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

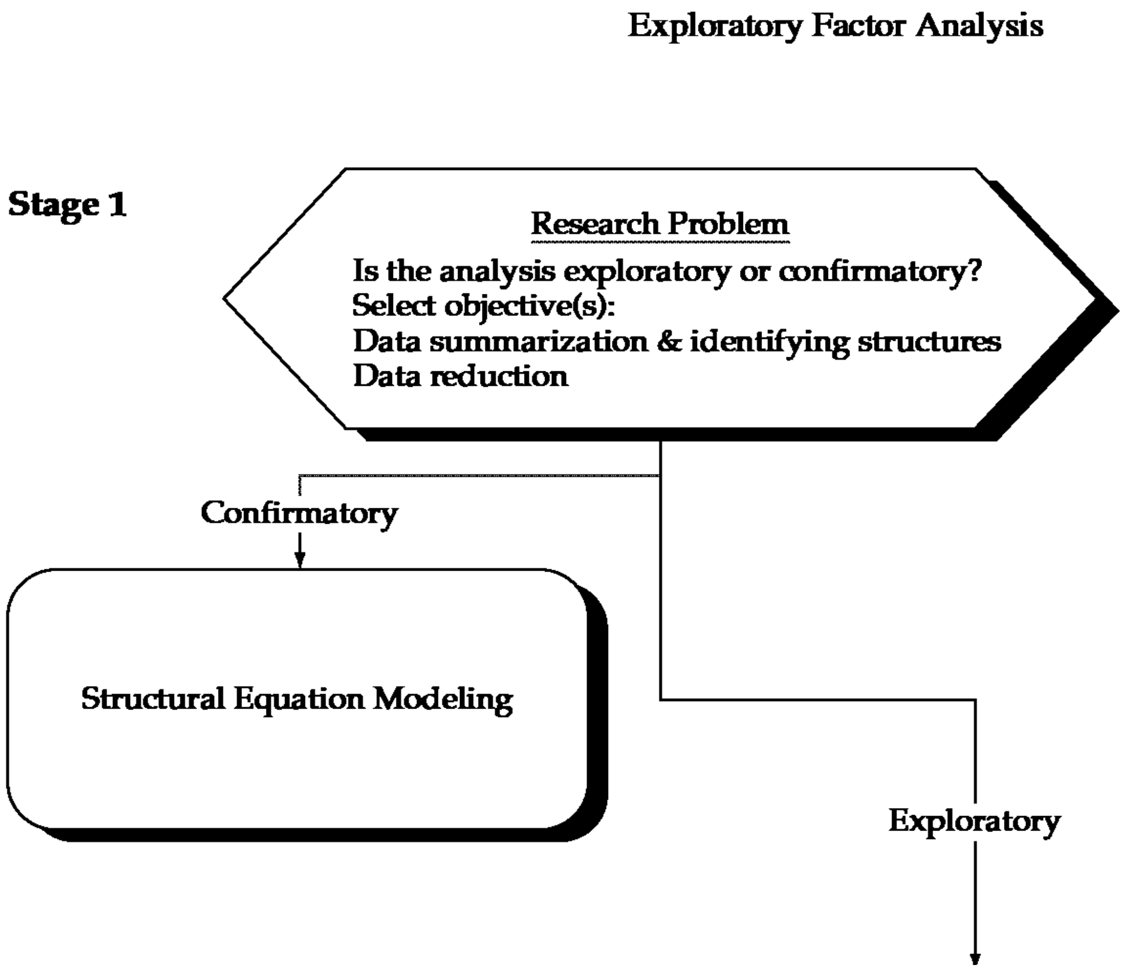
- **AN ILLUSTRATIVE EXAMPLE OF FACTOR ANALYSIS :**
  - Visual inspection of the original correlation matrix (Figure 1, part 1) does not easily reveal any specific pattern.
  - Among scattered high correlations, variable groupings are not apparent. The application of factor analysis results in the grouping 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

