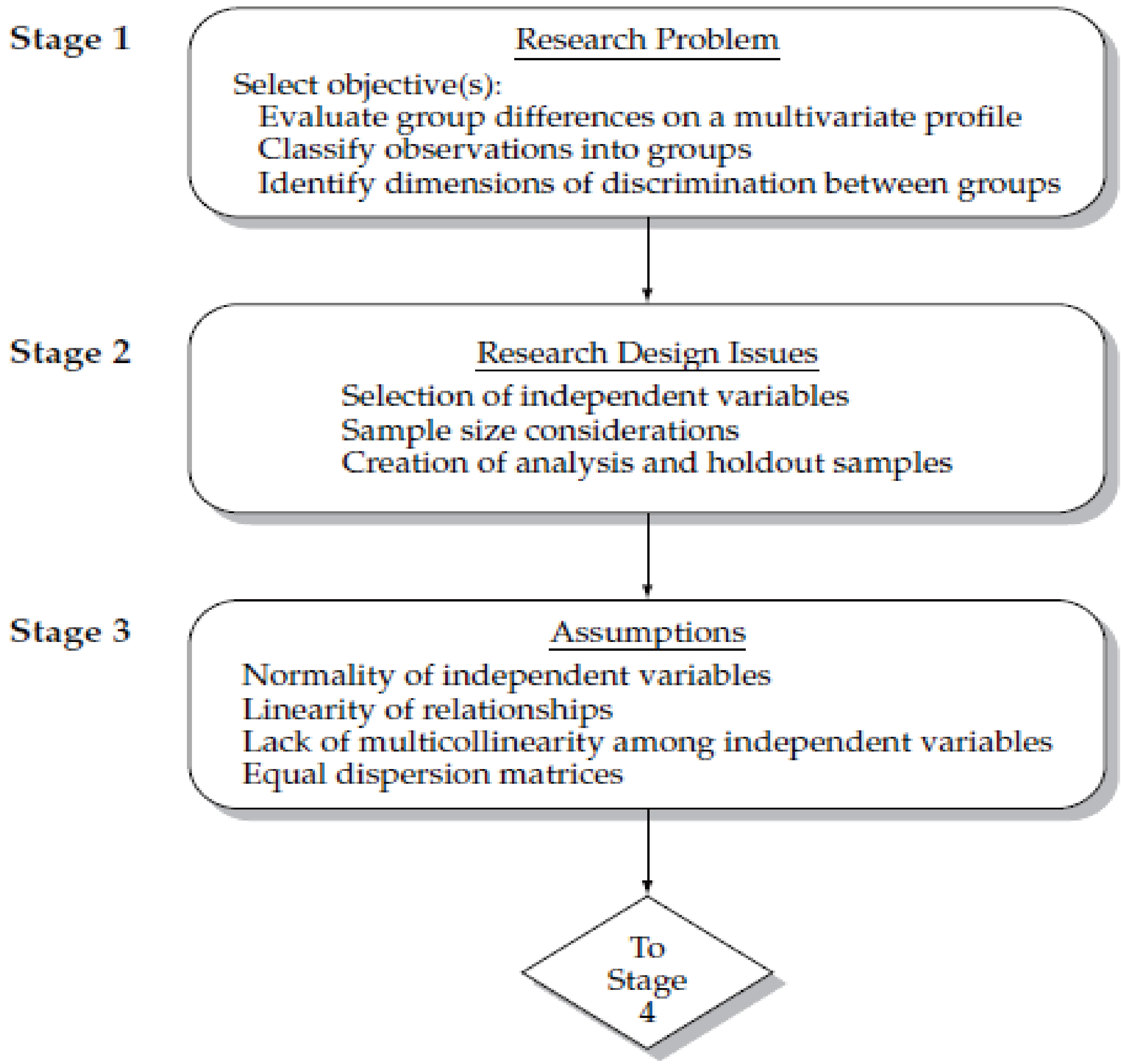
- **HYPOTHETICAL EXAMPLE OF DISCRIMINANT ANALYSIS**
- **CALCULATING A DISCRIMINANT FUNCTION**
- The first discriminant function contains just $X1$, equating the value of $X1$ to the discriminant Z score (also implying a weight of 1.0 for $X1$ and weights of zero for all other variables).
- As shown earlier, use of only $X1$, the best discriminator, results in the misclassification of subject 5 as shown in Table 2, where four out of five subjects in group 1 (all but subject 5) and five of five subjects in group 2 are correctly classified (i.e., lie on the diagonal of the classification matrix).
- The percentage correctly classified is thus 90 percent (9 out of 10 subjects).
- • Because $X2$ provides discrimination for subject 5, we can form a second discriminant function by equally combining $X1$ and $X2$ (i.e., implying weights of 1.0 for $X1$ and $X2$ and a weight of 0.0for $X3$) to utilize each variable's unique discriminatory powers. Using a cutting score of 11 with this new discriminant function (see Table 2) achieves a perfect classification of the two groups (100% correctly classified).
- Thus, $X1$ and $X2$ in combination are able to make better predictions of group membership than either variable separately.
- • The third discriminant function in Table 2 represents the actual estimated discriminant function ($Z = -4.53 + .476X1 + .359X2$). Based on a cutting score of 0, this third function also achieves a 100 percent correct classification rate with the maximum separation possible between groups.

Discriminant Analysis

- **A Geometric Representation of the Two-Group Discriminant Function**
- A graphical illustration of another two-group analysis will help to further explain the nature of discriminant analysis.



THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS



THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 1: OBJECTIVES OF DISCRIMINANT ANALYSIS**
- A review of the objectives for applying discriminant analysis should further clarify its nature.
- Discriminant analysis can address any of the following research objectives:
- **1.** Determining whether statistically significant differences exist between the average score profiles on a set of variables for two (or more) *a priori* defined groups
- **2.** Determining which of the independent variables most account for the differences in the average score profiles of the two or more groups
- **3.** Establishing the number and composition of the dimensions of discrimination between groups formed from the set of independent variables
- **4.** Establishing procedures for classifying objects (individuals, firms, products, etc.) into groups on the basis of their scores on a set of independent variables
- Discriminant analysis, therefore, can be considered either a type of profile analysis or an analytical predictive technique. In either case, the technique is most appropriate in situations with a single categorical dependent variable and several metrically scaled independent variables.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 1: OBJECTIVES OF DISCRIMINANT ANALYSIS**
- As a ***profile analysis,*** discriminant analysis provides an objective assessment of differences between groups on a set of independent variables. In this situation, discriminant analysis is quite similar to multivariate analysis of variance. For understanding group differences, discriminant analysis lends insight into the role of individual variables as well as defining combinations of these variables that represent dimensions of discrimination between groups.
- These dimensions are the collective effects of several variables that work jointly to distinguish between the groups. The use of sequential estimation methods also allows for identifying subsets of variables with the greatest discriminatory power.
- • For ***classification purposes,*** discriminant analysis provides a basis for classifying not only the sample used to estimate the discriminant function but also any other observations that can have values for all the independent variables. In this way, the discriminant analysis can be used to classify other observations into the defined groups.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 2: RESEARCH DESIGN FOR DISCRIMINANT ANALYSIS**
- Successful application of discriminant analysis requires consideration of several issues. These issues include the selection of both dependent and independent variables, the sample size needed for estimation of the discriminant functions, and the division of the sample for validation purposes.
- **Selecting Dependent and Independent Variables**
- To apply discriminant analysis, the researcher first must specify which variables are to be independent measures and which variable is to be the dependent measure.
- **THE DEPENDENT VARIABLE**
- The researcher should focus on the dependent variable first.
- The number of dependent variable groups (categories) can be two or more, but these groups must be mutually exclusive and exhaustive.
- In other words, each observation can be placed into only one group. In some cases, the dependent variable may involve two groups (dichotomous), such as good versus bad.
- In other cases, the dependent variable may involve several groups (multichotomous), such as the occupations of physician, attorney, or professor.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 2: RESEARCH DESIGN FOR DISCRIMINANT ANALYSIS**
- Theoretically, discriminant analysis can handle an unlimited number of categories in the dependent variable.
- In addition to being mutually exclusive and exhaustive, the dependent variable categories should be distinct and unique on the set of independent variables chosen.
- Discriminant analysis assumes that each group *should* have a unique profile on the independent variables used and thus develops the discriminant functions to maximally separate the groups based on these variables.
- Discriminant analysis does not, however, have a means of accommodating or combining categories that are not distinct on the independent variables.
- If two or more groups have quite similar profiles, discriminant analysis will not be able to uniquely profile each group, resulting in poorer explanation and classification of the groups as a whole.
- As such, the researcher must select the dependent variables and its categories to reflect differences in the independent variables.
- An example will help illustrate this issue.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 2: RESEARCH DESIGN FOR DISCRIMINANT ANALYSIS**
- Assume the researcher wishes to identify differences among occupational categories based on a number of demographic characteristics (e.g., income, education, household characteristics).
- If occupations are represented by a small number of categories (e.g., blue-collar, white-collar, clerical/staff, and professional/upper management), then we would expect unique differences between the groups and that discriminant analysis would be best able to develop discriminant functions that would explain the group differences and successfully classify individuals into their correct category.
- The researcher should also strive, all other things equal, for a smaller rather than larger number of categories in the dependent measure. It may seem more logical to expand the number of categories in search of more unique groupings, but expanding the number of categories represents more complexities in the profiling and classification tasks of discriminant analysis.
- If discriminant analysis can estimate up to $NG - 1$ (number of groups minus one) discriminant functions, then increasing the number of groups expands the number of possible discriminant functions, increasing the complexity in identifying the underlying dimensions of discrimination reflected by each discriminant function as well as representing the overall effect of each independent variable.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 2: RESEARCH DESIGN FOR DISCRIMINANT ANALYSIS**
- **Converting Metric Variables.**
- In some situations, however, discriminant analysis is appropriate even if the dependent variable is not a true nonmetric (categorical) variable. We may have a dependent variable that is an ordinal or interval measurement that we wish to use as a categorical dependent variable. In such cases, we would have to create a categorical variable, and two approaches are the most commonly used:
- The most common approach is to establish categories using the metric scale.
- For example, if we had a variable that measured the average number of cola drinks consumed per day, and the individuals responded on a scale from zero to eight or more per day, we could create an artificial trichotomy (three groups) by simply designating those individuals who consumed none, one, or two cola drinks per day as light users, those who consumed three, four, or five per day as medium users, and those who consumed six, seven, eight, or more as heavy users.
- Such a procedure would create a three-group categorical variable in which the objective would be to discriminate among light, medium, and heavy users of colas.
- Any number of categorical groups can be developed. Most frequently, the approach would involve creating two, three, or four categories.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 2: RESEARCH DESIGN FOR DISCRIMINANT ANALYSIS**
- **Converting Metric Variables.**
- When three or more categories are created, the possibility arises of examining only the extreme groups in a two-group discriminant analysis. The **polar extremes approach** involves comparing only the extreme two groups and excluding the middle group from the discriminant analysis.
- For example, the researcher could examine the light and heavy users of cola drinks and exclude the medium users. This approach can be used any time the researcher wishes to examine only the extreme groups.
- **THE INDEPENDENT VARIABLES**
- After a decision has been made on the dependent variable, the researcher must decide which independent variables to include in the analysis.
- Independent variables usually are selected in two ways.
- The first approach involves identifying variables either from previous research or from the theoretical model that is the underlying basis of the research question.
- The second approach is intuition—utilizing the researcher’s knowledge and intuitively selecting variables for which no previous research or theory exists but that logically might be related to predicting the groups for the dependent variable.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 2: RESEARCH DESIGN FOR DISCRIMINANT ANALYSIS**
- In both instances, the most appropriate independent variables are those that differ across at least two of the groups of the dependent variable.
- Remember that the purpose of any independent variable is to present a unique profile of at least one group as compared to others.
- Variables that do not differ across the groups are of little use in discriminant analysis.
- **Sample Size**
- Discriminant analysis, like the other multivariate techniques, is affected by the size of the sample being analyzed.
- Very small samples have so much sampling error that identification of all but the largest differences is improbable.
- Moreover, very large sample sizes will make all differences statistically significant, even though these same differences may have little or no managerial relevance.
- In between these extremes, the researcher must consider the impact of sample sizes on discriminant analysis, both at the overall level and on a group-by-group basis.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 2: RESEARCH DESIGN FOR DISCRIMINANT ANALYSIS**
- **OVERALL SAMPLE SIZE**
- The first consideration involves the overall sample size.
- Discriminant analysis is quite sensitive to the ratio of sample size to the number of predictor variables.
- As a result, many studies suggest a ratio of 20 observations for each predictor variable.
- Although this ratio may be difficult to maintain in practice, the researcher must note that the results become unstable as the sample size decreases relative to the number of independent variables.
- The minimum size recommended is five observations per independent variable.
- **SAMPLE SIZE PER CATEGORY**
- In addition to the overall sample size, the researcher also must consider the sample size of each category.
- At a minimum, the smallest group size of a category must exceed the number of independent variables. As a practical guideline, each category should have at least 20 observations.
- Even when all categories exceed 20 observations, however, the researcher must also consider the relative sizes of the categories.
- Wide variations in the groups' size will impact the estimation of the discriminant function and the classification of observations.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 2: RESEARCH DESIGN FOR DISCRIMINANT ANALYSIS**
- **Division of the Sample**
- The preferred means of validating a discriminant analysis is to divide the sample into two subsamples, one used for estimation of the discriminant function and another for validation purposes.
- In terms of sample size considerations, it is essential that each subsample be of adequate size to support conclusions from the results.
- As such, all of the considerations discussed in the previous section apply not only to the total sample, but also to each of the two subsamples.
- No hard-and-fast rules have been established, but it seems logical that the researcher would want at least 100 in the total sample to justify dividing it into the two groups.
- **CREATING THE SUBSAMPLES**
- A number of procedures have been suggested for dividing the sample into subsamples.
- The usual procedure is to divide the total sample of respondents randomly into two subsamples.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 2: RESEARCH DESIGN FOR DISCRIMINANT ANALYSIS**
- One of these subsamples, the **analysis sample**, is used to develop the discriminant function.
- The second, the **holdout sample**, is used to test the discriminant function.
- This method of validating the function is referred to as the **split-sample validation** or **cross-validation.**
- No definite guidelines have been established for determining the relative sizes of the analysis and holdout (or validation) subsamples.
- The most popular approach is to divide the total sample so that one-half of the respondents are placed in the analysis sample and the other half are placed in the holdout sample.
- However, no hard-and-fast rule has been established, and some researchers prefer a 60–40 or even 75–25 split between the analysis and the holdout groups, depending on the overall sample size.
- If the original groups are unequal, the sizes of the estimation and holdout samples should be proportionate to the total sample distribution.
- For instance, if a sample consists of 50 males and 50 females, the estimation and holdout samples would have 25 males and 25 females.
- If the sample contained 70 females and 30 males, then the estimation and holdout samples would consist of 35 females and 15 males each.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 3: ASSUMPTIONS OF DISCRIMINANT ANALYSIS**
- As with all multivariate techniques, discriminant analysis is based on a number of assumptions.
- These assumptions relate to both the statistical processes involved in the estimation and classification procedures and issues affecting the interpretation of the results.
- **Impacts on Estimation and Classification**
- The key assumptions for deriving the discriminant function are multivariate normality of the independent variables and unknown (but equal) dispersion and covariance structures (matrices) for the groups as defined by the dependent variable.
- Although the evidence is mixed regarding the sensitivity of discriminant analysis to violations of these assumptions, the researcher must always understand the impacts on the results that can be expected.
- Moreover, if the assumptions are violated and the potential remedies are not acceptable or do not address the severity of the problem, the researcher should consider alternative methods (e.g., logistic regression).

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 3: ASSUMPTIONS OF DISCRIMINANT ANALYSIS**
- **IDENTIFYING ASSUMPTION VIOLATIONS**
- Achieving univariate normality of individual variables will many times suffice to achieve multivariate normality.
- A number of tests for normality are available to the researcher, along with the appropriate remedies, those most often being transformations of the variables.
- The issue of equal dispersion of the independent variables (i.e., equivalent covariance matrices) is similar to homoscedasticity between individual variables.
- The most common test is the **Box's M** test assessing the significance of differences in the matrices between the groups.
- Here the researcher is looking for a *non-significant* probability level which would indicate that there were not differences between the group covariance matrices.
- We should use very conservative levels of significant differences (e.g., .01 rather than .05) when assessing whether differences are present.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 3: ASSUMPTIONS OF DISCRIMINANT ANALYSIS**
- **IMPACT ON ESTIMATION**
- Data not meeting the multivariate normality assumption can cause problems in the estimation of the discriminant function.
- Remedies may be possible through transformations of the data to reduce the disparities among the covariance matrices.
- However, in many instances these remedies are ineffectual. In these situations, the models should be thoroughly validated.
- If the dependent measure is binary, logistic regression should be used if at all possible.
- **IMPACT ON CLASSIFICATION**
- Unequal covariance matrices also negatively affect the classification process.
- If the sample sizes are small and the covariance matrices are unequal, then the statistical significance of the estimation process is adversely affected.
- The more likely case is that of unequal covariances among groups of adequate sample size, whereby observations are overclassified into the groups with larger covariance matrices.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

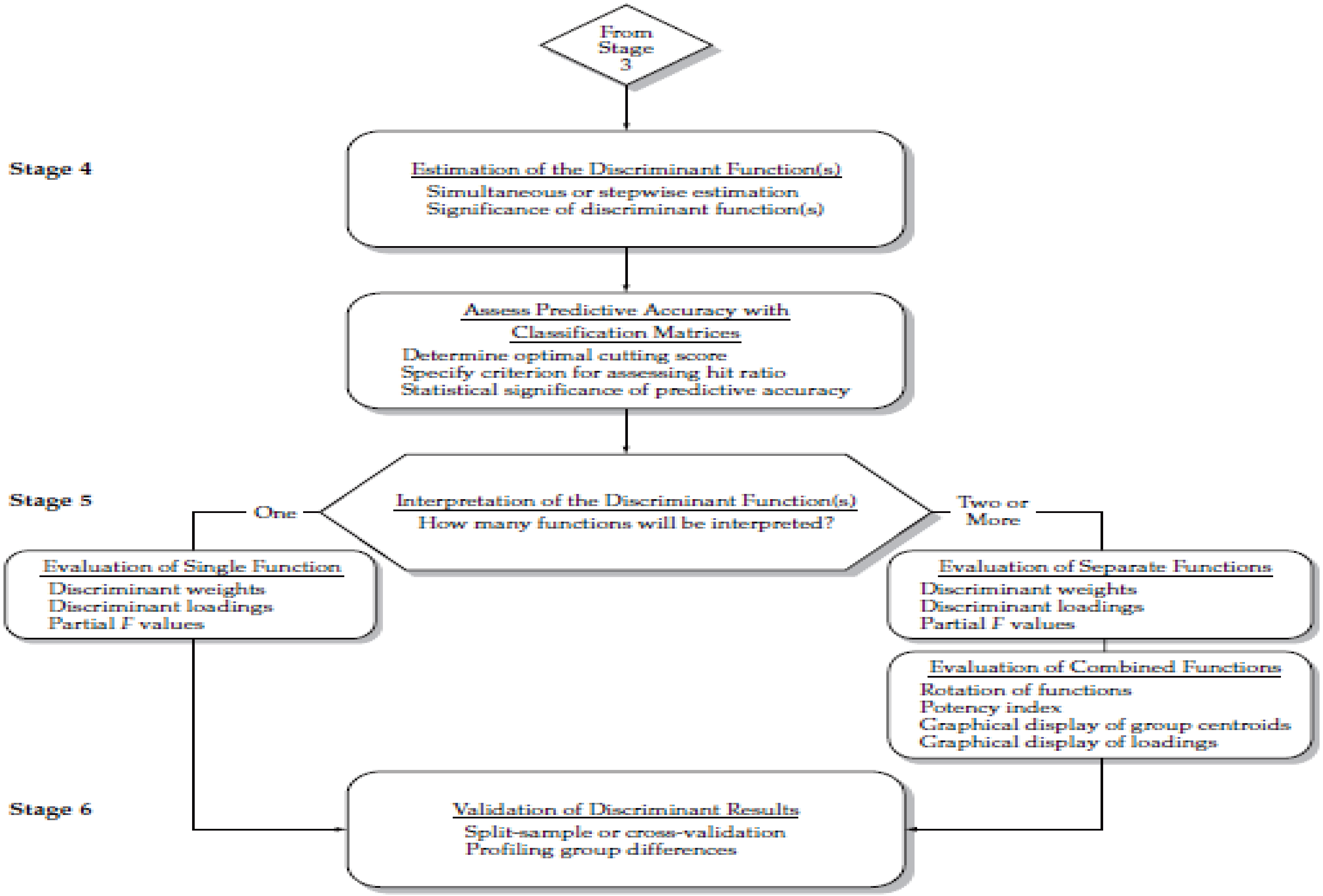
- **STAGE 3: ASSUMPTIONS OF DISCRIMINANT ANALYSIS**
- This effect can be minimized by increasing the sample size and also by using the group-specific covariance matrices for classification purposes, but this approach mandates cross-validation of the discriminant results.
- Finally, quadratic classification techniques are available in many of the statistical programs if large differences exist between the covariance matrices of the groups and the remedies do not minimize the effect.
- **Impacts on Interpretation**
- Another characteristic of the data that affects the results is multicollinearity among the independent variables.
- Multicollinearity, measured in terms of **tolerance**, denotes that two or more independent variables are highly correlated, so that one variable can be highly explained or predicted by the other variable(s) and thus it adds little to the explanatory power of the entire set.
- As with any of the multivariate techniques employing a variate, an implicit assumption is that all relationships are linear.
- Nonlinear relationships are not reflected in the discriminant function unless specific variable transformations are made to represent nonlinear effects.
- Finally, outliers can have a substantial impact on the classification accuracy of any discriminant analysis results.
- The researcher is encouraged to examine all results for the presence of outliers and to eliminate true outliers if needed.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- To derive the discriminant function, the researcher must decide on the method of estimation and then determine the number of functions to be retained.
- With the functions estimated, overall model fit can be assessed in several ways.
- First, **discriminant *Z* scores**, also known as the ***Z* scores**, can be calculated for each object.
- Comparison of the group means (centroids) on the *Z* scores provides one measure of discrimination between groups.
- Predictive accuracy can be measured as the number of observations classified into the correct groups, with a number of criteria available to assess whether the classification process achieves practical or statistical significance.
- Finally, casewise diagnostics can identify the classification accuracy of each case and its relative impact on the overall model estimation.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**



THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Selecting an Estimation Method**
- The first task in deriving the discriminant function(s) is to choose the estimation method.
- The two methods available are the simultaneous (direct) method and the stepwise method, each discussed next.
- **SIMULTANEOUS ESTIMATION**
- **Simultaneous estimation** involves computing the discriminant function so that all of the independent variables are considered concurrently.
- Thus, the discriminant function is computed based upon the entire set of independent variables, regardless of the discriminating power of each independent variable.
- The simultaneous method is appropriate when, for theoretical reasons, the researcher wants to include all the independent variables in the analysis and is not interested in seeing intermediate results based only on the most discriminating variables.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **STEPWISE ESTIMATION**
- **Stepwise estimation** is an alternative to the simultaneous approach.
- It involves entering the independent variables into the discriminant function one at a time on the basis of their discriminating power.
- The stepwise approach follows a sequential process of adding or deleting variables in the following manner:
- **1.** Choose the single best discriminating variable.
- **2.** Pair the initial variable with each of the other independent variables, one at a time, and select the variable that is best able to improve the discriminating power of the function in combination with the first variable.
- **3.** Select additional variables in a like manner. Note that as additional variables are included, some previously selected variables may be removed if the information they contain about group differences is available in some combination of the other variables included at later stages.
- **4.** Consider the process completed when either all independent variables are included in the function or the excluded variables are judged as not contributing significantly to further discrimination.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Statistical Significance**
- After estimation of the discriminant function(s), the researcher must assess the level of significance for the collective discriminatory power of the discriminant function(s) as well as the significance of each separate discriminant function.
- Evaluating the overall significance provides the researcher with the information necessary to decide whether to continue on to the interpretation of the analysis or if respecification is necessary.
- If the overall model is significant, then evaluating the individual functions identifies the function(s) that should be retained and interpreted.
- **OVERALL SIGNIFICANCE**
- In assessing the statistical significance of the overall model, different statistical criteria are applicable for simultaneous versus stepwise estimation procedures.
- In both situations, the statistical tests relate to the ability of the discriminant function(s) to derive discriminant *Z* scores that are significantly different between the groups.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Simultaneous Estimation.**
- When a simultaneous approach is used, the measures of Wilks' lambda, Hotelling's trace, and Pillai's criterion all evaluate the statistical significance of the discriminatory power of the discriminant function(s).
- Roy's greatest characteristic root evaluates only the first discriminant function.
- **Stepwise Estimation.**
- If a stepwise method is used to estimate the discriminant function, the Mahalanobis *D*2 and Rao's *V* measures are most appropriate. Both are measures of generalized distance.
- The Mahalanobis *D*2 procedure is based on generalized squared Euclidean distance that adjusts for unequal variances.
- The major advantage of this procedure is that it is computed in the original space of the predictor variables rather than as a collapsed version used in other measures.
- The Mahalanobis *D*2 procedure becomes particularly critical as the number of predictor variables increases, because it does not result in any reduction in dimensionality.
- A loss in dimensionality would cause a loss of information because it decreases variability of the independent variables.
- In general, Mahalanobis *D*2 is the preferred procedure when the researcher is interested in the maximal use of available information in a stepwise process.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **SIGNIFICANCE OF INDIVIDUAL DISCRIMINANT FUNCTIONS**
- If the number of groups is three or more, then the researcher must decide not only whether the discrimination between groups overall is statistically significant but also whether each of the estimated discriminant functions is statistically significant.
- As discussed earlier, discriminant analysis estimates one less discriminant function than there are groups.
- If three groups are analyzed, then two discriminant functions will be estimated; for four groups, three functions will be estimated; and so on.
- The conventional significance criterion of .05 or beyond is often used, yet some researchers extend the required significance level (e.g., .10 or more) based on the trade-off of cost versus the value of the information.
- If the higher levels of risk for including nonsignificant results (e.g., significance levels > .05) are acceptable, discriminant functions may be retained that are significant at the .2 or even the .3 level.
- If one or more functions are deemed not statistically significant, the discriminant model should be reestimated with the number of functions to be derived limited to the number of significant functions.
- In this manner, the assessment of predictive accuracy and the interpretation of the discriminant functions will be based only on significant functions.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Assessing Overall Model Fit**
- Once the significant discriminant functions have been identified, attention shifts to ascertaining the overall fit of the retained discriminant function(s).
- This assessment involves three tasks:
- **1.** Calculating discriminant *Z* scores for each observation
- **2.** Evaluating group differences on the discriminant *Z* scores
- **3.** Assessing group membership prediction accuracy
- The discriminant *Z* score is calculated for each discriminant function for every observation in the sample.
- The discriminant score acts as a concise and simple representation of each discriminant function, simplifying the interpretation process and the assessment of the contribution of independent variables.
- Groups can be distinguished by their discriminant scores and, as we will see, the discriminant scores can play an instrumental role in predicting group membership.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Assessing Overall Model Fit**
- **CALCULATING DISCRIMINANT *Z* SCORES**
- With the retained discriminant functions defined, the basis for calculating the discriminant *Z* scores has been established.
- As discussed earlier, the discriminant *Z* score of any discriminant function can be calculated for each observation by the following formula:

$$Z_{jk} = a + W_1 X_{1k} + W_2 X_{2k} + \ldots + W_n X_{nk}$$

where

Z_{jk} = discriminant Z score of discriminant function j for object k
a = intercept
W_i = discriminant weight for independent variable i
X_{ik} = independent variable i for object k

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- The discriminant *Z* score, a metric variable, provides a direct means of comparing observations on each function.
- Observations with similar *Z* scores are assumed more alike on the variables constituting this function than those with disparate scores.
- The discriminant function can be expressed with either standardized or unstandardized weights and values.
- The standardized version is more useful for interpretation purposes, but the unstandardized version is easier to use in calculating the discriminant *Z* score.
- **EVALUATING GROUP DIFFERENCES**
- Once the discriminant *Z* scores are calculated, the first assessment of overall model fit is to determine the magnitude of differences between the members of each group in terms of the discriminant *Z* scores.
- A summary measure of the group differences is a comparison of the group centroids, the average discriminant *Z* score for all group members.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- A measure of success of discriminant analysis is its ability to define discriminant function(s) that result in significantly different group centroids.
- The differences between centroids are measured in terms of Mahalanobis *D*2 measure, for which tests are available to determine whether the differences are statistically significant.
- The researcher should ensure that even with significant discriminant functions, significant differences occur between each of the groups.
- Group centroids on each discriminant function can also be plotted to demonstrate the results from a graphical perspective.
- Plots are usually prepared for the first two or three discriminant functions (assuming they are statistically significant functions).
- The values for each group show its position in reduced discriminant space. The researcher can see the differences between the groups on each function;

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **ASSESSING GROUP MEMBERSHIP PREDICTION ACCURACY**
- Given that the dependent variable is nonmetric, it is not possible to use a measure such as $R2$, as is done in multiple regression, to assess predictive accuracy.
- Rather, each observation must be assessed as to whether it was correctly classified. In doing so, several major considerations must be addressed:
- • The statistical and practical rationale for developing classification matrices
- • Classifying individual cases
- • Construction of the classification matrices
- • Standards for assessing classification accuracy

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Why Classification Matrices Are Developed.**
- The statistical tests for assessing the significance of the discriminant function(s) only assess the degree of difference between the groups based on the discriminant *Z* scores, but do not indicate how well the function(s) predicts.
- These statistical tests suffer the same drawbacks as the classical tests of hypotheses.
- For example, suppose the two groups are deemed significantly different beyond the .01 level. Yet with sufficiently large sample sizes, the group means (centroids) could be virtually identical and still have statistical significance.
- To determine the predictive ability of a discriminant function, the researcher must construct classification matrices.
- The **classification matrix** procedure provides a perspective on practical significance rather than statistical significance.
- With multiple discriminant analysis, the **percentage correctly classified**, also termed the **hit ratio**, reveals how well the discriminant function classified the objects.
- With a sufficiently large sample size in discriminant analysis, we could have a statistically significant difference between the two (or more) groups and yet correctly classify only 53 percent (when chance is 50%, with equal group sizes) .
- In such instances, the statistical test would indicate statistical significance, yet the hit ratio would allow for a separate judgment to be made in terms of practical significance.
- Thus, we must use the classification matrix procedure to assess predictive accuracy beyond just statistical significance.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Classifying Individual Observations.**
- The development of classification matrices requires that each observation be classified into one of the groups of the dependent variable based on the discriminant function(s).
- The objective is to characterize each observation on the discriminant function(s) and then determine the extent to which observation in each group can be consistently described by the discriminant functions.
- There are two approaches to classifying observations, one employing the discriminant scores directly and another developing a specific function for classification.
- ***Cutting Score Calculation***
- Using the discriminant functions deemed significant, we can develop classification matrices by calculating the **cutting score** (also called the *critical Z value*) for each discriminant function.
- The cutting score is the criterion against which each object's discriminant score is compared to determine into which group the object should be classified.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- The cutting score represents the dividing point used to classify observations into groups based on their discriminant function score.
- The calculation of a cutting score between any two groups is based on the two group centroids (group mean of the discriminant scores) and the relative size of the two groups.
- The group centroids are easily calculated and provided at each stage of the stepwise process.
- ***Developing a Classification Function***
- As noted earlier, using the discriminant function is only one of two possible approaches to classification.
- The second approach employs a **classification function**, also known as **Fisher's linear discriminant function**.
- The classification functions, one for each group, are used strictly for classifying observations.
- In this method of classification, an observation's values for the independent variables are inserted in the classification functions and a classification score for each group is calculated for that observation.
- The observation is then classified into the group with the highest classification score.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Defining Prior Probabilities.**
- The impact and importance of each group's sample size in the classification process is many times overlooked, yet is critical in making the appropriate assumptions in the classification process.
- Here we are concerned about the representativeness of the sample as it relates to representation of the relative sizes of the groups in the actual in the actual population, which can be stated as prior probabilities (i.e., the relative proportion of each group to the total sample).
- If, however, the sample was conducted randomly and the researcher feels that the group sizes are representative of the population, then the researcher can specify prior probabilities to be based on the estimation sample.
- Thus, the actual group sizes are assumed representative and used directly in the calculation of the cutting score.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- For example, consider a holdout sample consisting of 200 observations, with group sizes of 60 and 140 that relate to prior probabilities of 30 percent and 70 percent, respectively.
- If the sample is assumed representative, then the sample sizes of 60 and 140 are used in calculating the cutting score. If, however, the sample is deemed not representative, the researcher must specify the prior probabilities.
- If they are specified as equal (50% and 50%), sample sizes of 100 and 100 would be used in the cutting score calculation rather than the actual sample sizes.
- Specifying other values for the prior probabilities would result in differing sample sizes for the two groups.
- ***Calculating the Optimal Cutting Score***
- The importance of the prior probabilities can be illustrated in the calculation of the “optimal” cutting score, which takes into account the prior probabilities through the use of group sizes.
- The basic formula for computing the **optimal cutting score** between any two groups is:

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**

$$Z_{CS} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B}$$

where

Z_{CS} = optimal cutting score between groups A and B
N_A = number of observations in group A
N_B = number of observations in group B
Z_A = centroid for group A
Z_B = centroid for group B

- With unequal group sizes, the optimal cutting score for a discriminant function is now the weighted average of the group centroids.
- The cutting score is weighted toward the smaller group, hopefully making for a better classification of the larger group.
- If the groups are specified to be of equal size (prior probabilities defined as equal), then the optimum cutting score will be halfway between the two group centroids and becomes simply the average of the two centroids:

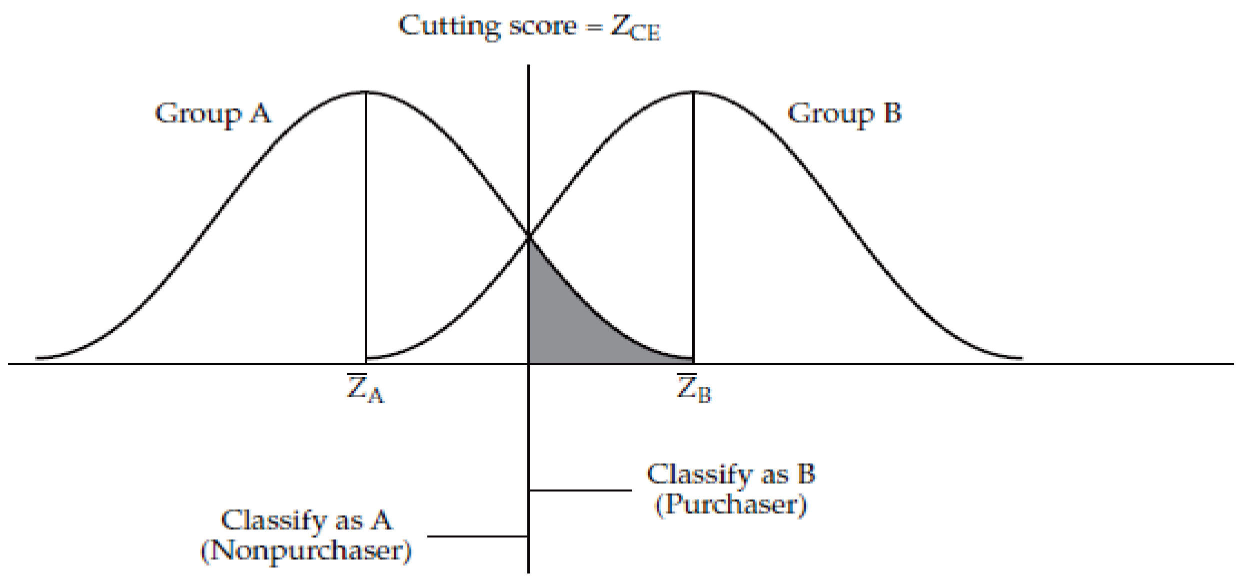
THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**

$$Z_{CE} = \frac{Z_A + Z_B}{2}$$

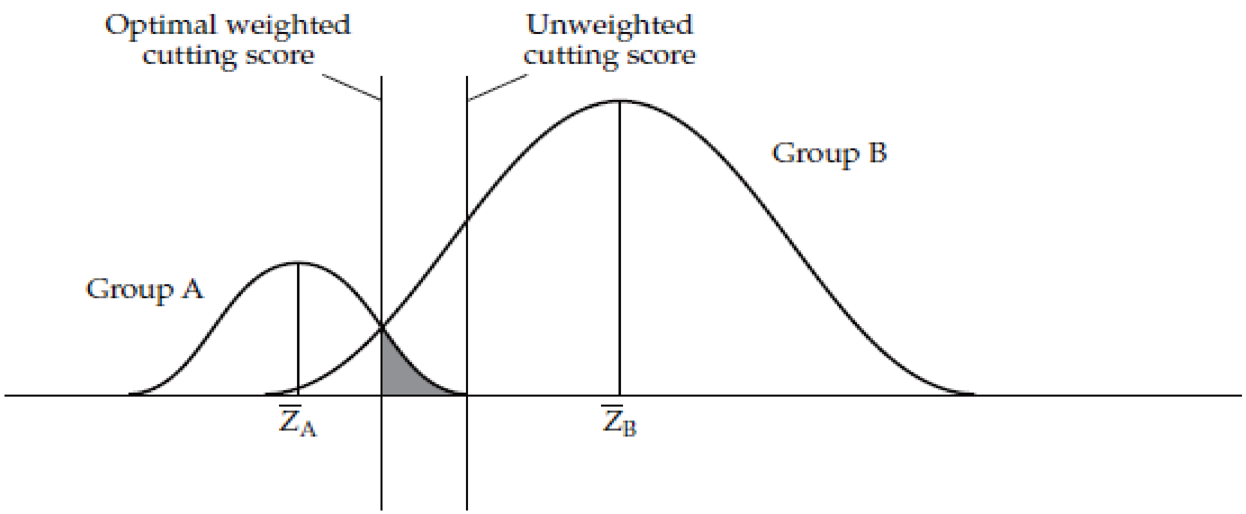
where

Z_{CE} = critical cutting score value for equal group sizes
Z_A = centroid for group A
Z_B = centroid for group B



THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**



- Both of the formulas for calculating the optimal cutting score assume that the distributions are normal and the group dispersion structures are known.
- Both the weighted and unweighted cutting scores are shown.
- It is apparent that if group A is much smaller than group B, the optimal cutting score will be closer to the centroid of group A than to the centroid of group B.
- Also, if the unweighted cutting score was used, none of the objects in group A would be misclassified, but a substantial portion of those in group B would be misclassified.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Costs of Misclassification**
- The optimal cutting score also must consider the cost of misclassifying an object into the wrong group.
- If the costs of misclassifying are approximately equal for all groups, the optimal cutting score will be the one that will misclassify the fewest number of objects across all groups.
- If the misclassification costs are unequal, the optimum cutting score will be the one that minimizes the costs of misclassification.
- **Constructing Classification Matrices**
- To validate the discriminant function through the use of classification matrices, the sample should be randomly divided into two groups.
- One of the groups (the analysis sample) is used to compute the discriminant function.
- The other group (the holdout or validation sample) is retained for use in developing the classification matrix.
- The classification of each observation can be accomplished through either of the classification approaches discussed earlier.
- For the Fisher's approach, an observation is classified into the group with the largest classification function score.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- When using the discriminant scores and the optimal cutting score, the procedure is as follows:
  - Classify an individual into group A if $Z_n < Z_{ct}$
  - or
  - Classify an individual into group B if $Z_n > Z_{ct}$
- Where
  - Zn = discriminant Z score for the nth individual
  - Zct = critical cutting score value
- The results of the classification procedure are presented in matrix form, as shown in Table 4.
- The entries on the diagonal of the matrix represent the number of individuals correctly classified.
- The numbers off the diagonal represent the incorrect classifications.
- The entries under the column labelled "Actual Group Size" represent the number of individuals actually in each of the two groups.
- The entries at the bottom of the columns represent the number of individuals assigned to the groups by the discriminant function.
- The percentage correctly classified for each group is shown at the right side of the matrix, and the overall percentage correctly classified, also known as the hit ratio, is shown at the bottom.
- In our example, the number of individuals correctly assigned to group 1 is 22, whereas 3 members of group 1 are incorrectly assigned to group 2. Similarly, the number of correct classifications to group 2 is 20, and the number of incorrect assignments to group 1 is 5.
- Thus, the classification accuracy percentages of the discriminant function for the actual groups 1 and 2 are 88 and 80 percent, respectively. The overall classification accuracy (hit ratio) is 84 percent.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**

