Chapter 3
Exploratory Factor Analysis

Chapter 3 Exploratory Factor Analysis

LEARNING OBJECTIVES

Upon completing this chapter, you should be able to do the following:

1. Differentiate factor analysis techniques from other multivariate techniques.
2. Distinguish between exploratory and confirmatory uses of factor analytic techniques.
3. Understand the seven stages of applying factor analysis.
4. Distinguish between R and Q factor analysis.
5. Identify the differences between component analysis and common factor analysis models.

Chapter 3 Exploratory Factor Analysis

LEARNING OBJECTIVES continued . . .

Upon completing this chapter, you should be able to do the following:

6. Tell how to determine the number of factors to extract.

7. Explain the concept of rotation of factors.

8. Describe how to name a factor.
9. Explain the additional uses of factor analysis.
10. State the major limitations of factor analytic techniques.

Exploratory Factor Analysis Defined

Exploratory factor analysis . . . is an interdependence technique whose primary purpose is to define the underlying structure among the variables in the analysis.

What is Exploratory Factor Analysis?

Exploratory Factor Analysis . . .

- Examines the interrelationships among a large number of variables and then attempts to explain them in terms of their common underlying dimensions.
- These common underlying dimensions are referred to as factors.
- A summarization and data reduction technique that does not have independent and dependent variables, but is an interdependence technique in which all variables are considered simultaneously.

Correlation Matrix for Store Image Elements

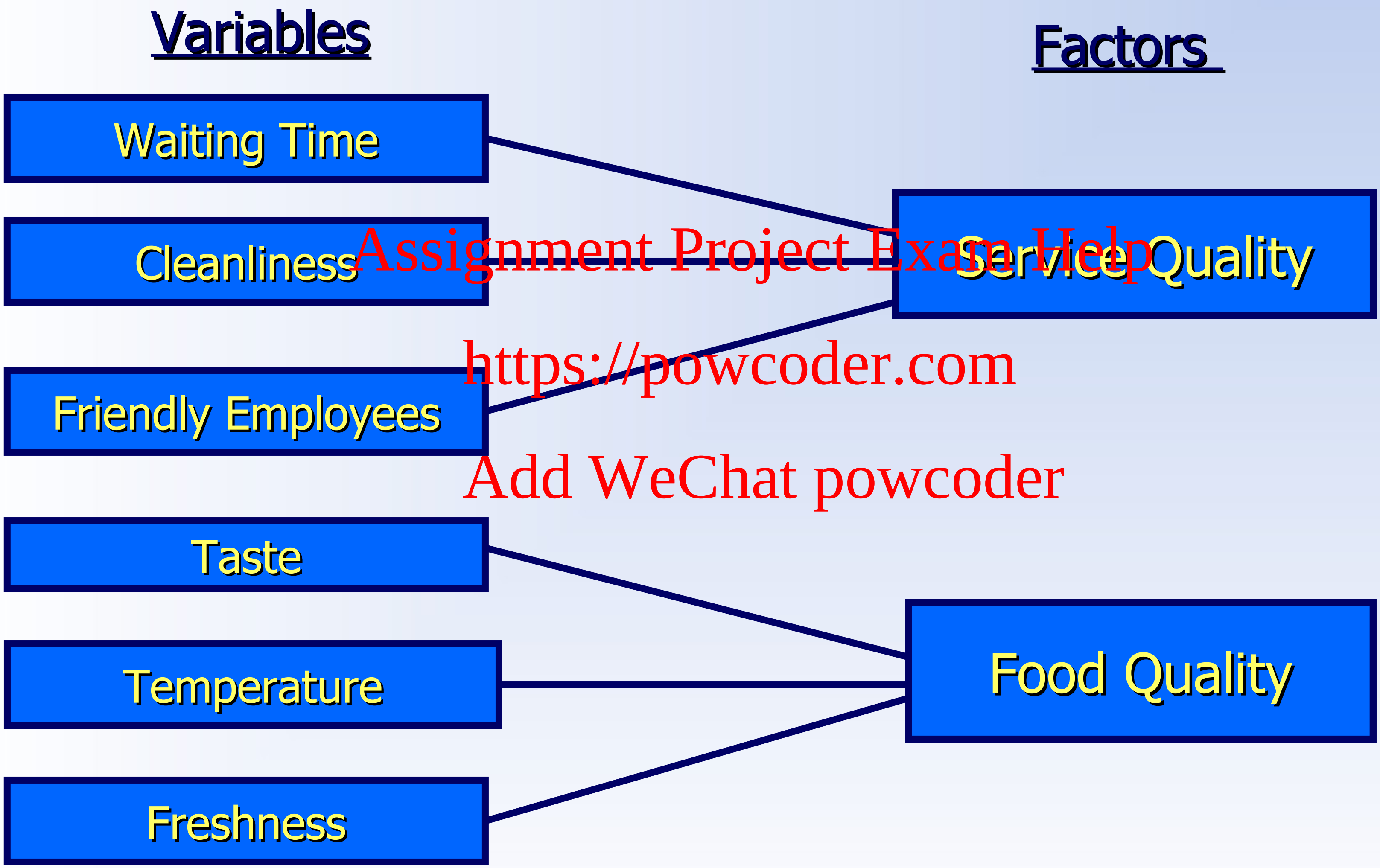
	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
V_1 Price Level	1.00								
V_2 Store Personnel	.427	1.00							
V_3 Return Policy	.302	.771	1.00						
V_4 Product Availability	.470	.497	.427	1.00					
V_5 Product Quality	.765	.406	.307	.472	1.00				
V_6 Assortment Depth	.281	.445	.423	.713	.325	1.00			
V_7 Assortment Width	.354	.490	.471	.719	.378	.724	1.00		
V_8 In-Store Service	.242	.719	.733	.428	.240	.311	.435	1.00	
V_9 Store Atmosphere	.372	.737	.774	.479	.326	.429	.466	.710	1.00