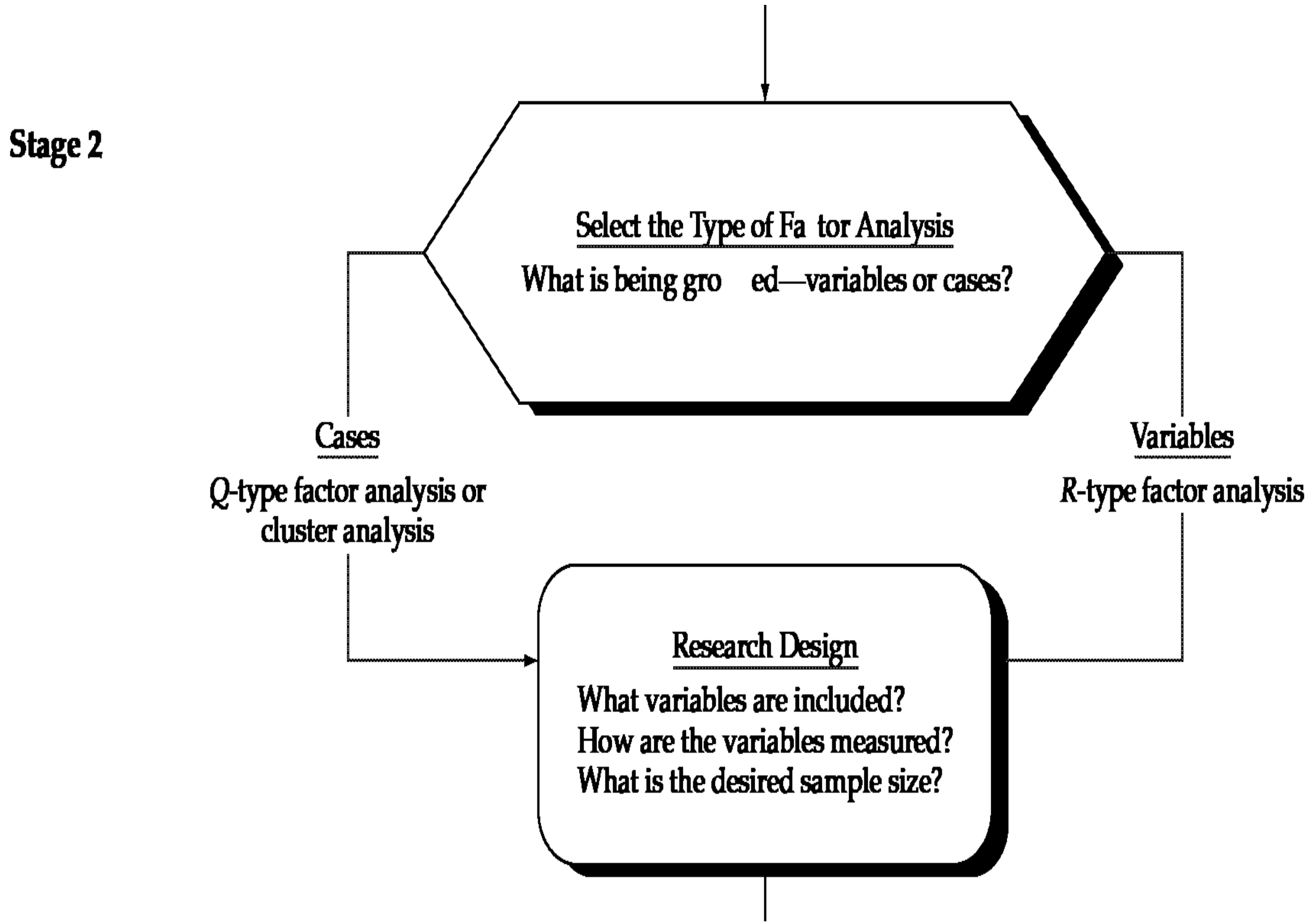
- **Using Factor Analysis with Other Multivariate Techniques**
  - Factor analysis, by providing insight into the interrelationships among variables and the underlying structure of the data, is an excellent starting point for many other multivariate techniques.
  - Variables determined to be highly correlated 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.