TABLE 4 Classification Matrix for Two-Group Discriminant Analysis

	Predicted Group			
Actual Group	**1**	**2**	**Actual Group Size**	**Percentage Correctly Classified**
1	22	3	25	88
2	5	20	25	80
Predicted group size	27	23	50	84[a]

[a]Percent correctly classified = (Number correctly classified/Total number of observations) × 100
= [(22 + 20)/50] × 100
= 84%

- One final topic regarding classification procedures is the *t* test available to determine the level of significance for the classification accuracy.
- The formula for a two-group analysis (equal sample size) is

$$t = \frac{p - .5}{\sqrt{\frac{.5(1.0 - .5)}{N}}}$$

where

p = proportion correctly classified
N = sample size

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Establishing Standards of Comparison for the Hit Ratio.**
- As noted earlier, the predictive accuracy of the discriminant function is measured by the hit ratio, which is obtained from the classification matrix.
- ***Standards of Comparison for the Hit Ratio for Equal Group Sizes***
- When the sample sizes of the groups are equal, the determination of the chance classification is rather simple; it is obtained by dividing 1 by the number of groups.
- The formula is: CEQUAL = 1 / Number of groups
- For instance, for a two-group function the chance probability would be .50; for a three-group function the chance probability would be .33; and so forth.
- ***Standards of Comparison for the Hit Ratio for Unequal Group Sizes***
- The determination of the chance classification for situations in which the group sizes are unequal is somewhat more involved.
- Let us assume that we have a total sample of 200 observations divided into holdout and analysis samples of 100 observations each. In the holdout sample, 75 subjects belong to one group and 25 to the other.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- Referred to as the **maximum chance criterion**, we could arbitrarily assign all the subjects to the largest group.
- The maximum chance criterion should be used when the sole objective of the discriminant analysis is to maximize the percentage correctly classified.
- It is also the most conservative standard because it will generate the highest standard of comparison.
- In our simple example of a sample with two groups (75 and 25 people each), using this method would set a 75 percent classification accuracy, what would be achieved by classifying everyone into the largest group without the aid of any discriminant function.
- It could be concluded that unless the discriminant function achieves a classification accuracy higher than 75 percent, it should be disregarded because it has not helped us improve the prediction accuracy we could achieve without using any discriminant analysis at all.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- When group sizes are unequal and the researcher wishes to correctly identify members of all of the groups, not just the largest group, the **proportional chance criterion** is deemed by many to be the most appropriate. The formula for this criterion is

$$CPRO = p2 + (1 - p)2$$

- Where
  - p = proportion of individuals in group 1
  - 1 - p = proportion of individuals in group 2
- Using the group sizes from our earlier example (75 and 25), we see that the proportional chance criterion would be 62.5 percent [.752 + (1.0 – .75)2 = .625] compared with 75 percent.
- Therefore, in this instance, the actual prediction accuracy of 75 percent may be acceptable because it is above the 62.5 percent proportional chance criterion.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Overall Versus Group-Specific Hit Ratios.**
- To this point, we focused on evaluating the overall hit ratio across all groups in assessing the predictive accuracy of a discriminant analysis.
- The researcher also must be concerned with the hit ratio (percent correctly classified) for each separate group.
- If you focus solely on the overall hit ratio, it is possible that one or more groups, particularly smaller groups, may have unacceptable hit ratios while the overall hit ratio is acceptable.
- The researcher should evaluate each group's hit ratio and assess whether the discriminant analysis provides adequate levels of predictive accuracy both at the overall level as well as for each group.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Statistically Based Measures of Classification Accuracy Relative to Chance.**
- A statistical test for the discriminatory power of the classification matrix when compared with a chance model is **Press's *Q* statistic**.
- This simple measure compares the number of correct classifications with the total sample size and the number of groups.
- The calculated value is then compared with a critical value (the chi-square value for 1 degree of freedom at the desired confidence level).
- If it exceeds this critical value, then the classification matrix can be deemed statistically better than chance.
- The *Q* statistic is calculated by the following formula:
  - Press's Q = [N - (nK)]2/N(K - 1)
- Where
  - K = number of groups
  - n = number of observations correctly classified
  - N = total sample size
- Press's Q = [50 - (42 * 2)]/50(2 - 1) = 23.12

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Statistically Based Measures of Classification Accuracy Relative to Chance.**
- The critical value at a significance level of .01 is 6.63. Thus, we would conclude that in the example the predictions were significantly better than chance, which would have a correct classification rate of 50 percent.
- This simple test is sensitive to sample size; large samples are more likely to show significance than small sample sizes of the same classification rate.
- For example, if the sample size is increased to 100 in the example and the classification rate remains at 84 percent, the *Q* statistic increases to 46.24.
- Thus, examine the *Q* statistic in light of the sample size because increases in sample size will increase the *Q* statistic even for the same overall classification rate.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **Casewise Diagnostics**
- The final means of assessing model fit is to examine the predictive results on a case-by-case basis.
- Similar to the analysis of residuals in multiple regression, the objective is to understand which observations (1) have been misclassified and (2) are not representative of the remaining group members.
- Although the classification matrix provides overall classification accuracy, it does not detail the individual case results.
- Also, even if we can denote which cases are correctly or incorrectly classified, we still need a measure of an observation’s similarity to the remainder of the group.
- **MISCLASSIFICATION OF INDIVIDUAL CASES**
- When analyzing residuals from a multiple regression analysis, an important decision involves setting the level of residual considered substantive and worthy of attention.
- In discriminant analysis, this issue is somewhat simpler because an observation is either correctly or incorrectly classified.
- All computer programs provide information that identifies which cases are misclassified and to which group they were misclassified.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- **ANALYZING MISCLASSIFIED CASES**
- The purpose of identifying and analyzing the misclassified observations is to identify any characteristics of these observations that could be incorporated into the discriminant analysis for improving predictive accuracy.
- This analysis may take the form of profiling the misclassified cases on either the independent variables or other variables not included in the model.
- **Profiling on the Independent Variables.**
- Examining these cases on the independent variables may identify nonlinear trends or other relationships or attributes that led to the misclassification.
- Several techniques are particularly appropriate in discriminant analysis:
- A graphical representation of the observations is perhaps the simplest yet effective approach for examining the characteristics of observations, especially the misclassified observations.
- The most common approach is to plot the observations based on their discriminant *Z* scores and portray the overlap among groups and the misclassified cases. If two or more functions are retained, the optimal cutting points can also be portrayed to give what is known as a **territorial map** depicting the regions corresponding to each group.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 4: ESTIMATION OF THE DISCRIMINANT MODEL AND ASSESSING OVERALL FIT**
- Plotting the individual observations along with the group centroids, as discussed earlier, shows not only the general group characteristics depicted in the centroids, but also the variation in the group members.
- A direct empirical assessment of the similarity of an observation to the other group members can be made by evaluating the Mahalanobis *D*2 distance of the observation to the group centroid.
- Based on the set of independent variables, observations closer to the centroid have a smaller Mahalanobis *D*2 and are assumed more representative of the group than those farther away.
- The empirical measure should be combined with a graphical analysis, however, because although a large Mahalanobis *D*2 value does indicate observations that are quite different from the group centroids, it does not always indicate misclassification.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 5: INTERPRETATION OF THE RESULTS**
- If the discriminant function is statistically significant and the classification accuracy is acceptable, the researcher should focus on making substantive interpretations of the findings.
- This process involves examining the discriminant functions to determine the relative importance of each independent variable in discriminating between the groups.
- Three methods of determining the relative importance have been proposed:
- **1.** Standardized discriminant weights
- **2.** Discriminant loadings (structure correlations)
- **3.** Partial *F* values
- **Discriminant Weights**
- The traditional approach to interpreting discriminant functions examines the sign and magnitude of the standardized **discriminant weight** (also referred to as a **discriminant coefficient**) assigned to each variable in computing the discriminant functions.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 5: INTERPRETATION OF THE RESULTS**
- When the sign is ignored, each weight represents the relative contribution of its associated variable to that function.
- Independent variables with relatively larger weights contribute more to the discriminating power of the function than do variables with smaller weights. The sign denotes only that the variable makes either a positive or a negative contribution.
- **Discriminant Loadings**
- **Discriminant loadings**, referred to sometimes as **structure correlations**, are increasingly used as a basis for interpretation because of the deficiencies in utilizing weights.
- Measuring the simple linear correlation between each independent variable and the discriminant function, the discriminant loadings reflect the variance that the independent variables share with the discriminant function.
- In that regard they can be interpreted like factor loadings in assessing the relative contribution of each independent variable to the discriminant function.
- One unique characteristic of loadings is that loadings can be calculated for all variables, whether they were used in the estimation of the discriminant function or not.
- This aspect is particularly useful when a stepwise estimation procedure is employed and some variables are not included in the discriminant function.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 5: INTERPRETATION OF THE RESULTS**
- In either simultaneous or stepwise discriminant analysis, variables that exhibit a loading of +/-.40 or higher are considered substantive.
- With stepwise procedures, this determination is supplemented because the technique prevents nonsignificant variables from entering the function.
- However, multicollinearity and other factors may preclude a variable from entering the equation, which does not necessarily mean that it does not have a substantial effect.
- **Partial *F* Values**
- As discussed earlier, two computational approaches—simultaneous and stepwise—can be utilized in deriving discriminant functions.
- When the stepwise method is selected, an additional means of interpreting the relative discriminating power of the independent variables is available through the use of partial *F* values.
- It is accomplished by examining the absolute sizes of the significant *F* values and ranking them.
- Large *F* values indicate greater discriminatory power. In practice, rankings using the *F* values approach are the same as the ranking derived from using discriminant weights, but the *F* values indicate the associated level of significance for each variable.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 5: INTERPRETATION OF THE RESULTS**
- **Interpretation of Two or More Functions**
- In cases of two or more significant discriminant functions, we are faced with additional problems of interpretation.
- These problems are found both in measuring the total discriminating effects across functions and in assessing the role of each variable in profiling each function separately.
- We address these two questions by introducing the concepts of rotation of the functions, the potency index, and stretched vectors representations.
- **ROTATION OF THE DISCRIMINANT FUNCTIONS**
- After the discriminant functions are developed, they can be rotated to redistribute the variance.
- Basically, rotation preserves the original structure and the reliability of the discriminant solution while making the functions easier to interpret substantively.
- In most instances, the VARIMAX rotation is employed as the basis for rotation.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 5: INTERPRETATION OF THE RESULTS**
- **POTENCY INDEX**
- Previously, we discussed using the standardized weights or discriminant loadings as measures of a variable's contribution to a discriminant function.
- When two or more functions are derived, however, a composite or summary measure is useful in describing the contributions of a variable across all significant functions.
- The **potency index** is a relative measure among all variables and is indicative of each variable's discriminating power.
- It includes both the contribution of a variable to a discriminant function (its discriminant loading) and the relative contribution of the function to the overall solution (a relative measure among the functions based on eigenvalues).
- The composite is simply the sum of the individual potency indices across all significant discriminant functions.
- Interpretation of the composite measure is limited, however, by the fact that it is useful only in depicting the relative position (such as the rank order) of each variable, and the absolute value has no real meaning.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 5: INTERPRETATION OF THE RESULTS**
- **POTENCY INDEX**
- The potency index is calculated by a two-step process:
- **Step 1:** *Calculate a potency value of each variable for each significant function.* In the first step, the discriminating power of a variable, represented by the squared value of the unrotated discriminant loading, is "weighted" by the relative contribution of the discriminant function to the overall solution. First, the relative eigenvalue measure for each significant discriminant function is calculated simply as:

$$\text{Relative eigenvalue of discriminant function } j = \frac{\text{Eigenvalue of discriminant function } j}{\text{Sum of eigenvalues across all significant functions}}$$

The potency value of each variable on a discriminant function is then:

$$\text{Potency value of variable } i \text{ on function } j = (\text{Discriminant loading}_{ij})^2 \times \text{Relative eigenvalue of function } j$$

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 5: INTERPRETATION OF THE RESULTS**
- **POTENCY INDEX**
- **Step 2:** *Calculate a composite potency index across all significant functions.* Once a potency value has been calculated for each function, the composite potency index for each variable is calculated as:
- **Composite potency of variable i = Sum of potency values of variable i across all significant discriminant functions**
- The potency index now represents the total discriminating effect of the variable across all of the significant discriminant functions. It is only a relative measure, however, and its absolute value has no substantive meaning.
- **GRAPHICAL DISPLAY OF DISCRIMINANT SCORES AND LOADINGS**
- To depict group differences on the predictor variables, the researcher can use two different approaches to graphical display.
- The territorial map plots the individual cases on the significant discriminant functions to enable the researcher to assess the relative position of each observation based on the discriminant function scores.
- The second approach is to plot the discriminant loadings to understand the relative grouping and magnitude of each loading on each function.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 5: INTERPRETATION OF THE RESULTS**
- **Territorial Map.**
- The most common graphical method is the territorial map, where each observation is plotted in a graphical display based on the discriminant function *Z* scores of the observations.
- For example, assume that a three-group discriminant analysis had two significant discriminant functions.
- A territorial map is created by plotting each observation's discriminant *Z* scores for the first discriminant function on the X axis and the scores for the second discriminant function on the Y axis.
- As such, it provides several perspectives on the analysis:
- • Plotting each group's members with differing symbols allows for an easy portrayal of the distinctiveness of each group as well as its overlap with each other group.
- • Plotting each group's centroids provides a means for assessing each group member relative to its group centroid. This procedure is particularly useful when assessing whether large Mahalanobis *D*2 measures lead to misclassification.
- • Lines representing the cutting scores can also be plotted, denoting boundaries depicting the ranges of discriminant scores predicted into each group. Any group's members falling outside these boundaries are misclassified.
- Denoting the misclassified cases allows for assessing which discriminant function was most responsible for the misclassification as well as the degree to which a case is misclassified.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 5: INTERPRETATION OF THE RESULTS**
- **Vector Plot of Discriminant Loadings.**
- The simplest graphical approach to depicting discriminant loadings is to plot the actual rotated or unrotated loadings on a graph.
- The preferred approach would be to plot the rotated loadings. Similar to the graphical portrayal of factor loadings, this method depicts the degree to which each variable is associated with each discriminant function.
- An even more accurate approach, however, involves plotting the loadings as well as depicting vectors for each loading and group centroid. A **vector** is merely a straight line drawn from the origin (center) of a graph to the coordinates of a particular variable's discriminant loadings or a group centroid.
- With a **stretched vector** representation, the length of each vector becomes indicative of the relative importance of each variable in discriminating among the groups.
- The plotting procedure proceeds in three steps:
- **1.** *Selecting variables:* All variables, whether included in the model as significant or not, may be plotted as vectors. In this way, the importance of collinear variables that are not included, such as in a stepwise solution, can still be portrayed.
- **2.** *Stretching the vectors:* Each variable's discriminant loadings are stretched by multiplying the discriminant loading (preferably after rotation) by its respective univariate *F* value. We note that vectors point toward the groups having the highest mean on the respective predictor and away from the groups having the lowest mean scores.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 5: INTERPRETATION OF THE RESULTS**
- **3.** *Plotting the group centroids:* The group centroids are also stretched in this procedure by multiplying them by the approximate *F* value associated with each discriminant function. If the loadings are stretched, the centroids must be stretched as well to plot them accurately on the same graph.
- The approximate *F* values for each discriminant function are obtained by the following formula:

$$F \text{ value}_{\text{Function}_i} = \text{Eigenvalue}_{\text{Function}_i}\left(\frac{N_{\text{Estimation Sample}} - NG}{NG - 1}\right)$$

where

$$N_{\text{Estimation Sample}} = \text{sample size of estimation sample}$$

- As an example, assume that the sample of 50 observations was divided into three groups. The multiplier of each eigenvalue would be (50 - 3) ÷ (3 - 1) = 23.5.
- When completed, the researcher has a portrayal of the grouping of variables on each discriminant function, the magnitude of the importance of each variable (represented by the length of each vector), and the profile of each group centroid (shown by the proximity to each vector).
- Although this procedure must be done manually in most instances, it provides a complete portrayal of both discriminant loadings and group centroids.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 6: VALIDATION OF THE RESULTS**
- The final stage of a discriminant analysis involves validating the discriminant results to provide assurances that the results have external as well as internal validity.
- *With the propensity of discriminant analysis to inflate the hit ratio if evaluated only on the analysis sample, validation is an essential step.*
- In addition to validating the hit ratios, the researcher should use group profiling to ensure that the group means are valid indicators of the conceptual model used in selecting the independent variables.
- **Validation Procedures**
- Validation is a critical step in any discriminant analysis because many times, especially with smaller samples, the results can lack generalizability (external validity).
- The most common approach for establishing external validity is the assessment of hit ratios.
- Validation can occur either with a separate sample (holdout sample) or utilizing a procedure that repeatedly processes the estimation sample.
- External validity is supported when the hit ratio of the selected approach exceeds the comparison standards that represent the predictive accuracy expected by chance.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 6: VALIDATION OF THE RESULTS**
- **UTILIZING A HOLDOUT SAMPLE**
- Most often the validation of the hit ratios is performed by creating a holdout sample, also referred to as the **validation sample**.
- The purpose of utilizing a holdout sample for validation purposes is to see how well the discriminant function works on a sample of observations not used to derive the discriminant function.
- This process involves developing a discriminant function with the analysis sample and then applying it to the holdout sample.
- The justification for dividing the total sample into two groups is that an upward bias will occur in the prediction accuracy of the discriminant function if the individuals used in developing the classification matrix are the same as those used in computing the function; that is, the classification accuracy will be higher than is valid when applied to the estimation sample.
- Other researchers have suggested that even greater confidence could be placed in the validity of the discriminant function by following this procedure several times.
- Instead of randomly dividing the total sample into analysis and holdout groups once, the researcher would randomly divide the total sample into analysis and holdout samples several times, each time testing the validity of the discriminant function through the development of a classification matrix and a hit ratio.
- Then the several hit ratios would be averaged to obtain a single measure.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 6: VALIDATION OF THE RESULTS**
- **CROSS-VALIDATION**
- The cross-validation approach to assessing external validity is performed with multiple subsets of the total sample.
- The most widely used approach is the jackknife method.
- Cross-validation is based on the "leave-one-out" principle.
- The most prevalent use of this method is to estimate $k - 1$ subsamples, eliminating one observation at a time from a sample of k cases.
- A discriminant function is calculated for each subsample and then the predicted group membership of the eliminated observation is made with the discriminant function estimated on the remaining cases.
- After all of the group membership predictions have been made, one at a time, a classification matrix is constructed and the hit ratio calculated.
- Cross-validation is quite sensitive to small sample sizes. Guidelines suggest that it be used only when the smallest group size is at least three times the number of predictor variables, and most researchers suggest a ratio of 5:1.
- However, cross-validation may represent the only possible validation approach in instances where the original sample is too small to divide into analysis and holdout samples but still exceeds the guidelines already discussed.
- Cross-validation is also becoming more widely used as major computer programs provide it as a program option.

THE DECISION PROCESS FOR DISCRIMINANT ANALYSIS

- **STAGE 6: VALIDATION OF THE RESULTS**
- **Profiling Group Differences**
- Another validation technique is to profile the groups on the independent variables to ensure their correspondence with the conceptual bases used in the original model formulation.
- After the researcher identifies the independent variables that make the greatest contribution in discriminating between the groups, the next step is to profile the characteristics of the groups based on the group means.
- This profile enables the researcher to understand the character of each group according to the predictor variables.
- For example, referring to the KitchenAid survey data presented in Table 1, we see that the mean rating on “durability” for the “would purchase” group is 7.4, whereas the comparable mean rating on “durability” for the “would not purchase” group is 3.2. Thus, a profile of these two groups shows that the “would purchase” group rates the perceived durability of the new product substantially higher than the “would not purchase” group.
- Another approach is to profile the groups on a separate set of variables that should mirror the observed group differences.
- This separate profile provides an assessment of external validity in that the groups vary on both the independent variable(s) and the set of associated variables.

CONTENTS

2m

16m

16m

Chapter 1

Operations Research (O.R.)

(Resource Management Techniques)

1.1 Introduction

Operations Research is the study of optimisation techniques. It is applied decision theory. The existence of optimisation techniques can be traced at least to the days of Newton and Lagrange. Rapid development and invention of new techniques occurred since the World War II essentially, because of the necessity to win the war with the limited resources available. Different teams had to do ***research*** on military ***operations*** in order to invent techniques to ***manage*** with available ***resources*** so as to obtain the desired objective. Hence the nomenclature ***Operations Research or Resource Management Techniques.***

Scope or Uses or Applications of O.R (Some O.R. Models)

O.R. is useful for solving

(1) Resource allocation problems

(2) Inventory control problems

(3) Maintenance and Replacement Problems

(4) Sequencing and Scheduling Problems

(5) Assignment of jobs to applicants to maximise total profit or minimize total cost.

(6) Transportation Problems

(7) Shortest route problems like travelling sales person problems

(8) Marketing Management problems

(9) Finance Management problems

(10) Production, Planning and control problems

(11) Design Problems

(12) Queuing problems, *etc.* to mention a few.

1.2 Role of operations research in Business and Management

1. Marketing Management Operations research techniques have definitely a role to play in

(a) Product selection

(b) Competitive strategies

(c) Advertising strategy *etc.*

2. Production Management

The O.R techniques are very useful in the following areas of production management.

(a) Production scheduling

(b) Project scheduling

(c) Allocation of resources

(d) Location of factories and their sizes

(e) Equipment replacement and Maintenance

(f) Inventory policy etc.

3. Finance Management

The techniques O.R are applied to Budgeting and Investment areas and especially to

(a) Cash flow analysis

(b) Capital requirement

(c) Credit policies

(d) Credit risks *etc.*

4. Personal Management

(a) Recruitment policies and

(b) Assignment of jobs are some of the areas of personnel management where O.R techniques are useful.

5. Purchasing and Procurement

(a) Rules for purchasing

(b) Determining the quantity

(c) Determining the time of purchaser are some of the areas where O.R. techniques can be applied

6. Distribution

In determining

(a) location of warehouses

(b) size of the warehouses

(c) rental outlets

(d) transportation strategies O.R. techniques are useful.

1.3 Role of O.R in Engineering

1. Optimal design of water resources systems
2. Optimal design of structures
3. Production, Planning, Scheduling and control
4. Optimal design of electrical networks
5. Inventory control
6. Planning of maintenance and replacement of equipment
7. Allocation of resources of services to maximise the benefit
8. Design of material handling
9. Optimal design of machines
10. Optimum design of control systems
11. Optimal selection of sites for an industry to mention a few.

1.4 Classification of Models

The first thing one has to do to use O.R. techniques after formulating a practical problem is to construct a suitable model to represent the practical problem. A model is a reasonably simplified representation of a real–world situation. It is an abstraction of reality. The models can broadly be classified as

Iconic (Physical) Models

Analogue Models

Mathematical Models

Also as

Static Models

Dynamic Models and, in addition, as

Deterministic Models

Stochastic Models

Models can further be subdivided as

Descriptive Models

Prescriptive Models

Predictive Models

Analytic Models

Simulation Models

Iconic Model :

This is a physical, or pictorial representation of various aspects of a system.

Example : Toy, Miniature model of a building, scaled up model of a cell in biology *etc.*

Analogue or Schematic model :

This uses one set of properties to represent another set of properties which a system under study has

Example : A network of water pipes to represent the flow of current in an electrical network or graphs, organisational charts etc.

Mathematical model or Symbolic model :

This uses a set of mathematical symbols (letters, numbers etc) to represent the decision variables of a system under consideration. These variables related by mathematical equations or inequalities which describes the properties of the system.

Example : A linear programming model, A system of equations representing an electrical network or differential equations representing dynamic systems etc.

Static Model :

This is a model which does not take time into account. It assumes that the values of the variables do not change with time during a certain period of time horizon.

Example : A linear programming problem, an assignment problem, transportation problem *etc.*

Dynamic model is a model which considers time as one of the important variables.

Example: A dynamic programming problem, A replacement problem

Deterministic model is a model which does not take uncertainty into account.

Example : A linear programming problem, an assignment problem *etc.*

Stochastic model is a model which considers uncertainty as an important aspect of the problem.

Example : Any stochastic programming problem, stochastic inventory models *etc.*

Descriptive model is one which just describes a situation or system.

Example : An opinion poll, any survey.

Predictive model is one which predicts something based on some data. Predicting election results before actually the counting is completed.

Prescriptive model is one which prescribes or suggests a course of action for a problem.

Example : Any programming (linear, nonlinear, dynamic, geometric *etc.*) problem.

Analytic model is a model in which exact solution is obtained by mathematical methods in closed form.

Simulation model is a representation of reality through the use of a model or device which will react in the same manner as reality under a given set of conditions. Once a simulation model is designed, it takes only a little time, in general, to run a simulation on a computer.

It is usually less mathematical and less time consuming and generally least expensive as well, in many situations.

Example : queuing problems, inventory problems.

1.5 Some characteristics of a good Model

(1) It should be reasonably simple.

(2) A good model should be capable of taking into account new changes in the situation affecting its frame significantly with ease *i.e.,* updating the models should be as simple and easy as possible,

(3) Assumptions made to simplify the model should be as small as possible.

(4) Number of variables used should be as small in number as possible.

(5) The model should be open to parametric treatment.

1.6 Principles of Modelling

(1) Do not build up a complicated model while a simple one will suffice.

(2) Beware of moulding the problems to fit a (favourite !) technique

(3) Deductions must be made carefully.

(4) Models should be validated prior to implementation.

(5) A model should neither be pressed to do nor criticised for failing to do that for which it was never intended.

(6) Beware of overselling the model in cases where assumption made for the construction of the model can be challenged.

(7) The solution of a model cannot be more accurate than the accuracy of the information that goes into the construction of the model.

(8) Models are only aids in decision making.

(9) Model should be as accurate as possible.

1.7 General Methods for solving O.R models

(1) ***Analytic Procedure :*** Solving models by classical Mathematical techniques like differential calculus, finite differences etc. to obtain analytic solutions.

(2) ***Iterative Procedure :*** Starts with a trial solution and a set of rules for improving it by repeating the procedure until further improvement is not possible.

(3) ***Monte–carlo technique :*** Taking sample observations, computing probability distributions for the variable using random numbers and constructing some functions to determine values of the decision variables.

1.8 Main Phases of O.R.

(1) ***Formulation of the Problems :*** Identifying the objective, the decision variables involved and the constraints that arise involving the decision variables

(2) ***Construction of a Mathematical Model:*** Expressing the measure of effectiveness which may be total profit, total cost, utility etc., to be optimised by a Mathematical function called ***objective function.*** Representing the constraints like budget constraints, raw materials constraints, ***resource constraints,*** quality constraints etc., by means of mathematical equations or inequalities.

(3) ***Solving the Model constructed:*** Determining the solution by analytic or iterative or Monto–carlo method depending upon the structure of the mathematical model.

(4) ***Controlling and updating :*** A solution which is optimum today may not be so tomorrow. The values of the variables may change, new variables may emerge. The structural relationship between the variables may also undergo a change. All these are determined in updating.

A solution from a model remains a solution only so long as the uncontrollable variables retain their values and the relationship between the variables does not change. The solution itself goes "out of control" if the values of one or more controlled variables vary or relationship between the variables undergoes a change. Therefore controls must be established to indicate the limits within which the model and its solution can be considered as reliable. This is called controlling.

(5) ***Testing the model*** and ***its solution i.e., validating the model :*** checking as far as possible either from the past available data or by expertise and experience whether the model gives a solution which can be used in practice.

(6) ***Implementation :*** Implement using the solution to achieve the desired goal.

1.9 Limitation

Mathematical models which are the essence of OR do not take into account qualitative or emotional or some human factors which are quite real and influence the decision making. All such influencing factors find no place in O.R. This is the main limitation of O.R. Hence O.R. is only an aid in decision making.

EXERCISE

1. **What is O.R ?** ***[MU. BE. Oct 96]***
2. **What is the scope of O.R ?**
3. **What are the applications of O.R. ?**
4. **List the uses of O.R.**
5. **Write a short note on the importance of operations research in production management** ***[MU. MBA April, 97]***
6. **Write a short note on the role of operations research in marketing management.** ***[MU. MBA Nov.96, Nov. 97, April 98]***
7. **Write a short note on the role of operations research in production planning.** ***[MU. MBA April, 96]***
8. **What are the different phases of O.R.?** ***[MKU. BE. Nov 97]***
9. **What are the characteristics of an O.R. problem ?** ***[MKU. BE. Nov 97]***
10. **What is a model ?**
11. **What is a Mathematical Model ?**
12. **What are the different types of Models ?**

13. **What are the characteristics of a good model ?**

14. **What are the limitations of a mathematical model ?**

15. **What are the limitations of an O.R. Model ?**

[MKU. BE. Nov 97]

16. **Explain the general methods of solving O.R. Models.**

17. **Explain the principles of modelling.**

18. **State the different types of Models used in O.R.**

[MU. BE. Oct 97]

19. **What are the various phases in the study of operations research ?** *[BRU. BE. Apr 97]*

20. **Answer the following questions with examples wherever necessary.**

    (a) **Necessities of OR in industry,**

    (b) **Fields of application of OR in industry**

    (c) **Deterministic models,**

    (d) **Mention atleast eight mathematical models.**

[MKU. BE. Nov 97]

21. **What is an iconic model in the study of operations research ?**

[MU. BE. Oct. '96]

Chapter 2

Linear Programming Formulation and Graphical Method

(Formulation and Graphical Solution)

2.1 Introduction

Linear programming problems deal with determining optimal allocations of limited resources to meet given objectives. The resources may be in the form of men, raw materials, market demand, money and machines etc. The objective is usually maximizing profit, minimizing total cost, maximizing utility etc. There are certain restrictions on the total amount of each resource available and on the quantity or quality of each product made.

Linear programming problem deals with the optimization (Maximization or Minimization) of a function of decision variables (The variables whose values determine the solution of a problem are called ***decision variables*** of the problem) known as ***objective function,*** subject to a set of simultaneous linear equations (or inequalities) known as ***constraints***. The term ***linear*** means that all the variables occurring in the objective function and the constraints are of the first degree in the problems under consideration and the term ***programming*** means the process of determining a particular course of action.

Linear programming techniques are used in many industrial and economic problems. They are applied in product mix, blending, diet, transportation and assignment problems. Oil refineries, airlines, railways, textile industries, chemical industries, steel industries, food processing industries and defence establishments are also the users of this technique.

2.2 Requirements for employing LPP Technique :

[BRU. BE. Nov 96]

1. There must be a well defined objective function.
2. There must be alternative courses of action to choose.

3. At least some of the resources must be in limited supply, which give rise to constraints.
4. Both the objective function and constraints must be linear equations or inequalities.

2.3 Mathematical Formulation of L.P.P

If x_j ($j = 1, 2, \dots, n$) are the n decision variables of the problem and if the system is subject to m constraints, the general Mathematical model can be written in the form :

$$\text{Optimize } Z = f(x_1, x_2, \dots x_n)$$

$$\text{subject to } g_i(x_1, x_2, \dots x_n) \leq, =, \geq b_i, (i = 1,2, \dots, m)$$

(called structural constraints)

$$\text{and } x_1, x_2, \dots x_n \geq 0,$$

(called the non–negativity restrictions or constraints)

Procedure for forming a LPP Model :

Step 1 : Identify the unknown decision variables to be determined and assign symbols to them.

Step 2 : Identify all the restrictions or constraints (or influencing factors) in the problem and express them as linear equations or inequalities of decision variables.

Step 3 : Identify the objective or aim and represent it also as a linear function of decision variables.

Step 4 : Express the complete formulation of LPP as a general mathematical model.

We consider only those situations where this will help the reader to put proper inequalities in the formulation.

1. Usage of manpower, time, raw materials etc are always less than or equal to the availability of manpower, time, raw materials etc.

2. Production is always greater than or equal to the requirement so as to meet the demand.

Example 1 : A firm manufactures two types of products A and B and sells them at a profit of Rs. 2 on type A and Rs. 3 on type B. Each product is processed on two machines M_1 and M_2. Type A requires 1 minute of processing time on M_1 and 2 minutes on M_2. Type B requires 1 minute on M_1 and 1 minute on M_2. Machine M_1 is available for not more than 6 hours 40 minutes while machine M_2 is

available for 10 hours during any working day. Formulate the problem as a LPP so as to maximize the profit.

Solution : Let the firm decide to produce x_1 units of product A and x_2 units of product B to maximize its profit.

To produce these units of type A and type B products, it requires

$x_1 + x_2$ processing minutes on M_1

$2x_1 + x_2$ processing minutes on M_2

Since machine M_1 is available for not more than 6 hours and 40 minutes and machine B is available for 10 hours doing any working day, the constraints are

$$x_1 + x_2 \leq 400$$
$$2x_1 + x_2 \leq 600$$
$$\text{and } x_1, x_2 \geq 0.$$

Since the profit from type A is Rs. 2 and profit from type B is Rs. 3, the total profit is $2x_1 + 3x_2$. As the objective is to maximize the profit, the objective function is maximize $Z = 2x_1 + 3x_2$.

$\therefore$ The complete formulation of the LPP is

$$\text{Maximize } Z = 2x_1 + 3x_2$$

subject to the constraints

$$x_1 + x_2 \leq 400$$
$$2x_1 + x_2 \leq 600$$
$$\text{and } x_1, x_2 \geq 0.$$

Example 2 : (Production Allocation Problem)

A firm produces three products. These products are processed on three different machines. The time required to manufacture one unit of each of the three products and the daily capacity of the three machines are given in the table below :

Machine	Time per unit (minutes)			Machine capacity (Minutes/day)
	Product 1	Product 2	Product3	
M_1	2	3	2	440
M_2	4	–	3	470
M_3	2	5	–	430

It is required to determine the number of units to be manufactured for each product daily. The profit per unit for product 1,2 and 3 is Rs.4, Rs.3 and Rs.6 respectively. It is assumed that all the amounts produced are consumed in the market. Formulate the mathematical model for the problem. *[MU. B. Tech. Leather. Oct 96]*

Solution : Let x_1, x_2 and x_3 be the number units of products 1, 2 and 3 produced respectively.

To produce these amount of products 1, 2 and 3, it requires :

$$\begin{aligned} 2x_1 + 3x_2 + 2x_3 &\quad \text{minutes on } M_1 \\ 4x_1 + 3x_3 &\quad \text{minutes on } M_2 \\ 2x_1 + 5x_2 &\quad \text{minutes on } M_3. \end{aligned}$$

But the capacity of the machines M_1, M_2 and M_3 are 440,470 and 430 (minutes/day).

$\therefore$ The constraints are

$$\begin{aligned} 2x_1 + 3x_2 + 2x_3 &\leq 440 \\ 4x_1 + 3x_3 &\leq 470 \\ 2x_1 + 5x_2 &\leq 430 \\ \text{and } x_1, x_2, x_3 &\geq 0 \end{aligned}$$

Since the profit per unit for product 1, 2, and 3 is Rs.4, Rs. 3 and Rs.6 respectively, the total profit is $4x_1 + 3x_2 + 6x_3$. As the objective is to maximize the profit, the objective function is maximize $Z = 4x_1 + 3x_2 + 6x_3$.

$\therefore$ The complete formulation of the LPP is

$$\text{Maximize } Z = 4x_1 + 3x_2 + 6x_3$$

subject to the constraints

$$\begin{aligned} 2x_1 + 3x_2 + 2x_3 &\leq 440 \\ 4x_1 + 3x_3 &\leq 470 \\ 2x_1 + 5x_2 &\leq 430 \\ \text{and } x_1, x_2, x_3 &\geq 0. \end{aligned}$$

Example 3 : **(Blending Problem)**

A firm produces an alloy having the following specifications :

(i) **Specific gravity ≤ 0.98**

(ii) **Chromimum $\geq 8\%$**

(iii) **Melting point $\geq 450°C$**

Raw materials A, B and C having the properties shown in the table can be used to make the alloy.

Property	Raw material		
	A	B	C
Specific gravity	0.92	0.97	1.04
Chromium	7%	13%	16%
Melting point	440°C	490°C	480°C

Cost of the various raw materials per unit ton are : Rs. 90 for A, Rs. 280 for B and Rs. 40 for C. Find the proportions in which A, B and C be used to obtain an alloy of desired properties while the cost of raw materials is minimum.

Solution : Let x_1, x_2 and x_3 be the tons of raw materials A,B and C to be used for making the alloy.

From these raw materials, the firm requires :

$$\begin{aligned} 0.92\,x_1 + 0.97x_2 + 1.04\,x_3 &\quad \text{specific gravity.} \\ 7x_1 + 13x_2 + 16x_3 &\quad \text{chromium} \\ 440\,x_1 + 490\,x_2 + 480\,x_3 &\quad \text{melting point.} \end{aligned}$$

$\therefore$ By the given specifications, the constraints are

$$\begin{aligned} 0.92\,x_1 + 0.97\,x_2 + 1.04\,x_3 &\leq 0.98 \\ 7\,x_1 + 13\,x_2 + 16\,x_3 &\geq 8 \\ 440\,x_1 + 490\,x_2 + 480\,x_3 &\geq 450 \\ \text{and } x_1, x_2, x_3 &\geq 0. \end{aligned}$$

Since the cost of the various raw materials per unit ton are : Rs. 90 for A, Rs. 280 for B and Rs. 40 for C, the total cost is

$90x_1 + 280x_2 + 40x_3$. As the objective is to minimize the total cost, the objective function is

$$\text{Minimize } Z = 90\,x_1 + 280\,x_2 + 40\,x_3$$

$\therefore$ The complete formulation of the LPP is

$$\text{Minimize } Z = 90\,x_1 + 280\,x_2 + 40\,x_3$$

subject to

$$0.92\,x_1 + 0.97\,x_2 + 1.04\,x_3 \leq 0.98$$
$$7\,x_1 + 13\,x_2 + 16\,x_3 \geq 8$$
$$440\,x_1 + 490\,x_2 + 480\,x_3 \geq 450$$
$$\text{and } x_1, x_2, x_3 \geq 0.$$

Example 4 : (Diet Problem)

A person wants to decide the constituents of a diet which will fulfil his daily requirements of proteins, fats and carbohydrates at the minmum cost. The choice is to be made from four different types of foods. The yields per unit of these foods are given in the following table:

Food type	Yield/unit			cost/unit (Rs)
	Proteins	Fats	carbohydrates	
1	3	2	6	45
2	4	2	4	40
3	8	7	7	85
4	6	5	4	65
Minimum requirement	800	200	700	

Formulate the L.P model for the problem.

Solution : Let x_1, x_2, x_3 and x_4 be the units of food of type 1, 2, 3 and 4 used respectively.