Correlation Matrix of Variables After Grouping Using Factor Analysis

	V_3	V_8	V_9	V_2	V_6	V_7	V_4	V_1	V_5
V_3 Return Policy	1.00								
V_8 In-store Service	.733	1.00							
V_9 Store Atmosphere	.774	.710	1.00						
V_2 Store Personnel	.741	.719	.787	1.00					
V_6 Assortment Depth	.423	.311	.429	.445	1.00				
V_7 Assortment Width	.471	.435	.468	.490	.724	1.00			
V_4 Product Availability	.427	.428	.479	.497	.713	.719	1.00		
V_1 Price Level	.302	.242	.372	.427	.281	.354	.470	1. 00	
V_5 Product Quality	.307	.240	.326	.406	.325	.378	.472	.765	1.00

Shaded areas represent variables likely to be grouped together by factor analysis.

Application of Factor Analysis to a Fast-Food Restaurant



Factor Analysis Decision Process

Stage 1: Objectives of Factor Analysis

Stage 2: Designing a Factor Analysis

Stage 3: Assumptions in Factor Analysis

Stage 4: Deriving Factors and Assessing Overall Fit

Stage 5: Interpreting the Factors

Stage 6: Validation of Factor Analysis

Stage 7: Additional uses of Factor Analysis Results

Stage 1: Objectives of Factor Analysis

1. Is the objective exploratory or confirmatory?
2. Specify the unit of analysis.
3. Data summarization and/or reduction?
4. Using factor analysis with other techniques.

Factor Analysis Outcomes

1. Data summarization = derives underlying dimensions that, when interpreted and understood, describe the data in a much smaller number of concepts than the original individual variables.
2. Data reduction = extends the process of data summarization by deriving an empirical value (factor score or summated scale) for each dimension (factor) and then substituting this value for the original values.

Types of Factor Analysis

1. Exploratory Factor Analysis (EFA) = is used to discover the factor structure of a construct and examine its reliability. It is data driven.

2. Confirmatory Factor Analysis (CFA) = is used to confirm the fit of the hypothesized factor structure to the observed (sample) data. It is theory driven.

Stage 2: Designing a Factor Analysis

Three Basic Decisions:

1. Calculation of input data – R vs. Q analysis.

2. Design of study in terms of number of variables, measurement properties of variables, and the type of variables.

3. Sample size necessary.

Rules of Thumb 3–1

Factor Analysis Design

- Factor analysis is performed most often only on metric variables, although specialized methods exist for the use of dummy variables. A small number of "dummy variables" can be included in a set of metric variables that are factor analyzed.
- If a study is being designed to reveal factor structure, strive to have at least five variables for each proposed factor.
- For sample size:
  - the sample must have more observations than variables.
  - the minimum absolute sample size should be 50 observations.
- Maximize the number of observations per variable, with a minimum of five and hopefully at least ten observations per variable.

Stage 3: Assumptions in Factor Analysis

Three Basic Decisions . . .

1. Calculation of input data – R vs. Q analysis.
2. Design of study in terms of number of variables, measurement properties of variables, and the type of variables.
3. Sample size required.

Assumptions

- Multicollinearity
  - Assessed using MSA (measure of sampling adequacy).

The MSA is measured by the Kaiser-Meyer-Olkin (KMO) statistic. As a measure of sampling adequacy, the KMO predicts if data are likely to factor well based on correlation and partial correlation. KMO can be used to identify which variables to drop from the factor analysis because they lack multicollinearity.

There is a KMO statistic for each individual variable, and their sum is the KMO overall statistic. KMO varies from 0 to 1.0. Overall KMO should be .50 or higher to proceed with factor analysis. If it is not, remove the variable with the lowest individual KMO statistic value one at a time until KMO overall rises above .50, and each individual variable KMO is above .50.