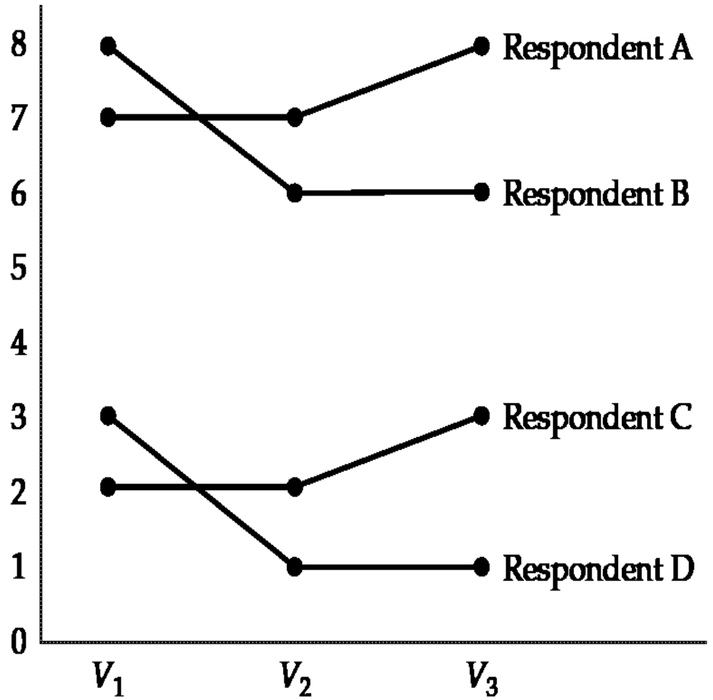
Factor Analysis Decision Process

- **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

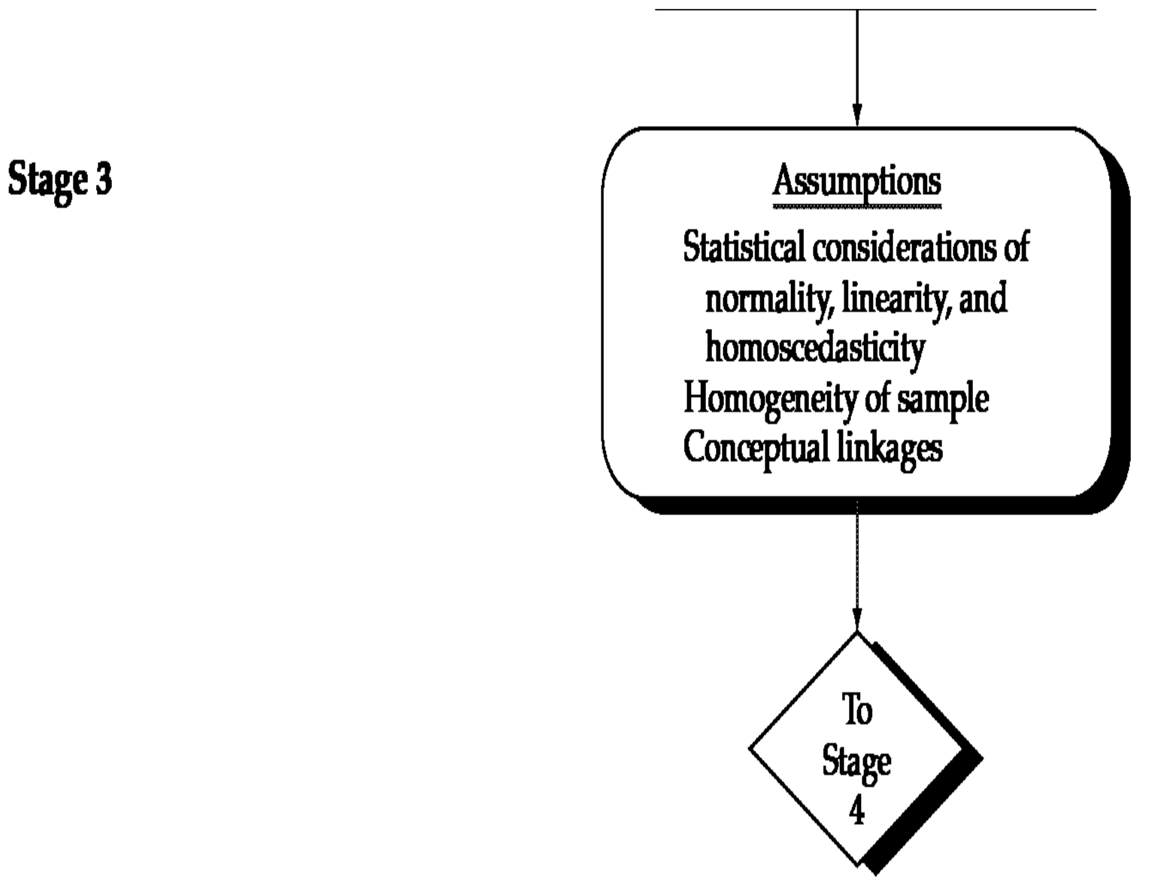
- **Stage 2 : Designing A Factor Analysis**
  - **Variable Selection and Measurement Issues :**
  - Two specific questions must be answered at this point:
    - (1) What type of variables can be used in factor analysis?
    - (2) How many variables should be included?
  - In terms of the types of variables 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

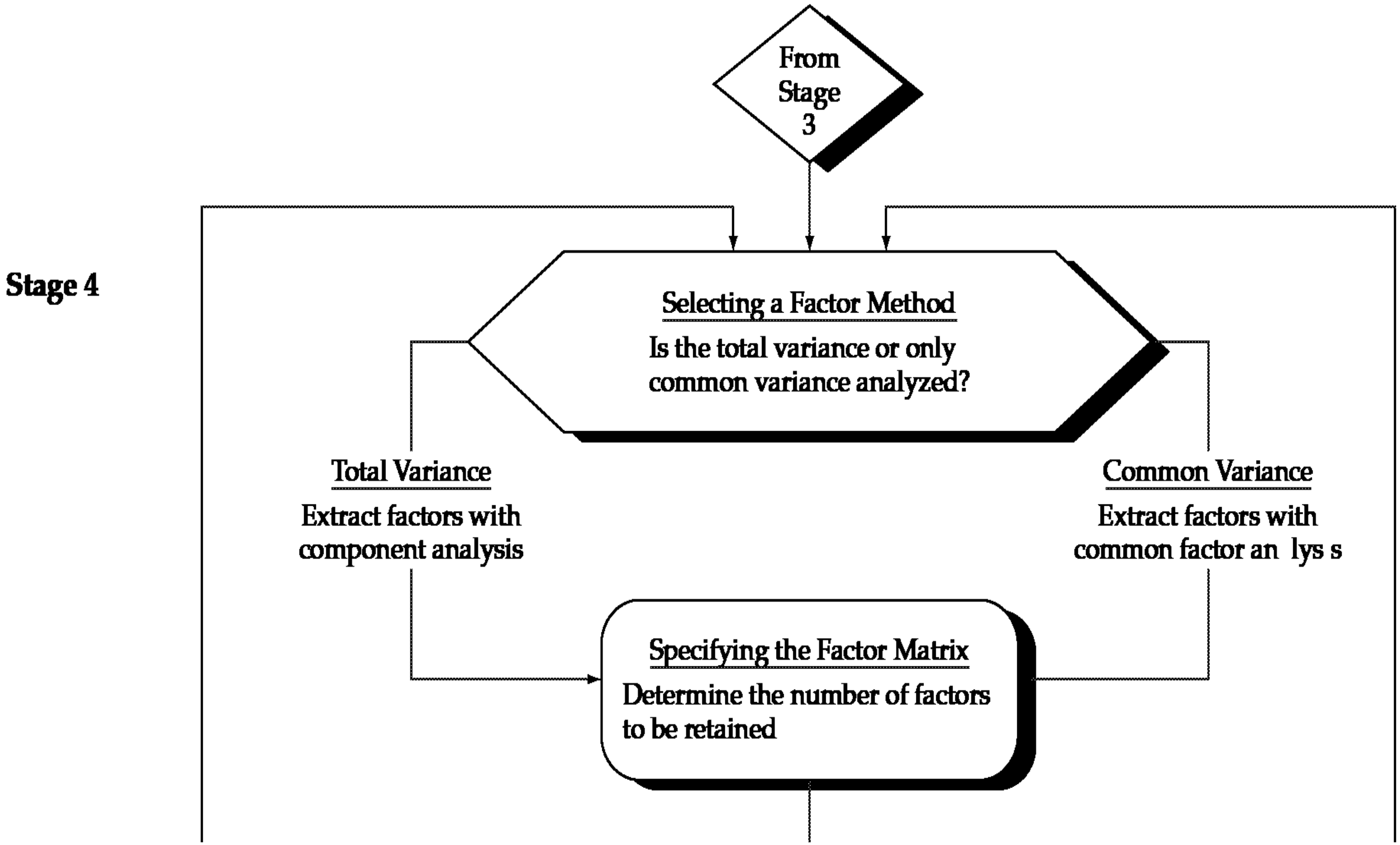
- **Stage 3 : Assumptions in factor analysis**
  - The MSA increases as
    - (1) the sample size increases,
    - (2) the average correlations increase,
    - (3) the number of variables increases, or
    - (4) the number of factors decreases.
  - The researcher should always have an overall MSA value 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