From these units of food of type 1, 2, 3 and 4 he requires

$$3x_1 + 4x_2 + 8x_3 + 6x_4 \quad \text{Proteins/day}$$
$$2x_1 + 2x_2 + 7x_3 + 5x_4 \quad \text{Fats / day}$$
$$6x_1 + 4x_2 + 7x_3 + 4x_4 \quad \text{Carbohydrates/day}$$

Since the minimum requirement of these proteins, fats and carbohydrates are 800, 200 and 700 respectively, the constraints are

$$3x_1 + 4x_2 + 8x_3 + 6x_4 \geq 800$$
$$2x_1 + 2x_2 + 7x_3 + 5x_4 \geq 200$$
$$6x_1 + 4x_2 + 7x_3 + 4x_4 \geq 700$$
$$\text{and } x_1, x_2, x_3, x_4 \geq 0.$$

Since, the costs of these food of type 1, 2, 3 and 4 are Rs. 45, Rs.40, Rs.85 and Rs. 65 per unit, the total cost is Rs.$45x_1 + 40x_2 + 85x_3 + 65x_4$. As the objective is to minimize the total cost, the objective function is

$$\text{Minimize } Z = 45x_1 + 40x_2 + 85x_3 + 65x_4$$

$\therefore$ The complete formulation of the L.P.P is

$$\text{Minimize } Z = 45x_1 + 40x_2 + 85x_3 + 65\,x_4$$

subject to

$$3x_1 + 4x_2 + 8x_3 + 6x_4 \geq 800$$
$$2x_1 + 2x_2 + 7x_3 + 5x_4 \geq 200$$
$$6x_1 + 4x_2 + 7x_3 + 4x_4 \geq 700$$
$$\text{and } x_1, x_2, x_3, x_4 \geq 0.$$

Example 5 : A farmer has 100 acre farm. He can sell all tomatoes, lettuce, or radishes he can raise. The price he can obtain is Rs.1.00 per kg for tomatoes, Rs. 0.75 a head for lettuce and Rs. 2.00 per kg for radishes. The average yield per - acre is 2,000 kgs of tomatoes, 3000 heads of lettuce, and 1000 kgs of radishes. Fertilizer is available at Rs.0.50 per kg and the amount required per acre is 100 kgs each for tomatoes and lettuce, and 50 kgs for radishes. Labour required for sowing, cultivating and harvesting per acre is 5 man-days for tomatoes and radishes, and 6 man-days for lettuce. A total of 400 man-days of labour are available at Rs. 20.00 per man-day

Formulate this problem as a L.P. model to maximise the farmer's total profit.

Solution : Let the farmer decide to allot x_1, x_2 and x_3 acre of his farm to grow tomatoes, lettuce, and radishes respectively to maximize his total profit.

Since the total area of the form is restricted to 100 acre and the total man-days labour is restricted to 400 man-days, the constraints are

$$x_1 + x_2 + x_3 \leq 100$$
$$5x_1 + 6x_2 + 5x_3 \leq 400$$
$$\text{and } x_1, x_2 \geq 0$$

The Farmer will produce 2000 x_1 kgs of tomatoes, 3000 x_2 heads of lettuce, and 1000 x_3 kgs of radishes.

$\therefore$ The total sale of farmer will be

$$= \text{Rs } [1 \infty 2000\,x_1 + 0.75 \infty 3000\,x_2 + 2 \infty 1000\,x_3]$$

Fertilizer expenditure will be

$$= \text{Rs } 0.50\,[100\,x_1 + 100\,x_2 + 50\,x_3]$$

Labour expenditure will be $=$ Rs 20 $[5x_1 + 6x_2 + 5x_3]$

$\therefore$ Farmer's net profit will be

$$= \text{Rs [Sale – total expenditure]}$$
$$= \text{Rs } [2000\,x_1 + 2250\,x_2 + 2000\,x_3 - 50\,x_1 - 50\,x_2 - 25x_3 - 100\,x_1 - 120\,x_2 - 100\,x_3]$$
$$= \text{Rs } [1850\,x_1 + 2080\,x_2 + 1875\,x_3]$$

$\therefore$ The objective function is

$$\text{Maximize } Z = 1850\,x_1 + 2080\,x_2 + 1875\,x_3$$

$\therefore$ The complete formulation of the L.P.P is

$$\text{Maximize } Z = 1850\,x_1 + 2080\,x_2 + 1875\,x_3$$

subject to the constraints

$$x_1 + x_2 + x_3 \le 100$$
$$5x_1 + 6x_2 + 5x_3 \le 400$$
$$\text{and } x_1, x_2, x_3 \ge 0.$$

Example 6 : **Old hens can be bought at Rs. 2 each and young ones at Rs.5 each. The old hens lay 3 eggs per week and the young ones lay 5 eggs per week, each egg being worth 30 paise. A hen costs Rs. 1 per week to feed. A person has only Rs.80 to spend for hens. How many of each kind should he buy to give a profit of more than Rs.6 per week, assuming that he cannot house more than 20 hens. Formulate this as a L.P.P.** *[MU. BE. 89]*

Solution : The person decides to buy x_1 old hens and x_2 young hens to maximize his profit.

Since he has only Rs. 80 to spend for hens and old hen costs Rs. 2 and young hen costs Rs. 5 each,

$$2x_1 + 5x_2 \le 80$$

Also, since he can not house more than 20 hens,

$$x_1 + x_2 \le 20$$

The total sale of eggs will be

$$= \text{Rs. } 0.3\,(3x_1 + 5x_2)$$

Expenditure on feeding will be

$$= \text{Rs.}1\,(x_1 + x_2)$$

$\therefore$ The net profit is $= \text{Rs. } [0.3(3x_1 + 5x_2) - 1(x_1 + x_2)]$

$$= \text{Rs. } (0.5x_2 - 0.1x_1)$$

$\therefore$ $0.5x_2 - 0.1\,x_1 \ge 6.$

$\therefore$ The constraints are

$$2x_1 + 5x_2 \le 80$$
$$x_1 + x_2 \le 20$$
$$0.5x_2 - 0.1\,x_1 \ge 6.$$
$$\text{and } x_1, x_2 \ge 0.$$

$\therefore$ The complete formulation of the L.P.P is

$$\text{Maximze } Z = 0.5x_2 - 0.1x_1$$

subject to the constraints

$$2x_1 + 5x_2 \le 80$$
$$x_1 + x_2 \le 20$$
$$0.5x_2 - 0.1\,x_1 \ge 6.$$
$$\text{and } x_1, x_2 \ge 0.$$

Example 7 : **A production planner in a soft drink plant has two bottling machines A and B. A is designed for 8–ounce bottles and B for 16-ounce bottles. However, each can be used on both types with some loss of efficiency. The following data is available :**

Machine	8–ounce bottles	16–ounce bottles
A	100/minute	40/minute
B	60/minute	75/minute

The machine can be run 8–hours per day, 5 days per week. Profit on 8–ounce bottle is 15 paise and on 16–ounce bottle is 25 paise. Weekly production of the drink cannot exceed 3,00,000 ounces and the market can absorb 25,000 eight–ounce bottles and 7000 sixteen–ounce bottles per week. The planner wishes to maximize his profit. Formulate this as a L.P.P. *[BRU. BE. Nov 96]*

Solution : Let the production planner decide to produce x_1 number of units of 8–ounce bottles and x_2 number of units of 16–ounce bottles to maximize his profit.

To produce these x_1 and x_2 units of 8–ounce and 16–ounce bottles, he requires

$$\frac{x_1}{100} + \frac{x_2}{40} \quad \text{minutes on machine A}$$
$$\frac{x_1}{60} + \frac{x_2}{75} \quad \text{minutes on machine B}$$

Since the machines A and B are available for 8–hours per day, 5–days per week, we have

$$\frac{x_1}{100} + \frac{x_2}{40} \le 2400 \Rightarrow 2x_1 + 5x_2 \le 4{,}80{,}000$$

$$\frac{x_1}{60} + \frac{x_2}{75} \le 2400 \Rightarrow 5x_1 + 4x_2 \le 7{,}20{,}000$$

Since the weekly production of the drink should not exceed 3,00,000 ounces and the market can absorb only upto 25,000 eight-ounce bottles and 7,000 sixteen–ounce bottles per week, we have

$$8x_1 + 16x_2 \le 3{,}00{,}000$$

$$x_1 \le 25{,}000 \text{ and } x_2 \le 7000$$

∴ The constraints are :

$$\begin{aligned} 2x_1 + 5x_2 &\le 4{,}80{,}000 \\ 5x_1 + 4x_2 &\le 7{,}20{,}000 \\ 8x_1 + 16x_2 &\le 3{,}00{,}000 \\ x_1 &\le 25000 \\ x_2 &\le 7000 \\ \text{and } x_1, x_2 &\ge 0. \end{aligned}$$

Since the profit on 8–ounce bottle is Rs. 0.15 and on 16-ounce bottle is Rs.0.25, the total profit is Rs.$0.15x_1 + 0.25x_2$. As the objective is to maximize the profit, the objective function is

$$\text{Maximize } Z = 0.15x_1 + 0.25x_2$$

∴ The complete formulation of the L.P.P is

$$\text{Maximize } Z = 0.15x_1 + 0.25\,x_2$$

subject to the constraints

$$\begin{aligned} 2x_1 + 5x_2 &\le 4{,}80{,}000 \\ 5x_1 + 4x_2 &\le 7{,}20{,}000 \\ 8x_1 + 16x_2 &\le 3{,}00{,}000 \\ x_1 &\le 25000 \\ x_2 &\le 7000 \\ \text{and } x_1, x_2 &\ge 0. \end{aligned}$$

EXERCISE

1. A company makes two types of leather products A and B. Product A is of high quality and product B is of lower quality. The respective profits are Rs.4 and Rs.3 per product. Each product A requires twice as much time as product B and if all products were of type B, the company could make 1000 per day. The supply of leather is sufficient for only 800 products per day (Both A and B combined). Product A requires a special spare part and only 400 per day are available. There are only 700 special spare parts a day available for product B. Formulate this as a L.P.P.

2. A firm engaged in producing two models A and B performs three operations–painting, Assembly and testing. The relevant data are as follows:

Model	Unit sale Price	Hours required for each unit		
		Assembly	Painting	Testing
A	Rs.50	1.0	0.2	0.0
B	Rs.80	1.5	0.2	0.1

Total number of hours available are : Assembly 600, painting 100, testing 30. Determine weekly production schedule to maximize the profit.

3. A farmer has 1,000 acres of land on which he can grow corn, wheat or soyabeans. Each acre of corn costs Rs.100 for preparation, requires 7 man-days of work and yields a profit Rs. 30. An acre of wheat costs Rs.120 to prepare, requires 10 man-days of work and yields a profit Rs.40. An acre of soyabeans costs Rs.70 to prepare, requires 8 man-days of work and yields a profit Rs.20. The farmer has Rs.1,00,000 for preparation and 8000 man-days of work. Formulate this as a L.P.P

4. A television company operates two assembly sections, section A and section B. Each section is used to assemble the components of three types of televisions : Colour, Standard and Economy. The expected daily production on each section is as follows :

T.V. Model	Section A	Section B
Colour	3	1
Standard	1	1
Economy	2	6

The daily running costs for two sections average Rs. 6000 for section A and Rs.4000 for section B. It is given that the company must produce atleast 24 colours, 16 standard and 40 Economy TV sets for which an order is pending. Formulate this as a L.P.P so as to minimize the total cost.

5. A transistor Radio company manufactures four models A, B, C and D which have profit contributions of Rs.8, Rs.15 and Rs.25 on models A, B and C respectively and a loss of Rs.1 on model D. Each type of radio requires a certain amount of time for the manufacturing of components for assembling and for packing. Specially a dozen units of model A require 1 hour of manufacturing, 2 hours for assembling and 1 hour for packing. The corresponding figures for a dozen units of model B are 2, 1 and 2 and for a dozen units of model C are 3, 5 and 1, while a dozen units of model D require 1 hour of packing only. During the forthcoming week, the company will be able to make available 15 hours of manufacturing, 20 hours of assembling and 10 hours of packing time. Formulate this as a L.P.P.

6. A soft drinks firm has two bottling plants, one located at P and the other at Q. Each plant produces three different soft drinks A, B and C. The capacities of the two plants in number of bottles per day, are as follows :

Products	Plants	
	P	Q
A	3000	1000
B	1000	1000
C	2000	6000

A market survey indicates that during the month of April, there will be a demand for 24000 bottles of A, 16000 bottles of B and 48000 bottles of C. The operating costs per day of running plants P and Q respectively are Rs.6000 and Rs. 4000. How many days should the firm run each plant in April so that the production cost is minimized?

7. Consider two different types of food stuffs A and B. Assume that these food stuffs contain vitamins V_1, V_2 and V_3 respectively. Minimum daily requirements of these vitamins are 1 mg. of V_1, 50 mg. of V_2 and 10 mg. of V_3. Suppose that the food stuff A contains 1 mg. of V_1, 100 mg. of V_2 and 10 mg. of V_3, whereas foodstuff B contains 1 mg. of V_1, 10 mg. of V_2 and 100 mg. of V_3. Cost of one unit of food stuff A is Rs.1 and that of B is Rs.1.5. Find the minimum cost diet that would supply the body atleast the minimum requirements of each vitamin.

8. A company produces refrigerators in Unit I and heaters in Unit II. The two products are produced and sold on a weekly basis. The weekly production cannot exceed 25 in Unit I and 36 in Unit II, due to constraints 60 workers are employed. A refrigerator requires 2 man – week of labour, while a heater requires 1 man–week of labour. The profit available is Rs. 600 per refrigerator and Rs.400 per heater. Formulate the LPP problem. *[BNU. BE. Nov 96]*

ANSWERS

1. Maximize $Z = 4x_1 + 3x_2$

subject to

$$\begin{aligned} 2x_1 + x_2 &\le 1000 \\ x_1 + x_2 &\le 800 \\ x_1 &\le 400 \\ x_2 &\le 700 \\ \text{and } x_1, x_2 &\ge 0. \end{aligned}$$

2. Maximize $Z = 50x_1 + 80x_2$

subject to

$$\begin{aligned} x_1 + 1.5x_2 &\le 600 \\ 0.2x_1 + 0.2x_2 &\le 100 \\ 0.1x_2 &\le 30 \\ \text{and } x_1, x_2 &\ge 0. \end{aligned}$$

3. Maximize $Z = 30x_1 + 40x_2 + 20x_3$

subject to

$$\begin{aligned} 10x_1 + 12x_2 + 7x_3 &\le 10000 \\ 7x_1 + 10x_2 + 8x_3 &\le 8000 \\ x_1 + x_2 + x_3 &\le 1000 \\ \text{and } x_1, x_2, x_3 &\ge 0. \end{aligned}$$

4. Minimize $Z = 6000x_1 + 4000x_2$

subject to

$$3x_1 + x_2 \geq 24$$
$$x_1 + x_2 \geq 16$$
$$2x_1 + 6x_2 \geq 40$$
$$\text{and } x_1, x_2 \geq 0.$$

5. Maximize $Z = 8x_1 + 15x_2 + 25x_3 - x_4$

subject to

$$x_1 + 2x_2 + 3x_3 \leq 15$$
$$2x_1 + x_2 + 5x_3 \leq 20$$
$$x_1 + 2x_2 + x_3 + x_4 \leq 10$$
$$\text{and } x_1, x_2, x_3, x_4 \geq 0.$$

6. Minimize $Z = 6000x_1 + 4000x_2$

subject to

$$3x_1 + x_2 \geq 24$$
$$x_1 + x_2 \geq 16$$
$$x_1 + 3x_2 \geq 24$$
$$\text{and } x_1, x_2 \geq 0.$$

7. Minimize $Z = x_1 + 1.5x_2$

subject to

$$x_1 + x_2 \geq 1$$
$$100x_1 + 10x_2 \geq 50$$
$$10x_1 + 100x_2 \geq 10$$
$$\text{and } x_1, x_2 \geq 0.$$

8. Maximize $Z = 600x_1 + 400x_2$

subject to

$$2x_1 + x_2 \leq 60$$
$$x_1 \leq 25$$
$$x_2 \leq 36$$
$$\text{and } x_1, x_2 \geq 0.$$

2.4 Basic Assumptions

[MU. MCA. Nov 98]

The linear programming problems are formulated on the basis of the following assumptions.

1. Proportionality : The contribution of each variable in the objective function or its usage of the resources is directly proportional to the value of the variable. *i.e.*, if resource availability increases by some percentage, then the output shall also increase by the same percentage.

2. Additivity : Sum of the resources used by different activities must be equal to the total quantity of resources used by each activity for all the resources individually or collectively.

3. Divisibility : The variables are not restricted to integer values.

4. Certainty or Deterministic : Co–efficients in the objective function and constraints are completely known and do not change during the period under study in all the problems considered.

5. Finiteness : Variables and constraints are finite in number.

6. Optimality : In a linear programming problem we determine the decision variables so as to extremise (optimize) the objective function of the LPP.

7. The problem involves only one objective namely profit maximization or cost minimization.

2.5 Graphical method of the solution of a L.P.P

Linear Programming problems involving only two variables can be effectively solved by a graphical method which provides a pictorial representation of the problems and its solutions and which gives the basic concepts used in solving general L.P.P. which may involve any finite number of variables. This method is simple to understand and easy to use. The redundant constraints are automatically eliminated from the system. Multiple solutions, unbounded solutions and infeasible solutions get highlighted very clearly in graphical analysis. Sensitivity analysis can be illustrated easily by drawing the graph of the changes.

Graphical method is not a powerful tool of linear programming as most of the practical situations do involve more than two variables. But the method is really useful to explain the basic concepts of L.P.P to the persons who are not familiar with this. Though graphical method can deal with any number of constraints but since each constraint is shown as a line on a graph, a large number of lines make the graph difficult to read.

Working procedure for Graphical method :

Given a L.P.P, optimize $Z = f(x_i)$ subject to the constraints $g_j(x_i) \leq, =, \geq b_j$ and the non-negativity restrictions $x_i \geq 0$, $i = 1, 2$; $j = 1,2,3\ldots\ldots m$

Step 1 : Consider the inequality constraints as equalities. Draw the straight lines in the XOY plane corresponding to each equality and non–negativity restrictions.

Step 2 : Find the permissible region (feasible region or solution space or convex region) for the values of the variables which is the region bounded by the lines drawn in step 1.

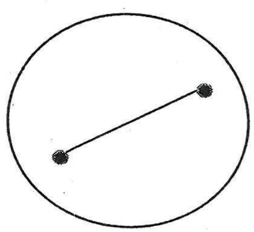
Step 3 : Find the points of intersection of the bounded lines (vertices of the permissible region) by solving the equations of the corresponding lines.

Step 4 : Find the values of Z at all vertices of the permissible region.

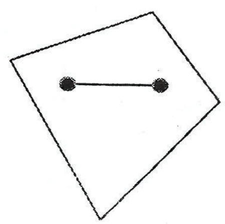
Step 5 : *(i)* For maximization problem, choose the vertex for which Z is maximum. *(ii)* For minimization problem, choose vertex for which Z is minimum.

Note : A ***region*** or a ***set*** of points is said to be ***convex*** if the line joining any two of its points lies completely within the region (or the set).

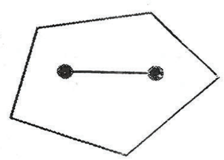
Example :



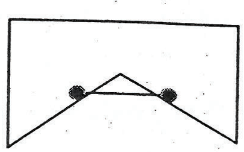
Convex



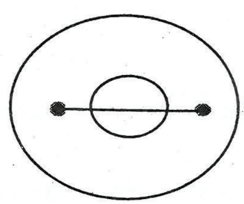
Convex



Convex



Not convex



Not convex

Example 1 : Solve the following L.P.P by the graphical method

$$\text{Max } Z = 3x_1 + 2x_2$$

subject to

$$-2x_1 + x_2 \leq 1$$
$$x_1 \leq 2$$
$$x_1 + x_2 \leq 3$$
$$\text{and } x_1, x_2 \geq 0$$

[MU. BE. Nov 94]

Solution : First consider the inequality constraints as equalities.

$$-2x_1 + x_2 = 1 \quad (1)$$
$$x_1 = 2 \quad (2)$$
$$x_1 + x_2 = 3 \quad (3)$$
$$\text{and } x_1 = 0 \quad (4)$$
$$x_2 = 0 \quad (5)$$

For the line, $-2x_1 + x_2 = 1$,

put $x_1 = 0 \Rightarrow x_2 = 1 \Rightarrow (0,1)$

put $x_2 = 0 \Rightarrow -2x_1 = 1 \Rightarrow x_1 = -0.5 \Rightarrow (-0.5, 0)$

So, the line (1) passes through the points (0,1) and (–0.5,0). The points on this line satisfy the equation $-2x_1 + x_2 = 1$ Now origin (0,0), on substitution, gives $-0 + 0 = 0 < 1$; hence it also satisfies the inequality $-2x_1 + x_2 \leq 1$. Thus all points on the origin side and on this line satisfy the inequality $-2x_1 + x_2 \leq 1$. Similarly interpreting the other constraints we get the feasible region OABCD. The feasible region is also known as solution space of the L.P.P. Every point within this area satisfies all the constraints.

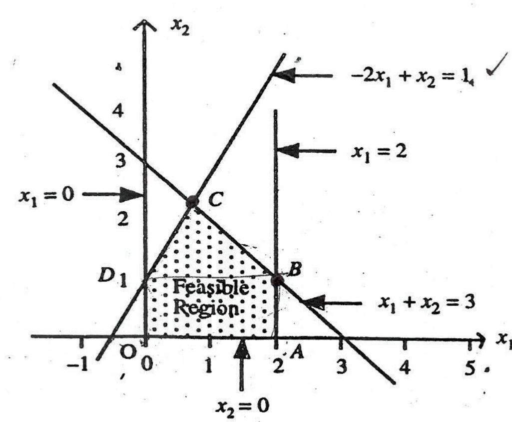


Fig. 2.1

Now our aim is to find the vertices of the solution space. B is the point of intersection of $x_1 = 2$ and $x_2 + x_2 = 3$. Solving these two equations, we have $x_1 = 2$, $x_2 = 1$. ∴ We have the vertex B(2,1). Similarly, C is the intersectin of $-2x_1 + x_2 = 1$ and $x_1 + x_2 = 3$. Solving these we have C $\left(\frac{2}{3}, \frac{7}{3}\right)$.

∴ The vertices of the solution space are O (0,0), A (2,0), B (2,1), C $\left(\frac{2}{3}, \frac{7}{3}\right)$ and D (0,1)

The values of Z at these vertices are given by ($\because Z = 3x_1 + 2x_2$)

Vertex	Value of Z
O (0, 0)	0
A(2,0)	6
B(2,1)	8
C $\left(\frac{2}{3}, \frac{7}{3}\right)$	$\frac{20}{3}$
D(0,1)	2

Since the problem is of maximization type, the optimum solution to the L.P.P is

$$\text{Maximum } Z = 8, x_1 = 2, x_2 = 1.$$

Example 2 : Solve the following L.P.P by the graphical method

$$\text{Minimize } Z = 3x_1 + 5x_2$$

subject to

$$-3x_1 + 4x_2 \le 12$$
$$x_1 \le 4$$
$$2x_1 - x_2 \ge -2$$
$$x_2 \ge 2$$
$$2x_1 + 3x_2 \ge 12 \text{ and } x_1, x_2 \ge 0.$$

Solution : Let us consider $-3x_1 + 4x_2 \le 12$ as the equation $-3x_1 + 4x_2 = 12$, then it gives a line in the $x_1 0 x_2$ plane. All the points on the origin side and on this line satisfy the inequality $-3x_1 + 4x_2 \le 12$. If we consider the constraint $2x_1 + 3x_2 \ge 12$ as the equation $2x_1 + 3x_2 = 12$, then it gives a line in the $x_1 0 x_2$ plane. All the points on the other side of the origin and on this line satisfy the inequality $2x_1 + 3x_2 \ge 12$. Similarly interpreting the other constraints, we get the feasible region ABCDE. Every point with in this area satisfies all the

constraints. The vertices of the solution space are A (3, 2), B (4, 2), C (4, 6) D $\left(\frac{4}{5}, \frac{18}{5}\right)$, and E $\left(\frac{3}{4}, \frac{7}{2}\right)$.

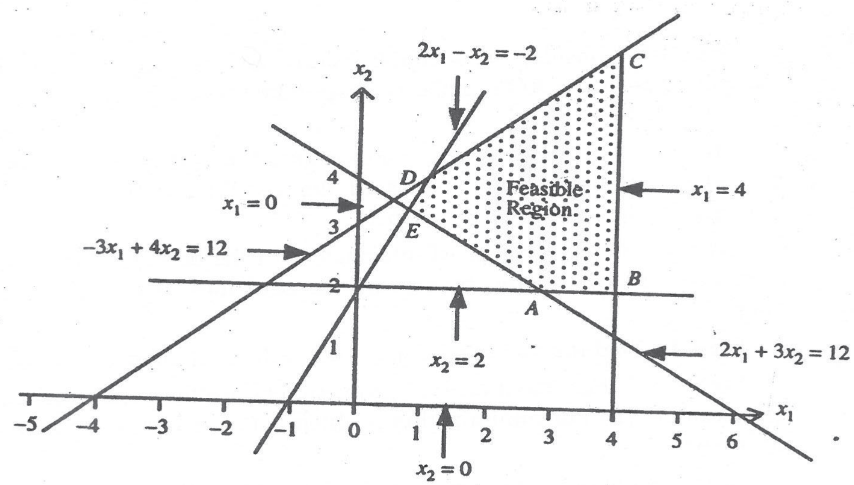


Fig 2.2

The values of Z at these vertices are given by ($\because Z = 3x_1 + 5x_2$)

Vertex	Value of Z
A (3, 2)	19
B (4, 2)	22
C (4, 6)	42
D $\left(\frac{4}{5}, \frac{18}{5}\right)$	$\frac{102}{5}$
E $\left(\frac{3}{4}, \frac{7}{2}\right)$	$\frac{79}{4}$

Since the problem is of minimization type, the optimum solution is

$$\text{Minimum } Z = 19, x_1 = 3, x_2 = 2.$$

Example 3 : A pineapple firm produces two products canned pineapple and canned juice. The specific amounts of material, labour and equipment required to produce each product and the availability of each of these resources are shown in the table given below :

	Canned Juice	Canned Pineapple	Available resources
Labour (Man hours)	3	2.0	12.0
Equipment (M/c hours)	1	2.3	6.9
Material (Unit)	1	1.4	4.9

Assuming one unit of canned juice and canned Pineapple has profit margins Rs.2 and Rs.1 respectively. Formulate this as a L.P.P. and solve it graphically also. *[MU. BE. Nov 91]*

Solution : Let x_1 be the number of units of canned juice and x_2 be the number of units of canned pineapple to be produced.

The constraints or restrictions in this problem are the labour, equipment and material.

$$\begin{aligned} \text{For labour, } 3x_1 + 2x_2 &\le 12 \\ \text{For Equipment, } x_1 + 2.3x_2 &\le 6.9 \\ \text{For material, } x_1 + 1.4x_2 &\le 4.9 \\ \text{and } x_1, x_2 &\ge 0. \end{aligned}$$

Here our objective is to maximize the profit.

$\therefore$ The objective function is Maximize $Z = 2x_1 + x_2$

$\therefore$ The complete formulation of the L.P.P.is Maximize $Z = 2x_1 + x_2$ subject to the constraints

$$\begin{aligned} 3x_1 + 2x_2 &\le 12 \\ x_1 + 2.3x_2 &\le 6.9 \\ x_1 + 1.4x_2 &\le 4.9 \text{ and } x_1, x_2 \ge 0. \end{aligned}$$

The solution space is given below with the shaded area with vertices O (0, 0), A (0, 3), B (1.8, 2.2), C (3.2, 1.2) and D (4, 0)

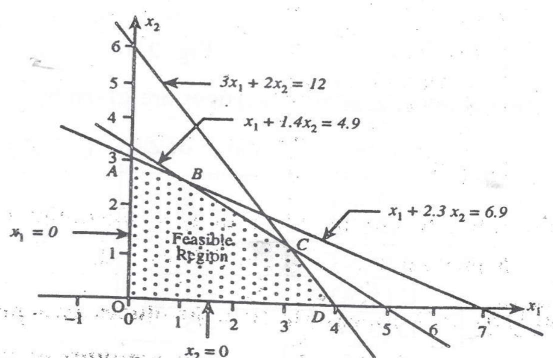


Fig. 2.3

The values of Z at these vertices are given by

Vertex	Value of Z
O (0, 0)	0
A (0, 3)	3
B (1.8, 2.2)	5.8
C (3.2, 1.2)	7.6
D (4,0)	8

$(\because Z = 2x_1 + x_2)$

Since the problem is of maximization type, the optimum solution is

$$\text{Maximum } Z = 8, x_1 = 4, x_2 = 0.$$

Example 4 : A Company manufactures 2 types of printed circuits. The requirements of transistors, resistors and capacitors for each type of printed circuits along with other data are given below :

	Circuit		Stock available
	A	B	
Transistor	15	10	180
Resistor	10	20	200
Capacitor	15	20	210
Profit	Rs.5	Rs.8	

How many circuits of each type should the company produce from the stock to earn maximum profit. *[MU. BE. Oct 95]*

Solution : Let x_1 be the number of type A circuits and x_2 be the number of type B circuits to be produced.

To produce these units of type A and type B circuits, the company requires

$$\begin{aligned} \text{Transistors} &= 15x_1 + 10x_2 \\ \text{Resistor} &= 10x_1 + 20x_2 \\ \text{Capacitors} &= 15x_1 + 20x_2 \end{aligned}$$

Since the availability of these transistors, resistors and capacitors are 180, 200 and 210 respectively, the constraints are

$$\begin{aligned} 15x_1 + 10x_2 &\le 180 \\ 10x_1 + 20x_2 &\le 200 \\ 15x_1 + 20x_2 &\le 210 \\ \text{and } x_1 \ge 0, \; x_2 &\ge 0. \end{aligned}$$

Since the profit from type A is Rs.5 and from type B is Rs.8 per units, the total profit is $5x_1 + 8x_2$

$\therefore$ The complete formulation of the L.P.P. is

$$\text{Maximize } Z = 5x_1 + 8x_2$$

subject to

$$15x_1 + 10x_2 \le 180$$
$$10x_1 + 20x_2 \le 200$$
$$15x_1 + 20x_2 \le 210$$
$$\text{and } x_1, x_2 \ge 0.$$

By using graphical method, the solution space is given below with shaded area OABCD with vertices O (0, 0), A (12, 0), B (10, 3), C (2, 9) and D (0, 10)

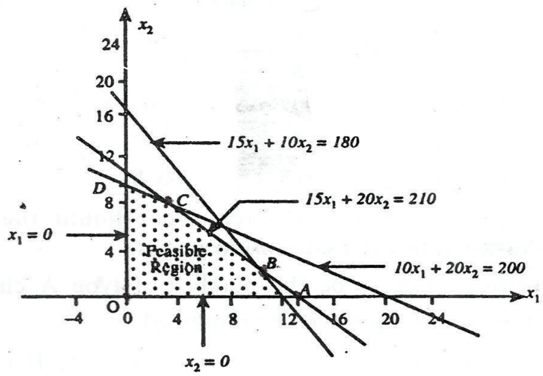


Fig. 2.4

The values of Z of these vertices are given by

Vertex	Value of Z
O (0, 0)	0
A (12, 0)	60
B (10, 3)	74
C (2, 9)	82
D (0,10)	80

($\because Z = 5x_1 + 8x_2$)

Since the problem is of maximization type, the optimum solution is

Maximum Z = 82, x_1 = 2, x_2 = 9.

Example 5 : **Apply graphical method to solve the LPP: Maximise $Z = x_1 - 2x_2$ subject to $-x_1 + x_2 \le 1$, $6x_1 + 4x_2 \ge 24$, $0 \le x_1 \le 5$ and $2 \le x_2 \le 4$.** ***[MU.MBA.Nov.96]***

Solution : By using graphical method, the solution space is given below with shaded area ABCDE with vertices A $\left(\frac{8}{3}, 2\right)$, B (5, 2), C (5, 4), D (3, 4) and E (2, 3)

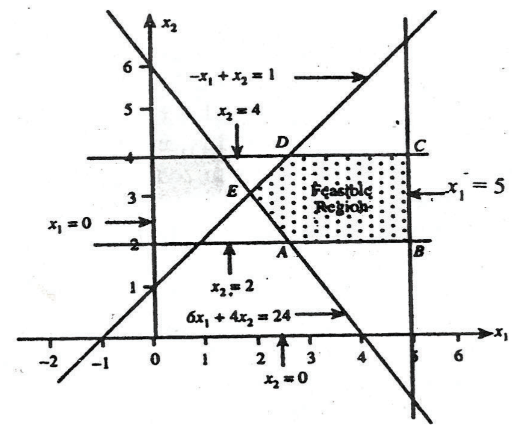


Fig. 2.5

The values of Z at these vertices are given by

Vertex	Value of Z
A $\left(\frac{8}{3}, 2\right)$	$-\frac{4}{3}$
B (5, 2)	1
C (5, 4)	– 3
D (3, 4)	– 5
E (2, 3)	– 4

($\because Z = x_1 - 2x_2$)

Since the problem is of maximization type, the optimal solution

Maximum Z = 1, $x_1 = 5$, $x_2 = 2$.

Example 6 : **A Company manufactures two types of cloth, using three different colours of wool. One yard length of type A cloth require 4 OZ of red wool, 5 OZ of green wool and 3 OZ of yellow wool. One yard length of type B cloth requires 5 OZ of red wool, 2 OZ of green wool and 8 OZ of yellow wool. The wool available for manufacturer is 1000 OZ of red wool, 1000 OZ of green wool and 1200 OZ of yellow wool. The manufacturer can make a profit of Rs. 5 on one yard of type A cloth and Rs.3 on one yard of type B cloth. Find the best combination of the quantities of type A and type B cloth which gives him maximum profit by solving the L.P.P. graphically.**

[MU. BE. Apr 92]

Solution : Let the manufacturer decide to produce x_1 yards of type A cloth and x_2 yards of type B cloth.

To produce these yards of type A and type B cloth, he requires

$$\begin{aligned} \text{red wool} &= 4x_1 + 5x_2 \text{ OZ} \\ \text{green wool} &= 5x_1 + 2x_2 \text{ OZ} \\ \text{yellow wool} &= 3x_1 + 8x_2 \text{ OZ} \end{aligned}$$

Since the availability of these red wool, green wool and yellow wool are 1000 OZ, 1000 OZ and 1200 OZ respectively, the constraints are

$$\begin{aligned} 4x_1 + 5x_2 &\le 1000 \\ 5x_1 + 2x_2 &\le 1000 \\ 3x_1 + 8x_2 &\le 1200 \\ \text{and } x_1 \ge 0, \; x_2 &\ge 0. \end{aligned}$$

Since the profit from one yard of type A cloth is Rs.5 and the profit from one yard of type B cloth is Rs. 3, the total profit is $5x_1 + 3x_2$

$\therefore$ The complete formulation of the L.P.P.is

$$\begin{aligned} \text{Maximize } Z &= 5x_1 + 3x_2 \\ \text{subject to} \quad 4x_1 + 5x_2 &\le 1000 \\ 5x_1 + 2x_2 &\le 1000 \\ 3x_1 + 8x_2 &\le 1200 \\ \text{and } x_1, x_2 &\ge 0. \end{aligned}$$

By using graphical method, the feasible region is given below with shaded area OABCD with vertices O (0, 0), A (200, 0), B $\left(\frac{3000}{17}, \frac{1000}{17}\right)$, C $\left(\frac{2000}{17}, \frac{1800}{17}\right)$ and D (0, 150)

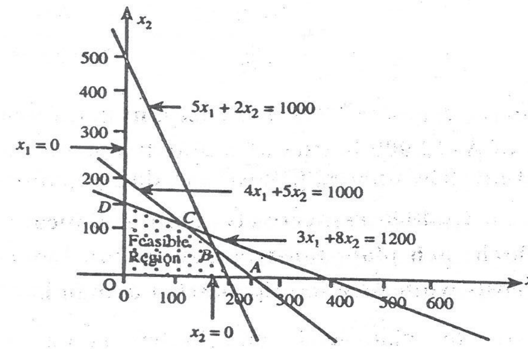


Fig 2.6

The values of Z at these vertices are given by $(\because \; Z = 5x_1 + 3x_2)$

Vertex	Value of Z
O (0, 0)	0
A (200, 0)	1000
B $\left(\frac{3000}{17}, \frac{1000}{17}\right)$	1058.8
C $\left(\frac{2000}{17}, \frac{1800}{17}\right)$	905.8
D (0, 150)	450

Since the problem is of maximization type, the optimum solution is

$$\text{Maximum } Z = 1058.8, \; x_1 = \frac{3000}{17}, \; x_2 = \frac{1000}{17}.$$

Note : In a given L.P.P, if any constraint does not affect the feasible region (or solution space), then the constraint is said to be a ***redundant constraint.***

Example 7 : **A Company making cold drinks has two bottling plants located at towns T_1 and T_2. Each plant produces three drinks A, B and C and their production capacity per day is given below :**

Cold drinks	Plant at	
	T_1	T_2
A	6000	2000
B	1000	2500
C	3000	3000

The marketing department of the company forecasts a demand of 80,000 bottles of A, 22,000 bottles of B and 40,000 bottles of C during the month of June. The operating costs per day of plants at T_1 and T_2 are Rs. 6000 and Rs.4000 respectively. Find graphically, the number of days for which each plant must be run in June so as to minimize the operating costs while meeting the market demand.

Solution : Let the plant at T_1 and T_2 be run for x_1 and x_2 days respectively.

Since the plants at T_1 and T_2 run for x_1 and x_2 days, they will produce,

$6000\,x_1 + 2000\,x_2$ bottles of A

$1000\,x_1 + 2500\,x_2$ bottles of B

$3000\,x_1 + 3000\,x_2$ bottles of C

Since the demand for the cold drinks A, B and C are 80,000, 22,000 and 40,000 respectively and the production is always greater than or equal to the demand, the constraints are

$$6000\,x_1 + 2000\,x_2 \geq 80000 \Rightarrow 6x_1 + 2x_2 \geq 80$$

$$\Rightarrow 3x_1 + x_2 \geq 40$$

$$1000\,x_1 + 2500\,x_2 \geq 22000 \Rightarrow x_1 + 2.5x_2 \geq 22$$

$$3000\,x_1 + 3000\,x_2 \geq 40000 \Rightarrow 3x_1 + 3x_2 \geq 40$$

$$\text{and } x_1, x_2 \geq 0.$$

Since the operating costs per day at T_1 is Rs. 6000 and at T_2 is Rs. 4000 and T_1, T_2 run for x_1 and x_2 days, the total operating cost is Rs. $6000x_1 + 4000x_2$.

Here our objective is to minimize the total operating cost. Therefore the objective function is minimize $Z = 6000\,x_1 + 4000\,x_2$.

$\therefore$ The complete formulation of the L.P.P.is

$$\text{Minimize } Z = 6000x_1 + 4000x_2$$

subject to

$$3x_1 + x_2 \geq 40$$
$$x_1 + 2.5x_2 \geq 22$$
$$3x_1 + 3x_2 \geq 40$$
$$\text{and } x_1, x_2 \geq 0.$$

By using graphical method, the feasible region is given below with shaded area with vertices A (22, 0), B (12, 4), C (0, 40)

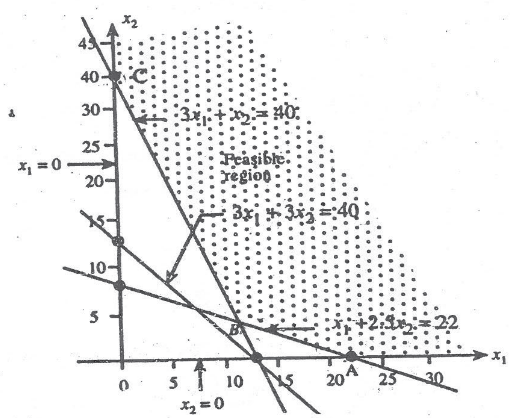


Fig 2.7

From the figure, we see that the constraint $3x_1 + 3x_2 \geq 40$ does not affect the solution space. So $3x_1 + 3x_2 \geq 40$ is a ***redundant constraint.*** Also from the direction of the arrows, we see that the solution space is unbounded above.