- Homogeneity of sample factor solutions

Rules of Thumb 3–2

Testing Assumptions of Factor Analysis

- There must be a strong conceptual foundation to support the assumption that a structure does exist before the factor analysis is performed.

- A statistically significant Bartlett's test of sphericity (sig. < .05) indicates that sufficient correlations exist among the variables to proceed.
- Measure of Sampling Adequacy (MSA) values must exceed .50 for both the overall test and each individual variable. Variables with values less than .50 should be omitted from the factor analysis one at a time, with the smallest one being omitted each time.

Stage 4: Deriving Factors and Assessing Overall Fit

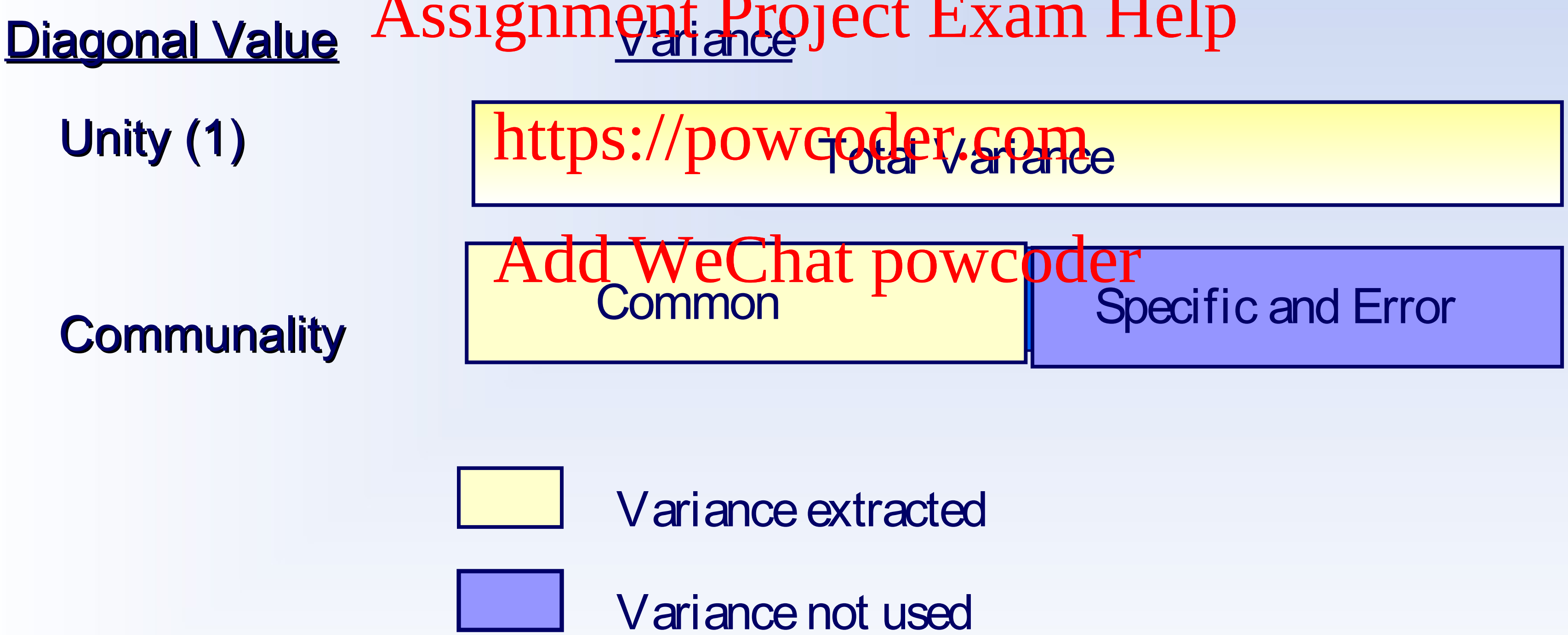
- Selecting the factor extraction method
  - common vs. component analysis.
- Determining the number of factors to represent the data.

Extraction Decisions

- Which method?
  - Principal Components Analysis
  - Common Factor Analysis

- How to rotate?
  - Orthogonal or Oblique rotation

Extraction Method Determines the Types of Variance Carried into the Factor Matrix



Principal Components vs. Common?

Two Criteria . . .

- Objectives of the factor analysis.

- Amount of prior knowledge about the variance in the variables.

Number of Factors?