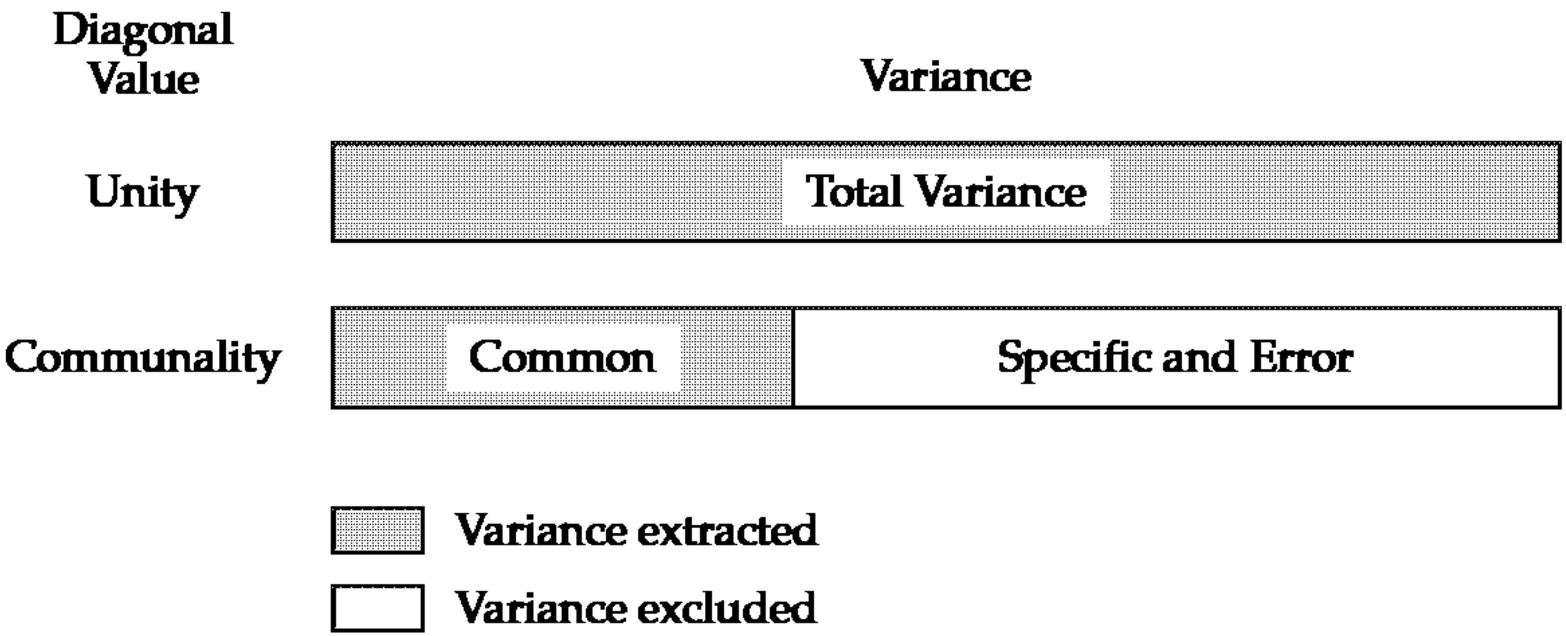
- **Stage 4 : Deriving factors and assessing overall fit**
  - Thus, the total variance of any variable is composed of its common, unique, and error variances.
  - As a variable is more highly correlated with one or more variables, the common variance (communality) increases.
  - However, if unreliable 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