The values of Z at these vertices A (22,0), B(12,4) and C(0,40) are given by

Vertex	Value of Z
A (22, 0)	1,32,000
B (12, 4)	88,000
C (0, 40)	1,60,000

$(\because\ Z = 6000x_1 + 4000x_2)$

Since the problem is of minimization type, the optimum solution is

$$\text{Minimum } Z = \text{Rs. } 88{,}000,\ x_1 = 12 \text{ days},\ x_2 = 4 \text{ days.}$$

Note : From the above examples, for problems involving two variables and having a finite solution, we observed that the optimal solution existed at a vertex of the feasible region. That is, *"if there exists an optimal solution of an L.P.P, it will be at one of the vertices of the feasible region"*.

2.6 Some more cases

The constraints generally, give region of feasible solution which may be bounded or unbounded. We discussed seven linear programming problems and the optimal solution for either of them was unique. However, it may not be true for every problem. In general, a linear programming problem may have :

(i) a unique optimal solution

(ii) an infinite number of optimal solutions

(iii) an unbounded solution

(iv) no solution

We now give a few examples to illustrate the cases.

Example 8 : **A firm manufactures two products A and B on which the profits earned per unit are Rs.3 and Rs.4 respectively. Each product is processed on two machines M_1 and M_2. Product A requires one minute of processing time on M_1 and two minutes on M_2 while B requires one minute on M_1 and one minute on M_2. Machine M_1 is available for not more than 7 hours 30 minutes while machine M_2 is available for 10 hours during any working day. Find the number of units of products A and B to be manufactured to get maximum profit. Formulate the above as a L.P.P. and solve by graphical method.** *[MU. BE. Apr 92, MSU. BE. Nov 96]*

Solution **:** Let the firm decide to manufacture x_1 units of product A and x_2 units of product B.

To produce these units of products A and B, it requires

$x_1 + x_2$ hours of processing times on M_1

$2x_1 + x_2$ hours of processing times on M_2

But the availability of these two machines M_1 and M_2 are 450 minutes and 600 minutes respectively, the constraints are

$$x_1 + x_2 \le 450$$
$$2x_1 + x_2 \le 600$$
$$\text{and } x_1, x_2 \ge 0$$

Since the profit from product A is Rs.3 per unit and from product B is Rs. 4 per unit, the total profit is Rs. $3x_1 + 4x_2$ and our objective is to maximize the profit

$\therefore$ The complete formulation of the L.P.P is

$$\text{Maximize } Z = 3x_1 + 4x_2$$

subject to

$$x_1 + x_2 \le 450 \quad \dots (i)$$
$$2x_1 + x_2 \le 600 \quad \dots (ii)$$
$$\text{and } x_1, x_2 \ge 0 \quad \dots (iii)$$

By graphical method, the solution space satisfying the constraints *(i)*, *(ii)* and meeting the non-negativity restriction *(iii)* is shown shaded in the following figure.

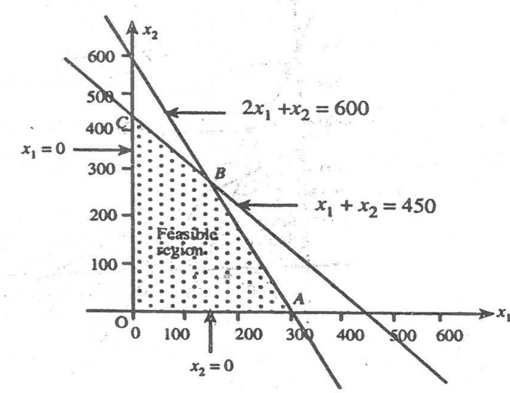


Fig. 2.8

The solution space is the region OABC. The vertices of this solution space are O (0, 0), A (300, 0), B (150, 300) and C (0, 450)

The values of Z at these vertices are given by

Vertex	Value of Z
O (0, 0)	0
A (300, 0)	900
B (150, 300)	1650
C (0, 450)	1800

($\because$ $Z = 3x_1 + 4x_2$)

Since the problem is of maximization type and the maximum value of Z is attained at a single vertex, this problem has a ***unique optimal solution.***

$\therefore$ The optimal solution is

$$\text{Maximum } Z = 1800, x_1 = 0, x_2 = 450.$$

Example 9 : Solve the following L.P.P. graphically.

$$\text{Maximize } Z = 100x_1 + 40x_2$$

subject to

$$5x_1 + 2x_2 \leq 1000$$
$$3x_1 + 2x_2 \leq 900$$
$$x_1 + 2x_2 \leq 500$$
$$\text{and } x_1, x_2 \geq 0.$$

Solution : By using graphical method, the solution space OABC shown shaded in the following figure.

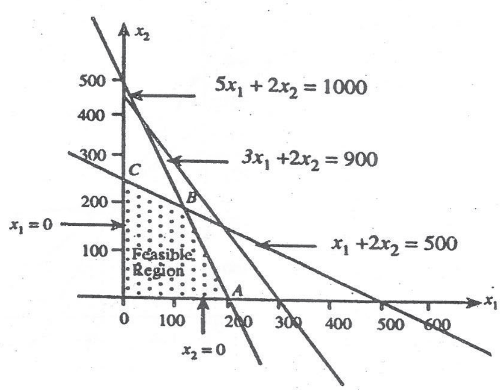


Fig. 2.9

The vertices of this convex region are O (0, 0), A (200, 0), B (125, 187.5) and C (0, 250)

The values of Z at these vertices are given by

Vertex	Value of Z
O (0, 0)	0
A (200, 0)	20,000
B (125, 187.5)	20,000
C (0, 250)	10,000

($\because$ $Z = 100x_1 + 40x_2$)

Here the maximum value of Z occurs at two vertices A and B.

Any point on the line joining A and B will also give the same maximum value of Z.

Since, there are infinite number of points between any ponits, there are infinite number of points on the line joining A and B gives the same maximum value of Z.

Thus, there are ***infinite number of optimal solutions*** for this L.P.P.

Note : An L.P.P having more than one optimal solution is said to have ***alternative*** or ***multiple optimal solutions.*** That is, the resources can be combined in more than one way to maximize the profit.

Example 10 : Using graphical method, solve the following L.P.P.

$$\textbf{Maximize } Z = 2x_1 + 3x_2$$

subject to

$$x_1 - x_2 \leq 2$$
$$x_1 + x_2 \geq 4 \text{ and } x_1, x_2 \geq 0.$$

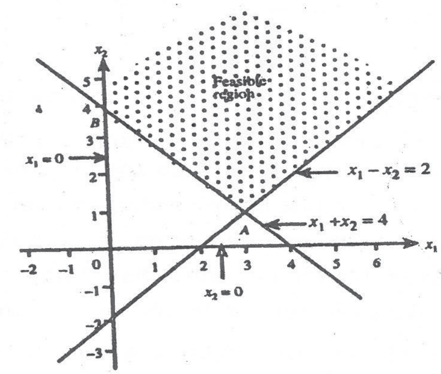


Fig. 2.10

Solution : By using graphical method, the solution space is shaded in the following figure.

Here the solution space is unbounded. The vertices of the feasible region (in the finite plane) are A (3,1) and B (0,4)

Valule of the objective function $Z = 2x_1 + 3x_2$ at these vertices are $Z(A) = 9$ and $Z(B) = 12$.

But there are points in this convex region for which Z will have much higher values. In fact, the maximum value of Z occurs at infinity. Hence this problem has an ***unbounded solution.***

Example 11 : Solve graphically the following L.P.P.:

$$\text{Maximize } Z = 4x_1 + 3x_2$$

subject to

$$x_1 - x_2 \le -1 \quad \ldots (i)$$

$$-x_1 + x_2 \le 0 \quad \ldots (ii)$$

$$\text{and } x_1, x_2 \ge 0 \quad \ldots (iii)$$

Solution : Any point satisfying the non-negativity restrictions *(iii)* lies in the first quadrant only. The two solution spaces, one satisfying *(i)*, and other satisfying *(ii)* are shown shaded in the following figure.

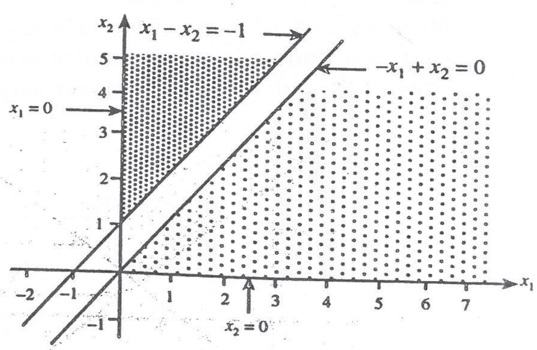


Fig. 2.11

There being no point (x_1, x_2) common to both the shaded regions. That is, we can not find a convex region for this problem. So the problem cannot be solved. Hence the problem have ***no feasible solution.***

2.7 Advantage of Linear Programming :

1. It provides an insight and perspective in to the problem environment. This generally results in clear picture of the true problem.
2. It makes a scientific and mathematical analysis of the problem situations.
3. It gives an opportunity to the decision maker to formulate his strategies consistent with the constraints and the objectives.
4. It deals with changing situations. Once a plan is arrived through the linear programming it can also be revaluated for changing conditions.
5. By using linear programming the decision maker makes sure that he is considering the best solution.

2.8 Limitations of Linear Programming : *[MU. MCA. Nov 98]*

1. The major limitation of linear programming is that it treats all relationships as linear. But it is not true in many real life situations.
2. The decision variables in some LPP would be meangingful only if they have integer values. But some times we get fractional values to the optimal solution, where only integer values are meaningful.
3. All the parameters in the linear programming model are assumed to be known constants. But in real life they may not be known completely or they may be probabilistic and they may be liable for changes from time to time.
4. The problems are complex if the number of variables and constraints are quite large.
5. Linear Programming deals with only a single objective problems, whereas in real life situations, there may be more than one objective.

EXERCISE

1. **Explain the essential characteristics and limitations of Linear Programming Problem.** ***[MU. MCA. Nov 98]***
2. **What is feasibility region in an LP problem ? Is it necessary that it should always be a convex set ?** ***[MU. MCA. Nov. 97]***
3. **What is a redundant constraint ? What does it imply ? Does it affect the optimal solution to an LPP ?** ***[MU. MCA. Nov. 97]***
4. **Explain the advantages of linear programming problems.**
5. **State the limitations of the graphical method of solving a LPP.** ***[BRU. BE. Apr. 97, Nov. 97]***

6. Solve the following by graphical method :

Maximize $x_1 - 3x_2$

subject to
$$\begin{aligned} x_1 + x_2 &\le 300 \\ x_1 - 2x_2 &\le 200 \\ 2x_1 + x_2 &\ge 100 \\ x_2 &\le 200 \\ \text{and } x_1, x_2 &\ge 0. \end{aligned}$$
[MU. BE. Apr 94]

7. By graphical method solve the following problem :

Maximize $Z = 3x_1 + 4x_2$

subject to
$$\begin{aligned} 5x_1 + 4x_2 &\le 200 \\ 3x_1 + 5x_2 &\le 150 \\ 5x_1 + 4x_2 &\ge 100 \\ 8x_1 + 4x_2 &\ge 80 \\ \text{and } x_1, x_2 &\ge 0. \end{aligned}$$
[MU. BE. Nov 93]

8. Solve the following problem by graphical method :

Maximize $5x + 8y$

subject to the constraints
$$\begin{aligned} 3x + 2y &\le 36 \\ x + 2y &\le 20 \\ 3x + 4y &\le 42 \\ \text{and } x &\ge 0 \\ y &\ge 0 \end{aligned}$$
[MU. BE. Apr 93]

9. Maximize $2x_1 + x_2$

subject to
$$\begin{aligned} 3x_1 + 2x_2 &\le 12.0 \\ x_1 + 2.3x_2 &\le 6.9 \\ x_1 + 1.4x_2 &\le 4.9 \\ \text{and } x_1, x_2 &\ge 0 \end{aligned}$$
Using graphical method. *[MU. BE. Apr 91]*

10. Using graphical method to solve L.P.P :

Minimize $Z = 3x_1 + 2x_2$

subject to
$$\begin{aligned} 5x_1 + x_2 &\ge 10 \\ x_1 + x_2 &\ge 6 \\ x_1 + 4x_2 &\ge 12 \\ \text{and } x_1, x_2 &\ge 0 \end{aligned}$$
[MU. BE. Nov 93]

11. Solve the following problem graphically.

Maximize $Z = 2x_1 + x_2$

subject to
$$\begin{aligned} 3x_1 + 2x_2 &\le 12.0 \\ x_1 + 2x_2 &\le 7 \\ x_1 + x_2 &\le 5 \\ \text{and } x_1, x_2 &\ge 0 \end{aligned}$$
[MU. BE. Apr 90, Apr 91]

12. Use graphical method to Maximize $Z = 6x_1 + 4x_2$

subject to
$$\begin{aligned} -2x_1 + x_2 &\le 2 \\ x_1 - x_2 &\le 2 \\ 3x_1 + 2x_2 &\le 9 \\ x_1 \ge 0,\ x_2 &\ge 0 \end{aligned}$$
[MU. BE. Apr 90]

13. Minimize $Z = x - 3y$

subject to the constraints
$$\begin{aligned} x + y &\le 300 \\ x - 2y &\le 200 \\ 2x + y &\ge 100 \\ y &\ge 200 \\ \text{and } x, y &\ge 0 \end{aligned}$$
by graphical method. *[MU. BE. Apr 93]*

14. Egg contains 6 units of vitamin A per gram and 7 units of vitamin B per gram and cost 12 paise per gram. Milk contains 8 units of vitamin A per gram and 12 units of vitamin B per gram, and costs 20 paise per gram. The daily minimum requirement of vitamin A and vitamin B are 100 units and 120 units respectively. Find the optimal product mix. *[MU. BE. Nov 93]*

15. Solve graphically the L.P.P

Maximize $Z = 3x_1 + 4x_2$

subject to the constraints
$$\begin{aligned} 2x_1 + 5x_2 &\le 120 \\ 4x_1 + 2x_2 &\le 80 \\ \text{and } x_1, x_2 &\ge 0 \end{aligned}$$
[MU. BE. Nov 92]

16. A company produces 2 types of hats. Each hat A require twice as much labour time as the second hat B. If all are of hat B only, the company can produce a total of 500 hats a day. The market limits daily sales of the hat A and hat B to 150 and 250 hats. The profits on hat A and B are Rs. 8 and Rs.5 respectively. Solve graphically to get the optimal solution . *[MU. BE. Apr 95]*

4

17. Solve graphically the following L.P.P :

$$\text{Maximize } Z = 5x_1 + 3x_2$$

subject to constraints

$$\begin{aligned} 3x_1 + 5x_2 &\leq 15 \\ 5x_1 + 2x_2 &\leq 10 \\ \text{and } x_1, x_2 &\geq 0 \end{aligned}$$

[MU. BE. Apr 93]

18. Solve graphically the following L.P.P.

$$\text{Minimize } Z = 20x_1 + 10x_2$$

subject to

$$\begin{aligned} x_1 + 2x_2 &\leq 40 \\ 3x_1 + x_2 &\geq 30 \\ 4x_1 + 3x_2 &\geq 60 \\ \text{and } x_1, x_2 &\geq 0 \end{aligned}$$

19. Solve graphically the following L.P.P.

$$\text{Max } Z = 3x + 2y$$

subject to

$$\begin{aligned} -2x + 3y &\leq 9 \\ x - 5y &\geq -20 \\ \text{and } x, y &\geq 0 \end{aligned}$$

20. Solve graphically the following L.P.P :

$$\text{Minimize } Z = -6x_1 - 4x_2$$

subject to

$$\begin{aligned} 2x_1 + 3x_2 &\geq 30 \\ 3x_1 + 2x_2 &\leq 24 \\ x_1 + x_2 &\geq 3 \\ \text{and } x_1, x_2 &\geq 0 \end{aligned}$$

21. Solve graphically the following L.P.P.

$$\text{Maximize } Z = 3x_1 - 2x_2$$

subject to

$$\begin{aligned} x_1 + x_2 &\leq 1 \\ 2x_1 + 2x_2 &\geq 4 \\ \text{and } x_1, x_2 &\geq 0 \end{aligned}$$

22. Solve graphically the following L.P.P.

$$\text{Maximize } Z = x_1 + x_2$$

subject to

$$\begin{aligned} x_1 - x_2 &\geq 0 \\ -3x_1 + x_2 &\geq 3 \\ \text{and } x_1, x_2 &\geq 0 \end{aligned}$$

23. A manufacturer of furniture makes two products, chairs and tables. Processing of these products is done on two machines A and B. A chair requires 2 hours on machine A and 6 hours on machines B. A table requires 5 hours on machine A and no time on machine B. There are 16 hours of time per day available on machine A and 30 hours on machine B. Profit gained by the manufacturer from a chair and a table is Rs. 2 and Rs. 10 respectively. What should be the daily production of each of the products.

24. A person requires 10, 12 and 12 units of chemicals A, B and C respectively for his garden. A liquid product contains 5, 2 and 1 units of A, B and C respectively per Jar. A dry product contains 1,2 and 4 units of A, B and C per carton. How many of each should he purchase in order to minimize the cost and meet the requirement.

25. Solve the following problem graphically,

$$\text{Maximize } Z = 40x_1 + 100x_2$$

subject to

$$\begin{aligned} 12x_1 + 6x_2 &\leq 3000 \\ 4x_1 + 10x_2 &\leq 2000 \\ 2x_1 + 3x_2 &\leq 900 \\ \text{and } x_1, x_2 &\geq 0 \end{aligned}$$

[MU. B. Tech. Leather Oct 96]

26. A garment manufacturing company can make two products, Prima and Seconda. Each of the products requires time on a cutting machine and a finishing Machine. Relevant data are

	Product	
	Prima	seconda
Cutting hours (per unit)	2	1
Finishing hours (per unit)	3	3
Unit cost Rs.	128	120
Selling price Rs.	134	129
Maximum sales (units per week)	200	200

The number of cutting hours available per week is 390 and the number of finishing hours available per week is 810. How much should each product be produced in order to maximize the profit ?

[BRU. BE. Apr 94]

27. A firm manufactures refrigerators and air coolers. Production takes place in two separate departments. Refrigerators are produced in department I and air coolers are produced in department II. The firm's two products are produced and sold on a weekly basis. The weekly pro-duction cannot exceed 25 refrigerators in department I and 35 air coolers in department II, because of the limited available facilities in these two departments. The firm regularly employs a total

of 60 workers in two departments. A refrigerator requires 2 man–weeks of labour, while an air coolers requires 1 man–week of labour. The firm receives a profit margin of Rs.300 and Rs.200 per refrigerator and cooler respectively. Determine the product mix in order to maximize profit. *[BRU. BE. Nov 94]*

28. The production and planning department of a soft drink plant faces the following problem. The bottling plant has two bottling machines A and B. A is designed for 80cc bottle and B for 160 cc bottles. However each can be used on both types with loss of efficiency. The following data is available.

Machine	80 cc bottle	160 cc bottle
A	100 per min	40 per min
B	60 per min	75 per min

The machine can run 8 hour per day 5 days per week. Profit on 80 cc bottle is 15 paise and 160 cc bottle is 25 paise. Weekly production on the drink cannot exceed 3 million cc and the market can absorb 250000 of 80 cc bottles and 70000 of 160 cc bottles per week. The department wishes to maximize the profit subject to production and marketing restrictions. Solve the problem by graphical method or by simplex method. *[BRU. BE. Nov 96]*

29. A company manufactures two products A and B. Each unit of B takes twice as long to produce as one unit of A and if the company were to produce only A it would have time to produce 2000 units per day. The availability of the raw material is sufficient to produce 1500 units per day of both A and B combined. Product B requiring a special ingredient only 600 units can be made per day. If A fetches a profit of Rs. 2 per unit and B a profit of Rs. 4 per unit find the optimum product mix by graphical method. *[BRU. BE. Apr 97]*

30. A company produces two different products A and B. The company makes a profit of Rs. 40 and Rs. 30 per unit on A and B respectively. The production process has a capacity of 30,000 man hours. It takes 3 hours to produce one unit of A and one hour to produce one unit of B. The market survey indicates that the maximum number of units A that can be sold is 8,000 and those of B is 12,000 units. Formulate the problem and solve it by graphical method to get maximum profit. *[MU. BE. Apr 97]*

31. Apply graphical method to find non–negative values of x_1 and x_2 which minimise $z = 10x_1 + 25x_2$ subject to $x_1 + x_1 \geq 50$, $x_1 \geq 20$, and $x_2 \leq 40$. *[M.U. M.B.A. Apr '97]*

ANSWERS

6. Max Z $= 200, x_1 = 200,\ x_2 = 0.$

7. Max Z $= \frac{1800}{13}, x_1 = \frac{400}{13}, x_2 = \frac{150}{13}$

8. Max Z $= 82, x = 2,\ y = 9.$

9. Max Z $= 8, x_1 = 4,\ x_2 = 0.$

10. Min Z $= 13, x_1 = 1,\ x_2 = 5.$ 11. Max Z $= 8, x_1 = 4,\ x_2 = 0.$

12. An infinite number of solutions with Max Z = 18,

    (i) $x_1 = \frac{13}{5}$, $x_2 = \frac{3}{5}$ *(ii)* $x_1 = \frac{5}{7}$, $x_2 = \frac{24}{7}$ etc.

13. Min Z $= -600, x = 0,\ y = 200.$

14. Min Z $= 12x_1 + 20x_2$

$$\text{subject to} \quad \begin{aligned} 6x_1 + 8x_2 &\geq 100 \\ 7x_1 + 12x_2 &\geq 120 \\ \text{and } x_1, x_2 &\geq 0. \end{aligned}$$

    Also Min Z $= 205,\ x_1 = 15,\ x_2 = 1.25$

15. Max Z $= 110,\ x_1 = 10, x_2 = 20$

16. Max Z $= 8x_1 + 5x_2$

$$\text{subject to} \quad \begin{aligned} 2x_1 + x_2 &\leq 500 \\ x_1 &\leq 150 \\ x_2 &\leq 250 \\ \text{and } x_1, x_2 &\geq 0. \end{aligned}$$

    Also Max Z $= 2250, x_1 = 125,\ x_2 = 250$

17. Max Z $= \frac{235}{19}, x_1 = \frac{20}{19}, x_2 = \frac{45}{19}$

18. Min Z $= 240,\ x_1 = 6, x_2 = 12$

19. Unbounded solution

20. An infinite number of optimal solutions with Min Z = –48

    (i) $x_1 = 8,\ x_2 = 0$

    (ii) $x_1 = \frac{12}{5}$, $x_2 = \frac{42}{5}$, *etc ......*

21. No feasible solution

22. No feasible solution

23. Max $Z = 2x_1 + 10x_2$

subject to
$$2x_1 + 5x_2 \le 16$$
$$6x_1 \le 30$$
$$\text{and } x_1, x_2 \ge 0.$$

Also Max $Z = 32, x_1 = 0, x_2 = 3.2.$

24. Min $Z = 3x_1 + 2x_2$

subject to
$$5x_1 + x_2 \ge 10$$
$$2x_1 + 2x_2 \ge 12$$
$$x_1 + 4x_2 \ge 12$$
$$\text{and } x_1, x_2 \ge 0.$$

Also Min $Z = 13, x_1 = 1, x_2 = 5.$

25. An infinite number of optimal solutions with Max $Z = 20{,}000$

(i) $x_1 = 0, x_2 = 200$

(ii) $x_1 = \frac{375}{2}, x_2 = 125,$ *etc.*

26. Max $Z = 6x_1 + 9x_2$

subject to
$$2x_1 + x_2 \le 390$$
$$3x_1 + 3x_2 \le 810$$
$$x_1 \le 200$$
$$x_2 \le 200 \text{ and } x_1, x_2 \ge 0.$$

Also Max $Z = 2220, x_1 = 70, x_2 = 200.$

27. Max $Z = 300x_1 + 200x_2$

subject to
$$2x_1 + x_2 \le 60$$
$$x_1 \le 25$$
$$x_2 \le 35 \quad \text{and } x_1, x_2 \ge 0.$$

Also Max $Z = \text{Rs.}10750, x_1 = \frac{25}{2}, x_2 = 35.$

28. Max $Z = 0.15x_1 + 0.25x_2$

subject to
$$2x_1 + 5x_2 \le 4{,}80{,}000$$
$$5x_1 + 4x_2 \le 7{,}20{,}000$$

$$80x_1 + 160x_2 \le 3{,}00{,}000$$
$$x_1 \le 2{,}50{,}000$$
$$x_2 \le 70{,}000 \quad \text{and } x_1, x_2 \ge 0.$$

Also Max $Z = 562.50, x_1 = 3750, x_2 = 0.$

29. Max $Z = 2x_1 + 4x_2$

subject to
$$x_1 + 2x_2 \le 2000$$
$$x_1 + x_2 \le 1500$$
$$x_2 \le 600 \quad \text{and } x_1, x_2 \ge 0.$$

Also Max $Z = 4000$ with

(i) $x_1 = 800, x_2 = 600,$ *(ii)* $x_1 = 1000, x_2 = 500$ etc.

30. Max $Z = 40x_1 + 30x_2$

subject to
$$3x_1 + x_2 \le 30{,}000$$
$$x_1 \le 8000$$
$$x_2 \le 12000 \quad \text{and } x_1, x_2 \ge 0.$$

Also Max $Z = 6{,}00{,}000, x_1 = 6000, x_2 = 12000$

31. Min $Z = 500, x_1 = 50, x_2 = 0.$

Chapter 3

General Linear Programming Problems – Simplex Methods

3.1.1 General Linear Programming Problem

The linear programming involving more than two variables may be expressed as follows :

Maximize (or) Minimize $Z = c_1x_1 + c_2x_2 + c_3x_3 + \ldots\ldots c_nx_n$

subject to the constraints

$$a_{11}x_1 + a_{12}x_2 + \ldots\ldots\ldots\ldots\ldots + a_{1n}x_n \leq \text{ or } = \text{ or } \geq b_1$$
$$a_{21}x_1 + a_{22}x_2 + \ldots\ldots\ldots\ldots\ldots + a_{2n}x_n \leq \text{ or } = \text{ or } \leq b_2$$
$$a_{31}x_1 + a_{32}x_2 + \ldots\ldots\ldots\ldots\ldots + a_{3n}x_n \leq \text{ or } = \text{ or } \leq b_3$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_{m1}x_1 + a_{m2}x_2 + \ldots\ldots\ldots\ldots\ldots + a_{mn}x_n \leq \text{ or } = \text{ or } \geq b_m$$

and the non-negativity restrictions

$$x_1, x_2, x_3, \ldots\ldots\ldots\ldots\ldots, x_n \geq 0.$$

Note : Some of the constraints may be equalities, some others may be inequalities of ($\leq$) type and remaining ones inequalities of ($\geq$) type or all of them are of same type.

Definition (1) : A set of values $x_1, x_2 \ldots\ldots x_n$ which satisfies the constraints of the LPP is called its ***solution.***

Definition (2) : Any solution to a LPP which satisfies the non-negativity restrictions of the LPP is called its ***feasible solution.***

Definition (3) : Any feasible solution which optimizes (maximizes or minimizes) the objective function of the LPP is called its ***optimum solution*** or ***optimal solution.***

Definition (4) : If the constraints of a general LPP be

$$\sum_{j=1}^{n} a_{ij}\, x_j \leq b_i \qquad (i = 1, 2, 3, \ldots k) \qquad \ldots(1)$$

then the non-negative variables s_i which are introduced to convert the inequalities (1) to the equalities

$$\sum_{j=1}^{n} a_{ij}\, x_j + s_i = b_i \qquad (i = 1, 2, 3, \ldots k)$$

are called ***slack variables.*** The value of these variables can be interpreted as the amount of unused resource.

Definition (5) : If the constraints of a general LPP be

$$\sum_{j=1}^{n} a_{ij}\, x_j \geq b_i \qquad (i = k, k+1, \ldots) \qquad \ldots(1)$$

then the non-negative variables s_i which are introduced to convert the inequalities (1) to the equalities

$$\sum_{j=1}^{n} a_{ij}\, x_j - s_i = b_i \qquad (i = k, k+1, \ldots)$$

are called ***surplus variables.*** The value of these variables can be interpreted as the amount over and above the required level.

3.1.2 Canonical and Standard forms of LPP :

After the formulation of LPP, the next step is to obtain its solution. But before any method is used to find its solution, the problem must be presented in a suitable from. Two forms are dealt with here, the canonical form and the standard form.

The canonical form : The general linear programming problem can always be expressed in the following form :

Maximize $Z = c_1x_1 + c_2x_2 + c_3x_3 + \ldots\ldots c_nx_n$

subject to the constraints

$$a_{11}x_1 + a_{12}x_2 + \ldots\ldots\ldots\ldots\ldots + a_{1n}x_n \leq b_1$$
$$a_{21}x_1 + a_{22}x_2 + \ldots\ldots\ldots\ldots\ldots + a_{2n}x_n \leq b_2$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_{m1}x_1 + a_{m2}x_2 + \ldots\ldots\ldots\ldots\ldots + a_{mn}x_n \leq b_m$$

and the non-negativity restrictions

$$x_1, x_2, \ldots\ldots\ldots\ldots\ldots, x_n \geq 0.$$

This form of LPP is called the ***canonical form*** of the LPP.

In matrix notation the canonical form of LPP can be expressed as :

$$\begin{aligned} \text{Maximize } Z &= CX \text{ (objective function)} \\ \text{subject to } AX &\leq b \text{ (constraints)} \\ \text{and } X &\geq 0 \text{ (non-negativity restrictions)} \\ \text{where } C &= (c_1\ c_2 \ldots\ldots c_n), \end{aligned}$$

Chapter 7

Transportation Model

7.1 Introduction

Transportation deals with the transportation of a commodity (single product) from 'm' sources (origins or supply or *capacity* centres) to 'n' destinations (sinks or demand or requirement centres). It is assumed that

i) Level of supply at each source and the amount of demand at each destination and

ii) The unit transportation cost of commodity from each source to each destination are known [given].

It is also assumed that the cost of transportation is linear.

The objective is to determine the amount to be shifted from each source to each destination such that the total transportation cost is minimum.

Note : The transportation model also can be modified to account for multiple commodities.

I. Mathematical Formulation of a Transportation Problem :

Let us assume that there are m sources and n destinations.

Let a_i be the supply (capacity) at source i, b_j be the demand at destination j, c_{ij} be the unit transportation cost from source i to destination j and x_{ij} be the number of units shifted from source i to destination j.

Then the transportation problem can be expressed mathematically as

$$\text{Minimize } Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}\, x_{ij}$$

subject to the constraints

$$\sum_{j=1}^{n} x_{ij} = a_i, \qquad i = 1, 2, 3 \ldots m.$$

$$\sum_{i=1}^{m} x_{ij} = b_j, \qquad j = 1, 2, 3 \ldots n.$$

$$\text{and } x_{ij} \geq 0, \text{ for all } i \text{ and } j.$$

Note 1 : The two sets of constraints will be *consistent* if

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$$

$$\textit{(total supply)} \qquad \textit{(total demand)}$$

which is the necessary and sufficient condition for a transportation problem to have a feasible solution. Problems satisfying this condition are called ***balanced transportation problems.***

Note 2 : If $\sum a_i \neq \sum b_j$, then the transportation problem is said to be ***unbalanced.***

Note 3 : For any transportation problem, the coefficients of all x_{ij} in the constraints are unity.

Note 4 : The objective function and the constraints being all linear, the transportation problem is a special class of linear programming problem. Therefore it can be solved by simplex method. But the number of variables being large, there will be too many calculations. So we can look for some other technique which would be simpler than the usual simplex method.

Standard transportation table :

Transportation problem is explicitly represented by the following transportation table.

Destination

		D_1	D_2	D_3	...	D_j	...	D_n	Supply
	S_1	c_{11}	c_{12}	c_{13}		c_{1j}		c_{1n}	a_1
Source	S_2	c_{21}	c_{22}	c_{23}		c_{2j}		c_{2n}	a_2
									$\vdots$
	S_i	c_{i1}	c_{i2}			c_{ij}		c_{in}	$\vdots$
	S_m	c_{m1}	c_{m2}			c_{mj}		c_{mn}	a_m
Demand		b_1	b_2	b_3	...	...	...	b_n	$\boxed{\Sigma a_i = \Sigma b_j}$

The mn squares are called ***cells***. The unit transportation cost c_{ij} from the i^{th} source to the j^{th} destination is displayed in the ***upper left side of the (i, j)th cell.*** Any feasible solution is shown in the table by entering the value of x_{ij} ***in the centre of the*** $(i,j)^{th}$ cell. The various a's and b's are called ***rim requirements***. The feasibility of a solution can be verified by summing the values of x_{ij} along the rows and down the columns.

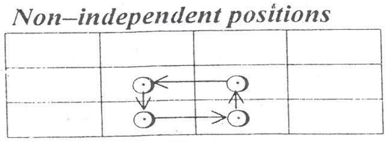
Definition 1: A set of non-negative values x_{ij}, $i = 1,2, \ldots m$; $j = 1,2 \ldots n$. that satisfies the constraints (rim conditions and also the non-negativity restrictions) is called a ***feasible solution*** to the transportation problem.

Note : A balanced transportation problem will always have a feasible solution.

Definition 2: A feasible solution to a $(m \times n)$ transportation problem that contains no more than $m + n - 1$ non-negative allocations is called a ***basic feasible solution*** (BFS) to the transportation problem.

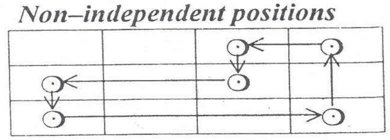
The allocations are said to be in ***independent positions*** if it is impossible to increase or decrease any allocation without either changing the position of the allocation or violating the rim requirements. A simple rule for allocations to be in independent positions is that it is impossible to travel from any allocation, back to itself by a series of horizontal and vertical jumps from one occupied cell to another, without a direct reversal of the route. Example

Non–independent positions



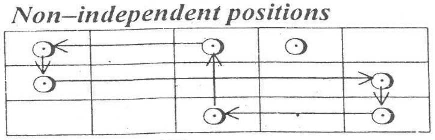
(i)

Non–independent positions



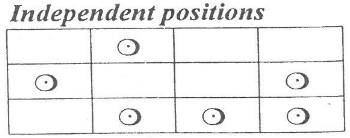
(ii)

Non–independent positions



(i)

Independent positions



(ii)

Definition 3 : A basic feasible solution to a $(m \times n)$ transportation problem is said to be a ***non-degenerate basic feasible solution*** if it contains exactly $m + n - 1$ non-negative allocations in independent positions.

Definition 4: A basic feasible solution that contains less than $m + n - 1$ non-negative allocations is said to be a ***degenerate basic*** feasible solution.

Definition 5: A feasible solution (not necessarily basic) is said to be an ***optimal solution*** if it minimizes the total transportation cost.

Note : The number of basic variables in an $m \times n$ balanced transportation problem is atmost $m + n - 1$.

Note : The number of non–basic variables in an $m \times n$ balanced transportation problem is atleast $mn - (m + n - 1)$

II. Methods for finding initial basic feasible solution

The transportation problem has a solution if and only if the problem is balanced. Therefore before starting to find the initial basic feasible solution, check whether the given transportation problem is balanced. If not one has to balance the transportation problem first. The way of doing this is discussed in section 7.4 page 7.40 In this section all the given transportation problems are balanced.

Method 1 : North west Corner Rule :

Step 1 : The first assignment is made in the cell occupying the upper left–hand (north–west) corner of the transportation table. The maximum possible amount is allocated there. That is $x_{11} = \min \{a_1, b_1\}$.

Case (i) : If $\min \{a_1, b_1\} = a_1$, then put $x_{11} = a_1$, decrease b_1 by a_1 and move vertically to the 2nd row (*i.e.,*) to the cell (2,1) cross out the first row.

Case (ii) : If $\min \{a_1, b_1\} = b_1$, then put $x_{11} = b_1$, and decrease a_1 by b_1 and move horizontally right (*i.e.,*) to the cell (1,2) cross out the first column

Case (iii) : If $\min \{a_1, b_1\} = a_1 = b_1$ then put $x_{11} = a_1 = b_1$ and move diagonally to the cell (2,2) cross out the first row and the first column.

Step 2: Repeat the procedure until all the rim requirements are satisfied.

Method 2 : Least Cost method (or) Matrix minima method (or) Lowest cost entry method :

Step 1 : Identify the cell with smallest cost and allocate $x_{ij} = \text{Min} \{a_i, b_j\}$

Case (i) : If $\min \{a_i, b_j\} = a_i$, then put $x_{ij} = a_i$, cross out the i^{th} row and decrease b_j by a_i, Go to step (2).

Case (ii) : If min $\{a_i, b_j\} = b_j$ then put $x_{ij} = b_j$ cross out the j^{th} column and decrease a_i by b_j Go to step (2).

Case (iii) : If min $\{a_i, b_j\} = a_i = b_j$, then put $x_{ij} = a_i = b_j$, cross out either i^{th} row or j^{th} column but not both, Go to step (2).

Step 2 : Repeat step (1) for the resulting reduced transportation table until all the rim requirements are satisfied.

Method 3: Vogel's approximation method *(VAM) (or) Unit cost penalty method :* *[MU. MBA. Nov 96, Apr 95, Apr 97]*

Step 1 : Find the difference (penalty) between the smallest and next smallest costs in each row (column) and write them in brackets against the corresponding row (column).

Step 2 : Identify the row (or) column with largest penalty. If a tie occurs, break the tie arbitrarily. Choose the cell with smallest cost in that selected row or column and allocate as much as possible to this cell and cross out the satisfied row or column and go to step (3).

Step 3 : Again compute the column and row penalties for the reduced transportation table and then go to step (2). Repeat the procedure until all the rim requirements are satisfied.

Example 1 : Determine basic feasible solution to the following transportation problem using North West Corner Rule :

Sink

		A	B	C	D	E	Supply
	P	2	11	10	3	7	4
Origin	Q	1	4	7	2	1	8
	R	3	9	4	8	12	9
Demand		3	3	4	5	6	

[MU. BE. Apr 94]

Solution: Since $\sum a_i = \sum b_j = 21$, the given problem is balanced.

$\therefore$ There exists a feasible solution to the transportation problem.

2 3	11	10	3	7	4
1	4	7	2	1	8
3	9	4	8	12	9
3	3	4	5	6	

Following North West Corner rule, the first allocation is made in the cell (1,1).

Here x_{11} = min $\{a_1, b_1\}$ = min {4, 3} = 3

$\therefore$ Allocate 3 to the cell (1,1) and decrease 4 by 3 *ie.,* 4 – 3 = 1

As the first column is satisfied, we cross out the first column and the resulting reduced Transportation table is

11 1	10	3	7	1
4	7	2	1	8
9	4	8	12	9
3	4	5	6	

Here the North West Corner cell is (1,2).

So Allocate x_{12} = min {1,3} = 1 to the cell (1,2) and move vertically to cell (2,2). The resulting reduced transportation table is

4 2	7	2	1	8
9	4	8	12	9
2	4	5	6	

Allocate x_{22} = min {8, 2} = 2 to the cell (2, 2) and move horizontally to the cell (2,3). The resulting transportation table is

7 4	2	1	6
4	8	12	9
4	5	6	

Allocate x_{23} = min {6,4} = 4 and move horizontally to the cell (2,4).

The resulting reduced transportation table is

2 **2**	1	2
8	12	9
5	6	

Allocate x_{24} = {2,5} = 2 and move vertically to the cell (3,4). The resulting reduced transportation table is

8 **3**	12	9
3	6	

Allocate x_{34} = min {9, 3} = 3 and move horizontally to the cell (3, 5), which is

12 **6**	6
6	

Allocate x_{35} = min {6,6} = 6

Finally the initial basic feasible solution is as shown in the following table.