- A Priori Criterion
- Latent Root Criterion
- Percentage of Variance
- Scree Test Criterion

Eigenvalue Plot for Scree Test Criterion

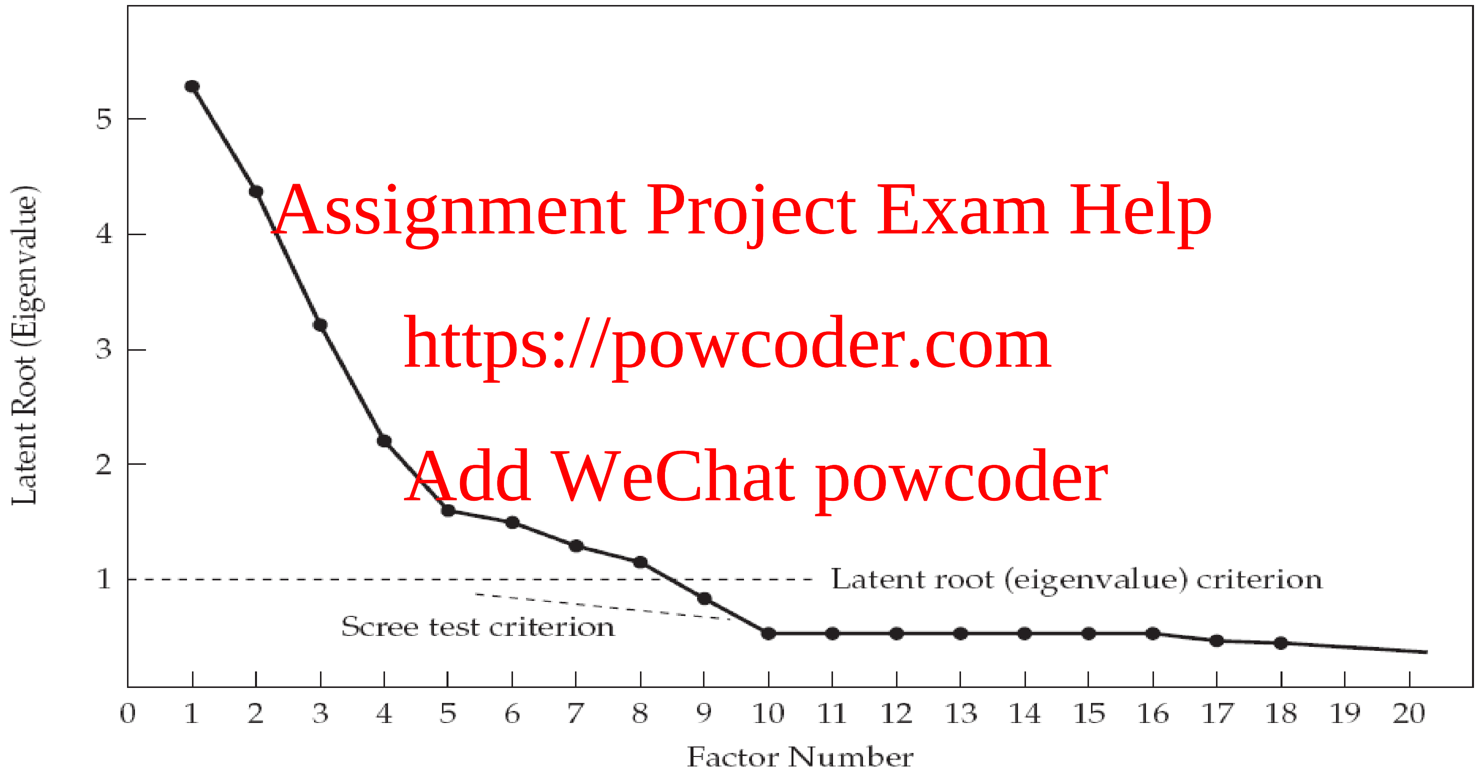


FIGURE 3-6 Eigenvalue Plot for Scree Test Criterion

Rules of Thumb 3–3

Choosing Factor Models and Number of Factors

- Although both component and common factor analysis models yield similar results in common research settings (30 or more variables or communalities of .60 for most variables):
  - ✓ the component analysis model is most appropriate when data reduction is paramount.
  - ✓ the common factor model is best in well-specified theoretical applications.
- Any decision on the number of factors to be retained should be based on several considerations:
  - ✓ use of several stopping criteria to determine the initial number of factors to retain.
  - ✓ Factors With Eigenvalues greater than 1.0.
  - ✓ A pre-determined number of factors based on research objectives and/or prior research.
  - ✓ Enough factors to meet a specified percentage of variance explained, usually 60% or higher.
  - ✓ Factors shown by the scree test to have substantial amounts of common variance (i.e., factors before inflection point).
  - ✓ More factors when there is heterogeneity among sample subgroups.
- Consideration of several alternative solutions (one more and one less factor than the initial solution) to ensure the best structure is identified.