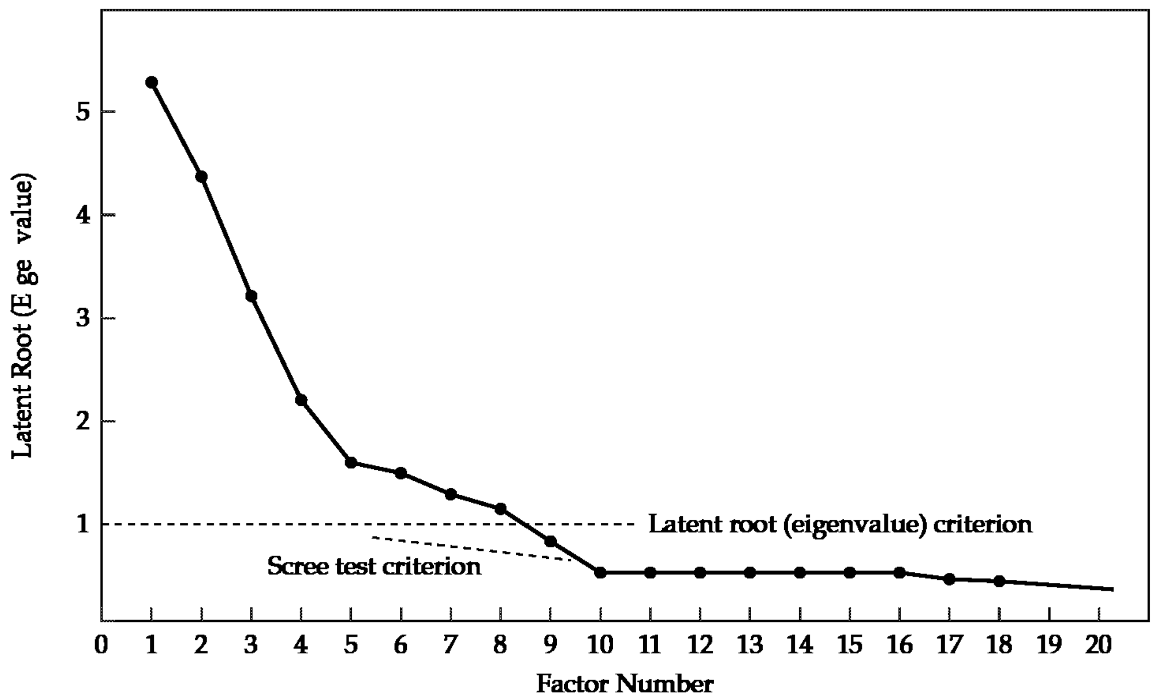
- **Stage 4 : Deriving factors and assessing overall fit**
  - **A PRIORI CRITERION :**
  - The a priori criterion is a simple yet reasonable criterion under certain circumstances.
  - When applying it, we already knows how many factors to extract before undertaking the factor analysis.
  - The researcher 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.