2 **3**	11 **1**	10	3	7
1	4 **2**	7 **4**	2 **2**	1
3	9	4	8 **3**	12 **6**

From this table we see that the number of positive independent allocations is equal to $m + n - 1 = 3 + 5 - 1 = 7$. This ensures that the solution is non degenerate basic feasible.

$\therefore$ The initial transportation cost $= \text{Rs. } 2 \times 3 + 11 \times 1 + 4 \times 2 + 7 \times 4 + 2 \times 2 + 8 \times 3 + 12 \times 6$

= Rs.153/-

Example 2 : Find the initial basic feasible solution for the following transportation problem by Least Cost Method.

	To				Supply
From	**1**	**2**	**1**	**4**	**30**
	3	**3**	**2**	**1**	**50**
	4	**2**	**5**	**9**	**20**
Demand	**20**	**40**	**30**	**10**	

[MU. BE. Apr 95, MSU. BE. Nov 96]

Solution : Since $\Sigma a_i = \Sigma b_j = 100$, the given TPP is balanced.

$\therefore$ There exists a feasible solution to the transportation problem.

1 **20**	2	1	4	30
3	3	2	1	50
4	2	5	9	20
20	40	30	10	

By least cost method, min $c_{ij} = c_{11} = c_{13} = c_{24} = 1$

Since more than one cell having the same minimum c_{ij}, break the tie.

Let us choose the cell (1,1) and allocate x_{11} = min $\{a_1, b_1\}$ = min {30, 20} = 20 and cross out the satisfied column and decrease 30 by 20.

The resulting reduced transportation table is

2	1 **10**	4	10
3	2	1	50
2	5	9	20
40	30	10	

Here min $c_{ij} = c_{13} = c_{24} = 1$

Choose the cell (1,3) and allocate $x_{13} = \min \{a_1, b_3\} = \min \{10, 30\} = 10$ and cross out the satisfied row.

The resulting reduced transportation table is

3	2	1 **10**	50
2	5	9	20
40	20	10	

Here min $c_{ij} = c_{24} = 1$,

$\therefore$ Allocate $x_{24} = \min \{a_2, b_4\} = \min (50, 10) = 10$ and cross out the satisfied column.

The resulting transportation table is

3	2 **20**	40
2	5	20
40	20	

Here min $c_{ij} = c_{23} = c_{32} = 2$. Choose the cell (2,3) and allocate $x_{23} = \min \{a_2, b_3\} = \min (40, 20) = 20$ and cross out the satisfied column.

The resulting reduced transportation table is

3	20
2 **20**	20
40	

Here min $c_{ij} = c_{32} = 2$. Choose the cell (3, 2) and allocate $x_{32} = \min \{a_3, b_2\} = \min (20, 40) = 20$ and cross out the satisfied row.

The resulting reduced transportation table is

3 **20**	20
20	

Finally the initial basic feasible solution is as shown in the following table.

1 **20**	2	1 **10**	4
3	3 **20**	2 **20**	1 **10**
4	2 **20**	5	9

From this table we see that the number of positive independent allocations is equal to $m + n - 1 = 3 + 4 - 1 = 6$. This ensures that the solution is non degenerate basic feasible.

$\therefore$ The initial transportation cost $= \text{Rs. } 1 \times 20 + 1 \times 10 + 3 \times 20 + 2 \times 20 + 1 \times 10 + 2 \times 20$

$= 20 + 10 + 60 + 40 + 10 + 40$

$=$ Rs.180/-

Example 3 : **Find the initial basic feasible solution for the following transportation problem by VAM.**

Distribution Centres

		D_1	D_2	D_3	D_4	Availability
	S_1	11	13	17	14	250
Origin	S_2	16	18	14	10	300
	S_3	21	24	13	10	400
Requirements		200	225	275	250	

Solution : Since $\Sigma a_i = \Sigma b_j = 950$, the given problem is balanced.

$\therefore$ There exists a feasible solution to this problem.

11 200	13	17	14	250	(2)
16	18	14	10	300	(4)
21	24	13	10	400	(3)
200	225	275	250		
(5)	(5)	(1)	(0)		

First let us find the difference (penalty) between the smallest and next smallest costs in each row and colum and write them in brackets against the respective rows and columns.

The largest of these differences is (5) and is associated with the first two columns of the transportation table. We choose the first column arbitrarily.

In this selected column, the cell (1,1) has the minimum unit transportation cost $c_{11} = 11$.

$\therefore$ Allocate x_{11} = min {250, 200} = 200 to this cell (1,1) and decrease 250 by 200 and cross out the satisfied column.

The resulting reduced transportation table is

13 50	17	14	50	(1)
18	14	10	300	(4)
24	13	10	400	(3)
225	275	250		
(5)	(1)	(0)		

The row and column differences are now computed for this reduced transportation table. The largest of these is (5) which is associated with the second column. Since $c_{12} = 13$ is the minimum cost, we allocate x_{12} = min {50,225} = 50 to the cell (1,2) and decrease 225 by 50 and cross out the satisfied row.

Continuing in this manner, the subsequent reduced transportation tables and the differences for the surviving rows and columns are shown below :

18 175	14	10	300	(4)
24	13	10	400	(3)
175	275	250		
(6)	(1)	(0)		

(i)

14	10 125	125	**(4)**
13	10	400	(3)
	250		
(1)	(0)		

(ii)

13	10 125	400	(3)
275	125		

(iii)

13 275	275
275	

(iv)

Finally the initial basic feasible solution is as shown in the following table.

11 **200**	13 **50**	17	14
16	18 **175**	14	10 **125**
21	24	13 **275**	10 **125**

From this table we see that the number of positive independent allocations is equal to $m + n - 1 = 3 + 4 - 1 = 6$. This ensures that the solution is non degenerate basic feasible.

∴ The initial transportation cost = Rs. $11 \times 200 + 13 \times 50 + 18 \times 175$

$+ 10 \times 125 + 13 \times 275 + 10 \times 125$

= Rs. 12075/-

Example 4 : Find the starting solution of the following transportation model

1	2	6	7
0	4	2	12
3	1	5	11
10	10	10	

using *(i) North West Corner rule*
(ii) Least Cost method
(iii) Vogel's approximation method.

Solution : Since $\Sigma a_i = \Sigma b_j = 30$, the given Transportation problem is balanced. Hence there exists a basic feasible solution to this problem.

(i) North West Corner rule : Using this method, the allocations are shown in the tables below :

1 7	2	6	7
0	4	2	12
3	1	5	11
10	10	10	

(i)

0 3	4	2	12
3	1	5	11
3	10	10	

(ii)

4 8	2	9
1	5	11
10	10	

(iii)

1 1	5	11
1	10	

(iv)

5 10	10
10	

(v)

The starting solution is as shown in the following table

1 **7**	2	6
0 **3**	4 **9**	2
3	1 **1**	5 **10**

∴ The initial transportation cost $= \text{Rs. } 1 \times 7 + 0 \times 3 + 4 \times 9 + 1 \times 1 + 5 \times 10$

$= \text{Rs. } 94/-$

(ii) Least Cost Method : Using this method, the allocations are as shown in the table below :

1	2	6 7	7
0 10	4	2	12
3	1	5	11
10	10	10	

(i)

2	6	7
4	2	2
1 10	5	11
10	10	

(ii)

6	7
2 2	2
5	1
10	

(iii)

6	7
5 **1**	1
8	

(iv)

6 7	7
7	

(v)

The starting solution is as shown in the following table :

1	2	6 **7**
0 **10**	4	2 **2**
3	1 **10**	5 **1**

∴ The initial transportation cost $= \text{Rs. } 6 \times 7 + 0 \times 10 + 2 \times 2 + 1 \times 10 + 5 \times 1$

$= \text{Rs. } 61/-$

(iii) Vogel's Approximation Method: Using this method, the allocations are shown in the tables below :

1	2	6	7	(1)
0	4	2 10	12	(2)
3	1	5	11	(2)
10	10	10		
(1)	(1)	(3)		

(i)

1	2	7	(1)
0 2	4	2	**(4)**
3	1	11	(2)
10	10		
(1)	(1)		

(ii)

1	2	7	(1)
3	1 10	11	**(2)**
8	10		
(2)	(1)		

(iii)

1 7	7
3	1
8	

(iv)

3 1	1
1	

(v)

The starting solution is as shown in the following table :

1 7	2	6
0 2	4	2 10
3 1	1 10	5

$\therefore$ The initial transportation cost $=$ Rs. $1 \times 7 + 0 \times 2 + 2 \times 10 + 3 \times 1 + 1 \times 10$

$=$ Rs. 40/-

Note : For the above problem, the number of positive allocations in independent positions is $(m + n - 1)$ (*i.e.*, $m + n - 1 = 3 + 3 - 1 = 5$). This ensures that the given problem has a non-degenerate basic feasible solution by using all the three methods. This need not be the case in all the problems.

7.2 Transportation Algorithm (or) MODI Method (modified distribution method) (Test for optimal solution).

[MU. MBA. Apr 96, Apr 97]

Step 1 : Find the initial basic feasible solution of the given problem by Northwest Corner rule (or) Least Cost method or VAM.

Step 2 : Check the number of occupied cells. If these are less than $m + n - 1$, there exists degeneracy and we introduce a very small positive assignment of $\in$ (≈ 0) in suitable independent positions, so that the number of occupied cells is exactly equal to $m + n - 1$.

Step 3 : Find the set of values u_i, v_j $(i = 1,2,3, \ldots m \; ; j = 1, 2, 3 \ldots n)$ from the relation $c_{ij} = u_i + v_j$ for each occupied cell (i, j), by starting initially with $u_i = 0$ or $v_j = 0$ preferably for which the corresponding row or column has maximum number of individual allocations.

Step 4 : Find $u_i + v_j$ for each unoccupied cell (i, j) and enter at the upper right corner of the corresponding cell (i, j).

Step 5 : Find the cell evaluations $d_{ij} = c_{ij} - (u_i + v_j)$ (d_{ij} = upper left – upper right) for each unoccupied cell (i, j) and enter at the lower right corner of the corresponding cell (i, j).

Step 6 : Examine the cell evaluations d_{ij} for all unoccupied cells (i, j) and conclude that

(i) if all $d_{ij} > 0$, then the solution under the test is optimal and unique.

(ii) if all $d_{ij} > 0$, with atleast one $d_{ij} = 0$, then the solution under the test is optimal and an alternative optimal solution exists.

(iii) if atleast one $d_{ij} < 0$, then the solution is not optimal. Go to the next step.

Step 7 : Form a new B.F.S by giving maximum allocation to the cell for which d_{ij} is most negative by making an occupied cell empty. For that draw a closed path consisting of horizontal and vertical lines beginning and ending at the cell for which d_{ij} is most negative and having its ***other corners at some allocated cells.*** Along this closed loop indicate $+\theta$ and $-\theta$ alternatively at the corners. Choose minimum of the allocations from the cells having $-\theta$. Add this minimum allocation to the cells with $+\theta$ and subtract this minimum allocation from the allocation to the cells with $-\theta$.

Step 8: Repeat steps (2) to (6) to test the optimality of this new basic feasible solution.

Step 9: Continue the above procedure till an optimum solution is attained.

Note: The Vogels approximation method (VAM) takes into account not only the least cost c_{ij} but also the costs that just exceed the least cost c_{ij} and therefore yields better initial solution than obtained from other methods in general. This can be justified by the above example (4). So to find the initial solution, give preference to VAM unless otherwise specified.

Example 1 **: Solve the transportation problem :**

	1	2	3	4	Supply
I	21	16	25	13	11
II	17	18	14	23	13
III	32	27	18	41	19
Demand	6	10	12	15	

[MU. BE. Apr 91, Apr 92, Apr 93, Apr 97, MSU. BE. Apr 97]

Solution : Since $\sum a_i = \sum b_j = 43$, the given transportation problem is balanced. $\therefore$ There exists a basic feasible solution to this problem.

By Vogel's approximation method, the initial solution is as shown in the following table:

21	16	25	13 **11**	(3)	–	–	–
17 **6**	18 **3**	14	23 **4**	(3)	(3)	(3)	(4)
32	27 **7**	18 **12**	41	(9)	(9)	(9)	(9)
(4)	(2)	(4)	**(10)**				
(15)	(9)	(4)	**(18)**				
(15)	(9)	(4)	–				
–	**(9)**	(4)	–				

That is

21	16	25	13 **11**
17 **6**	18 **3**	14	23 **4**
32	27 **7**	18 **12**	41

From this table, we see that the number of non-negative independent allocations is $(m + n - 1) = (3 + 4 - 1) = 6$. Hence the solution is non-degenerate basic feasible.

$\therefore$ The initial transportation cost

= Rs. $13 \times 11 + 17 \times 6 + 18 \times 3 + 23 \times 4 + 27 \times 7 + 18 \times 12$

= Rs. 796/-

To find the optimal solution

Consider the above transportation table. Since $m + n - 1 = 6$, we apply MODI method,

Now we determine a set of values u_i and v_j for each occupied cell (i, j) by using the relation $c_{ij} = u_i + v_j$. As the 2nd row contains maximum number of allocations, we choose $u_2 = 0$.

Therefore

$$c_{21} = u_2 + v_1 \Rightarrow 17 = 0 + v_1 \Rightarrow v_1 = 17$$
$$c_{22} = u_2 + v_2 \Rightarrow 18 = 0 + v_2 \Rightarrow v_2 = 18$$

$$c_{24} = u_2 + v_4 \Rightarrow 23 = 0 + v_4 \Rightarrow v_4 = 23$$
$$c_{14} = u_1 + v_4 \Rightarrow 13 = u_1 + 23 \Rightarrow u_1 = -10$$
$$c_{32} = u_3 + v_2 \Rightarrow 27 = u_3 + 18 \Rightarrow u_3 = 9$$
$$c_{33} = u_3 + v_3 \Rightarrow 18 = 9 + v_3 \Rightarrow v_3 = 9$$

Thus we have the following transportation table :

21	16	25	13 **11**	$u_1 = -10$
17 **6**	18 **3**	14	23 **4**	$u_2 = 0$
32	27 **7**	18 **12**	41	$u_3 = 9$
$v_1 = 17$	$v_2 = 18$	$v_3 = 9$	$v_4 = 23$	

Now we find $u_i + v_j$ for each unoccupied cell (i, j) and enter at the upper right corner of the corresponding unoccupied cell (i, j).

Then we find the cell evaluations $d_{ij} = c_{ij} - (u_i + v_j)$ (*ie.*, upper left corner – upper right corner) for each unoccupied cell (i, j) and enter at the lower right corner of the corresponding unoccupied cell (i, j)

Thus we get the following table :

21 7 14	16 8 8	25 –1 26	13 **11**	$u_1 = -10$
17 **6**	18 **3**	14 9 5	23 **4**	$u_2 = 0$
32 26 6	27 **7**	18 **12**	41 32 9	$u_3 = 9$
$v_1 = 17$	$v_2 = 18$	$v_3 = 9$	$v_4 = 23$	

Since all $d_{ij} > 0$, the solution under the test is optimal and unique.

$\therefore$ The optimum allocation schedule is given by $x_{14} = 11$, $x_{21} = 6$, $x_{22} = 3$, $x_{24} = 4$, $x_{32} = 7$, $x_{33} = 12$ and the optimum (minimum) transportation cost

= Rs. $13 \times 11 + 17 \times 6 + 18 \times 3 + 23 \times 4 + 27 \times 7 + 18 \times 12$

= Rs. 796/-

Example 2 : Obtain on optimum basic feasible solution to the following transportation problem :

	To			Available
From	7	3	2	2
	2	1	3	3
	3	4	6	5
Demand	4	1	5	10

[MU. MCA. May 93]

Solution : Since $\Sigma\, a_i = \Sigma\, b_j = 10$, the given transportation problem is balanced. ∴ There exists a basic feasible solution to this problem.

By Vogel's approximation method, the initial solution is as shown in the following table :

7	3	2 **2**	(1)	**(5)**	–
2	1 **1**	3 2	(1)	(1)	(1)
3 **4**	4	6 **1**	(1)	(3)	**(3)**
(1) (1) (1)	**(2)** – –	(1) (1) (3)			

That is

7	3	2 **2**
2	1 **1**	3 **2**
3 **4**	4	6 **1**

From this table we see that the number of non-negative allocations is $m + n - 1 = (3 + 3 - 1) = 5$.

Hence the solution is non-degenerate basic feasible

∴ The initial transportation cost

$$= \text{Rs. } 2 \times 2 + 1 \times 1 + 3 \times 2 + 3 \times 4 + 6 \times 1$$
$$= \text{Rs. } 29/-$$

For optimality : Since the number of non-negative independent allocations is $m + n - 1$, we apply MODI method.

Since the third column contains maximum number of allocations, we choose $v_3 = 0$.

Now we determine a set of values u_i and v_j by using the occupied cells and the relation $c_{ij} = u_i + v_j$.

That is,

7 −1	3 0	2 **2**	$u_1 = 2$
2	1 **1**	3 **2**	$u_2 = 3$
3 **4**	4	6 **1**	$u_3 = 6$
$v_1 = -3$	$v_2 = -2$	$v_3 = 0$	

Now we find $u_i + v_j$ for each unoccupied cell (i,j) and enter at the upper right corner of the corresponding unoccupied cell (i,j).

Then we find the cell evaluations $d_{ij} = c_{ij} - (u_i + v_j)$ for each unoccupied cell (i,j) and enter at the lower right corner of the corresponding unoccupied cell (i,j).

Thus we get the following table

7 −1 8	3 0 3	2 **2**	$u_1 = 2$
2 0 2	1 **1**	3 **2**	$u_2 = 3$
3 **4**	4 4 0	6 **1**	$u_3 = 6$
$v_1 = -3$	$v_2 = -2$	$v_3 = 0$	

Since all $d_{ij} > 0$, with $d_{32} = 0$, the current solution is optimal and there exists an alternative optimal solution.

$\therefore$ The optimum allocation schedule is given by $x_{13} = 2$, $x_{22} = 1$, $x_{23} = 2$, $x_{31} = 4$, $x_{33} = 1$ and the optimum (minimum) transportation cost

$= \text{Rs. } 2 \times 2 + 1 \times 1 + 3 \times 2 + 3 \times 4 + 6 \times 1 = \text{Rs. } 29/-$

Example 3 : Find the optimal transportation cost of the following matrix using least cost method for finding the critical solution.

Market

	A	B	C	D	E	Available
P	4	1	2	6	9	100
Factory Q	6	4	3	5	7	120
R	5	2	6	4	8	120
Demand	40	50	70	90	90	

[MU. MCA. Apr 93]

Solution :

Since $\Sigma a_i = \Sigma b_j = 340$, the given transportation problem is balanced.
$\therefore$ There exists a basic feasible solution to this problem.

By using Least cost method, the initial solution is as shown in the following table :

4	1 **50**	2 **50**	6	9
6 **10**	4	3 **20**	5	7 **90**
5 **30**	2	6	4 **90**	8

$\therefore$ The initial transportation cost

$= \text{Rs. } 1 \times 50 + 2 \times 50 + 6 \times 10 + 3 \times 20 + 7 \times 90 + 5 \times 30 + 4 \times 90$

$= \text{Rs. } 1410/-$

For optimality : Since the number of non-negative independent allocations is $(m + n - 1)$, we apply MODI method :

That is

4 5 −1	1 **50**	2 **50**	6 4 2	9 6 3	$u_1 = -1$
6 **10**	4 2 2	3 **20**	5 5 0	7 **90**	$u_2 = 0$
5 **30**	2 1 1	6 2 4	4 **90**	8 6 2	$u_3 = -1$
$v_1 = 6$	$v_2 = 2$	$v_3 = 3$	$v_4 = 5$	$v_5 = 7$	

Since $d_{11} = -1 < 0$, the current solution is not optimal.

Now let us form a new basic feasible solution by giving maximum allocation to the cell (i,j) for which d_{ij} is most negative by making an occupied cell empty. Here the cell (1,1) having the negative value $d_{11} = -1$. We draw a closed loop consisting of horizontal and vertical lines beginning and ending at this cell (1,1) and having its other corners at some occupied cells. Along this closed loop indicate $+\theta$ and $-\theta$ alternatively at the corners, we have

4 $+\theta$	1 **50**	2 **50** $-\theta$	6	9
6 **10** $-\theta$	4	3 **20** $+\theta$	5	7 **90**
5 **30**	2	6	4 **90**	8

From the two cells (1,3), (2,1) having $-\theta$, we find that the minimum of the allocations 50,10 is 10. Add this 10 to the cells with $+\theta$ and subtract this 10 to the cells with $-\theta$.

Hence the new basic feasible solution is displayed in the following table :

4 10	1 50	2 40	6	9
6	4	3 30	5	7 90
5 30	2	6	4 90	8

We see that the above table satisfies the rim conditions with $(m + n - 1)$ non-negative allocations at independent positions. So we apply MODI method.

4 10	1 50	2 40	6 3 3	9 6 3	$u_1 = 0$
6 5 1	4 2 2	3 30	5 4 1	7 90	$u_2 = 1$
5 30	2 2 0	6 3 3	4 90	8 7 1	$u_3 = 1$
$v_1 = 4$	$v_2 = 1$	$v_3 = 2$	$v_4 = 3$	$v_5 = 6$	

Since all $d_{ij} > 0$, with $d_{32} = 0$, the current solution is optimal and there exists an alternative optimal solution.

$\therefore$ The optimum allocation schedule is given by $x_{11} = 10$, $x_{12} = 50$, $x_{13} = 40$, $x_{23} = 30$, $x_{25} = 90$, $x_{31} = 30$, $x_{34} = 90$ and the optimum (minimum) transportation cost

$= \text{Rs. } 4 \times 10 + 1 \times 50 + 2 \times 40 + 3 \times 30 + 7 \times 90 + 5 \times 30 + 4 \times 90$

$= \text{Rs. } 1400/-$

7.3 Degeneracy in Transportation Problems

In a transportation problem, whenever the number of non-negative independent allocations is less than $m + n - 1$, the transportation problem is said to be a ***degenerate*** one. Degeneracy may occur either at the initial stage or at an intermediate stage at some subsequent iteration.

To resolve degeneracy, we allocate an extremely small amount (close to zero) to one or more empty cells of the transportation table (generally minimum cost cells if possible), so that the total number of occupied cells becomes $(m + n - 1)$ at independent positions. We denote this small amount by $\in$ (epsilon) satisfying the following conditions :

(i) $0 < \in < x_{ij}$, for all $x_{ij} > 0$

(ii) $x_{ij} \pm \in = x_{ij}$, for all $x_{ij} > 0$

The cells containing $\in$ are then treated like other occupied cells and the problem is solved in the usual way. The $\in$'s are kept till the optimum solution is attained. Then we let each $\in \to 0$.

Example 1 : **Find the non-degenerate basic feasible solution for the following transportation problem using**

(i) **North west corner rule**

(ii) **Least cost method**

(iii) **Vogel's approximation method.**

	To				Supply
	10	20	5	7	10
	13	9	12	8	20
From	4	5	7	9	30
	14	7	1	0	40
	3	12	5	19	50
Demand	60	60	20	10	

[MU. MCA. Apr 93]

Solution : Since $\Sigma\, a_i = \Sigma\, b_j = 150$, the given transportation problem is balanced.

$\therefore$ There exists a basic feasible solution to this problem.

***(i)* The starting solution by NWC rule is as shown in the following table.**

10 **10**	20	5	7
13 **20**	9	12	8
4 **30**	5	7	9
14	7 **40**	1	0
3	12 **20**	5 **20**	19 **10**

Since the number of non-negative allocations at independent positions is 7 which is less than $(m + n - 1) = (5 + 4 - 1) = 8$, this basic feasible solution is a degenerate one.

To resolve this degeneracy, we allocate a very small quantity $\in$ to the unoccupied cell (5,1) so that the number of occupied cells becomes $(m + n - 1)$. Hence the non-degenerate basic feasible solution is as shown in the following table.

10 **10**	20	5	7
13 **20**	9	12	8
4 **30**	5	7	9
14	7 **40**	1	0
3 $\in$	12 **20**	5 **20**	19 **10**

The initial transportation cost $= \text{Rs. } 10 \times 10 + 13 \times 20 + 4 \times 30 + 7 \times 40 + 3 \times \in + 12 \times 20 + 5 \times 20 + 19 \times 10$

$= \text{Rs.}(1290 + 3\in)$

$= \text{Rs. } 1290/-, \text{ as } \in \to 0.$

***(ii)* Least cost method :** Using this method the starting solution is as shown in the following table :

10	20 **10**	5	7
13	9 **20**	12	8
4 **10**	5 **20**	7	9
14	7 **10**	1 **20**	0 **10**
3 **50**	12	5	19

m+n-1

5+4-1 = 8

8 balanced

No. of allocations

Since the number of non-negative allocations at independent positions is $(m + n - 1) = 8$, the solution is non-degenerate basic feasible.

The initial transportation cost $= \text{Rs. } 20 \times 10 + 9 \times 20 + 4 \times 10 + 5 \times 20 + 7 \times 10 + 1 \times 20 + 0 \times 10 + 3 \times 50$

$= \text{Rs. } 760/-$

***(iii)* Vogel's approximation Method :** The starting solution by this method is as shown in the following table :

10 **10**	20	5	7
13	9 **20**	12	8
4	5 **30**	7	9
14	7 **10**	1 **20**	0 **10**
3 **50**	12	5	19

Since the number of non-negative allocations is 7 which is less than $(m + n - 1) = (5 + 4 - 1) = 8$, this basic solution is a degenerate one.

To resolve this degeneracy, we allocate a very small quantity $\in$ to the unoccupied cell (5, 2) so that the number of occupied cells becomes $(m + n - 1)$. Hence the non-degenerate basic feasible solution is as shown in the following table.

10 **10**	20	5	7
13	9 **20**	12	8
4	5 **30**	7	9
14	7 **10**	1 **20**	0 **10**
3 **50**	12 $\in$	5	19

$\therefore$ The initial transportation cost

$= \text{Rs. } 10 \times 10 + 9 \times 20 + 5 \times 30 + 7 \times 10 + 1 \times 20 + 0 \times 10 + 3 \times 50 + 12 \times \in$

$= \text{Rs. } (670 + 12 \in)$

$= \text{Rs. } 670/- = \text{ as } \in \to 0.$ *[Please refer note in page 7.17]*

$\boxed{\textbf{Example 2}}$ **: Solve the following transportation problem using Vogel's method.**

Warehouse

		A	B	C	D	E	F	Available
	1	9	12	9	6	9	10	5
Factory	2	7	3	7	7	5	5	6
	3	6	5	9	11	3	11	2
	4	6	8	11	2	2	10	9
Requirement		4	4	6	2	4	2	

[MU. MCA. Apr 92]

Solution : Since $\Sigma a_i = \Sigma b_j = 22$, the given transportation problem is balanced. $\therefore$ There exists a basic feasible solution to this problem. By Vogel's approximation method, the initial solution is as shown in the following table :

9	12	9 **5**	6	9	10
7	3 **4**	7	7	5	5 **2**
6 **1**	5	9 **1**	11	3	11
6 **3**	8	11	2 **2**	2 **4**	10

Since the number of non-negative allocations is 8 which is less than $(m + n - 1) = (4 + 6 - 1) = 9$, this basic solution is degenerate one.

To resolve degeneracy, we allocate a very small quantity $\in$ to the cell (3,2), so that the number of occupied cells becomes $(m + n - 1)$. Hence the non-degenerate basic feasible solution is as shown in the following table.

9	12	9 **5**	6	9	10	u_1
7	3 **4**	7	7	5	5 **2**	u_2
6 **1**	5 $\in$	9 **1**	11	3	11	$u_3 = 0$
6 **3**	8	11	2 **2**	2 **4**	10	$u_4 = 0$
v_1	v_2	v_3	v_4	v_5	v_6	

The initial transportation cost $= \text{Rs. } 9 \times 5 + 3 \times 4 + 5 \times 2 + 6 \times 1 + 5 \times \in + 9 \times 1 + 6 \times 3 + 2 \times 2 + 2 \times 4$

$= \text{Rs. } (112 + 5 \in) = \text{Rs. } 112/- , \in \to 0.$

To find the optimal solution

Now the number of non-negative allocations at independent positions is $(m+n-1)$. We apply the MODI method.

9 6 3	12 5 7	9 **5**	6 2 4	9 2 7	10 7 3	$u_1 = 0$
7 4 3	3 **4**	7 7 0	7 0 7	5 0 5	5 **2**	$u_2 = -2$
6 **1**	5 $\in$	9 **1**	11 2 9	3 2 1	11 7 4	$u_3 = 0$
6 **3**	8 5 3	11 9 2	2 **2**	2 **4**	10 7 3	$u_4 = 0$
$v_1 = 6$	$v_2 = 5$	$v_3 = 9$	$v_4 = 2$	$v_5 = 2$	$v_6 = 7$	

Since all $d_{ij} > 0$ with $d_{23} = 0$, the solution under the test is optimal and an alternative optimal solution is also exists.

$\therefore$ The optimum allocation schedule is given by $x_{13} = 5$, $x_{22} = 4$, $x_{26} = 2$, $x_{31} = 1$, $x_{32} = \in$, $x_{33} = 1$, $x_{41} = 3$, $x_{44} = 2$, $x_{45} = 4$ and the optimum (minimum) transportation cost is

$= \text{Rs. } 9 \times 5 + 3 \times 4 + 5 \times 2 + 6 \times 1 + 5 \times \in + 9 \times 1 + 6 \times 3 + 2 \times 2 + 2 \times 4$

$= \text{Rs. } (112 + 5 \in)$

$= \text{Rs. } 112$ as $\in \to 0$.

Example 3 : **Solve the transportation problem:**

	To				
					Supply
From	1	2	3	4	6
	4	3	2	0	8
	0	2	2	1	10
Demand	4	6	8	6	

[MU. MCA. Apr 87]

Solution : Since $\sum a_i = \sum b_j = 24$, the given transportation problem is balanced. $\therefore$ There exists a basic feasible solution.

By using Vogel's approximation method, the initial solution is as shown in the following table :

1	2 **6**	3	4
4	3	2 **2**	0 **6**
0 **4**	2	2 **6**	1

4

Since the number of non-negative allocations at independent positions is 5, which is less than $(m+n-1) = (3+4-1) = 6$, this basic feasible solution is degenerate.

To resolve degeneracy, we allocate a very small quantity $\in$ to the cell (3,2) so that the number of occupied cells becomes $(m+n-1)$. Hence the non-degenerate initial basic feasible solution is given by

1	2 **6**	3	4
4	3	2 **2**	0 **6**
0 **4**	2 $\in$	2 **6**	1

The initial transportation cost

$= \text{Rs. } 2 \times 6 + 2 \times 2 + 0 \times 6 + 0 \times 4 + 2 \times \in + 2 \times 6$

$= \text{Rs. } (28 + 2 \in)$

$= \text{Rs. } 28/-$, as $\in \to 0$.

To find the optimal solution

Now the number non-negative allocations at independent positions is $(m+n-1)$. We apply the MODI method.

1 0 1	2 **6**	3 2 1	4 0 4	$u_1 = 0$
4 0 4	3 2 1	2 **2**	0 **6**	$u_2 = 0$
0 **4**	2 $\in$	2 **6**	1 0 1	$u_3 = 0$
$v_1 = 0$	$v_2 = 2$	$v_3 = 2$	$v_4 = 0$	

Since all $d_{ij} > 0$ the solution under the test is optimal and unique.

$\therefore$ The optimum allocation schedule is given by $x_{12} = 6$, $x_{23} = 2$, $x_{24} = 6$, $x_{31} = 4$, $x_{32} = \in$, $x_{33} = 6$ and the optimum (minimum) transportation cost.

$= \text{Rs. } 2 \times 6 + 2 \times 2 + 0 \times 6 + 0 \times 4 + 2 \times \in + 2 \times 6$

$= \text{Rs. } (28 + 2 \in) = \text{Rs. } 28, \text{ as } \in \to 0.$

Example 4 : Find the optimal solution of the following problem

	Destination			
	X	**Y**	**Z**	**Supply**
P	**1**	**2**	**0**	**30**
Origin Q	**2**	**3**	**4**	**35**
R	**1**	**5**	**6**	**35**
Demand	**30**	**40**	**30**	

[MU. BE. Apr 95]

Solution : Since $\Sigma a_i = \Sigma b_j = 100$, the given transportation problem is balanced.

By using the Vogel's approximation method, the basic feasible solution is displayed in the following table.

1	2	0 **30**
2	3 **35**	4
1 **30**	5 **5**	6

Since the number of non-negative allocations at independent positions is 4 which is less than $(m + n - 1) = 3 + 3 - 1 = 5$, this initial solution is degenerate.

To resolve degeneracy we allocate a very small quantity $\in$ to the cell (3, 3), so that the number of occupied cells becomes $(m + n - 1)$. Hence the non-degenerate basic feasible solution is given by

1	2	0 **30**
2	3 **35**	4
1 **30**	5 **5**	6 $\in$

Now the number of non-negative allocations at independent positions is $(m + n - 1) = 5$. We apply MODI method.

1 −5 6	2 −1 3	0 **30**	$u_1 = -6$
2 −1 3	3 **35**	4 4 0	$u_2 = -2$
1 **30**	5 **5**	6 $\in$	$u_3 = 0$
$v_1 = 1$	$v_2 = 5$	$v_3 = 6$	

Since all $d_{ij} > 0$ with $d_{23} = 0$, the solution under the test is optimal and there exists an alternative optimal solution.

$\therefore$ The optimal allocation schedule is given by $x_{13} = 30$, $x_{22} = 35$, $x_{31} = 30$, $x_{32} = 5$, $x_{33} = \in$ and the optimum (minimum) transportation cost.

$= \text{Rs. } 0 \times 30 + 3 \times 35 + 1 \times 30 + 5 \times 5 + 6 \times \in$

$= \text{Rs. } (160 + 6\in)$

$= \text{Rs. } 160/- \text{ as } \in \to 0.$

Example 5 : Solve the following transportation problem to minimize the total cost of transportation.

		Destination				
		1	2	3	4	Supply
Origin	1	14	56	48	27	70
	2	82	35	21	81	47
	3	99	31	71	63	93
Demand		70	35	45	60	210

[BNU. BE. Nov 96]

Solution: Since $\Sigma a_i = \Sigma b_j = 210$, the given transportation problem is balanced. ∴ There exists a basic feasible solution to this problem.

By using Vogel's approximation method, the initial solution is as shown in the following table :

14 **70**	56	48	27
82	35	21 **45**	81 **2**
99	31 **35**	71	63 **58**

Since the number of non-negative allocations is 5, which is less than $(m + n - 1) = (3 + 4 - 1) = 6$, this basic feasible solution is degenerate.

To resolve degeneracy, we allocate a very small quantity ∈ to the cell (1,4), so that the number of occupied cells becomes $(m + n - 1)$. Hence the non-degenerate basic feasible solution is given in the following table

14 **70**	56	48	27 ∈
82	35	21 **45**	81 **2**
99	31 **35**	71	63 **58**

$u_3 + v_4 = 63$ $u_3 = 63$ $u_1 + v_2 = 3$ $v_2 = -32$ $u_2 + v_4 = 81$ $u_2 = 81$ $u_1 + v_4 = 4$ $u_1 = 27$ $u_2 + v_3 = 21$ $v_3 = -60$

To find the optimal solution :

Now the number of non-negative allocations at independent positions is $(m + n - 1) = 6$. We apply MODI method.

14 **70**	56 −5 61	48 −33 81	27 ∈	$u_1 = 27$
82 68 14	35 49 −14	21 **45**	81 **2**	$u_2 = 81$
99 50 49	31 **35**	71 3 68	63 **58**	$u_3 = 63$
$v_1 = -13$	$v_2 = -32$	$v_3 = -60$	$v_4 = 0$	

Since $d_{22} = -14 < 0$, the solution under the test is not optimal.

Now let us form a new basic feasible solution by giving maximum allocation to the cell (2,2) by making an occupied cell empty. We draw a closed loop consisting of horizontal and vertical lines beginning and ending at this cell (2,2) and having its other corners at some occupied cells. Along this closed loop, indicate +θ and −θ alternatively at the corners.

14 **70**	56	48	27 ∈
82	35 +θ	21 **45**	81 **2** −θ
99	31 **35** −θ	71	63 +θ **58**

From the two cells (2,4), (3,2) having − θ we find that the minimum of the allocations 2,35 is 2. Add this 2 to the cells with +θ and subtract this 2 to the cells with −θ. Hence the new basic feasible solution is given by

14 **70**	56	48	27 ∈
82	35 **2**	21 **45**	81
99	31 **33**	71	63 **60**

We see that the above table satisfies the rim conditions with $(m + n - 1)$ non-negative allocations at independent positions. We apply MODI method for optimality.

14 **70**	56 −5 61	48 −19 67	27 ∈	$u_1 = -40$
82 54 28	35 **2**	21 **45**	81 67 14	$u_2 = 0$
99 50 49	31 **33**	71 17 54	63 **60**	$u_3 = -4$
$v_1 = 54$	$v_2 = 35$	$v_3 = 21$	$v_4 = 67$	

$u_2 + v_2 = 35$
$v_2 = 35$
$u_3 + v_2 = 31$
$u_3 = -4$
$u_3 + v_4 = 63$
$v_4 = 67$
$u_1 + v_4 = 27$
$u_1 = -40$
$u_3 + v_1 = ...$
$v_1 = 54$

Since all $d_{ij} > 0$, the solution under the test is optimal.

$\therefore$ The optimal allocation schedule is given by $x_{11} = 70$, $x_{14} = \epsilon$ $x_{22} = 2$, $x_{23} = 45$, $x_{32} = 33$, $x_{34} = 60$ and the optimum (minimum) transportation cost

$= \text{Rs. } 14 \times 70 + 27 \times \in + 35 \times 2 + 21 \times 45 + 31 \times 33 + 63 \times 60$

$= \text{Rs. } 6798/\text{-} \text{ as } \in \to 0.$

Example 6: Solve the following transportation problem, in which a_i is the availability at origin O_i and b_j is the requirement at the destination D_j and cell entries are unit costs of transportation from any origin to any destination :

	D_1	D_2	D_3	D_4	D_5	a_i
O_1	4	7	3	8	2	4
O_2	1	4	7	3	8	7
O_3	7	2	4	7	7	9
O_4	4	8	2	4	7	2
b_j	8	3	7	2	2.	

[BRU. BE. Nov 96]

Solution : Since $\Sigma a_i = \Sigma b_j = 22$, the given problem is balanced. $\therefore$ There exists a basic feasible solution to this problem.

By using Vogel's approximation method, the initial solution is as shown in the following table :

4 **1**	7	3 **1**	8	2 **2**
1 **7**	4	7	3	8
7	2 **3**	4 **6**	7	7
4	8	2 ∈	4 **2**	7

Since the number of non-negative allocations is 7, which is less than $(m + n - 1) = (4 + 5 - 1) = 8$, this basic feasible solution is degenerate.

To resolve this degeneracy, we allocate a very small quantity ∈ to the cell (4,3), so that the number of occupied cells becomes $(m + n - 1)$. Hence the non-degenerate basic feasible solution is given in the following table

4 **1**	7	3 **1**	8	2 **2**
1 **7**	4	7	3	8
7	2 **3**	4 **6**	7	7
4	8	2 ∈	4 **2**	7

: optimal solution : Now the number of non-negative ndependent positions is $(m + n - 1) = 8$. We apply MODI

	7 1 / 6	3 / **1**	8 5 / 3	2 / **2**	$u_1 = 0$
1 / **7**	4 −2 / 6	7 0 / 7	3 2 / 1	8 −1 / 9	$u_2 = -3$
7 5 / 2	2 / **3**	4 / **6**	7 6 / 1	7 3 / 4	$u_3 = 1$
4 3 / 1	8 0 / 8	2 / $\in$	4 / **2**	7 1 / 6	$u_4 = -1$
$v_1 = 4$	$v_2 = 1$	$v_3 = 3$	$v_4 = 5$	$v_5 = 2$	

Since all $d_{ij} > 0$, the solution under the test is optimal.