Processes of Factor Interpretation

- Estimate the Factor Matrix
- Factor Rotation
- Factor Interpretation
- Respecification of factor model, if needed, may involve . . .
  - Deletion of variables from analysis
  - Desire to use a different rotational approach
  - Need to extract a different number of factors
  - Desire to change method of extraction

Rotation of Factors

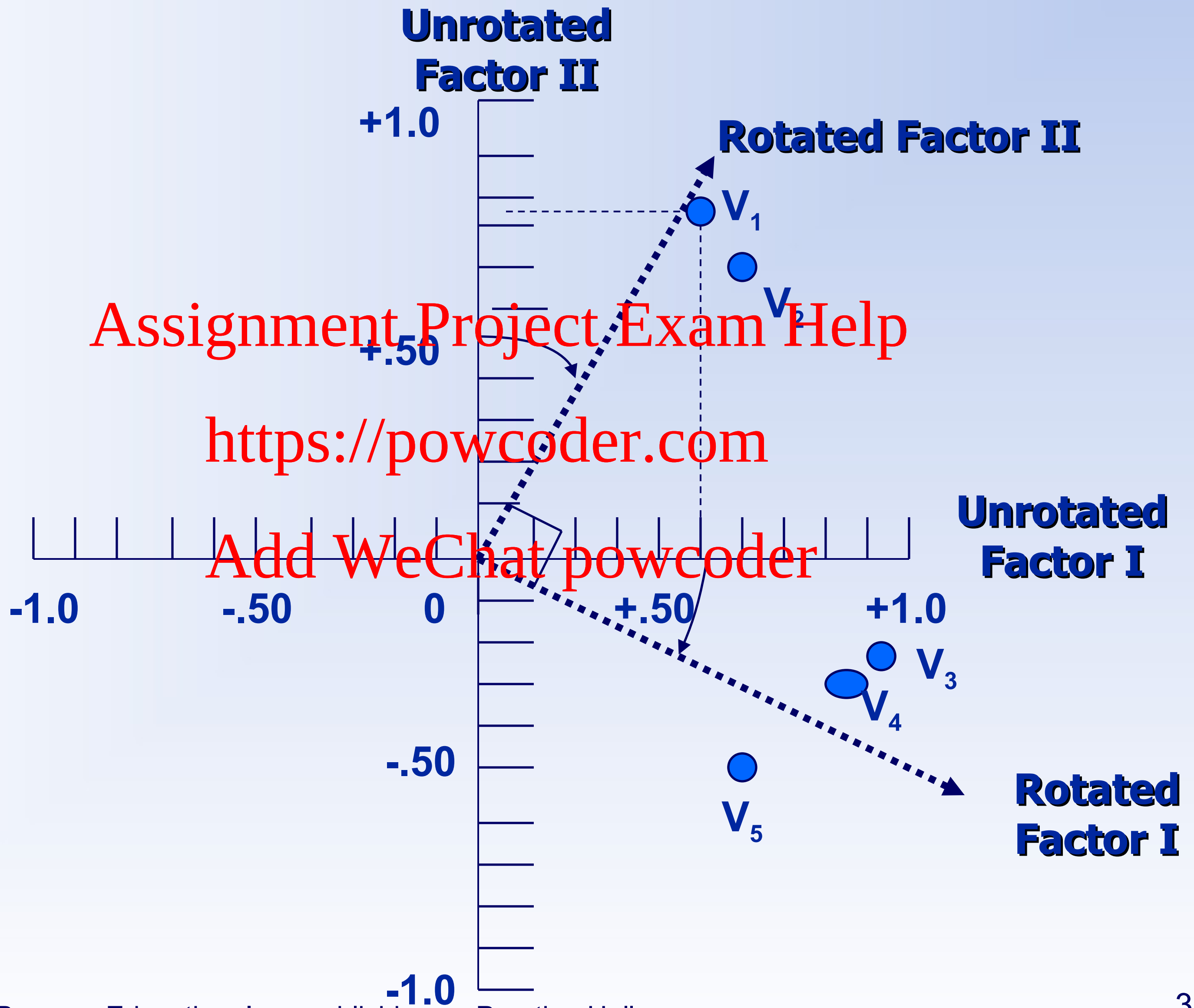
Factor rotation = the reference axes of the factors are turned about the origin until some other position has been reached. Since unrotated factor solutions extract factors based on how much variance they account for, with each subsequent factor accounting for less variance. The ultimate effect of rotating the factor matrix is to redistribute the variance from earlier factors to later ones to achieve a simpler, theoretically more meaningful factor pattern.