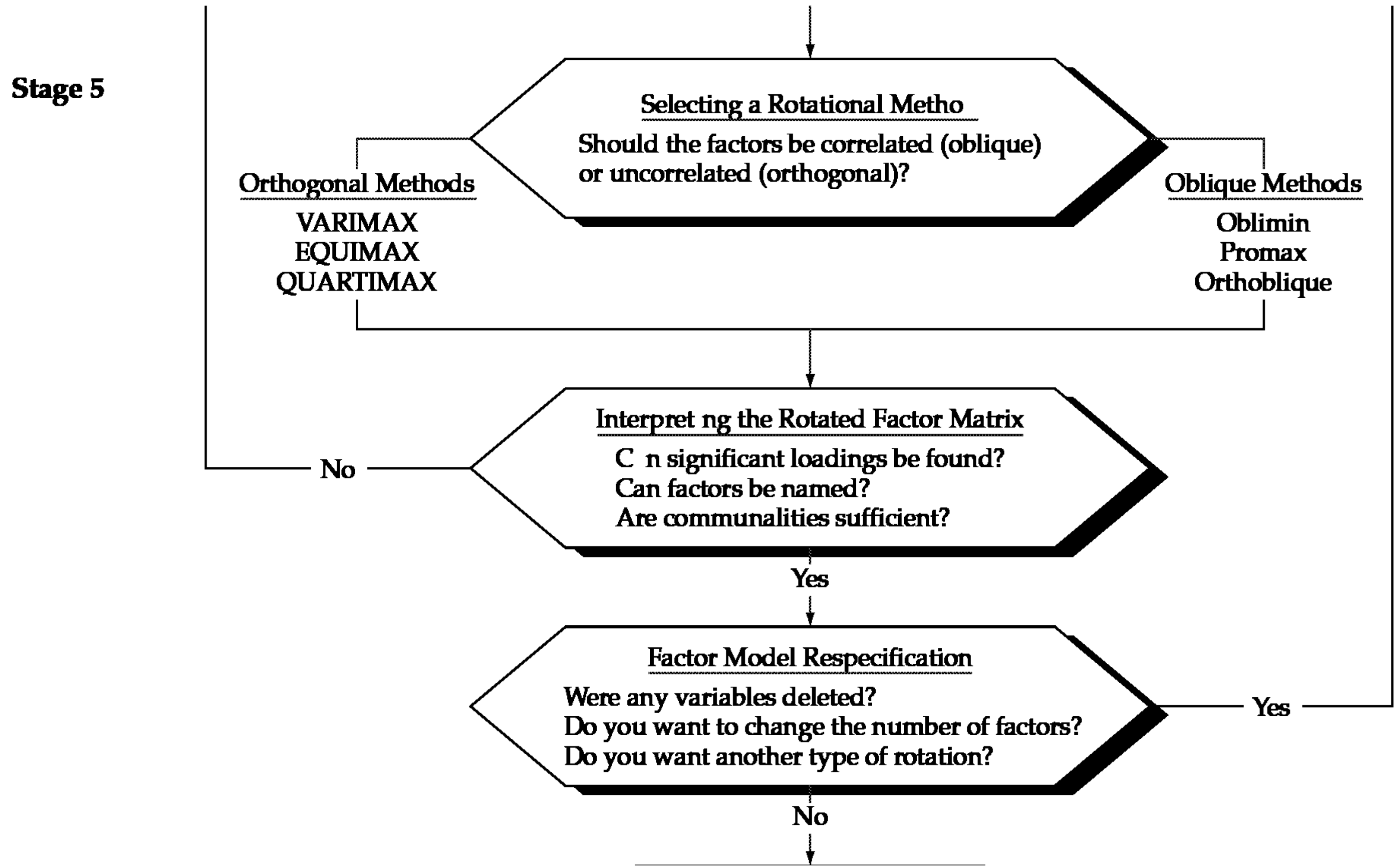
UNIT - II

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**

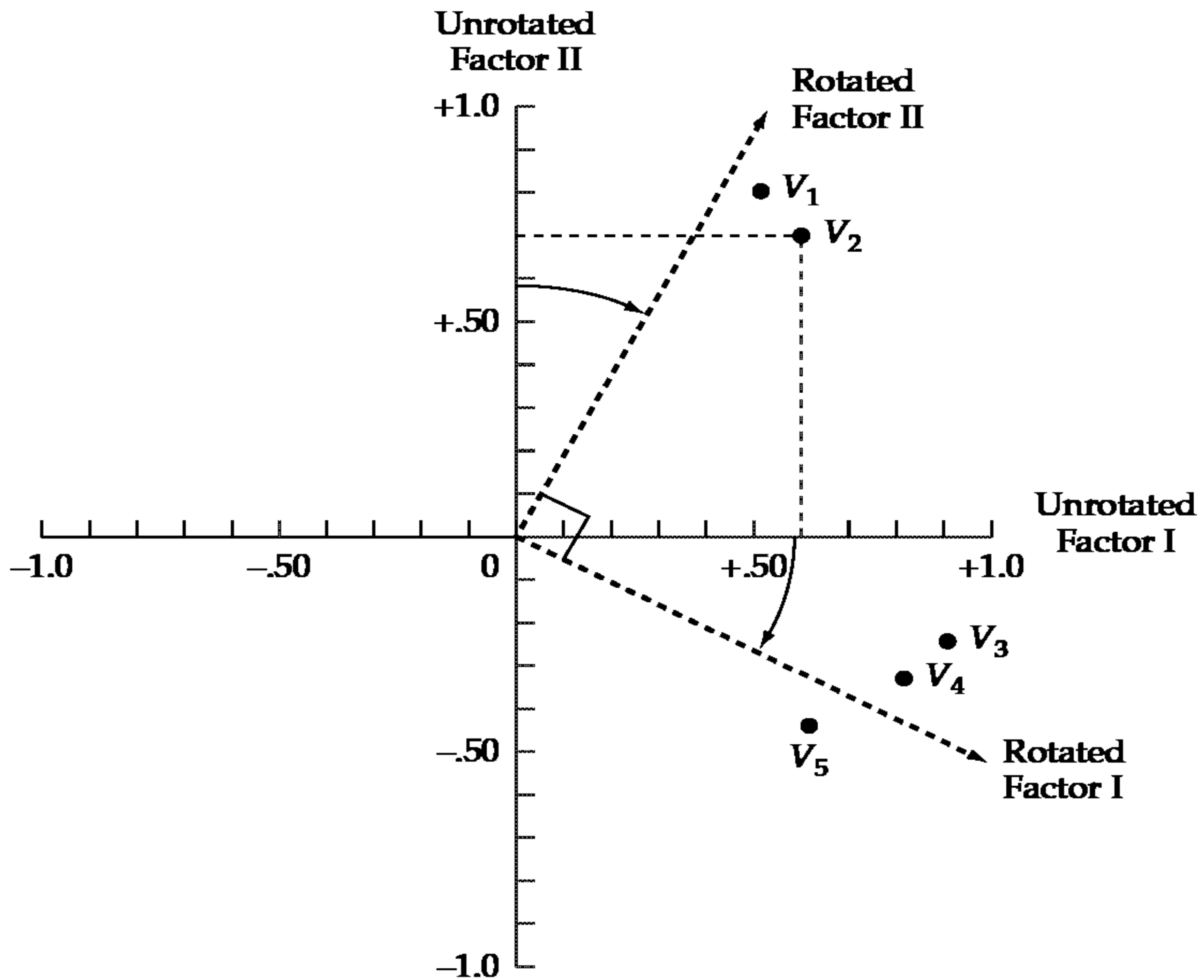


FIGURE 7 Orthogonal Factor Rotation

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

Variables	Unrotated Factor Loadings I	Unrotated Factor Loadings II	Rotated Factor Loadings I	Rotated Factor Loadings 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**