$\therefore$ The optimal allocation schedule is given by $x_{11} = 1$, $x_{13} = 1$, $x_{15} = 2$, $x_{21} = 7$, $x_{32} = 3$, $x_{33} = 6$, $x_{43} = \in$, $x_{44} = 2$ and the optimum (minimum) transportation cost

= Rs. $4 \times 1 + 3 \times 1 + 2 \times 2 + 1 \times 7 + 2 \times 3 + 4 \times 6 + 2 \times \in + 4 \times 2$

= Rs. $(56 + 2\in)$

= Rs. 56/- as $\in \to 0$.

7.4 Unbalanced Transportation Problems

If the given transportation problem is unbalanced one, *i.e.*, if $\sum a_i \neq \sum b_j$, then convert this into a balanced one by introducing a dummy source or dummy destination with zero cost vectors (zero unit transportation costs) as the case may be and then solve by usual method.

When the total supply is greater than the total demand, a dummy destination is included in the matrix with zero cost vectors. The excess supply is entered as a rim requirement for the dummy destination.

When the total demand is greater than the total supply, a dummy source is included in the matrix with zero cost vectors. The excess demand is entered as a rim requirement for the dummy source.

Example 1 : Solve the transportation problem

Destination

	A	B	C	D	Supply
Source 1	11	20	7	8	50
2	21	16	20	12	40
3	8	12	18	9	70
Demand :	30	25	35	40 / 130	160

Solution : Since the total supply ($\sum a_i = 160$) is greater than the total demand ($\sum b_j = 130$), the given problem is an unbalanced transportation problem. To convert this into a balanced one, we introduce a dummy destination E with zero unit transportation costs and having demand equal to $160 - 130 = 30$ units.

$\therefore$ The given problem becomes

Destination

	A	B	C	D	E	Supply
Source 1	11	20	7	8	0	50
2	21	16	20	12	0	40
3	8	12	18	9	0	70
	30	25	35	40	30	160

By using VAM the initial solution is as shown in the following table

11	20	7 **35**	8 **15**	0
21	16	20	12 **10**	0 **30**
8 **30**	12 **25**	18	9 **15**	0

$\therefore$ The initial transportation cost

= Rs. $7 \times 35 + 8 \times 15 + 12 \times 10 + 0 \times 30 + 8 \times 30 + 12 \times 25 + 9 \times 15$

= Rs. 1160/-

For optimality: Since the number non-negative allocations at independent positions is $(m + n - 1)$, we apply the MODI method.

11 7 4	20 11 9	7 **35**	8 **15**	0 −4 4	$u_1 = 8$
21 11 10	16 15 1	20 11 9	12 **10**	0 **30**	$u_2 = 12$
8 **30**	12 **25**	18 8 10	9 **15**	0 −3 3	$u_3 = 9$
$v_1 = -1$	$v_2 = 3$	$v_3 = -1$	$v_4 = 0$	$v_5 = -12$	

Since all $d_{ij} > 0$, the solution under the test is optimal and unique.

$\therefore$ The optimum allocation schedule is

$x_{13} = 35, x_{14} = 15, x_{24} = 10, x_{25} = 30, x_{31} = 30, x_{32} = 25, x_{34} = 15$

It can be noted that $x_{25} = 30$ means that 30 units are despatched from source 2 to the dummy destination E. In other words, 30 units are left undespatched from source 2.

The optimum (minimum) transportation cost

= Rs. $7 \times 35 + 8 \times 15 + 12 \times 10 + 0 \times 30 + 8 \times 30 + 12 \times 25 + 9 \times 15$

= Rs. 1160/-

Example 2 : **Solve the transportation problem with unit transportation costs, demands and supplies as given below :**

	Destination D1	D2	D3	D4	Supply
Source S_1	6	1	9	3	70
Source S_2	11	5	2	8	55
Source S_3	10	12	4	7	70
Demand	85	35	50	45	195

[MU MBA Apr 95]

Solution : Since the total demand ($\Sigma b_i = 215$) is greater than the total supply ($\Sigma a_j = 195$), the given problem is unbalanced transportation problem. To convert this into a balanced one, we introduce a dummy source S_4 with zero unit transportation costs and having supply equal to $215 - 195 = 20$ units. $\therefore$ The given problem becomes

	Destination D1	D2	D3	D4	Supply
Source S_1	6	1	9	3	70
Source S_2	11	5	2	8	55
Source S_3	10	12	4	7	70
Source S_4	0	0	0	0	20
	85	35	50	45	215

As this problem is balanced, there exists a basic feasible solution to this problem. By using Vogel's approximation method, the initial solution is as shown in the following table.

6 **65**	1 **5**	9	3
11	5 **30**	2 **25**	8
10	12	4 **25**	7 **45**
0 **20**	0	0	0

$\therefore$ The initial transportation cost

= Rs. $6 \times 65 + 1 \times 5 + 5 \times 30 + 2 \times 25 + 4 \times 25 + 7 \times 45 + 0 \times 20$

= Rs. 1010/-

For optimality : Since number of non-negative allocations at independent positions is $(m + n - 1)$, we apply the MODI method :

6 **65**	1 **5**	9 −2 11	3 1 2	$u_1 = 6$
11 10 1	5 **30**	2 **25**	8 5 3	$u_2 = 10$
10 12 −2	12 7 5	4 **25**	7 **45**	$u_3 = 12$
0 **20**	0 −5 5	0 −8 8	0 −5 5	$u_4 = 0$
$v_1 = 0$	$v_2 = -5$	$v_3 = -8$	$v_4 = -5$	

Since $d_{31} = -2 < 0$, the solution under the test is not optimal.

Now let us form a new basic feasible solution by giving maximum allocation to the cell (3,1) (since d_{31} is $-ve$) by making an occupied cell empty. For this, we draw a closed path consisting of horizontal and vertical lines beginning and ending at this cell (3,1) and having its other corners at some occupied cells. Along this closed loop, indicate $+\theta$ and $-\theta$ alternatively at the corners.

We have

6 65 $-\theta$	1 5 $+\theta$	9	3
11	5 30 $-\theta$	2 25 $+\theta$	8
10 $+\theta$	12	4 25 $-\theta$	7 45
0 20	0	0	0

From the three cells (1,1), (2,2), (3,3) having $-\theta$, we find that the minimum of the allocations 65, 30, 25 is 25. Add this 25 to the cells with $+\theta$ and subtract this 25 to the cells with $-\theta$. Finally, the new basic feasible solution is displayed in the following table.

6 40	1 30	9	3
11	5 5	2 50	8
10 25	12	4	7 45
0 20	0	0	0

We see that the above table satisfies the rim conditions with $(m+n-1)$ non-negative allocations at independent positions. Now we check for optimality

6 40	1 30	9 −2 11	3 3 0	$u_1 = 6$
11 10 1	5 5	2 50	8 7 1	$u_2 = 10$
10 25	12 5 7	4 2 2	7 45	$u_3 = 10$
0 20	0 −5 5	0 −8 8	0 −3 3	$u_4 = 0$
$v_1 = 0$	$v_2 = -5$	$v_3 = -8$	$v_4 = -3$	

Since all $d_{ij} > 0$ with $d_{14} = 0$, the solution under the test is optimal and an alternative optimal solution exists.

$\therefore$ The optimum allocation schedule is given by

$x_{11} = 40, x_{12} = 30, x_{22} = 5, x_{23} = 50, x_{31} = 25, x_{34} = 45, x_{41} = 20.$

It can be noted that $x_{41} = 20$ means that 20 units are despatched from the dummy source S_4 to the destination D_1. In other words, 20 units are not fulfilled for the destination D_1.

The optimum (minimum) transportation cost

= Rs. 6 × 40 +1 × 30 + 5 × 5 +2 × 50 +10 × 25 +7 × 45 + 0 × 20

= Rs. 960/-

Example 3 : Solve the transportation problem with unit transportation costs in rupees, demands and supplies as given below :

		Destination			
		D_1	D_2	D_3	Supply (units)
Origin	A	5	6	9	100
	B	3	5	10	75
	C	6	7	6	50
	D	6	4	10	75
Demand (units)		70	80	120	

[MU MBA Apr 98]

Solution : Since the total supply ($\Sigma a_i = 300$) is greater than the total demand ($\Sigma b_j = 270$), the given transportation problem is unbalanced.

To convert this into a balanced one, we introduce a dummy source D_4 with zero unit transportation costs and having demand equal to $300 - 270 = 30$ units. $\therefore$ The given problem becomes

	Destination D_1	D_2	D_3	D_4	**Supply**
Origin A	5	6	9	0	100
B	3	5	10	0	75
C	6	7	6	0	50
D	6	4	10	0	75
Demand	70	80	120	30	300

By using VAM the initial solution is given by

5	6	9 **100**	0
3 **70**	5 **5**	10	0
6	7	6 **20**	0 **30**
6	4 **75**	10	0

Since the number of non-negative allocations is 6, which is less than $(m + n - 1) = 4 + 4 - 1 = 7$, this basic feasible solution is degenerate.

To resolve this degeneracy, we allocate a very small quantity $\in$ to the cell (2, 4), so that the number of occupied cells becomes $(m + n - 1)$. Hence the non-degenerate basic feasible solution is given in the following table

5	6	9 **100**	0
3 **70**	5 **5**	10	0 $\in$
6	7	6 **20**	0 **30**
6	4 **75**	10	0

Now the number of non-negative allocations at independent positions is $(m + n - 1)$. We apply MODI method.

5 6 −1	6 8 −2	9 **100**	0 3 −3	$u_1 = 3$
3 **70**	5 **5**	10 6 4	0 $\in$	$u_2 = 0$
6 3 3	7 5 2	6 **20**	0 **30**	$u_3 = 0$
6 2 4	4 **75**	10 5 5	0 −1 1	$u_4 = -1$
$v_1 = 3$	$v_2 = 5$	$v_3 = 6$	$v_4 = 0$	

Since there are some $d_{ij} < 0$, the current solution is not optimal.

Since $d_{14} = -3$ is the most negative, let us form a new basic feasible solution by giving maximum allocation to the corresponding cell (1,4) by making an occupied cell empty. We draw a closed loop consisting of horizontal and vertical lines beginning and ending at this cell (1,4) and having its other corners at some occupied cells. Along this closed loop indicate $+\theta$ and $-\theta$ alternatively at the corners.

We have

5	6	9 $-\theta$ **100**	0 $+\theta$
3 **70**	5 **5**	10	0 $\in$
6	7	6 **20** $+\theta$	0 **30** $-\theta$
6	4 **75**	10	0

From the two cells (1,3), (3,4) having $-\theta$, we find that the minimum of the allocations 100,30 is 30. Add this 30 to the cells with $+\theta$ and subtract this 30 to the cells with $-\theta$. Hence the new basic feasible solution is given in the following table.

5	6	9 **70**	0 **30**
3 **70**	5 **5**	10	0 $\in$
6	7	6 **50**	0
6	4 **75**	10	0

We see that the above table satisfies the rim conditions with $(m + n - 1)$ non-negative allocations at independent positions. So we apply MODI method.

5 (3) [2]	6 (5) [1]	9 **70**	0 **30**	$u_1 = 0$
3 **70**	5 **5**	10 (9) [1]	0 $\in$	$u_2 = 0$
6 (0) [6]	7 (2) [5]	6 **50**	0 (−3) [3]	$u_3 = -3$
6 (2) [4]	4 **75**	10 (8) [2]	0 (−1) [1]	$u_4 = -1$
$v_1 = 3$	$v_2 = 5$	$v_3 = 9$	$v_4 = 0$	

Since all $d_{ij} > 0$, the current solution is optimal and unique.

The optimum allocation schedule is given by

$x_{13} = 70$, $x_{14} = 30$, $x_{21} = 70$, $x_{22} = 5$, $x_{24} = \in$, $x_{33} = 50$, $x_{42} = 75$ and the optimum (minimum) transportation cost

$= \text{Rs. } 9 \times 70 + 0 \times 30 + 3 \times 70 + 5 \times 5 + 0 \times \in + 6 \times 50 + 4 \times 75$

$= \text{Rs. } 1465/-$

7.5 Maximization case in Transportation Problems

So far we have discussed the transportation problems in which the objective has been to minimize the total transportation cost and algorithms have been designed accordingly.

If we have a transportation problem where the objective is to maximize the total profit, first we have to convert the maximization problem into a minimization problem by multiplying all the entries by −1 (or) by subtracting all the entries from the highest entry in the given transportation table. The modified minimization problem can be solved in the usual manner.

Example 1 : Solve the following transportation problem to maximize profit

Profits (Rs)/Unit

Destination

		A	B	C	D	Supply
	1	40	25	22	33	100
Source	2	44	35	30	30	30
	3	38	38	28	30	70
Demand		40	20	60	30	

Solution : Since the given problem is of maximization type, first convert this into a minimization problem by subtracting the cost elements (entries or c_{ij}) from the highest cost element ($c_{ij} = 44$) in the given transportation problem. Then the given problem becomes.

Destination

		A	B	C	D	Supply
	1	4	19	22	11	100
Source	2	0	9	14	14	30
	3	6	6	16	14	70
Demand		40	20	60	30	

This modified minimization problem is unbalanced ($\Sigma a_i = 200$, $\Sigma b_j = 150$ and $\Sigma a_i \neq \Sigma b_j$). To make it balanced, we introduce a dummy destination E with demand $(200 - 150) = 50$ units with zero costs c_{ij}. Hence the balanced minimization transportation problem becomes

Destination

		A	B	C	D	E	Supply
	1	4	19	22	11	0	100
Source	2	0	9	14	14	0	30
	3	6	6	16	14	0	70
Demand		40	20	60	30	50	200

Since $\Sigma a_i = \Sigma b_j = 200$, there exists a basic feasible solution to this problem and is displayed in the following table by using VAM. [Try least cost method]

4 10	19	22 60	11 30	0
0 30	9	14	14	0
6	6 20	16	14	0 50

Since the number of non-negative allocations at independent position is 6, which is less than $(m + n - 1) = (3 + 5 - 1) = 7$, this initial solution is degenerate.

To resolve degeneracy, we allocate a very small quantity $\in$ to the cell (3, 3), so that the number of occupied cells becomes $(m + n - 1)$. Hence the initial solution is given by

4 10	19	22 60	11 30	0
0 30	9	14	14	0
6	6 20	16 $\in$	14	0 50

Now the number of non-negative allocations at independent positions is $(m + n - 1)$. We apply MODI method for optimal solution.

4 10	19 12 7	22 60	11 30	0 6 −6	$u_1 = 0$
0 30	9 8 1	14 18 −4	14 7 7	0 2 −2	$u_2 = -4$
6 −2 8	6 20	16 $\in$	14 5 9	0 50	$u_3 = -6$
$v_1 = 4$	$v_2 = 12$	$v_3 = 22$	$v_4 = 11$	$v_5 = 6$	

Since d_{15}, d_{23}, d_{25} are less than zero, the current solution under the test is not optimal. Here $d_{15} = -6$ is the most negative value of d_{ij}.

Let us form a new basic feasible solution by giving maximum allocation to the cell (1,5) by making an occupied cell empty. For this, we draw a closed path consisting of horizontal and vertical lines beginning and ending at this cell (1,5) and having its other corners at some occupied cells. Along this closed loop, indicate $+\theta$ and $-\theta$ alternatively at the corners.

4 10	19	22 60 $-\theta$	11 30	0 $+\theta$
0 30	9	14	14	0
6	6 20	16 $+\theta$ $\in$	14	0 $-\theta$ 50

From the two cells (1,3), (3,5) having $-\theta$, we find that the minimum of 60, 50 is 50. Add this 50 to the cells with $+\theta$ and subtract this 50 to the cells with $-\theta$. Hence the new basic feasible solution is displayed in the following table.

4 10	19	22 10	11 30	0 50
0 30	9	14	14	0
6	6 20	16 50	14	0

We see that the above table satisfies the rim conditions with $(m + n - 1)$ non-negative allocations at independent positions.

Now we apply the MODI method for optimality.

4 **10**	19 12 7	22 **10**	11 **30**	0 **50**	$u_1 = 0$
0 **30**	9 8 1	14 18 −4	14 7 7	0 −4 4	$u_2 = -4$
6 −2 8	6 **20**	16 **50**	14 5 9	0 −6 6	$u_3 = -6$
$v_1 = 4$	$v_2 = 12$	$v_3 = 22$	$v_4 = 11$	$v_5 = 0$	

Since $d_{23} = -4 < 0$ the current solution is not optimal.

Let us form a new basic feasible solution by giving maximum allocation to the cell (2,3) by making an occupied cell empty. For this, we draw a closed loop consisting of horizontal and vertical lines beginning and ending at this cell (2,3) and having its other corner at some occupied cells. Along this closed loop, indicate $+\theta$ and $-\theta$ alternatively at the corners.

10

4 **20** $+\theta$	19	22 **10** $-\theta$	11 **30**	0 **50**
0 **30** $-\theta$	9	14 $+\theta$	14	0
6	6 **20**	16 **50**	14	0

From the two cells (1,3), (2,1) having $-\theta$, we find that the minimum of the allocations 10, 30 is 10. Add this 10 to the cells wih $+\theta$ and subtract this 10 to the cells with $-\theta$. Hence the new basic feasible solution is displayed in the following table.

4 **20**	19	22	11 **30**	0 **50**
0 **20**	9	14 **10**	14	0
6	6 **20**	16 **50**	14	0

Now the number non-negative allocations at independent positions is $(m + n - 1)$. We apply MODI method for the optimality

4 **20**	19 8 11	22 18 4	11 **30**	0 **50**	$u_1 = 0$
0 **20**	9 4 5	14 **10**	14 7 7	0 −4 4	$u_2 = -4$
6 2 4	6 **20**	16 **50**	14 9 5	0 −2 2	$u_3 = -2$
$v_1 = 4$	$v_2 = 8$	$v_3 = 18$	$v_4 = 11$	$v_5 = 0$	

Since all $d_{ij} > 0$, the current solution is optimal and unique.

$\therefore$ The optimum allocation schedule is given by

$x_{11} = 20, x_{14} = 30, x_{15} = 50, x_{21} = 20, x_{23} = 10, x_{32} = 20, x_{33} = 50.$

The optimum profit

= Rs. $40 \times 20 + 33 \times 30 + 0 \times 50 + 44 \times 20 + 30 \times 10 + 38 \times 20 + 28 \times 50$

= Rs. 5130/-.

Example 2: Solve the following transportation problem to maximize profit.

Destination

		A	**B**	**C**	**D**	**Supply**
	1	15	51	42	33	23
Source	**2**	80	42	26	81	44
	3	90	40	66	60	33
Demand		23	31	16	30	100

Solution : Since the given problem is of maximization type, we convert this in to minimization problem by multiplying the profit costs c_{ij} by −1.

Destination

	A	B	C	D	Supply
Source 1	−15	−51	−42	−33	23
Source 2	−80	−42	−26	−81	44
Source 3	−90	−40	−66	−60	33
Demand	23	31	16	30	100

Since $\Sigma a_i = \Sigma b_j = 100$, there exists a basic feasible solution to this problem and is displayed in the following table by using VAM.

−15	−51 **23**	−42	−33
−80 **6**	−42 **8**	−26	−81 **30**
−90 **17**	−40	−66 **16**	−60

Since the number of non-negative allocations at independent positions is $(m + n - 1) = 6$, we apply MODI method for optimal solution.

−15 −89 74	−51 **23**	−42 −65 23	−33 −90 57	$u_1 = -9$
−80 **6**	−42 **8**	−26 −56 30	−81 **30**	$u_2 = 0$
−90 **17**	−40 −52 12	−66 **16**	−60 −91 31	$u_3 = -10$
$v_1 = -80$	$v_2 = -42$	$v_3 = -56$	$v_4 = -81$	

Since all $d_{ij} > 0$, the current solution is optimal and unique.

∴ The optimum allocations are given by $x_{12} = 23$, $x_{21} = 6$, $x_{22} = 8$, $x_{24} = 30$, $x_{31} = 17$, $x_{33} = 16$

∴ The optimum profit

= Rs. 51 × 23 + 80 × 6 + 42 × 8 + 81 × 30 + 90 × 17 + 66 × 16

= Rs. 7005 /-

EXERCISE

1. **What do you mean by transportation model ?**
2. **Define : Feasible solution, Basic feasible solution, Degenerate basic feasible solution, Non-degenerate basic feasible solution and optimal solution of a transportation problem.**
3. **Explain, in brief, with examples**
   (i) **North West Corner rule.** ***(ii)*** **Lowest Cost entry method.** ***(iii)*** **Vogel's approximation method.**
   [MU. MBA. Nov. 96, Apr 95]
4. **What do you mean by balanced and unbalanced transportation problems? Explain how would you convert the unbalanced problem into a balanced one?**
5. **State all the constraints in a transportation problem and how they are different from linear programming problem.**
   [MU. BE. Apr 94]
6. **Write down the dual of a transportation problem. Explain how this helps us in identifying whether the current solution is optimal or not.** ***[MU. MCA. Nov 96]***
7. **Explain an algorithm for solving a transportation problem.**
   [BRU. BE. Apr 95]
8. **Describe the method of solving unbalanced transportation problem.** ***[MU. MCA. Apr 95]***
9. **State the classical transportation problem and write down its mathematical model.**
10. **Give mathematical formulation of a transportation problem.**
    [MU. BE. Nov 93, Nov 89]
11. **How the problem of degeneracy arises in a transportation problem ? Explain how does one over come it.**
    [MU. BE. Apr 92, BRU. BE. Apr 96]
12. **Describe a transportation problem and give a method of finding an initial feasible solution .** ***[MU. BE. Nov 91]***
13. **Obtain the initial (starting) solution for the following transportation problem.**

Destination

	A	B	C	Supply
Source 1	2	7	4	5
Source 2	3	3	1	8
Source 3	5	4	7	7
Source 4	1	6	2	14
Demand	7	9	18	34

(i) **North West Corner rule** ***(ii)*** **Least Cost method**
(iii) **Vogel's approximation method**

14. Solve the following transportation problem

To

From	A	B	C	Availability
I	50	30	220	1
II	90	45	170	3
III	250	200	50	4
Requirement	4	2	2	

[MU. BE. Nov 89]

15. Obtain an optimum basic feasible solution to the transportation problem :

Warehouse

Factory	W_1	W_2	W_3	W_4	Capacity
F_1	19	30	50	10	7
F_2	70	30	40	60	9
F_3	40	8	70	20	18
Requirement	5	8	7	14	

[MU. BE. Apr 90, B. Tech. Leather. Oct 96]

16. Obtain an optimum basic feasible solution to the following transportation problem :

To

From				Available
	7	3	4	2
	2	1	3	3
	3	4	6	5
Demand	4	1	5	10

[MU. BE. Nov 91]

17. Solve the transportation problem :

Destinations

Source	1	2	3	Capacities
1	2	2	3	10
2	4	1	2	15
3	1	3	1	40
Demands	20	15	30	

[MU. BE. Nov 92]

18. Solve the following transportation problem where the cell entries denote the unit transportation costs.

Destination

Origin	A	B	C	D	Available
P	5	4	2	6	20
Q	8	3	5	7	30
R	5	9	4	6	50
Required	10	40	20	30	

[MU. BE. Nov 91]

19. A company has 4 warehouses and 6 stores, the cost of shipping one unit from warehouse i to store j is c_{ij}

$$C_{ij} = \begin{matrix} 7 & 10 & 7 & 4 & 7 & 8 \\ 5 & 1 & 5 & 5 & 3 & 3 \\ 4 & 3 & 7 & 9 & 1 & 9 \\ 4 & 6 & 9 & 0 & 0 & 8 \end{matrix}$$

and the requirements of six stores are 4, 4, 6, 2, 4, 2 and quantities at warehouses are 5, 6, 2, 9, find the minimum cost solution.

[MU. BE. Nov 93]

20. A company has four warehouses *a, b, c, d*. It is required to deliver a product from these warehouses to three customers A, B and C. The warehouses have the following amounts in stock :

Ware house	:	*a*	*b*	*c*	*d*
No.of Units	:	15	16	12	13

and the customer's requirements are :

Customer	:	*A*	*B*	*C*
No. of Units	:	18	20	18

The table below shows the costs of transporting one unit from warehouse to the customer :

Warehouse

Customer	*a*	*b*	*c*	*d*
A	8	9	6	3
B	6	11	5	10
C	3	8	7	9

Find the optimal transportation routes. *[MU. BE. Nov 93]*

21. Explain the concept of degeneracy in transportation problems:

Obtain an optimum basic feasible solution to the following transportation problem

To

From				Supply
	7	3	4	2
	2	1	3	3
	3	4	6	3
Demand	4	1	3	8

[MU. MCA. May 92]

22. Solve the transportation problem with unit transportation costs, demands and supplies as given below :

Destination Centre

		D_1	D_2	D_3	D_4	Supply
	F_1	3	3	4	1	100
Factory	F_2	4	2	4	2	125
	F_3	1	5	3	2	75
Demand		120	80	75	25	

[MU. MBA Nov 95]

23. Find the minimum cost of transportation, given

Warehouses

	D_1	D_2	D_3	D_4	Supply
Factory F_1	19	30	50	10	7
F_2	70	30	40	60	9
F_3	40	8	70	20	18
Demand	5	8	7	14	

[MU. BE. Apr 90, Nov 94]

24. Compute the initial feasible solution to the following transportation problem, given the cost of transportation in rupees.

To

		P	Q	R	Supply
	A	5	1	7	10
From	B	6	4	6	80
	C	3	2	5	15
Demand		75	20	50	

[MU. BE. Oct 95]

25. Solve the transportation problem where the cell entries are transportation costs.

	D_1	D_2	D_3	D_4	Capacity
O_1	1	2	4	4	6
O_2	4	3	2	0	8
O_3	0	2	2	1	10
Required	4	6	8	6	

[MU. BE. Apr 88]

26. **Solve the transportation problem**

Demand point

		1	2	3	4	Supply
	1	2	3	11	7	6
Source	2	1	0	6	1	1
	3	5	8	15	9	10
Demand		7	5	3	2	

[MU. BE. Apr 89]

27. An automobile dealer is faced with the problem of determining the minimum cost policy for supplying dealers with the desired number of automobiles. The relevant data are given below. Obtain the minimum total cost of transportation

Dealers

		1	2	3	4	5	Supply
	A	1.2	1.7	1.6	1.8	2.4	300
Plant	B	1.8	1.5	2.2	1.2	1.6	400
	C	1.5	1.4	1.2	1.5	1.0	100
Requirement		100	50	300	150	200	

The cost unit is in 100 rupees.

[MU. BE. Apr 89]

28. Food packets have to be air lifted by three aircrafts from an airport and air–dropped to five villages. The quantities that can be carried in one trip by these aircrafts to the village are given below. The total number of trips per day an aircraft can make to the villages are also given. Find the number of trips each aircraft should make to each village so that the total quantity of food transported is maximum.

	V_1	V_2	V_3	V_4	V_5	Trips/day by air crafts
A_1	10	8	6	9	12	50
A_2	5	3	8	4	10	90
A_3	7	9	6	10	4	60
Trips/day to village	100	80	70	40	20	

29. A company produces a small component for an industrial products and distributes it to five whole salers at a fixed delivered price of Rs. 2.50 per unit. Sales forecasts indicate that monthly deliveries will be 3000, 3000, 10,000, 5000 and 4000 units to wholesalers 1,2,3,4, and 5 respectively. The monthly production capacities are 5000, 10000 and 12500 at plants 1,2 and 3 respectively. The direct costs of production of each unit are Rs.1.00, Rs.0.90 and Rs.0.80 at plants 1,2 and 3 respectively. The transportation cost of shipping a unit from a plant to a wholesaler are given below :

	wholesaler 1	2	3	4	5
Plant 1	0.05	0.07	0.10	0.15	0.15
Plant 2	0.08	0.06	0.09	0.12	0.14
Plant 3	0.10	0.09	0.08	0.10	0.15

Find how many components each plant supplies to each wholesaler in order to maximize its profit.

30. A company has three plants A, B, C and three warehouses X, Y, Z. The number of units available at the plants is 60,70, 80 and the demand at X, Y, Z are 50, 80, 80 respectively. The unit costs of transportation is given by the following table.

	X	Y	Z
A	8	7	3
B	3	8	9
C	11	3	5

Find the allocation so that the total transportation cost is minimum. *[MU. BE. Oct 96]*

31. An oil corporation has got three refineries R_1, R_2, R_3 and it has to send petrol to four different depots D_1, D_2, D_3 and D_4. The cost of supplying of one unit of petrol from each refinery to each depot is given below. The requirements of the depot and the available petrol at the refineries are also given. Find the minimum cost of shipping after obtaining the initial solution by Vogel's Approximation method.

	Depot D_1	D_2	D_3	D_4	Available
Refinery R_1	10	12	15	8	130
Refinery R_2	14	11	9	10	150
Refinery R_3	20	5	7	18	170
Required :	90	100	140	120	

[MU. BE. Oct 96]

32. Find the optimum solution to the following transportation problem in which cells contain the transportation cost in rupees.

Table

	W_1	W_2	W_3	W_4	W_5	Available
F_1	7	6	4	5	9	40
F_2	8	5	6	7	8	30
F_3	6	8	9	6	5	20
F_4	5	7	7	8	6	10
Required :	30	30	15	20	5	100 Total

[MU. B. Tech. Leather. Oct 96]

33. The following is a transportation problem relating three warehouses (A, B and C) and four customers (1,2,3 and 4). The capacities at the warehouses and the demands from the customers are shown around the perimeter. Per unit transportation costs are shown in the cells. Cost minimization is the objective.

	1	2	3	4	Cap
A	7	8	11	10	30
B	10	12	5	4	45
C	6	11	10	9	35
Dem	20	28	19	33	–

Find the optimal solution and the total cost of transportation.

[BRU. BE. Apr 94]

34. A fan manufacturing company has its plants at calcutta and Delhi. The company is having warehouses in Nagpur, Patna and Baroda. The following table indicates the maximum capacities of the plants and the demands of the warehouses. The cost in rupees of shipping one unit from the particular plant to the given warehouses are shown on the corner of the cells. Find the least cost shipping assignment.

	Warehouses			
Plants	Nagpur	Patna	Baroda	Capacity
Delhi	3	4	2	600
Calcutta	3	2	5	700
Demand	400	500	400	–

[BRU. BE. Nov 95]

35. The projects X, Y, Z require truck loads of 45,50 and 20 respectively per week. The availabilities in plants A,B,C are 40, 40 and 40 of truck loads respectively per week. The cost of transport per unit of truck load from plant to project is given below :

		Project		
		X	Y	Z
	A	5	20	5
Plant	B	10	30	8
	C	10	20	12

(i) Determine an initial feasible solution by VAM.

(ii) Obtain an optimal solution by MODI method. The objective is to minimize the total cost of transportation.

[BRU. BE. Nov 94, MSU. BE. Apr 97]

36. Production shops P, Q, R can produce 5 new products A, B, C, D, E with their excess production capacity. The unit costs are given below with sale potential and availability of the capacity. It is given that the production shop R cannot produce the fifth product E. Find the optimal production schedule. Start the solution by VAM.

		New Products					
		A	B	C	D	E	Availability
	P	20	19	14	21	16	40
Production shop	Q	15	20	13	19	16	60
	R	18	15	18	20	–	90
Sales potential		30	40	70	40	60	

Find the allocation so that the total transportation cost is minimum. *[BRU. BE. Nov 96]*

37. Solve the transportation problem with unit transportation costs in rupees, requirements and availability as given below :

		Distribution centre				
		D_1	D_2	D_3	D_4	Availability
	F_1	10	15	12	12	200
Factory	F_2	8	10	11	9	150
	F_3	11	12	13	10	120
Requirement		140	120	80	220	

[MU. MBA. Apr. 97]

38. Solve the transportation problem with unit transportation costs in rupees and units of demand and supply as given below:

		Destination				
		D_1	D_2	D_3	D_4	Availability
	S_1	10	13	11	8	65
Source	S_2	9	12	12	10	44
	S_3	13	9	11	9	41
Demand		60	40	55	45	

[MU. MBA. Nov. 97]

39. Solve the transportation problem with unit transportation costs, demands and supplies as given below:

		Destination			
		D_1	D_2	D_3	Supply
	S_1	4	1	7	80
Source	S_2	3	2	2	20
	S_3	5	3	4	50
Demand		60	40	35	

[MU. MBA. Apr. 96]

ANSWERS

13. (i) $x_{11} = 5, x_{21} = 2, x_{22} = 6, x_{32} = 3, x_{33} = 4, x_{41} = 14.$
and the transportation cost = Rs. 102/-
(ii) $x_{11} = 5, x_{21} = 2, x_{22} = 6, x_{32} = 3, x_{33} = 4, x_{41} = 14.$
and the transportation cost = Rs. 102/-
(iii) $x_{11} = 5, x_{21} = 2, x_{22} = 6, x_{32} = 3, x_{33} = 4, x_{41} = 14.$
and the transportation cost = Rs. 102/-

14. $x_{11} = 1, x_{21} = 3, x_{31} = \in, x_{32} = 2,$ and $x_{33} = 2$
and minimum T.P. cost = Rs. 743.

15. $x_{12} = 5, x_{14} = 2, x_{22} = 2, x_{23} = 7, x_{32} = 6$ and $x_{34} = 12.$
and minimum T.P. cost = Rs. 743.

16. $x_{13} = 2, x_{22} = 1, x_{23} = 2, x_{31} = 4, x_{33} = 1$ and the optimum transportation cost is Rs. 33.

17. $x_{11} = 10, x_{22} = 15, x_{31} = 10, x_{33} = 30,$ and the optimum transportation cost is Rs. 75/-.

18. $x_{12} = 10, x_{13} = 10, x_{22} = 30, x_{31} = 10, x_{33} = 10, x_{34} = 30$ and the optimum transportation cost is Rs. 420/-.

19. $x_{13} = 5, x_{22} = 4, x_{26} = 2, x_{31} = 1, x_{32} = \in, x_{33} = 1, x_{41} = 3,$ $x_{44} = 2, x_{45} = 4,$ and the optimal transportation cost = Rs. $(68 + 3\in)$ = Rs. 68/- , as $\in \to 0$.

20. $x_{12} = 5, x_{14} = 13, x_{22} = 8, x_{23} = 12, x_{31} = 15, x_{32} = 3,$
and the optimum transportation cost = Rs. 301/-.

21. $x_{13} = 2, x_{21} = 1, x_{22} = 1, x_{23} = 1, x_{31} = 3$
and the optimum transportation cost = Rs.23/-.

22. $x_{11} = 45, x_{13} = 30, x_{14} = 25, x_{22} = 80, x_{23} = 45, x_{31} = 75,$
and the optimum transportation cost = Rs.695/-.

23. $x_{11} = 5, x_{14} = 2, x_{22} = 2, x_{23} = 7, x_{32} = 6, x_{34} = 12,$
and the optimum transportation cost = Rs.743/-.

24. $x_{12} = 10, x_{21} = 60, x_{22} = 10, x_{23} = 10, x_{31} = 15, x_{43} = 40,$
and the optimum transportation cost = Rs. 515/-

25. $x_{12} = 6, x_{23} = 2, x_{24} = 6, x_{31} = 4, x_{33} = 6$
and the optimum transportation cost = Rs. 28/-

26. $x_{12} = 5, x_{23} = 1, x_{33} = 1, x_{13} = 1, x_{31} = 7, x_{34} = 2,$
and the optimum transportation cost = Rs. 100/-

27. $x_{11} = 100, x_{13} = 200, x_{22} = 50, x_{24} = 150, x_{25} = 200, x_{33} = 100,$
and the optimum transportation cost = Rs. 1,13,500/-

28. $x_{11} = 50, x_{23} = 70, x_{25} = 20, x_{32} = 20, x_{34} = 40, x_{41} = 50,$ $x_{42} = 60$ and the maximum quantity of transportation is 1840 units.

29. $x_{11} = 2500, x_{16} = 2500, x_{21} = 500, x_{22} = 3000, x_{23} = 2500,$ $x_{25} = 4000, x_{33} = 7500, x_{34} = 5000$ units respectively. The total transportation cost = Rs. 23,730/-. Total sale = Rs. 62,500/-. Total production cost = Rs. 23,600/-.
Therefore, the net maximum profit
= Total sale – (Total transportation cost + Total production cost)
= 62,500 – (23,730 + 23,600)
= Rs. 15,170/-

30. $x_{13} = 60, x_{21} = 50, x_{23} = 20, x_{32} = 80$
and the optimum transportation cost is Rs. 750/-

31. $x_{11} = 90, x_{14} = 40, x_{23} = 70, x_{24} = 80, x_{32} = 100, x_{33} = 70$
and the minimum (optimum) shipping cost is Rs. 3640/-

32. $x_{11} = 5, x_{13} = 15, x_{14} = 20, x_{21} = \in, x_{22} = 30, x_{31} = 15, x_{35} = 5,$ $x_{41} = 10$, and the optimum solution is Rs.510/- as $\in \to 0$.

33. $x_{12} = 28, x_{15} = 2, x_{23} = 19, x_{24} = 26, x_{31} = 20, x_{34} = 7, x_{35} = 8,$
and the optimum (minimum) transportation cost is Rs. 606/-.

34. $x_{11} = 200, x_{13} = 400, x_{21} = 200, x_{22} = 500,$
and the least shipping cost is Rs. 3000/-.

35. (i) $x_{11} = 40, x_{21} = 10, x_{22} = 15, x_{23} = 20, x_{32} = 35, x_{34} = 5,$
and the initial transportation cost is Rs.1560/-.
(ii) $x_{11} = 30, x_{12} = 10, x_{21} = 15, x_{23} = 20, x_{24} = 5, x_{32} = 40,$
and the optimum (minimum) transportation cost is Rs.1460/-

36. $x_{13} = 10, x_{15} = 30, x_{23} = 60, x_{31} = 30, x_{32} = 40, x_{34} = 20,$ $x_{44} = 20, x_{45} = 30$ and the optimum production cost is Rs. 2940/-.
Note that an alternative optimum production schedule also exists.

37. $x_{11} = 140, x_{13} = 60, x_{22} = 50, x_{24} = 100, x_{34} = 120, x_{42} = 70,$ $x_{43} = 20$. The minimum transportation cost is Rs. 4720/-.

38. $x_{11} = 16, x_{13} = 4, x_{14} = 45, x_{21} = 44, x_{32} = 40, x_{33} = 1,$ $x_{43} = 50$. The minimum transportation cost = Rs. 1331/-

39. $x_{11} = 40, x_{12} = 40, x_{23} = 20, x_{31} = 20, x_{33} = 15, x_{34} = 15$. The Minimum transportation cost = Rs. 400/-.

Chapter 8

Assignment Problem

8.1 Introduction

The assignment problem is a particular case of the transportation problem in which the objective is to assign a number of tasks (Jobs or origins or sources) to an equal number of facilities (machines or persons or destinations) at a minimum cost (or maximum profit).

Suppose that we have 'n' jobs to be performed on 'm' machines (one Job to one machine) and our objective is to assign the jobs to the machines at the minimum cost (or maximum profit) under the assumption that each machine can perform each job but with varying degree of efficiencies.

The assignment problem can be stated in the form of $m \times n$ matrix (c_{ij}) called a ***Cost matrix*** (or) ***Effectiveness matrix*** where c_{ij} is the cost of assigning i^{th} machine to the j^{th} job.