Two Rotational Approaches

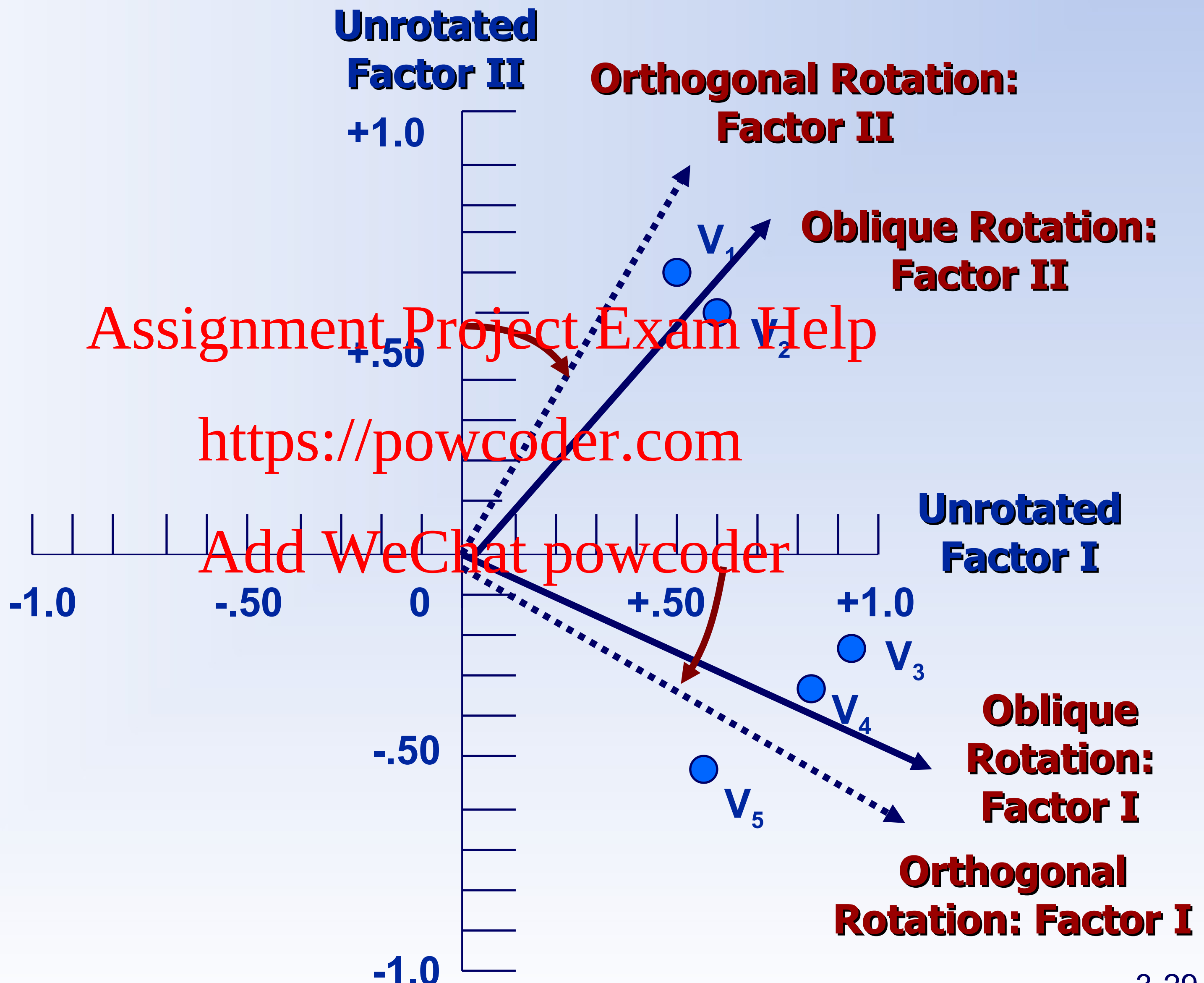
1. Orthogonal = axes are maintained at 90 degrees.

2. Oblique = axes are not maintained at 90 degrees.

Orthogonal Factor Rotation



Oblique Factor Rotation



Orthogonal Rotation Methods

- Quartimax (simplify rows)

- Varimax (simplify columns)

- Equimax (combination)

Rules of Thumb 3–4

Choosing Factor Rotation Methods

- Orthogonal rotation methods . . .
  - are the most widely used rotational methods.

  - are The preferred method when the research goal is data reduction to either a smaller number of variables or a set of uncorrelated measures for subsequent use in other multivariate techniques.

- Oblique rotation methods . . .
  - best suited to the goal of obtaining several theoretically meaningful factors or constructs because, realistically, very few constructs in the "real world" are uncorrelated.

Which Factor Loadings Are Significant?

- Customary Criteria = Practical Significance.

- Sample Size & Statistical Significance.

- Number of Factors ($\uparrow$ = >) and/or Variables ($\uparrow$ = <) .

Guidelines for Identifying Significant Factor Loadings Based on Sample Size

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

*Significance is based on a .05 significance level (a), a power level of 80 percent, and standard errors assumed to be twice those of conventional correlation coefficients.