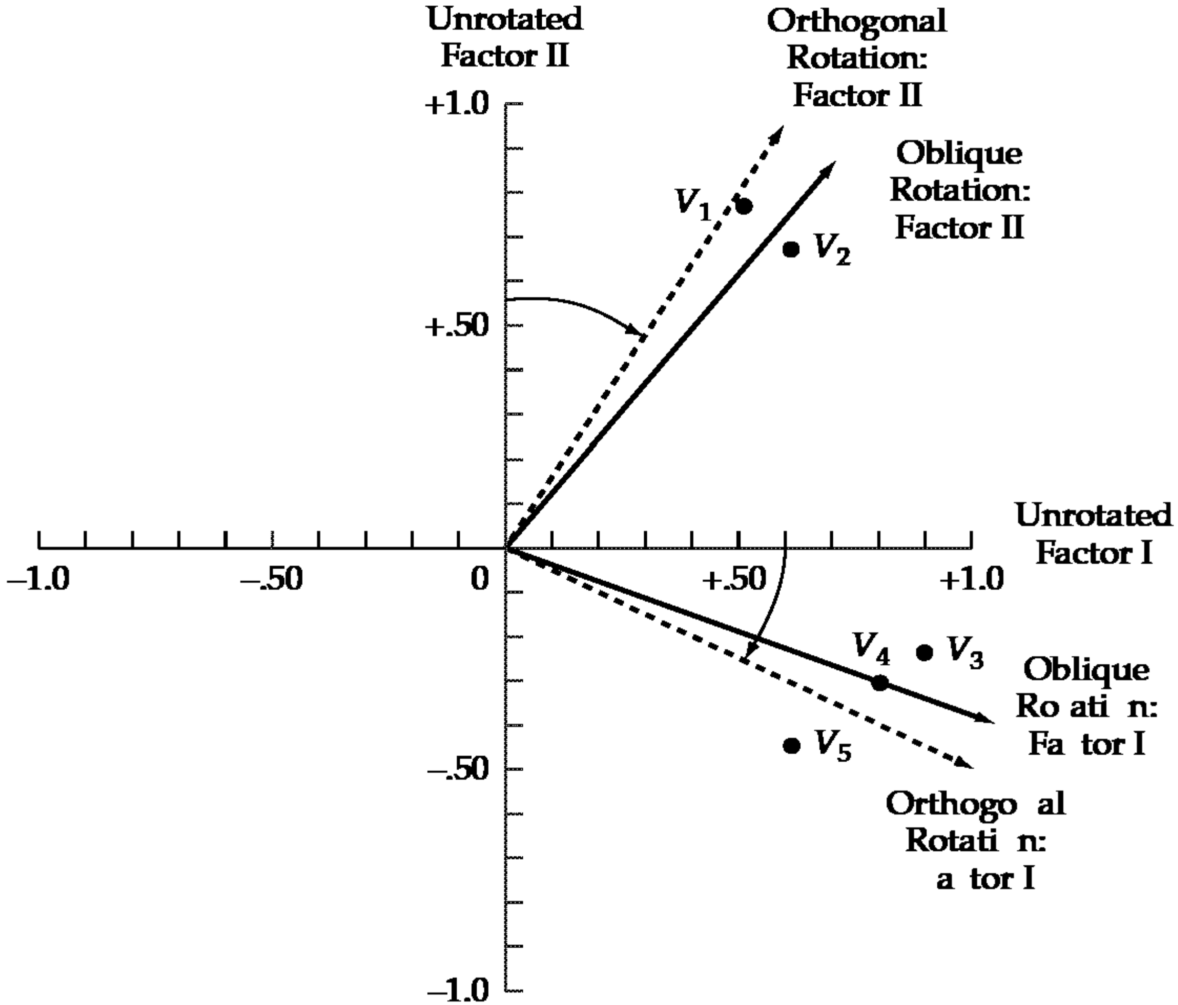


FIGURE 8 Oblique Factor Rotation

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

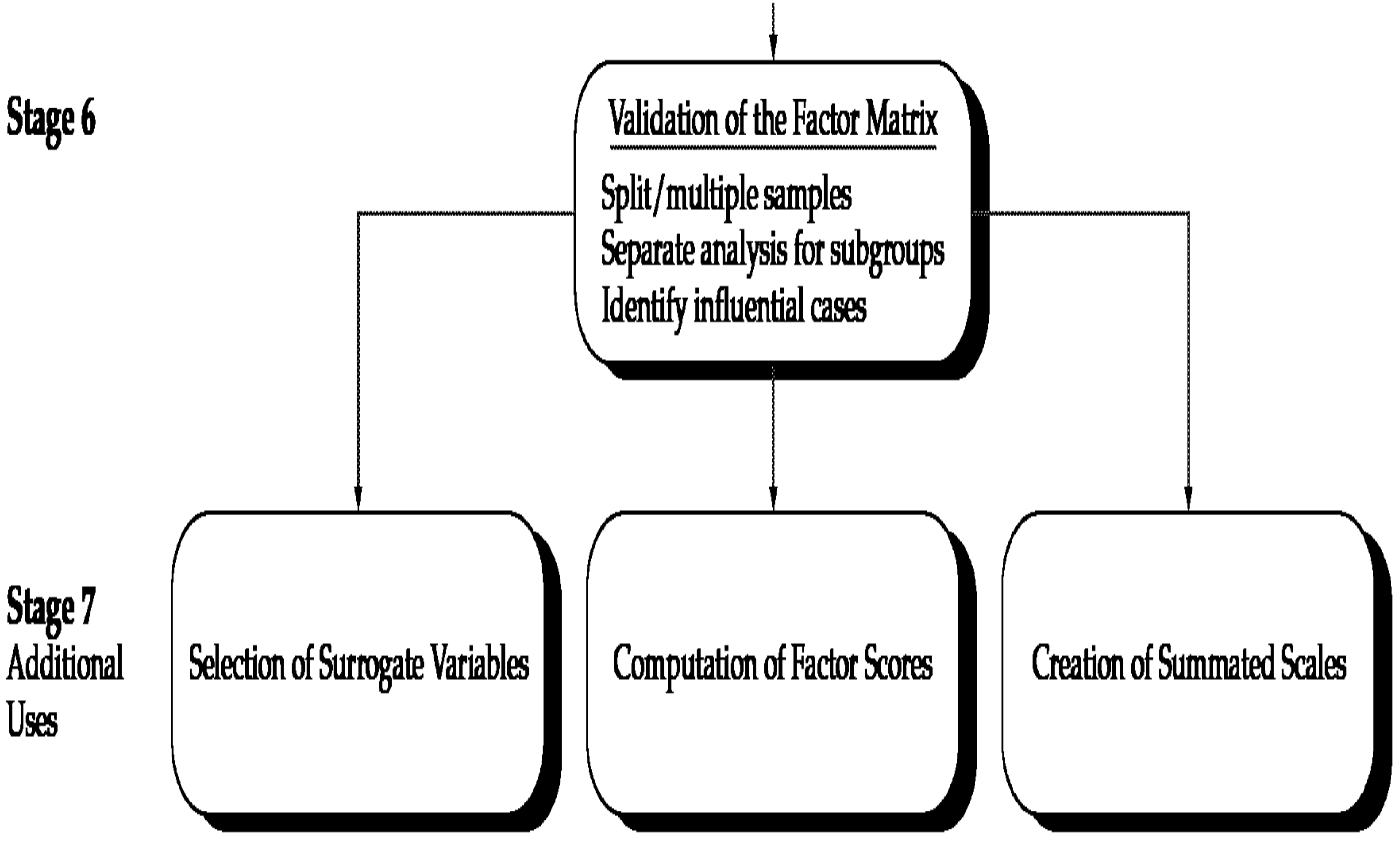
- **Stage 5 : Interpreting The Factors**
  - **Steps 1 and 2: Examine the Factor-Loading Matrix and Identify Significant Loadings.**
  - Given the sample size of 202, factor loadings of .40 and higher will be considered significant for interpretative purposes.
  - Using this threshold for the factor 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