		Jobs				
		1	2	3	....	n
	1	c_{11}	c_{12}	c_{13}	....	c_{1n}
	2	c_{21}	c_{22}	c_{23}	....	c_{2n}
Machines	3	c_{31}	c_{32}	c_{33}	....	c_{3n}
	:	...	...	...	....	...
	:	...	...	...	....	...
	:	...	...	...	....	...
	m	c_{m1}	c_{m2}	c_{m3}	....	c_{mn}

8.2 Mathematical formulation of an assignment problem.

[MU. BE. Apr 93, Oct 96, MU. MBA. Nov 96]

Consider an assignment problem of assigning n jobs to n machines (one job to one machine). Let c_{ij} be the unit cost of assigning i^{th} machine to the j^{th} job and

$$\text{let } x_{ij} = \begin{cases} 1, \text{ if } j^{th} \text{ job is assigned to } i^{th} \text{ machine} \\ 0, \text{ if } j^{th} \text{ job is not assigned to } i^{th} \text{ machine} \end{cases}$$

The assignment model is then given by the following LPP

$$\text{Minimize } Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}\, x_{ij}$$

subject to the constraints

$$\sum_{i=1}^{n} x_{ij} = 1, \; j = 1,2, \ldots n$$

$$\sum_{j=1}^{n} x_{ij} = 1, \; i = 1,2, \ldots n$$

$$\text{and } x_{ij} = 0 \text{ (or) } 1.$$

8.3 Comparison with Transportation Model

[MU. MBA Nov 95, Apr. 97]

The assignment problem may be considered as a special case of the transportation problem. Consider a transportation problem with 'n' sources and 'n' destinations.

		Destination					
		1	2	3	....	n	*Supply* (a_i)
	1	c_{11}	c_{12}	c_{13}	....	c_{1n}	a_1
	2	c_{21}	c_{22}	c_{23}	....	c_{2n}	a_2
Source	3	c_{31}	c_{32}	c_{33}	....	c_{3n}	a_3
	:	...	...	...	....	...	...
	:	...	...	...	....	...	...
	:	...	...	...	....	...	...
	n	c_{n1}	c_{n2}	c_{n3}	....	c_{nn}	a_n
Demand (b_j)		b_1	b_2	b_3	....	b_n	

We have to find x_{ij} (i, j = 1, 2, 3 ... n) for which the total transportation cost

$$Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}\, x_{ij}$$

is minimized.

subject to the constraints

$$\sum_{j=1}^{n} x_{ij} = a_i, \quad i = 1, 2, \dots n$$

$$\sum_{i=1}^{n} x_{ij} = b_j, \quad j = 1, 2, \dots n$$

$$\sum a_i = \sum b_j, \quad i, j = 1, 2, \dots n$$

$$\text{and } x_{ij} \geq 0, \quad i, j = 1, 2, 3, \dots n$$

Here the 'sources' represent 'facilities' or 'machines' and 'destinations' represent 'jobs'.

Suppose that the supply available at each source is 1 *i.e.,* $a_i = 1$ and the demand required at each destination is 1 *i.e.,* $b_j = 1$.

Let c_{ij} be the unit transportation cost from the i^{th} source to the j^{th} destination. Here it means the cost of assigning the i^{th} machine to the j^{th} job.

Let x_{ij} be the amount to be shipped from i^{th} source to the j^{th} destination. Here it means the assignment of the i^{th} machine to the j^{th} job. We can restrict the value of x_{ij} to be either 0 (or) 1. $x_{ij} = 0$ means that the i^{th} machine does not get the j^{th} job and $x_{ij} = 1$ means that the i^{th} machine gets the j^{th} job.

Since each machine should be assigned to only one job and each job requires only one machine, the total assignment value of the i^{th} machine is 1, (*i.e.,*) $\sum x_{ij} = 1$ and the total assignment value of the j^{th} job is 1, (*i.e.,*) $\sum x_{ij} = 1$.

Hence the assignment problem can be expressed as

$$\text{Minimize Z} = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}\, x_{ij}$$

where c_{ij} is the cost of assigning i^{th} machine to the j^{th} job subject to the constraints.

$$x_{ij} = \begin{cases} 1, \text{ if } i^{\text{th}} \text{ machine is assigned to the } j^{\text{th}} \text{ job} \\ 0, \text{ if } i^{\text{th}} \text{ machine is not assigned to the } j^{\text{th}} \text{ job} \end{cases}$$

$$i.e., x_{ij} = 0 \text{ or } 1 \Rightarrow x_{ij}(x_{ij} - 1) = 0 \Rightarrow x_{ij}^2 = x_{ij}$$

$$\sum_{j=1}^{n} x_{ij} = 1, \; i = 1, 2, \dots n \text{ and}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \; j = 1, 2, \dots n.$$

From this we see that assignment problem represents a transportation problem with all demands and supplies equal to 1.

The units available at each source and units demanded at each destination are equal to 1. It means exactly that there is only one occupied cell in each row and each column of the transportation table. *i.e.,* only 'n' occupied cells in place of the required $n + n - 1 = 2n - 1$ occupied cells. Hence ***an assignment problem is always a degenerate form of a transportation problem.***

But the transportation technique (or) simplex method can not be used to solve the assignment problem because of degeneracy. In fact a very convenient iterative procedure is available for solving an assignment problem.

The technique used for solving assignment problem makes use of the following two theorems.

Theorem 1 : The optimum assignment schedule remains unaltered if we add or subtract a constant from all the elements of the row or column of the assignment cost matrix.

Theorem 2 : If for an assignment problem all $c_{ij} > 0$, then an assignment schedule (x_{ij}) which satisfies $\sum c_{ij}\, x_{ij} = 0$, must be optimal.

8.4 Difference between the transportation problem and the assignment problem.

[MU. BE. Nov 94]

	Transportation Problem	*Assignment Problem*
(a)	Supply at any source may be any positive quantity a_i	Supply at any source (machine) will be 1 *i.e.,* $a_i = 1$.
(b)	Demand at any destination may be any positive quantity b_j	Demand at any destination (job) will be 1. *i.e.,* $b_j = 1$
(c)	One or more source to any number of destinations	One source (machine) to only one destination (job).

8.5 Assignment Algorithm (or) Hungarian Method.

[MU. MBA. Apr 95, BRU. BE. Nov 96, MU. BE. Apr 95, MU. MCA. Nov 95, Nov 98]

First check whether the number of rows is equal to the number of columns, If it is so, the assignment problem is said to be **balanced**. Then proceed to step 1. If it is not balanced, then it should be balanced before applying the algorithm. The method of balancing is discussed in sec 5.6 page 5.15.

Step 1 : Subtract the smallest cost element of each row from all the elements in the row of the given cost matrix. See that each row contains atleast one zero.

Step 2 : Subtract the smallest cost element of each column from all the elements in the column of the resulting cost matrix obtained by step 1.

Step 3 : (Assigning the zeros)

(a) Examine the rows successively until a row with exactly one unmarked zero is found. Make an assignment to this single unmarked zero by encircling it. Cross all other zeros in the column of this encircled zero, as these will not be considered for any future assignment. Continue in this way until all the rows have been examined.

(b) Examine the columns successively until a column with exactly one unmarked zero is found. Make an assignment to this single unmarked zero by encircling it and cross any other zero in its row. Continue until all the columns have been examined.

Step 4 : (Apply optimal Test)

(a) If each row and each column contain exactly one encircled zero, then the current assignment is optimal.

(b) If atleast one row/column is without an assignment (i.e., if there is atleast one row/column is without one encircled zero), then the current assignment is not optimal. Go to step 5.

Step 5 : Cover all the zeros by drawing a minimum number of straight lines as follows.

(a) Mark (✓) the rows that do not have assignment.

(b) Mark (✓) the columns (not already marked) that have zeros in marked rows.

(c) Mark (✓) the rows (not already marked) that have assignments in marked columns.

(d) Repeat *(b)* and *(c)* until no more marking is required.

(e) Draw lines through all unmarked rows and marked columns. If the number of these lines is equal to the order of the matrix then it is an optimum solution otherwise not.

Step 6 : Determine the smallest cost element not covered by the straight lines. Subtract this smallest cost element from all the uncovered elements and add this to all those elements which are lying in the intersection of these straight lines and do not change the remaining elements which lie on the straight lines.

Step 7 : Repeat steps (1) to (6), until an optimum assignment is attained.

Note 1 : In case some rows or columns contain more than one zero, encircle any unmarked zero arbitrarily and cross all other zeros in its column or row. Proceed in this way until no zero is left unmarked or encircled.

Note 2 : The above assignment algorithm is only for minimization problems.

Note 3 : If the given assignment problem is of maximization type, convert it to a minimization assignment problem by max Z = – min (–Z) and multiply all the given cost elements by –1 in the cost matrix and then solve by assignment algorithm.

Note 4 : Some times, a final cost matrix contains more than required number of zeros at independent positions. This implies that there is more than one optimal solution (multiple optimal solutions) with the same optimum assignment cost.

Example 1 : Consider the problem of assigning five jobs to five persons. The assignment costs are given as follows :

$$\begin{array}{cc} & \textbf{Job} \\ & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \textbf{Person} \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{pmatrix} 8 & 4 & 2 & 6 & 1 \\ 0 & 9 & 5 & 5 & 4 \\ 3 & 8 & 9 & 2 & 6 \\ 4 & 3 & 1 & 0 & 3 \\ 9 & 5 & 8 & 9 & 5 \end{pmatrix} \end{array}$$

Determine the optimum assignment schedule.

[MU. BE. Apr 90, Apr 91]

Solution : The cost matrix of the given assignment problem is

$$\begin{pmatrix} 8 & 4 & 2 & 6 & 1 \\ 0 & 9 & 5 & 5 & 4 \\ 3 & 8 & 9 & 2 & 6 \\ 4 & 3 & 1 & 0 & 3 \\ 9 & 5 & 8 & 9 & 5 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 13 & 49 & 2 & 10 \\ 0 & 35 & 29 & 7 & 20 \\ 13 & 0 & 63 & 9 & 17 \\ 47 & 15 & 0 & 22 & 12 \\ 25 & 0 & 46 & 11 & 14 \end{pmatrix}$$

Step 2 : Select the smallest cost element in each column and subtract this from all the elements of the corresponding column, we get the following reduced matrix

$$\begin{pmatrix} \cancel{0} & 13 & 49 & (0) & \cancel{0} \\ (0) & 35 & 29 & 5 & 10 \\ 13 & (0) & 63 & 7 & 7 \\ 47 & 15 & (0) & 20 & 2 \\ 25 & \cancel{0} & 46 & 9 & 4 \end{pmatrix}$$

Step 3 : Now we shall examine the rows successively. Second row contains a single unmarked zero, encircle this zero and cross all other zeros in its column. Third row contains a single unmarked zero, encircle this zero and cross all other zeros in its column. Fourth row contains a single unmarked zero, encircle this zero and cross all other zeros in its column. After this no row is with exactly one unmarked zero. So go for columns.

Examine the columns successively. Fourth column contains exactly one unmarked zero, encircle this zero and cross all other zeros in its row. After examining all the rows and columns, we get

$$\begin{array}{c|c} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} R1 \\ R2 \\ R3 \\ R4 \\ R5 \end{matrix} & \begin{pmatrix} \cancel{0} & 13 & 49 & (0) & \cancel{0} \\ (0) & 35 & 29 & 5 & 10 \\ 13 & (0) & 63 & 7 & 7 \\ 47 & 15 & (0) & 20 & 2 \\ 25 & \cancel{0} & 46 & 9 & 4 \end{pmatrix} \end{array}$$

2 3 4

Step 4 : Since the 5^{th} row and 5^{th} column do not have any assignment, the current assignment is not optimal.

Step 5 : Cover all the zeros by drawing a minimum number of straight lines as follows :

(a) Mark (✓) the rows that do not have assignment. The row 5 is marked.

(b) Mark (✓) the columns (not already marked) that have zeros in marked rows. Thus column 2 is marked.

(c) Mark (✓) the rows (not already marked) that have assignments in marked columns. Thus row 3 is marked.

(d) Repeat (b) and (c) until no more marking is required. In the present case this repetition is not necessary.

(e) Draw lines through all unmarked rows (rows 1, 2 and 4). and marked columns (column 2). We get

Step

$$\begin{pmatrix} 0 & 13 & 49 & 0 & 0 \\ 0 & 35 & 29 & 5 & 10 \\ 13 & (0) & 63 & 7 & 7 \\ 47 & 15 & 0 & 20 & 2 \\ 25 & 0 & 46 & 9 & (4) \end{pmatrix}$$

Min ✓

Sub ✓

Add intersc

Step 6 : Here 4 is the smallest element not covered by these straight lines. Subtract this 4 from all the uncovered elements and add this 4 to all those elements which are lying in the intersection of these straight lines and do not change the remaining elements which lie on these straight lines, we get the following matrix.

$$\begin{pmatrix} 0 & 17 & 49 & 0 & 0 \\ 0 & 39 & 29 & 5 & 10 \\ 9 & 0 & 59 & 3 & 3 \\ 47 & 19 & 0 & 20 & 2 \\ 21 & 0 & 42 & 5 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we examine the rows and columns successively, *i.e.*, repeat step 3 above, we get

$$\begin{array}{c|c} & \begin{matrix} M1 & M2 & M3 & M4 & M5 \end{matrix} \\ \begin{matrix} J1 \\ J2 \\ J3 \\ J4 \\ J5 \end{matrix} & \begin{pmatrix} \cancel{0} & 17 & 49 & (0) & \cancel{0} \\ (0) & 39 & 29 & 5 & 10 \\ 9 & (0) & 59 & 3 & 3 \\ 47 & 19 & (0) & 20 & 2 \\ 21 & \cancel{0} & 42 & 5 & (0) \end{pmatrix} \end{array}$$

20

Since the number of rows is equal to the number of columns in the cost matrix, the given assignment problem is balanced.

Step 1: Select the smallest cost element in each row and subtract this from all the elements of the corresponding row, we get the reduced matrix

$$\begin{pmatrix} 7 & 3 & 1 & 5 & 0 \\ 0 & 9 & 5 & 5 & 4 \\ 1 & 6 & 7 & 0 & 4 \\ 4 & 3 & 1 & 0 & 3 \\ 4 & 0 & 3 & 4 & 0 \end{pmatrix}$$

Step 2 : Select the smallest cost element in each column and subtract this from all the elements of the corresponding column, we get the reduced matrix.

$$\begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \begin{pmatrix} 7 & 3 & 0 & 5 & (0) \\ (0) & 9 & 4 & 5 & 4 \\ 1 & 6 & 6 & (0) & 4 \\ 4 & 3 & (0) & \cancel{0} & 3 \\ 4 & (0) & 2 & 4 & \cancel{0} \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make assignments in the reduced matrix.

Step 3 : Examine the rows successively until a row with exactly one unmarked zero is found. Since the 2nd row contains a single zero, encircle this zero and cross all other zeros of its column. The 3rd row contains exactly one unmarked zero, so encircle this zero and cross all other zeros in its column. The 4th row contains exactly one unmarked zero, so encircle this zero and cross all other zeros in its column. The 1st row contains exactly one unmarked zero, so encircle this zero and cross all other zeros in its column. Finally the last row contains exactly one unmarked zero, so encircle this zero and cross all other zeros in its column. Like wise examine the columns successively. The assignments in rows and columns in the reduced matrix is given by

$$\begin{pmatrix} 7 & 3 & \cancel{0} & 5 & (0) \\ (0) & 9 & 4 & 5 & 4 \\ 1 & 6 & 6 & (0) & 4 \\ 4 & 3 & (0) & \cancel{0} & 3 \\ 4 & (0) & 2 & 4 & \cancel{0} \end{pmatrix}$$

Step 4 : Since each row and each column contains exactly one assignment (*i.e.*, exactly one encircled zero) the current assignment is optimal.

∴ The optimum assignment schedule is given by A → 5, B → 1, C → 4, D → 3, E → 2.

The optimum (minimum) assignment cost = (1 + 0 + 2 + 1 + 5) cost units = 9 units of cost.

$\boxed{\textbf{Example 2}}$ **: The processing time in hours for the jobs when allocated to the different machines are indicated below. Assign the machines for the jobs so that the total processing time is minimum**

		Machines				
		M_1	M_2	M_3	M_4	M_5
	J_1	9	22	58	11	19
	J_2	43	78	72	50	63
Jobs	J_3	41	28	91	37	45
	J_4	74	42	27	49	39
	J_5	36	11	57	22	25

[MU. MBA. Nov 95]

Solution : The cost matrix of the given problem is

9	22	58	11	19
43	78	72	50	63
41	28	91	37	45
74	42	27	49	39
36	11	57	22	25

Since the number of rows is equal to the number of columns in the cost matrix, the given assignment problem is balanced.

Step 1 : Select the smallest cost element in each row and subtract this from all the elements of the corresponding row, we get the reduced matrix.

In the above matrix, each row and each column contains exactly one assignment (*i.e.*, exactly one encircled zero), therefore the current assignment is optimal.

$\therefore$ The optimum assignment schedule is $J_1 \rightarrow M_4$, $J_2 \rightarrow M_1$, $J_3 \rightarrow M_2$, $J_4 \rightarrow M_3$, $J_5 \rightarrow M_5$ and the optimum (minimum) processing time

$= 11 + 43 + 28 + 27 + 25$ hours

$= 134$ hours.

$\boxed{\text{Example 3}}$: Four different jobs can be done on four different machines. The set up and take down time costs are assumed to be prohibitively high for change overs. The matrix below gives the cost in rupees of processing job *i* on machine *j*.

		Machines			
		M_1	M_2	M_3	M_4
	J_1	5	7	11	6
Jobs	J_2	8	5	9	6
	J_3	4	7	10	7
	J_4	10	4	8	3

How should the jobs be assigned to the various machines so that the total cost is minimized? ***[MU. BE. Nov 92]***

Solution : The assignment problem is given by the cost matrix

$$\begin{pmatrix} 5 & 7 & 11 & 6 \\ 8 & 5 & 9 & 6 \\ 4 & 7 & 10 & 7 \\ 10 & 4 & 8 & 3 \end{pmatrix}$$

Since the number of rows is equal to the number of columns in the cost matrix, the given assignment problem is balanced.

Select the smallest cost element in each row and subtract this from all the elements of the corresponding row, we get the reduced matrix

$$\begin{pmatrix} 0 & 2 & 6 & 1 \\ 3 & 0 & 4 & 1 \\ 0 & 3 & 6 & 3 \\ 7 & 1 & 5 & 0 \end{pmatrix}$$

Select the smallest cost element in each column and subtract this from all the elements of the corresponding column, we get the following reduced matrix

$$\begin{pmatrix} 0 & 2 & 2 & 1 \\ 3 & 0 & 0 & 1 \\ 0 & 3 & 2 & 3 \\ 7 & 1 & 1 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make the assignment in rows and columns of this reduced matrix.

Examine the rows successively until a row with exactly one unmarked zero is found. The first row contains exactly one unmarked zero, encircle this zero and cross all other zeros of its column. The fourth row contains exactly one unmarked zero, encircle this zero and cross all other zeros in its column. The 2nd column contains exactly one unmarked zero, encircle this zero and cross all other zeros in its row. We get

$$\begin{pmatrix} (0) & 2 & 2 & 1 \\ 3 & (0) & \cancel{0} & 1 \\ \cancel{0} & 3 & 2 & 3 \\ 7 & 1 & 1 & (0) \end{pmatrix}$$

Since there are some rows and columns without assignment (*i.e.*, without encircled zero), the current assignment is not optimal.

Cover all the zeros by drawing a minimum number of straight lines.

We get

$$\begin{pmatrix} 0 & 2 & 2 & (1) \\ 3 & 0 & 0 & 1 \\ 0 & 3 & 2 & 3 \\ 7 & 1 & 1 & 0 \end{pmatrix}$$

Here 1 is the smallest cost element not covered by these straight lines. Add this 1 to those elements which lie in the intersection of these straight lines, subtract this 1 from all the uncovered elements and do not change the remaining elements which lie on the straight lines we get

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 4 & 0 & 0 & 1 \\ 0 & 2 & 1 & 2 \\ 8 & 1 & 1 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make assignment in the rows and columns of this reduced matrix. We get

$$\begin{pmatrix} \cancel{0} & 1 & 1 & \cancel{0} \\ 4 & (0) & \cancel{0} & 1 \\ (0) & 2 & 1 & 2 \\ 8 & 1 & 1 & (0) \end{pmatrix}$$

Since there are some rows and columns without assignment, the current assignment is not optimal.

Cover all the zeros by drawing a minimum number of straight lines. We get

$$\begin{array}{c} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 4 & 0 & 0 & 1 \\ 0 & 2 & 1 & 2 \\ 8 & (1) & 1 & 0 \end{pmatrix} \begin{array}{c} \checkmark \\ \\ \checkmark \\ \checkmark \end{array} \\ \checkmark \qquad\qquad\qquad \checkmark \end{array}$$

Here 1 is the smallest cost element not covered by these straight lines. Subtract this 1 from all the uncovered elements, add this 1 to those elements which lie in the intersection of these straight lines and do not change the remaining elements which lie on these straight lines. We get

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 \\ 8 & 0 & 0 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make the assignment in the rows and columns of this reduced matrix. We get

$$\begin{pmatrix} (0) & \cancel{0} & \cancel{0} & \cancel{0} \\ 5 & (0) & \cancel{0} & 2 \\ \cancel{0} & 1 & (0) & 2 \\ 8 & \cancel{0} & \cancel{0} & (0) \end{pmatrix}$$

Since each row and each column contains exactly one assignment (*i.e.*, exactly one encircled zero) the current assignment is optimal.

$\therefore$ The optimum assignment schedule is given by $J_1 \to M_1$, $J_2 \to M_2$, $J_3 \to M_3$, $J_4 \to M_4$ and the optimum (minimum) assignment cost

$= \text{Rs. } (5 + 5 + 10 + 3) = \text{Rs. } 23$

Example 4 : The assignment cost of assigning any one operator to any one machine is given in the following table

		Operators			
		I	II	III	IV
Machine	A	10	$\boxed{5}$	13	15
	B	$\boxed{3}$	9	18	3
	C	10	7	$\boxed{3}$	2
	D	5	11	9	$\boxed{7}$

Find the optimal assignment by Hungarian method.

[BNU. BE. Nov 96]

Solution : The cost matrix of the given assignment problem is

$$\begin{pmatrix} 10 & 5 & 13 & 15 \\ 3 & 9 & 18 & 3 \\ 10 & 7 & 3 & 2 \\ 5 & 11 & 9 & 7 \end{pmatrix}$$

Since the number of rows is equal to the number of columns in the cost matrix, the given assignment problem is balanced.

Select the smallest cost element in each row and subtract this from all the elements of the corresponding row, we get the reduced matrix.

$$\begin{pmatrix} 5 & 0 & 8 & 10 \\ 0 & 6 & 15 & 0 \\ 8 & 5 & 1 & 0 \\ 0 & 6 & 4 & 2 \end{pmatrix}$$

Select the smallest cost element in each column and subtract this from all the elements of the corresponding column, we get the reduced matrix

$$\begin{pmatrix} 5 & 0 & 7 & 10 \\ 0 & 6 & 14 & 0 \\ 8 & 5 & 0 & 0 \\ 0 & 6 & 3 & 2 \end{pmatrix}$$

Since each row and each column contain atleast one zero, we shall make the assignment in rows and columns of this reduced cost matrix.

$$\begin{pmatrix} 5 & \boxed{(0)} & 7 & 10 \\ \cancel{0} & 6 & 14 & \boxed{(0)} \\ 8 & 5 & \boxed{(0)} & \cancel{0} \\ \boxed{(0)} & 6 & 3 & 2 \end{pmatrix}$$

4 = 4//

Since each row and each column contain exactly one assignment (i.e., exactly one encircled zero), the current assignment is optimal.

The optimum assignment schedule is

A → II, B → IV, C → III, D → I

A – II D – A

and the optimum (minimum) assignment cost

= Rs. (5 + 3 + 3 + 5) = Rs. 16/-

B – IV

C – III

8.6 Unbalanced Assignment Models

If the number of rows is not equal to the number columns in the cost matrix of the given assignment problem, then the given assignment problem is said to be unbalanced.

First convert the unbalanced assignment problem in to a balanced one by adding dummy rows or dummy columns with zero cost elements in the cost matrix depending upon whether $m < n$ or $m > n$ and then solve by the usual method.

Example 1 : A company has four machines to do three jobs. Each job can be assigned to one and only one machine. The cost of each job on each machine is given in the following table.

		Machines			
		1	**2**	**3**	**4**
	A	**18**	**24**	**28**	**32**
Jobs	**B**	**8**	**13**	**17**	**19**
	C	**10**	**15**	**19**	**22**

What are job assignments which will minimize the cost?

[MU. BE. 1977]

Solution : The cost matrix of the given assignment problem is

$$\begin{pmatrix} 18 & 24 & 28 & 32 \\ 8 & 13 & 17 & 19 \\ 10 & 15 & 19 & 22 \end{pmatrix}$$

Since the number of rows is less than the number of columns in the cost matrix, the given assignment problem is unbalanced.

To make it a balanced one, add a dummy job D (row) with zero cost elements. The balanced cost matrix is given by

$$\begin{pmatrix} 18 & 24 & 28 & 32 \\ 8 & 13 & 17 & 19 \\ 10 & 15 & 19 & 22 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Now select the smallest cost element in each row (column) and subtract this from all the elements of the corresponding row (column), we get the reduced matrix

$$\begin{pmatrix} 0 & 6 & 10 & 14 \\ 0 & 5 & 9 & 11 \\ 0 & 5 & 9 & 12 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

In this reduced matrix, we shall make the assignment in rows and columns having single zero. We have

$$\begin{pmatrix} (0) & 6 & 10 & 14 \\ \cancel{0} & 5 & 9 & 11 \\ \cancel{0} & 5 & 9 & 12 \\ \cancel{0} & (0) & \cancel{0} & \cancel{0} \end{pmatrix}$$

Since there are some rows and columns without assignment, the current assignment is not optimal.

Cover the all zeros by drawing a minimum number of straight lines. Choose the smallest cost element not covered by these straight lines.

$$\begin{array}{cc} \begin{pmatrix} 0 & 6 & 10 & 14 \\ 0 & 5 & 9 & 11 \\ 0 & (5) & 9 & 12 \\ 0 & 0 & 0 & 0 \end{pmatrix} & \begin{matrix} \checkmark \\ \checkmark \\ \checkmark \\ \\ \end{matrix} \\ \begin{matrix} \checkmark & & & \end{matrix} & \end{array}$$

Here 5 is the smallest cost element not covered by these straight lines. Subtract this 5 from all the uncovered elements, add this 5 to those elements which lie in the intersection of these straight lines and do not change the remaining elements which lie on the straight lines. We get

$$\begin{pmatrix} 0 & 1 & 5 & 9 \\ 0 & 0 & 4 & 6 \\ 0 & 0 & 4 & 7 \\ 5 & 0 & 0 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make assignment in the rows and columns having single zero. We get

$$\begin{pmatrix} (0) & 1 & 5 & 9 \\ \cancel{0} & (0) & 4 & 6 \\ \cancel{0} & \cancel{0} & 4 & 7 \\ 5 & \cancel{0} & (0) & \cancel{0} \end{pmatrix}$$

Since there are some rows and columns without assignment, the current assignment is not optimal.

Cover all the zeros by drawing a minimum number of straight lines.

$$\begin{array}{c} \begin{pmatrix} 0 & 1 & 5 & 9 \\ 0 & 0 & 4 & 6 \\ 0 & 0 & (4) & 7 \\ 5 & 0 & 0 & 0 \end{pmatrix} \begin{matrix} \checkmark \\ \checkmark \\ \checkmark \\ \end{matrix} \\ \checkmark \quad \checkmark \qquad\qquad\qquad \end{array}$$

Choose the smallest cost element not covered by these straight lines, subtract this from all the uncovered elements, add this to those elements which are in the intersection of the lines and do not change the remaining elements which lie on these straight lines. Thus we get

$$\begin{pmatrix} 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 3 \\ 9 & 4 & 0 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make the assignment in the rows and columns having single zero. We get

$$\begin{pmatrix} (0) & 1 & 1 & 5 \\ \cancel{0} & (0) & \cancel{0} & 2 \\ \cancel{0} & \cancel{0} & (0) & 3 \\ 9 & 4 & \cancel{0} & (0) \end{pmatrix}$$

Since each row and each column contains exactly one assignment (*i.e.*, exactly one encircled zero) the current assignment is optimal.

$\therefore$ The optimum assignment schedule is given by $A \rightarrow 1$, $B \rightarrow 2$, $C \rightarrow 3$, $D \rightarrow 4$ and the optimum (minimum) assignment cost

$= (18 + 13 + 19 + 0)$ cost units $=$ 50/- units of cost

Note 1 : For this problem, the alternative optimum schedule is $A \rightarrow 1$, $B \rightarrow 3$, $C \rightarrow 2$, $D \rightarrow 4$, with the same optimum assignment cost = Rs. $(18 + 17 + 15 + 0)$ = 50/- units of cost.

Note 2 : Here the assignment $D \rightarrow 4$ means that the dummy Job D is assigned to the 4th Machine. It means that machine 4 is left without any assignment.

Example 2 : Assign four trucks 1,2,3 and 4 to vacant spaces A, B, C, D, E and F so that the distance travelled is minimized. The matrix below shows the distance.

	1	2	3	4
A	4	7	3	7
B	8	2	5	5
C	4	9	6	9
D	7	5	4	8
E	6	3	5	4
F	6	8	7	3

Solution : The matrix of the assignment problem is

$$\begin{pmatrix} 4 & 7 & 3 & 7 \\ 8 & 2 & 5 & 5 \\ 4 & 9 & 6 & 9 \\ 7 & 5 & 4 & 8 \\ 6 & 3 & 5 & 4 \\ 6 & 8 & 7 & 3 \end{pmatrix}$$

Since the number of rows is more than the number of columns, the given assignment problem is unbalanced. To make it balanced, let us introduce two dummy trucks (columns) with zero costs. We get

$$\begin{pmatrix} 4 & 7 & 3 & 7 & 0 & 0 \\ 8 & 2 & 5 & 5 & 0 & 0 \\ 4 & 9 & 6 & 9 & 0 & 0 \\ 7 & 5 & 4 & 8 & 0 & 0 \\ 6 & 3 & 5 & 4 & 0 & 0 \\ 6 & 8 & 7 & 3 & 0 & 0 \end{pmatrix}$$

Select the smallest cost element in each row (column) and subtract this from all the elements of the corresponding row (column). We get

$$\begin{pmatrix} 0 & 5 & 0 & 4 & 0 & 0 \\ 4 & 0 & 2 & 2 & 0 & 0 \\ 0 & 7 & 3 & 6 & 0 & 0 \\ 3 & 3 & 1 & 5 & 0 & 0 \\ 2 & 1 & 2 & 1 & 0 & 0 \\ 2 & 6 & 4 & 0 & 0 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we make the assignment in rows and columns having single zero. We get

$$\begin{pmatrix} \cancel{0} & 5 & (0) & 4 & \cancel{0} & \cancel{0} \\ 4 & (0) & 2 & 2 & \cancel{0} & \cancel{0} \\ (0) & 7 & 3 & 6 & \cancel{0} & \cancel{0} \\ 3 & 3 & 1 & 5 & (0) & \cancel{0} \\ 2 & 1 & 2 & 1 & \cancel{0} & (0) \\ 2 & 6 & 4 & (0) & \cancel{0} & \cancel{0} \end{pmatrix}$$

Since each row and each column contains exactly one assignment (*i.e.*, exactly one encircled zero), the current assignment is optimal.

$\therefore$ The optimum assignment schedule is given by A $\to$ 3, B $\to$ 2, C $\to$ 1, D $\to$ 5, E $\to$ 6, F $\to$ 4, and the optimum (minimum) distance.

$$= (3 + 2 + 4 + 0 + 0 + 3)$$

units of distance $=$ 12/- units of distance.

Example 3 : A batch of 4 jobs can be assigned to 5 different machines. The set up time (in hours) for each job on various machines is given below :

		Machine				
		1	**2**	**3**	**4**	**5**
	1	**10**	**11**	**4**	**2**	**8**
	2	**7**	**11**	**10**	**14**	**12**
Job	**3**	**5**	**6**	**9**	**12**	**14**
	4	**13**	**15**	**11**	**10**	**7**

Find an optimal assignment of jobs to machines which will minimize the total set up time.

[BRU. BE. Nov 96, BNU. BE. Nov 96, MSU. BE. Nov 97]

Solution : The matrix of the given assignment problem is

$$\begin{pmatrix} 10 & 11 & 4 & 2 & 8 \\ 7 & 11 & 10 & 14 & 12 \\ 5 & 6 & 9 & 12 & 14 \\ 13 & 15 & 11 & 10 & 7 \end{pmatrix}$$

Since the number of rows is less than the number of columns in the cost matrix, the given assignment problem is unbalanced.

To make it a balanced one, add a dummy job 5 (row) with zero cost elements. The balanced cost matrix is given by

$$\begin{pmatrix} 10 & 11 & 4 & 2 & 8 \\ 7 & 11 & 10 & 14 & 12 \\ 5 & 6 & 9 & 12 & 14 \\ 13 & 15 & 11 & 10 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Now select the smallest cost element in each row (column) and subtract this from all the elements of the corresponding row (column), we get the reduced cost matrix.

$$\begin{pmatrix} 8 & 9 & 2 & 0 & 6 \\ 0 & 4 & 3 & 7 & 5 \\ 0 & 1 & 4 & 7 & 9 \\ 6 & 8 & 4 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Since each row and each columns contains atleast one zero, we shall make the assignment in rows and columns of this reduced cost matrix

$$\begin{pmatrix} 8 & 9 & 2 & (0) & 6 \\ (0) & 4 & 3 & 7 & 5 \\ \cancel{0} & 1 & 4 & 7 & 9 \\ 6 & 8 & 4 & 3 & (0) \\ \cancel{0} & (0) & \cancel{0} & \cancel{0} & \cancel{0} \end{pmatrix}$$

Since there are some rows and columns with out assignment, the current assignment is not optimal.

Cover all the zeros by drawing a minimum number of straight lines.

$$\begin{array}{cc} \begin{pmatrix} 8 & 9 & 2 & 0 & 6 \\ 0 & 4 & 3 & 7 & 5 \\ 0 & (1) & 4 & 7 & 9 \\ 6 & 8 & 4 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} \\ \checkmark \\ \checkmark \\ \\ \\ \end{array} \\ \begin{array}{ccccc} \checkmark & & & & \end{array} & \end{array}$$

Here 1 is the smallest cost element not covered by these straight lines. Subtract this 1 from all the uncovered elements, add this 1 to those elements which lie in the intersection of these straight lines and do not change the remaining elements which lie on the straight lines. We get

$$\begin{pmatrix} 9 & 9 & 2 & 0 & 6 \\ 0 & 3 & 2 & 6 & 4 \\ 0 & 0 & 3 & 6 & 8 \\ 7 & 8 & 4 & 3 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make the assignment in rows and columns of this reduced cost matrix

$$\begin{pmatrix} 9 & 9 & 2 & (0) & 6 \\ (0) & 3 & 2 & 6 & 4 \\ \cancel{0} & (0) & 3 & 6 & 8 \\ 7 & 8 & 4 & 3 & (0) \\ 1 & \cancel{0} & (0) & \cancel{0} & \cancel{0} \end{pmatrix}$$

Since each row and each column contains exactly one assignment (i.e., exactly one encircled zero), the current assignment is optimal.

$\therefore$ The optimum assignment schedule is given by Job 1 $\rightarrow$ M/c 4, Job 2 $\rightarrow$ M/c 1, Job 3 $\rightarrow$ M/c 2, Job 4 $\rightarrow$ M/c 5. M/c 3 is left without any assignment.

The optimum (minimum) total set up time

$$= 2 + 7 + 6 + 7 \text{ hours}$$
$$= 22 \text{ hours.}$$

8.7 Maximization case in Assignment Problems

In an assignment problem, we may have to deal with maximization of an objective function. For example, we may have to assign persons to jobs in such a way that the total profit is maximized. The maximization problem has to be converted into an equivalent minimization problem and then solved by the usual Hungarian Method.

The conversion of the maximization problem into an equivalent minimization problem can be done by any one of the following methods :

(i) Since max Z = – min (–Z), multiply all the cost elements c_{ij} of the cost matrix by –1.

(ii) Subtract all the cost elements c_{ij} of the cost matrix from the highest cost element in that cost matrix.

Example 1 : A company has a team of four salesmen and there are four districts where the company wants to start its business. After taking into account the capabilities of salesman and the nature of districts, the company estimates that the profit per day in rupees for each salesman in each district is as below :

		Districts			
		1	**2**	**3**	**4**
Salesmen	**A**	16	10	14	11
	B	14	11	15	15
	C	15	15	13	12
	D	13	12	14	15

Find the assignment of salesmen to various districts which will yield maximum profit. *[MU. BE. Nov 93]*

Solution : The cost matrix of the given assignment problem is

$$\begin{pmatrix} (16) & 10 & 14 & 11 \\ 14 & 11 & 15 & 15 \\ 15 & 15 & 13 & 12 \\ 13 & 12 & 14 & 15 \end{pmatrix}$$

Since this is a maximization problem, it can be converted into an equivalent minimization problem by subtracting all the cost elements in the cost matrix from the highest cost element 16 of this cost matrix. Thus the cost matrix of the equivalent minimization problem is

$$\begin{pmatrix} 0 & 6 & 2 & 5 \\ 2 & 5 & 1 & 1 \\ 1 & 1 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}$$

Select the smallest cost element in each row (column) and subtract this from all the cost elements of the corresponding row (column). We get the reduced cost matrix

$$\begin{pmatrix} 0 & 6 & 2 & 5 \\ 1 & 4 & 0 & 0 \\ 0 & 0 & 2 & 3 \\ 2 & 3 & 1 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make the assignment in rows and columns having single zero. We get

$$\begin{pmatrix} (0) & 6 & 2 & 5 \\ 1 & 4 & (0) & \cancel{0} \\ \cancel{0} & (0) & 2 & 3 \\ 2 & 3 & 1 & (0) \end{pmatrix}$$

Since each row and each column contains exactly one encircled zero, the current assignment is optimal.

$\therefore$ The optimum assignment schedule is given by A $\rightarrow$ 1, B $\rightarrow$ 3, C $\rightarrow$ 2, D $\rightarrow$ 4 and the optimum (maximum) profit

= Rs. (16 + 15 + 15 + 15)

= Rs. 61/-.

Example 2 : Solve the assignment problem for maximization given the profit matrix (profit in rupees).

Machines

		P	Q	R	S
Job	**A**	**51**	**53**	**54**	**50**
	B	**47**	**50**	**48**	**50**
	C	**49**	**50**	**60**	**61**
	D	**63**	**64**	**60**	**60**

[MU. BE. Apr 95]

Solution : The profit matrix of the given assignment problem is

$$\begin{pmatrix} 51 & 53 & 54 & 50 \\ 47 & 50 & 48 & 50 \\ 49 & 50 & 60 & 61 \\ 63 & (64) & 60 & 60 \end{pmatrix}$$

Since this is a maximization problem, it can be converted into an equivalent minimization problem by subtracting all the profit elements in the profit matrix from the highest profit element 64 of this profit matrix. Thus the cost matrix of the equivalent minimization problem is

$$\begin{pmatrix} 13 & 11 & 10 & 14 \\ 17 & 14 & 16 & 14 \\ 15 & 14 & 4 & 3 \\ 1 & 0 & 4 & 4 \end{pmatrix}$$

Select the smallest cost in each row and subtract this from all the cost elements of the corresponding row. We get

$$\begin{pmatrix} 3 & 1 & 0 & 4 \\ 3 & 0 & 2 & 0 \\ 12 & 11 & 1 & 0 \\ 1 & 0 & 4 & 4 \end{pmatrix}$$

Select the smallest cost element in each column and subtract this from all the cost elements of the corresponding column. We get

$$\begin{pmatrix} 2 & 1 & 0 & 4 \\ 2 & 0 & 2 & 0 \\ 11 & 11 & 1 & 0 \\ 0 & 0 & 4 & 4 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make the assignment in rows and columns having single zero. We get

$$\begin{pmatrix} 2 & 1 & (0) & 4 \\ 2 & (0) & 2 & \cancel{0} \\ 11 & 11 & 1 & (0) \\ (0) & \cancel{0} & 4 & 4 \end{pmatrix}$$

Since each row and each column contains exactly one encircled zero, the current assignment is optimal.