Rules of Thumb 3–5

Assessing Factor Loadings

- While factor loadings of ±.30 to ±.40 are minimally acceptable, values greater than ± .50 are considered necessary for practical significance.
- To be considered significant:
  - A smaller loading is needed given either a larger sample size, or a larger number of variables being analyzed.
  - A larger loading is needed given a factor solution with a larger number of factors, especially in evaluating the loadings on later factors.
- Statistical tests of significance for factor loadings are generally very conservative and should be considered only as starting points needed for including a variable for further consideration.

Stage 5: Interpreting the Factors

- Selecting the factor extraction method – common vs. component analysis.
- Determining the number of factors to represent the data.

Interpreting a Factor Matrix:

1. Examine the factor matrix of loadings.
2. Identify the highest loading across all factors for each variable.
3. Assess communalities of the variables.
4. Label the factors.

Rules of Thumb 3–6

Interpreting The Factors

- **An optimal structure exists when all variables have high loadings only on a single factor.**
- **Variables that cross-load (load highly on two or more factors) are usually deleted unless theoretically justified or the objective is strictly data reduction.**
- **Variables should generally have communalities of greater than .50 to be retained in the analysis.**
- **Respecification of a factor analysis can include options such as:**
  - **deleting a variable(s),**
  - **changing rotation methods, and/or**
  - **increasing or decreasing the number of factors.**

Stage 6: Validation of Factor Analysis

- Confirmatory Perspective.
- Assessing Factor Structure Stability.
- Detecting Influential Observations.

Stage 7: Additional Uses of Factor Analysis Results

- Selecting Surrogate Variables

- Creating Summated Scales

- Computing Factor Scores

Rules of Thumb 3–7
Summated Scales

- A summated scale is only as good as the items used to represent the construct. While it may pass all empirical tests, it is useless without theoretical justification.
- Never create a summated scale without first assessing its unidimensionality with exploratory or confirmatory factor analysis.
- Once a scale is deemed unidimensional, its reliability score, as measured by Cronbach's alpha:
    - should exceed a threshold of .70, although a .60 level can be used in exploratory research.
    - the threshold should be raised as the number of items increases, especially as the number of items approaches 10 or more.
- With reliability established, validity should be assessed in terms of:
    - convergent validity = scale correlates with other like scales.
    - discriminant validity = scale is sufficiently different from other related scales.
    - nomological validity = scale "predicts" as theoretically suggested.

Rules of Thumb 3–8

Representing Factor Analysis In Other Analyses

- <u>The single surrogate variable</u>:
  - ✓ Advantages: simple to administer and interpret.
  - ✓ Disadvantages:
    1) does not represent all "facets" of a factor
    2) prone to measurement error.

- <u>Factor scores</u>:
  - ✓ Advantages:
    1) represents all variables loading on the factor,
    2) best method for complete data reduction.
    3) Are by default orthogonal and can avoid complications caused by multicollinearity.
  - ✓ Disadvantages:
    1) interpretation more difficult since all variables contribute through loadings
    2) Difficult to replicate across studies.

Rules of Thumb 3–8 Continued . . .

Representing Factor Analysis In Other Analyses

- Summated scales:
  - ✓ Advantages:
    1) compromise between the surrogate variable and factor score options.
    2) reduces measurement error.
    3) represents multiple facets of a concept.
    4) easily replicated across studies.
  - ✓ Disadvantages:
    1) includes only the variables that load highly on the factor and excludes those having little or marginal impact.
    2) not necessarily orthogonal.
    3) Require extensive analysis of reliability and validity issues.