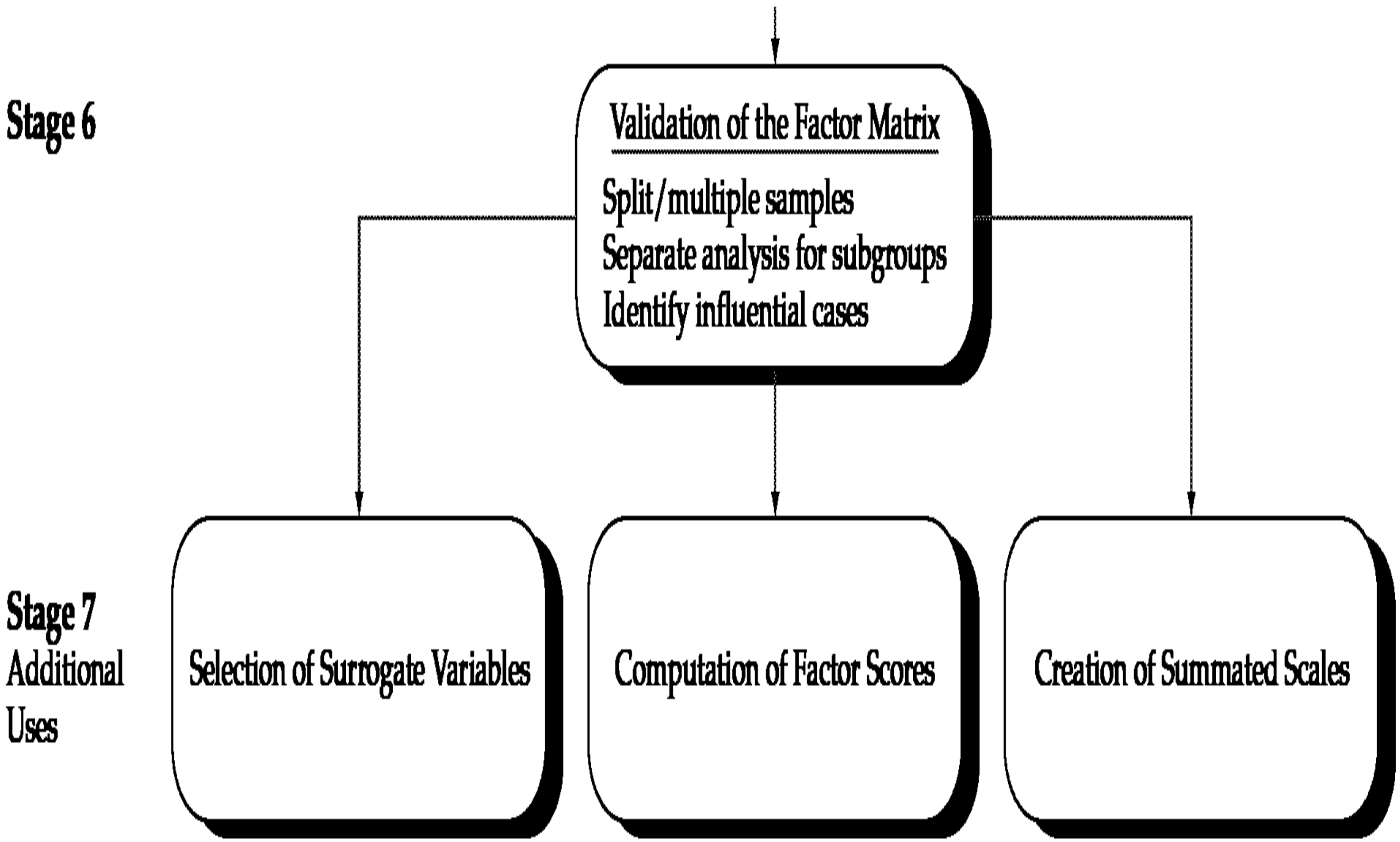
- **STAGE 6: VALIDATION OF FACTOR ANALYSIS**
  - The sixth stage involves assessing the degree of generalizability of the results to the population and the potential influence of individual cases or respondents on the overall results.
  - The issue of generalizability is critical for each of the multivariate 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.