$\therefore$ The optimum assignment schedule is given by $A \rightarrow R$, $B \rightarrow Q$, $C \rightarrow S$, $D \rightarrow P$ and the optimum (maximum) profit

$$= \text{Rs. } (54 + 50 + 61 + 63)$$
$$= \text{Rs. } 228/-$$

Example 3 : **A company is faced with the problem of assigning four different salesman to four territories for promoting its sales. Territories are not equally rich in their sales potential and the salesman also differ in their ability to promote sales. The following table gives the expected annual sales (in thousands of Rs) for each salesman if assigned to various territories. Find the assignment of salesman so as to maximize the annual sales.**

		Territories			
		1	2	3	4
Salesmen	1	60	50	40	30
	2	40	30	20	15
	3	40	20	35	10
	4	30	30	25	20

[BRU. BE. Apr 95]

Solution : The cost matrix of the given assignment problem is

$$\begin{pmatrix} (60) & 50 & 40 & 30 \\ 40 & 30 & 20 & 15 \\ 40 & 20 & 35 & 10 \\ 30 & 30 & 25 & 20 \end{pmatrix}$$

Since this is a maximization problem, it can be converted into an equivalent minimization problem by subtracting all the cost elements in the cost matrix from the highest cost element 60 of this cost matrix. Thus the cost matrix of the equivalent minimization problem is

$$\begin{pmatrix} 0 & 10 & 20 & 30 \\ 20 & 30 & 40 & 45 \\ 20 & 40 & 25 & 50 \\ 30 & 30 & 35 & 40 \end{pmatrix}$$

Select the smallest cost element in each row (column) and subtract this from all the cost elements of the corresponding row (column). We get the reduced cost matrix

$$\begin{pmatrix} 0 & 10 & 15 & 20 \\ 0 & 10 & 15 & 15 \\ 0 & 20 & 0 & 20 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we make the assignment in rows and columns of this reduced cost matrix

$$\begin{pmatrix} (0) & 10 & 15 & 20 \\ \cancel{0} & 10 & 15 & 15 \\ \cancel{0} & 20 & (0) & 20 \\ \cancel{0} & (0) & \cancel{0} & \cancel{0} \end{pmatrix}$$

Since there are some rows and columns without assignment, the current assignment is not optimal.

Cover all the zeros by drawing a minimum number of straight lines

$$\begin{array}{c c} \begin{pmatrix} 0 & 10 & 15 & 20 \\ 0 & (10) & 15 & 15 \\ 0 & 20 & 0 & 20 \\ 0 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} \checkmark \\ \checkmark \\ \\ \\ \end{array} \\ \checkmark \quad\quad\quad\quad\quad\quad\quad & \end{array}$$

Here 10 is the smallest cost element not covered by these straight lines. Subtract this 10 from all the uncovered elements, add this 10 to those elements which lie in the intersection of these straight lines and do not change the remaining elements which lie on these straight lines. We get

$$\begin{pmatrix} 0 & 0 & 5 & 10 \\ 0 & 0 & 5 & 5 \\ 10 & 20 & 0 & 20 \\ 10 & 0 & 0 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we make the assignment in rows and columns of this reduced cost matrix.

$$\begin{pmatrix} (0) & \cancel{0} & 5 & 10 \\ \cancel{0} & (0) & 5 & 5 \\ 10 & 20 & (0) & 20 \\ 10 & \cancel{0} & \cancel{0} & (0) \end{pmatrix}$$

Since each row and each column contains exactly one assignment (*i.e.*, exactly one encircled zero), the current assignment is optimal.

$\therefore$ The optimum assignment schedule is given by Salesman 1 $\rightarrow$ Territory 1, Salesman 2 $\rightarrow$ Territory 2, Salesman 3 $\rightarrow$ Territory 3, Salesman 4 $\rightarrow$ Territory 4.

The optimum (maximum) annual sales

= 60 + 30 + 35 + 20 (in thousand of rupees)

= 145 (in thousand of rupees)

= Rs. 1,45,000/-.

Note : For this problem, there exists alternative optimal assignment schedule with the same maximum sales Rs. 1,45,000/-.

8.8 Restrictions in Assignments

The assignment technique assumes that the problem is free from practical restrictions and any task could be assigned to any facility. But in some cases, it may not be possible to assign a particular task to a particular facility due to space, size of the task, process capability of the facility, technical difficulties or other restrictions. This can be overcome by assigning a very high processing time or cost element (it can be ∞) to the corresponding cell. This cell will be automatically excluded in the assignment because of the unused high time cost associated with it.

Example 1 : A machine shop purchased a drilling machine and two lathes of different capacities. The positioning of the machines among 4 possible locations on the shop floor is important from the standard of materials handling. Given the cost estimate per unit time of materials below, determine the optimal location of the machines

	Location			
	1	**2**	**3**	**4**
Lathe 1	12	9	12	9
Drill	15	not suitable	13	20
Lathe 2	4	8	10	6

[MU. BE. Apr 93]

Solution : Since the drilling machine is not suitable for location 2, the corresponding cost element should be taken as ∞. Thus the cost matrix of the given assignment problem is

$$\begin{pmatrix} 12 & 9 & 12 & 9 \\ 15 & \infty & 13 & 20 \\ 4 & 8 & 10 & 6 \end{pmatrix}$$

Since the number of rows is less than the number of columns, we add a dummy row (a dummy drilling machine or a dummy lathe 3) with zero cost elements. The cost matrix for the balanced assignment problem is

$$\begin{pmatrix} 12 & 9 & 12 & 9 \\ 15 & \infty & 13 & 20 \\ 4 & 8 & 10 & 6 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Select the smallest cost in each row (column) and subtract this from all the cost elements of the corresponding row (column). We get the reduced matrix

$$\begin{pmatrix} 3 & 0 & 3 & 0 \\ 2 & \infty & 0 & 7 \\ 0 & 4 & 6 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make the assignment in rows and columns having single zero. We get

$$\begin{pmatrix} 3 & (0) & 3 & \cancel{0} \\ 2 & \infty & (0) & 7 \\ (0) & 4 & 6 & 2 \\ \cancel{0} & \cancel{0} & \cancel{0} & (0) \end{pmatrix}$$

Since each row and each column contains exactly one encircled zero, the current assignment is optimal.

$\therefore$ The optimum assignment schedule is given by

Lathe 1 → Location 2, Drill → Location 3,
Lathe 2 → Location 1, Dummy drill → Location 4

and the optimum (minimum) assignment cost

$$= (9 + 13 + 4 + 0) \text{ unit of cost}$$
$$= 26/\text{- units of cost.}$$

Note : For this the alternate optimum assignment is

Lathe 1 → Location 4, Drill → Location 3,
Lathe 2 → Location 1, Dummy drill → Location 2.

with the same optimum (minimum) assignment cost

$$= (9 + 13 + 4 + 0) \text{ units of cost}$$
$$= 26/\text{- units of cost.}$$

Example 2 : Five workers are available to work with the machines and the respective costs (in rupees) associated with each worker - machine assignment is given below. A sixth machine is available to replace one of the existing machines and the associated costs are also given below :

Machines

		M_1	M_2	M_3	M_4	M_5	M_6
	W_1	12	3	6	–	5	8
	W_2	4	11	–	5	–	3
Workers	W_3	8	2	10	9	7	5
	W_4	–	7	8	6	12	10
	W_5	5	8	9	4	6	–

(i) **Determine whether the new machine can be accepted ?**

(ii) **Determine also optimal assignment and the associated saving in cost.**

Solution : The cost matrix of the given assignment problem is

$$\begin{pmatrix} 12 & 3 & 6 & \infty & 5 & 8 \\ 4 & 11 & \infty & 5 & \infty & 3 \\ 8 & 2 & 10 & 9 & 7 & 5 \\ \infty & 7 & 8 & 6 & 12 & 10 \\ 5 & 8 & 9 & 4 & 6 & \infty \end{pmatrix}$$

Since the number of rows is less than the number of columns, the given assignment problem is unbalanced. Add a dummy worker W_6 (dummy row) with zero cost elements.

Thus the cost matrix of the balanced assignment problem is

$$\begin{pmatrix} 12 & 3 & 6 & \infty & 5 & 8 \\ 4 & 11 & \infty & 5 & \infty & 3 \\ 8 & 2 & 10 & 9 & 7 & 5 \\ \infty & 7 & 8 & 6 & 12 & 10 \\ 5 & 8 & 9 & 4 & 6 & \infty \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Select the smallest cost in each row and column and subtract this from all the cost elements of the corresponding row and column of the cost matrix. We get

$$\begin{pmatrix} 9 & 0 & 3 & \infty & 2 & 5 \\ 1 & 8 & \infty & 2 & \infty & 0 \\ 6 & 0 & 8 & 7 & 5 & 3 \\ \infty & 1 & 2 & 0 & 6 & 4 \\ 1 & 4 & 5 & 0 & 2 & \infty \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Since each row and each column contains atleast one zero, we shall make the assignment in rows and columns having single zero. We get

$$\begin{pmatrix} 9 & (0) & 3 & \infty & 2 & 5 \\ 1 & 8 & \infty & 2 & \infty & (0) \\ 6 & \cancel{0} & 8 & 7 & 5 & 3 \\ \infty & 1 & 2 & (0) & 6 & 4 \\ 1 & 4 & 5 & \cancel{0} & 2 & \infty \\ (0) & \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} & \cancel{0} \end{pmatrix}$$

Since some rows and columns are without assignment, the current assignment is not optimal.

Cover all the zeros by drawing minimum number of straight lines.

$$\begin{array}{c c}
\left(\begin{array}{cccccc}
9 & 0 & 3 & \infty & 2 & 5 \\
1 & 8 & \infty & 2 & \infty & 0 \\
6 & 0 & 8 & 7 & 5 & 3 \\
\infty & 1 & 2 & 0 & 6 & 4 \\
(1) & 4 & 5 & 0 & 2 & \infty \\
0 & 0 & 0 & 0 & 0 & 0
\end{array}\right) & \begin{array}{c} \checkmark \\ \\ \checkmark \\ \checkmark \\ \checkmark \\ \\ \end{array} \\
\begin{array}{cccccc} & \checkmark & & \checkmark & & \end{array} &
\end{array}$$

Choose the smallest cost element not covered by these straight lines. Here 1 is such element. Subtract this 1 from all the uncovered elements, add this 1 to those elements which are in the intersection of the straight lines and do not change the remaining elements which lie on these straight lines. We get

$$\begin{pmatrix}
8 & 0 & 2 & \infty & 1 & 4 \\
1 & 9 & \infty & 3 & \infty & 0 \\
5 & 0 & 7 & 7 & 4 & 2 \\
\infty & 1 & 1 & 0 & 5 & 3 \\
0 & 4 & 4 & 0 & 1 & \infty \\
0 & 1 & 0 & 1 & 0 & 0
\end{pmatrix}$$

Now we shall make the assignment in rows and columns having single zero.

$$\begin{pmatrix}
8 & (0) & 2 & \infty & 1 & 4 \\
1 & 9 & \infty & 3 & \infty & (0) \\
5 & \cancel{0} & 7 & 7 & 4 & 2 \\
\infty & 1 & 1 & (0) & 5 & 3 \\
(0) & 4 & 4 & \cancel{0} & 1 & \infty \\
0 & 1 & (0) & 1 & \cancel{0} & \cancel{0}
\end{pmatrix}$$

Since some rows and columns are without assignment, the current assignment is not optimal.

Cover all the zeros by drawing minimum number of straight lines.

$$\begin{array}{c c}
\left(\begin{array}{cccccc}
8 & 0 & 2 & \infty & (1) & 4 \\
1 & 9 & \infty & 3 & \infty & 0 \\
5 & 0 & 7 & 7 & 4 & 2 \\
\infty & 1 & 1 & 0 & 5 & 3 \\
0 & 4 & 4 & 0 & 1 & \infty \\
0 & 1 & 0 & 1 & 0 & 0
\end{array}\right) & \begin{array}{c} \checkmark \\ \\ \checkmark \\ \\ \\ \\ \end{array} \\
\begin{array}{cccccc} & \checkmark & & & & \end{array} &
\end{array}$$

Subtract the smallest uncovered element 1 from all the uncovered elements, add this 1 to those elements which are in the intersection of these straight lines and do not change the remaining elements which lie on these straight lines. We get

$$\begin{pmatrix}
7 & 0 & 1 & \infty & 0 & 3 \\
1 & 10 & \infty & 3 & \infty & 0 \\
4 & 0 & 6 & 6 & 3 & 1 \\
\infty & 2 & 1 & 0 & 5 & 3 \\
0 & 5 & 4 & 0 & 1 & \infty \\
0 & 2 & 0 & 1 & 0 & 0
\end{pmatrix}$$

Now we shall make the assignment in rows and columns having single zeros.

$$\begin{pmatrix}
7 & \cancel{0} & 1 & \infty & (0) & 3 \\
1 & 10 & \infty & 3 & \infty & (0) \\
4 & (0) & 6 & 6 & 3 & 1 \\
\infty & 2 & 1 & (0) & 5 & 3 \\
(0) & 5 & 4 & \cancel{0} & 1 & \infty \\
\cancel{0} & 2 & (0) & 1 & \cancel{0} & \cancel{0}
\end{pmatrix}$$

Since each row and each column contains exactly one encircled zero, the current assignment is optimal.

$\therefore$ The optimum assignment schedule is given by

$W_1 \to M_5$, $W_2 \to M_6$, $W_3 \to M_2$, $W_4 \to M_4$, $W_5 \to M_1$, $W_6 \to M_3$,

and the optimum (minimum) assignment cost according to this schedule is

$$= \text{Rs. } (5+3+2+6+5+0)$$
$$= \text{Rs. } 21/-$$

Now, if the sixth machine M_6 is not assigned to any of the workers, the given problem reduces to balanced one (deleting the sixth column). Applying the assignment algorithm to this balanced problem (reduced problem), the optimal assignment schedule is given by

$W_1 \to M_5$, $W_2 \to M_1$, $W_3 \to M_2$, $W_4 \to M_3$, $W_5 \to M_4$,

and the optimum (minimum) assignment cost according to this schedule is

$$= \text{Rs. } (5+4+2+8+4)$$
$$= \text{Rs. } 23/-$$

It is clear from the above that the minimum cost is more when there are only five machines. Hence, the sixth machine should be accepted. By accepting this sixth machine the associated saving cost will be Rs. (23 – 21) = Rs. 2.

8.9 Travelling Salesman Problem

A salesman normally must visit a number of cities starting from his head quarters. The distance (or time or cost) between every pair of cities are assumed to be known. The problem of finding the shortest distance (or minimum time or minimum cost) if the salesman starts from his headquarters and passes through each city under his Jurisdiction exactly once and returns to the headquarters is called the ***Travelling salesman problem or A Travelling salesperson problem.***

A travelling salesman problem is very similar to the assignment problem with the additional constraints.

(a) The salesman should go through every city exactly once except the starting city (headquarters).

(b) The salesman starts from one city (head quarters) and comes back to that city (headquarters).

(c) Obviously going from any city to the same city directly is not allowed (*i.e.,* no assignments should be made along the diagonal line).

Note 1 : Conditions *(a)* and *(b)* are usually called ***route conditions.***

Note 2 : If a salesman has to visit n cities, then he will have a total of $(n-1)!$ possible round trips.

Therefore, the necessary basic steps to solve a travelling salesman problem are :

(i) Assigning an infinitely large element (∞) in each of the squares along the diagonal line in the cost matrix.

(ii) Solving the problem as a routine assignment problem.

(iii) Scrutinizing the solution obtained under *(ii)* to see if the "route" conditions are satisfied.

(iv) If not, making adjustments in assignments to satisfy the condition with minimum increase in total cost (*i.e.,* to satisfy route condition "next best solution" may require to be considered).

Example 1 : Solve the following travelling salesman problem

		To			
		A	**B**	**C**	**D**
	A	–	46	16	40
From	**B**	41	–	50	40
	C	82	32	–	60
	D	40	40	36	–

[MU. BE. Apr 93]

Solution : The cost matrix of the given travelling salesman problem is

$$\begin{pmatrix} \infty & 46 & 16 & 40 \\ 41 & \infty & 50 & 40 \\ 82 & 32 & \infty & 60 \\ 40 & 40 & 36 & \infty \end{pmatrix}$$

Solve this as a routine assignment problem

Subtract the smallest cost element in each row from all the elements of the corresponding row. We get.

$$\begin{pmatrix} \infty & 30 & 0 & 24 \\ 1 & \infty & 10 & 0 \\ 50 & 0 & \infty & 28 \\ 4 & 4 & 0 & \infty \end{pmatrix}$$

Subtract the smallest cost element in each column from all the elements of the corresponding column. We get

$$\begin{pmatrix} \infty & 30 & 0 & 24 \\ 0 & \infty & 10 & 0 \\ 49 & 0 & \infty & 28 \\ 3 & 4 & 0 & \infty \end{pmatrix}$$

Now we shall make the assignment in rows and columns having single zero. We get

$$\begin{pmatrix} \infty & 30 & (0) & 24 \\ (0) & \infty & 10 & \cancel{0} \\ 49 & (0) & \infty & 28 \\ 3 & 4 & \cancel{0} & \infty \end{pmatrix}$$

Since some rows and columns are with out assignment, the current assignment is not optimal.

Cover all the zeros by drawing a minimum number of straight lines.

$$\begin{pmatrix} \infty & 30 & 0 & 24 \\ 0 & \infty & 10 & 0 \\ 49 & 0 & \infty & 28 \\ (3) & 4 & \cancel{0} & \infty \end{pmatrix}$$

Subtract the smallest uncovered cost element 3 from all uncovered elements, add this 3 to those elements which are in the intersection of these straight lines and do not change the remaining elements which lie on these straight lines. We have

$$\begin{pmatrix} \infty & 27 & 0 & 21 \\ 0 & \infty & 13 & 0 \\ 49 & 0 & \infty & 28 \\ 0 & 1 & 0 & \infty \end{pmatrix}$$

Now we shall make the assignment in rows and columns having single zeros. We get

$$\begin{pmatrix} \infty & 27 & (0) & 21 \\ \cancel{0} & \infty & 13 & (0) \\ 49 & (0) & \infty & 28 \\ (0) & 1 & \cancel{0} & \infty \end{pmatrix}$$

Since each row and each column contains exactly one encircled zero, the current assignment is optimal for the assignment problem.

$\therefore$ The optimum assignment schedule is given by

$A \to C, \quad B \to D, \quad C \to B, \quad D \to A,$

i.e., $A \to C, \quad C \to B, \quad B \to D, \quad D \to A,$

i.e., $A \to C \to B \to D \to A$

Check whether the route conditions are satisfied.

$A \to C \to B \to D \to A$ satisfies the route condition.

$\therefore$ The required minimum costs.

$$= (16 + 32 + 40 + 40) \text{ units of cost.}$$
$$= 128/- \text{ units of cost.}$$

Example 2 : Solve the following travelling salesman problem so as to minimize the cost per cycle.

		To				
		A	**B**	**C**	**D**	**E**
	A	–	3	6	2	3
From	**B**	3	–	5	2	3
	C	6	5	–	6	4
	D	2	2	6	–	6
	E	3	3	4	6	–

[MU. BE. 85, Nov 93]

Solution : The cost matrix of the given travelling salesman problem is

$$\begin{pmatrix} \infty & 3 & 6 & 2 & 3 \\ 3 & \infty & 5 & 2 & 3 \\ 6 & 5 & \infty & 6 & 4 \\ 2 & 2 & 6 & \infty & 6 \\ 3 & 3 & 4 & 6 & \infty \end{pmatrix}$$

Subtract the smallest cost element in each row from all the elements of the corresponding row. We get

$$\begin{pmatrix} \infty & 1 & 4 & 0 & 1 \\ 1 & \infty & 3 & 0 & 1 \\ 2 & 1 & \infty & 2 & 0 \\ 0 & 0 & 4 & \infty & 4 \\ 0 & 0 & 1 & 3 & \infty \end{pmatrix}$$

Subtract the smallest cost element in each column from all the elements of the corresponding column. We get

$$\begin{pmatrix} \infty & 1 & 3 & 0 & 1 \\ 1 & \infty & 2 & 0 & 1 \\ 2 & 1 & \infty & 2 & 0 \\ 0 & 0 & 3 & \infty & 4 \\ 0 & 0 & 0 & 3 & \infty \end{pmatrix}$$

Now we shall make the assignment in rows and columns having single zeros. We get

$$\begin{pmatrix} \infty & 1 & 3 & (0) & 1 \\ 1 & \infty & 2 & \cancel{0} & 1 \\ 2 & 1 & \infty & 2 & (0) \\ \cancel{0} & (0) & 3 & \infty & 4 \\ \cancel{0} & \cancel{0} & (0) & 3 & \infty \end{pmatrix}$$

Since some rows and columns are without assignment, the current assignment is not optimal.

Cover all the zeros by drawing a minimum number of straight lines.

$$\begin{pmatrix} \infty & (1) & 3 & 0 & 1 \\ 1 & \infty & 2 & 0 & 1 \\ 2 & 1 & \infty & 2 & 0 \\ 0 & 0 & 3 & \infty & 4 \\ 0 & 0 & 0 & 3 & \infty \end{pmatrix} \begin{matrix} \checkmark \\ \checkmark \\ \\ \\ \\ \end{matrix}$$

Subtract the smallest uncovered cost element 1 from all uncovered elements, add this 1 to those elements which are in the intersection of these straight lines and do not change the remaining elements which lie on these straight lines. We have

$$\begin{pmatrix} \infty & 0 & 2 & 0 & 0 \\ 0 & \infty & 1 & 0 & 0 \\ 2 & 1 & \infty & 3 & 0 \\ 0 & 0 & 3 & \infty & 4 \\ 0 & 0 & 0 & 4 & \infty \end{pmatrix}$$

Now we shall make the assignment in rows and columns having single zero. We get

$$\begin{pmatrix} \infty & \cancel{0} & 2 & (0) & \cancel{0} \\ (0) & \infty & 1 & \cancel{0} & \cancel{0} \\ 2 & 1 & \infty & 3 & (0) \\ \cancel{0} & (0) & 3 & \infty & 4 \\ \cancel{0} & \cancel{0} & (0) & 4 & \infty \end{pmatrix}$$

Since each row and each column contains exactly one encircled zero, the current assignment is optimal.

$\therefore$ The optimal assignment schedule is given by

$$A \to D, \quad B \to A, \quad C \to E, \quad D \to B, \quad E \to C$$

i.e., $\quad A \to D \to B \to A, \quad C \to E \to C$

and the corresponding optimum (minimum) assignment cost

$= (2 + 3 + 4 + 2 + 4)$ units of cost

$=$ 15/- units of cost.

But this assignment schedule does not provide the solution of this travelling salesman problem, because it does not satisfy the 'route' condition.

We try to find the next best solution which satisfies the route condition also. The next minimum (non-zero) cost element in the cost matrix is 1. So we try to bring 1 in to the solution. But the 1 occurs at two places. We shall consider all the cases separately until the acceptable solution is reached.

We start with making an assignment at (2, 3) instead of zero assignment at (2, 1). The resulting feasible solution will then be

$$\begin{pmatrix} \infty & \cancel{0} & 2 & (0) & \cancel{0} \\ \cancel{0} & \infty & (1) & \cancel{0} & \cancel{0} \\ 2 & 1 & \infty & 3 & (0) \\ \cancel{0} & (0) & 3 & \infty & 4 \\ (0) & \cancel{0} & \cancel{0} & 4 & \infty \end{pmatrix}$$

$\therefore$ The optimum assignment is given by

$$A \to D, \quad B \to C, \quad C \to E, \quad D \to B, \quad E \to A,$$

i.e., $\quad A \to D \to B \to C \to E \to A$

Also, when an assignment is made at (3, 2) instead of zero assignment at (3, 5), the resulting feasible solution will be

$$\begin{pmatrix} \infty & \cancel{0} & 2 & \cancel{0} & (0) \\ \cancel{0} & \infty & 1 & (0) & \cancel{0} \\ 2 & (1) & \infty & 3 & \cancel{0} \\ (0) & \cancel{0} & 3 & \infty & 4 \\ \cancel{0} & \cancel{0} & (0) & 4 & \infty \end{pmatrix}$$

$\therefore$ The optimum assignment is given by

$A \to E, \quad B \to D, \quad C \to B, \quad D \to A, \quad E \to C,$

i.e., $\quad A \to E \to C \to B \to D \to A$

$\therefore$ For the given travelling salesman problem, the optimum assignment schedule is given by

$A \to D \to B \to C \to E \to A$, (or)

$A \to E \to C \to B \to D \to A$

In both cases, the optimum (minimum) assignment cost is 16/- units of cost.

EXERCISE

1. **What are assignment problems? Describe Mathematical formulation of an assignment problem?**

   [MU. BE. Apr 93, Oct 96, BRU. BE. Nov 96, MU. MBA. Nov 96,]

2. **Distinguish between transportation model and assignment model.** ***[MU. BE. Nov 94]***

3. **Explain how the assignment problem can be treated as a particular case of transportation problem? Why this method is not preferred ?** ***[MU. MBA Nov. 95, Apr.97]***

4. **Explain the steps in the Hungarian Method used for solving assignment problems.**

   [MU. MBA. Apr 95, BRU. BE. Nov 96, MU. MCA. Nov 95]

5. **Define an unbalanced assignment problem and describe the steps involved in solving it.**

6. **Explain how maximization problems are solved using assignment model technique ?**

7. **What do you understand by restricted assignments? Explain how should one overcome it ?**

8. **Enumerate the steps to solve an unbalanced profit maximization problem containing one or more restricted assignments ?**

9. **Is it possible to have more than one optimal solution to an assignment problem ? How is the presence of an alternate solution established ?**

10. **What is the difference between assignment problem and travelling salesman problem?**

11. **What is the objective of the travelling salesman problem ?** ***[MU. BE. Oct 96]***

12. **What is travelling salesman problem ?** ***[BRU. BE. Nov 96]***

13. **Solve the assignment problem**

	A	B	C	D
I	1	4	6	3
II	9	7	10	9
III	4	5	11	7
IV	8	7	8	5

[MU. BE. Apr 91]

14. **Consider the problem of assigning five jobs to five persons. The assignment costs are given as follows :**

Jobs

Person	1	2	3	4	5
A	8	5	2	6	1
B	0	9	5	5	4
C	3	8	9	2	6
D	4	3	1	0	3
E	9	5	8	9	5

Determine the optimum assignment schedule. ***[MU. BE. Apr 90, Apr 91]***

15. **A department has four subordinates and four tasks are to performed. The subordinates differ in efficiency and tasks differ in their intrinsic difficulties. The estimate of time (in hours) each man would take to perform each task is given by**

	Tasks I	II	III	IV
Subordinate 1	8	26	17	11
2	13	28	4	26
3	38	19	18	15
4	19	26	24	10

Find out how the tasks be allotted to each subordinate so as to optimize the total man-hours. *[MU. BE. Nov 91]*

16. Solve the assignment problem

Machine	Job P	Q	R	S
A	18	26	17	11
B	13	28	14	26
C	38	19	18	15
D	19	26	24	10

[MU. BE. Oct 95]

17. A department head has four tasks to be performed by three subordinates, the subordinates differing in efficiency. The estimates of the time, each subordinate would take to perform, is given below in the matrix. How should he allocate the tasks one to each man, so as to minimize the total man-hour ?

Tasks	Men 1	2	3
I	9	26	15
II	13	27	6
III	35	20	15
IV	18	30	20

[MU. BE. Nov 92]

18. A company has four machines on which to do three jobs. Each job can be assigned to one and only one machine. The cost of each job on each machine is given in the following table :

Job	Machine P	Q	R	S
A	18	24	28	32
B	8	13	17	19
C	10	15	19	22

What are the job assignments which will minimize the cost ? *[MU. BE. Apr 87, MBA. Apr 95]*

19. Solve the following unbalanced problem of assigning four jobs to three different men (only one job to each man). The time to perform the job by different men is given in the following table :

Men	Job J_1	J_2	J_3	J_4
M_1	7	5	8	4
M_2	5	6	7	4
M_3	8	7	9	8

[MU. BE. Nov 92]

20. Solve the following assignment problem to find the maximum total expected sale.

Salesman	Area I	II	III	IV
A	42	35	28	21
B	30	25	20	15
C	30	25	20	15
D	24	20	16	12

[MU. BE. Nov 90]

21. A sales manager has to assign salesmen to four territories. He has 4 candidates of varying experience and capabilities and assesses the possible profit for each salesman in each territory as given below. Find the assignment which maximises the profit.

Salesman	Territory A	B	C	D
1	35	27	28	37
2	28	34	29	40
3	35	24	32	28
4	24	32	25	28

[MU. BE. Nov 92]

22. A company has 5 jobs to be done. The following data shows the return (in rupees) by assigning the ith machine to the jth job. Using Hungarian method, assign the 5 jobs to the 5 machines so as to maximize the total expected profit.

		Job				
		1	2	3	4	5
	A	62	78	50	101	82
	B	71	84	61	73	59
Machine	C	87	92	111	70	81
	D	45	64	87	77	80
	E	60	70	98	66	83

[MU. BE. Apr 91, Apr 98, BRU. ME. 81, MKU. BE. Nov 97]

23. A company is faced with the problem of assigning 4 machines to 6 different jobs (one machine to one job only). The profits are estimated as follows :

		Machine			
		A	B	C	D
	1	3	6	2	6
	2	7	1	4	4
Job	3	3	8	5	8
	4	6	4	3	7
	5	5	2	4	3
	6	5	7	6	4

Solve the problem to maximize the total profit.

[MU. BE. Apr 90, Nov 94]

24. Five operators have to be assigned to five machines. The assignment costs are given in the table below :

		Machine				
		I	II	III	IV	V
	A	5	5	–	2	6
Operator	B	7	4	2	3	4
	C	9	3	5	–	3
	D	7	2	6	7	2
	E	6	5	7	9	1

Operator A cannot operate machine III and operator C cannot operate machine IV. Find the optimal assignment schedule.

25. Solve the following assignment problem

		Task				
		A	B	C	D	E
	M_1	4	6	10	5	6
Machine	M_2	7	4	Not suitable	5	4
	M_3	Not suitable	6	9	6	2
	M_4	9	3	7	2	3

26. Solve the following assignment problem

		Machine				
		1	2	3	4	5
	A	7	7	∞	4	8
	B	9	6	4	5	6
Task	C	11	5	7	∞	5
	D	9	4	8	9	4
	E	8	7	9	11	3

27. Given the following matrix of setup costs show how to sequence production so as to minimize setup cost per cycle.

		To				
		A	B	C	D	E
	A	–	2	5	7	1
From	B	6	–	3	8	2
	C	8	7	–	4	7
	D	12	4	6	–	5
	E	1	3	2	8	–

28. Solve the following travelling salesman problem so as to minimize cost per cycle :

To city

From city	1	2	3	4	5
1	∞	10	25	25	10
2	1	∞	10	15	2
3	8	9	∞	20	10
4	14	10	24	∞	15
5	10	8	25	27	∞

29. A salesman has to visit five cities A, B, C, D, and E. The distances (in hundred km) between the five cities are as follows :

To

From	A	B	C	D	E
A	–	7	6	8	4
B	7	–	8	5	6
C	6	8	–	9	7
D	8	5	9	–	8
E	4	6	7	8	–

If the salesman starts from city A and has to come back to city A, which route he should select so that the total distance travelled by him is minimized.

30. Solve the following travelling salesman problem :

	A	B	C	D	E	F
A	∞	5	12	6	4	8
B	6	∞	10	5	4	3
C	8	7	∞	6	3	11
D	5	4	11	∞	5	8
E	5	2	7	8	∞	4
F	6	3	11	5	4	∞

31. Solve the travelling salesman problem given by the following data.

$c_{12} = 20,\quad c_{13} = 4,\quad c_{14} = 10,\quad c_{23} = 5,\quad c_{34} = 6,$

$c_{25} = 10,\quad c_{35} = 6,\quad c_{45} = 20,\quad$ where $c_{ij} = c_{ji}$

and there is no route between cities i and j if a value for c_{ij} is not shown above.

32. A company has five jobs to be done one five machines ; any job can be done on any machine. The costs of doing the jobs in different machines are given below. Assign the jobs for different machines so as to minimize the total cost.

Machines

Jobs	A	B	C	D	E
I	13	8	16	18	19
II	9	15	24	9	12
III	12	9	4	4	4
IV	6	12	10	8	13
V	15	17	18	12	20

[MU. BE. Oct 96]

33. The owner of a small machine shop has four machinists available to assign to jobs for the day. Five jobs are offered with expected profit for each machinist on each job as follows:

Jobs

Machinist	A	B	C	D	E
M_1	12	28	0	51	32
M_2	12	34	11	23	9
M_3	37	42	61	21	31
M_4	0	14	37	27	30

Assign machinists to jobs which results in overall maximum profit. Which job should be declined ?

[MU. MBA. Nov 96]

34. **A machine tool company decides to make four sub assemblies through four contractors. Each contractor is to receive only one subassembly. The cost of each subassembly is determined by the bids submitted by each contractor and is shown in table given below in hundreds of rupees. Assign the different subassemblies to contractors so as to minimize the total cost.**

Table

		Contractors			
		1	2	3	4
	1	15	13	14	17
Subassemblies	2	11	12	15	13
	3	13	12	10	11
	4	15	17	14	16

Assign machinists to jobs which results in overall maximum profit. Which job should be declined ?

[MU. B. Tech. Leather. Oct 96]

35. **The R and D company has recently requested the skill testing agency to test four applicants for the three jobs that are available at this time. Each job has a primary skill and R and D's objective is to pick the three applicants whose aptitude test scores will maximize R and D's total performance. Only one applicant can be assigned to each job. Their aptitude test scores is listed below :**

		Job		
		A	B	C
	1	95	110	103
Applicant	2	89	95	100
	3	120	132	118
	4	107	119	112

Determine the three best applicants for the three jobs. What are their total aptitude test scores ? ***[BRU. BE. Apr 96]***

ANSWERS

13. $1 \longrightarrow A$, $2 \longrightarrow C$, $3 \longrightarrow B$, $4 \longrightarrow D$,
minimum cost Rs. 21.

14. $A \longrightarrow 5$, $B \longrightarrow 1$, $C \longrightarrow 4$, $D \longrightarrow 3$, $E \longrightarrow 2$,
minimum cost Rs. 9.

15. $1 \longrightarrow I$, $2 \longrightarrow III$, $3 \longrightarrow II$, $4 \longrightarrow IV$,
minimum time 41 hours.

16. $A \longrightarrow R$, $B \longrightarrow P$, $C \longrightarrow Q$, $D \longrightarrow S$,
minimum cost Rs. 59.

17. $I \longrightarrow 1$, $II \longrightarrow 3$, $III \longrightarrow 2$, $IV \longrightarrow 4$
minimum time 35 hours.

18. $A \longrightarrow P$, $B \longrightarrow Q$, $C \longrightarrow R$, $D \longrightarrow S$,
minimum cost Rs. 50.

19. $M_1 \longrightarrow J_4$, $M_2 \longrightarrow J_1$, $M_3 \longrightarrow J_2$, $M_4 \longrightarrow J_3$,
minimum cost Rs. 16.

20. $A \longrightarrow I$, $B \longrightarrow II$, $C \longrightarrow III$, $D \longrightarrow IV$, (or)
$A \longrightarrow I$, $B \longrightarrow III$, $C \longrightarrow II$, $D \longrightarrow IV$,
maximum total expected sale = Rs. 99000.

21. $1 \longrightarrow A$, $2 \longrightarrow D$, $3 \longrightarrow C$, $4 \longrightarrow B$,
maximum profit Rs. 139.

22. $A \longrightarrow 4$, $B \longrightarrow 2$, $C \longrightarrow 1$, $D \longrightarrow 5$, $E \longrightarrow 3$,
maximum profit is Rs. 450.

23. $2 \longrightarrow A$, $3 \longrightarrow B$, $4 \longrightarrow D$, $6 \longrightarrow C$
maximum profit is Rs. 28.

24. $A \longrightarrow IV$, $B \longrightarrow III$, $C \longrightarrow II$, $D \longrightarrow I$, $E \longrightarrow V$, (or)
$A \longrightarrow IV$, $B \longrightarrow III$, $C \longrightarrow V$, $D \longrightarrow II$, $E \longrightarrow I$,
minimum cost Rs. 15.

25. $M_1 \longrightarrow A, M_2 \longrightarrow B, M_3 \longrightarrow E, M_4 \longrightarrow D, M_5 \longrightarrow C$

minimum assignment cost is Rs. 12.

26. $A \longrightarrow 4, B \longrightarrow 3, C \longrightarrow 2, D \longrightarrow 1, E \longrightarrow 5$

minimum assignment cost is Rs. 25

27. $A \longrightarrow B \longrightarrow C \longrightarrow D \longrightarrow A$

minimum cost is Rs. 15.

28. $1 \longrightarrow 5 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4 \longrightarrow 1$

minimum cost is Rs. 62.

29. $A \longrightarrow E \longrightarrow B \longrightarrow D \longrightarrow C \longrightarrow A$

minimum distance is Rs. 30 (in hundred km).

30. $A \longrightarrow B \longrightarrow F \longrightarrow E \longrightarrow C \longrightarrow D \longrightarrow A$

minimum cost is Rs. 30.

31. $A \longrightarrow C \longrightarrow B \longrightarrow D \longrightarrow A$

minimum cost is Rs. 19.

32. $I \longrightarrow B, II \longrightarrow E, III \longrightarrow C, IV \longrightarrow A, V \longrightarrow D$

minimum cost is Rs. 42/-.

33. $M_1 \longrightarrow D, M_2 \longrightarrow B, M_3 \longrightarrow C, M_4 \longrightarrow E, M_5 \longrightarrow A$

The maximum profit is Rs. 176/-. Job A should be declined.

34. $1 \longrightarrow 2$, $2 \longrightarrow 1$, $3 \longrightarrow 4$, $4 \longrightarrow 3$, and the minimum assignment cost is Rs. 49/-.

35. Applicant 1 $\rightarrow$ Job 3, Applicant 3 $\rightarrow$ Job 1, Applicant 4 $\rightarrow$ Job 2, and their total aptitude test scores = 103 + 120 + 119 = 342.

Chapter 9

Integer Programming

9.1 Introduction

A linear programming problem in which some or all of the variables in the optimal solution are restricted to assume non-negative integer values is called an ***integer programming problem*** [or ***I.P.P*** or ***integer linear programming***].

In a linear programming problem, if all the variables in the optimal solution are restricted to assume non-negative integer values, then it is called the ***Pure (all) integer programming problem [Pure I.P.P]***.

In a linear programming problem, if only some of the variables in the optimal solution are restricted to assume non-negative integer values, while the remaining variables are free to take any non-negative values, then it is called a ***Mixed integer programming problem [Mixed I.P.P]***

Further, if all the variables in the optimum solution are allowed to take values either 0 or 1 as in 'do' or 'not to do' type decisions, then the problem is called the ***Zero–one programming problem (or) standard discrete programming problem.***

The general integer programming problem is given by

$$\text{Maximize } Z = CX$$

subject to the constraints

$$AX \leq b,$$
$$X \geq 0 \text{ and some or all variables are integers.}$$

Importance of integer Programming :

In linear programming problem, all the decision variables were allowed to take any non-negative real (continuous or fractional) values, as it is quite possible and appropriate to have fractional values in many situations. For example, it is quite possible to use 6.38 kg of raw material, or 5.62 machine hours etc. However in many situations, especially in business and industry, these decision variables make sense only if they have integer values in the optimal solution. For example, it is meaningless to produce 8.13 chairs or 6.85 tables, or to open 3.83 branches of a bank