Description of HBAT Primary Database Variables

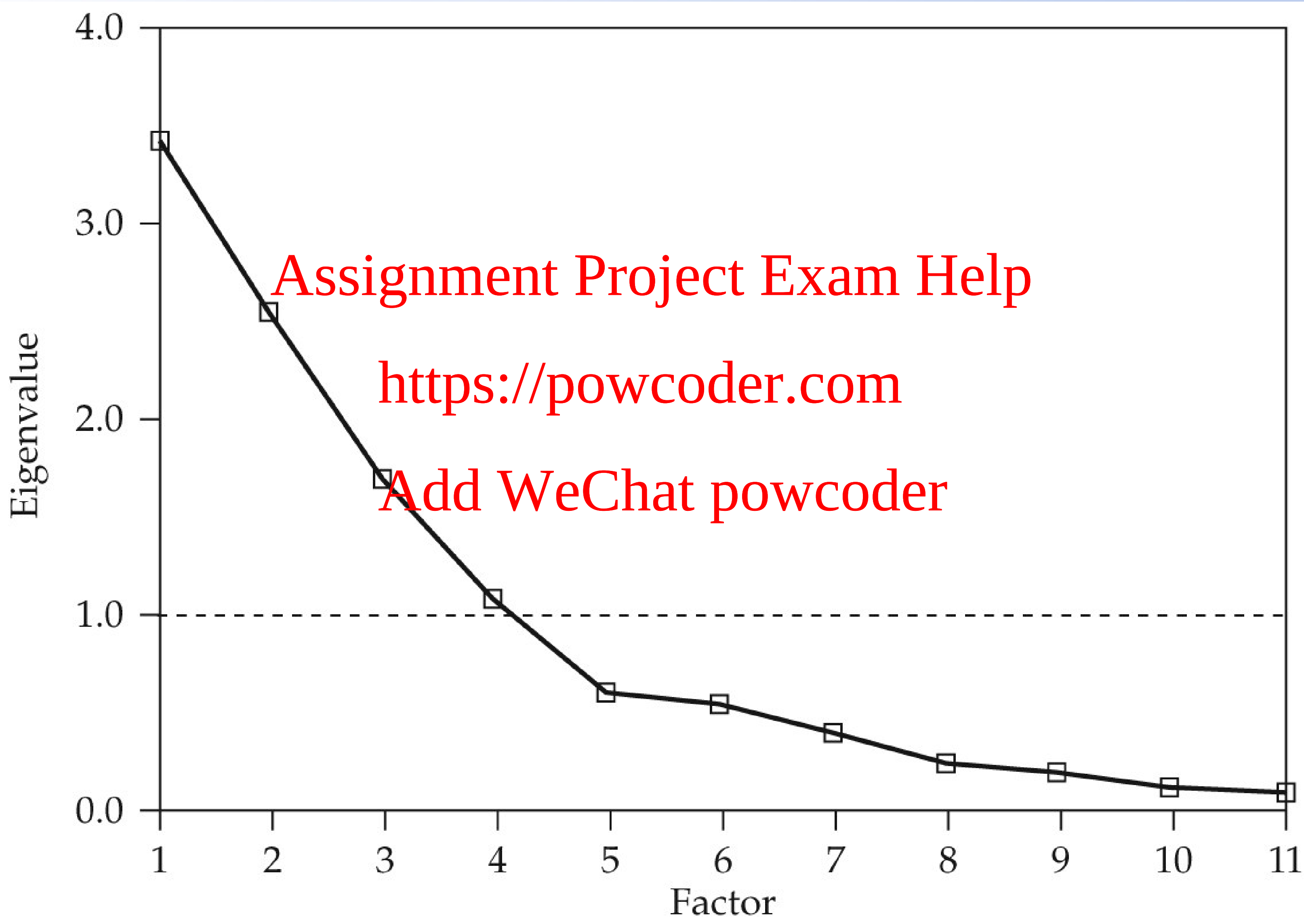
	Variable Description	Variable Type
Data Warehouse Classification Variables		
X1	Customer Type	nonmetric
X2	Industry Type	nonmetric
X3	Firm Size	nonmetric
X4	Region	nonmetric
X5	Distribution System	nonmetric
Performance Perceptions Variables		
X6	Product Quality	metric
X7	E-Commerce Activities/Website	metric
X8	Technical Support	metric
X9	Complaint Resolution	metric
X10	Advertising	metric
X11	Product Line	metric
X12	Salesforce Image	metric
X13	Competitive Pricing	metric
X14	Warranty & Claims	metric
X15	New Products	metric
X16	Ordering & Billing	metric
X17	Price Flexibility	metric
X18	Delivery Speed	metric
Outcome/Relationship Measures		
X19	Satisfaction	metric
X20	Likelihood of Recommendation	metric
X21	Likelihood of Future Purchase	metric
X22	Current Purchase/Usage Level	metric
X23	Consider Strategic Alliance/Partnership in Future	nonmetric

Rotated Component Matrix
"Reduced Set" of HBAT Perceptions Variables

	Component 1	2	3	4	Communality
X9 – Complaint Resolution .933				.890	
X18 – Delivery Speed	.931				.894
X16 – Order & Billing	.886				.806
X12 – Salesforce Image		.898			.860
X7 – E-Commerce Activities		.868			.780
X10 – Advertising	.743			.585	
X8 – Technical Support			.940		.894
X14 – Warranty & Claims			.933		.891
X6 – Product Quality				.892	.798
X13 – Competitive Pricing			-.730	.661	
Sum of Squares	2.589	2.216	1.846	1.406	8.057
Percentage of Trace	25.893	22.161	18.457	14.061	80.572

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax.

Scree Test for HBAT Component Analysis



Factor Analysis Learning Checkpoint

1. What are the major uses of factor analysis?
2. What is the difference between component analysis and common factor analysis?

3. Is rotation of factors necessary?

4. How do you decide how many factors to extract?
5. What is a significant factor loading?

6. How and why do you name a factor?
7. Should you use factor scores or summated ratings in follow-up analyses?