Social Media Marketing Research
社會媒體行銷研究

Exploratory Factor Analysis

1002SMMR08
TMIXM1A
Thu 7,8 (14:10-16:00) L511

Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授
Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系

http://mail. tku.edu.tw/myday/

2012-04-26

課程大綱（Syllabus)

週次	日期	內容（Subject/Topics）
1	101/02/16	Course Orientation of Social Media Marketing Research
2	101/02/23	Social Media: Facebook, Youtube, Blog, Microblog
3	101/03/01	Social Media Marketing
4	101/03/08	Marketing Research
5	101/03/15	Marketing Theories
6	101/03/22	Measuring the Construct
7	101/03/29	Measurement and Scaling
8	101/04/05	教學行政觀摩日 (--No Class--)
9	101/04/12	Paper Reading and Discussion

課程大綱 (Syllabus)

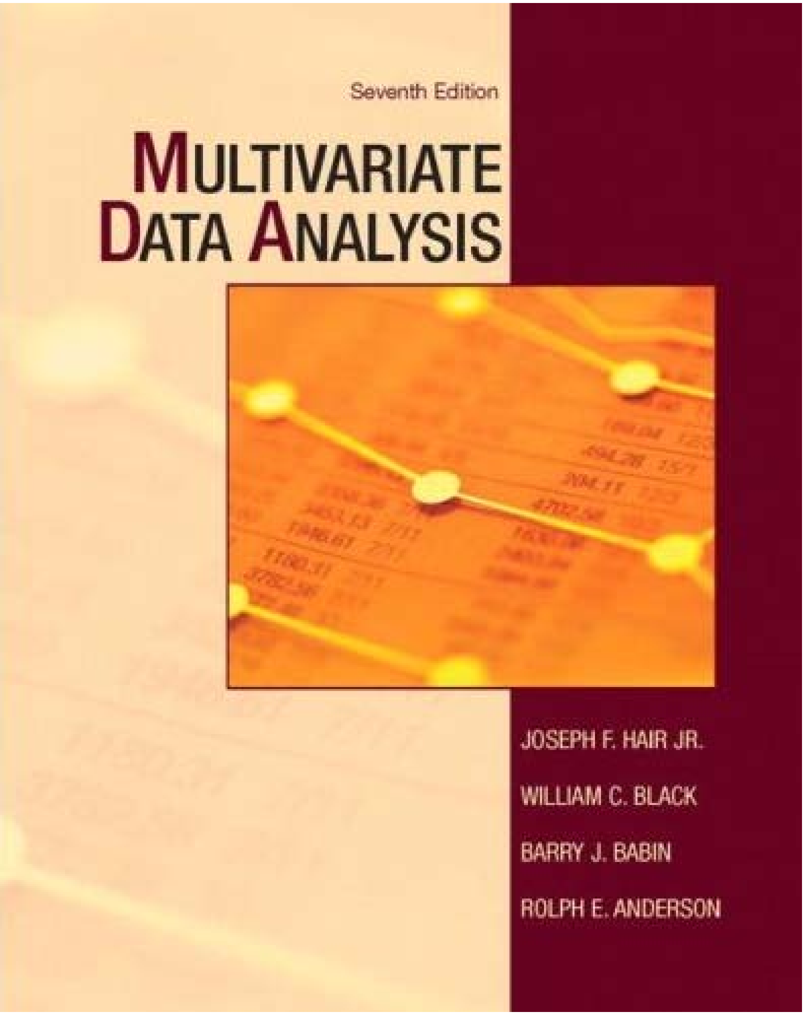
週次	日期	內容（Subject/Topics）
10	101/04/19	Midterm Presentation
11	101/04/26	Exploratory Factor Analysis
12	101/05/03	Paper Reading and Discussion
13	101/05/10	Confirmatory Factor Analysis
14	101/05/17	Paper Reading and Discussion
15	101/05/24	Communicating the Research Results
16	101/05/31	Paper Reading and Discussion
17	101/06/07	Term Project Presentation 1
18	101/06/14	Term Project Presentation 2

Outline

- Seven stages of applying factor analysis
- Exploratory Factor Analysis (EFA) vs. Confirmatory Factor Analysis (CFA)
- Identify the differences between component analysis and common factor analysis models
- How to determine the number of factors to extract
- How to name a factor

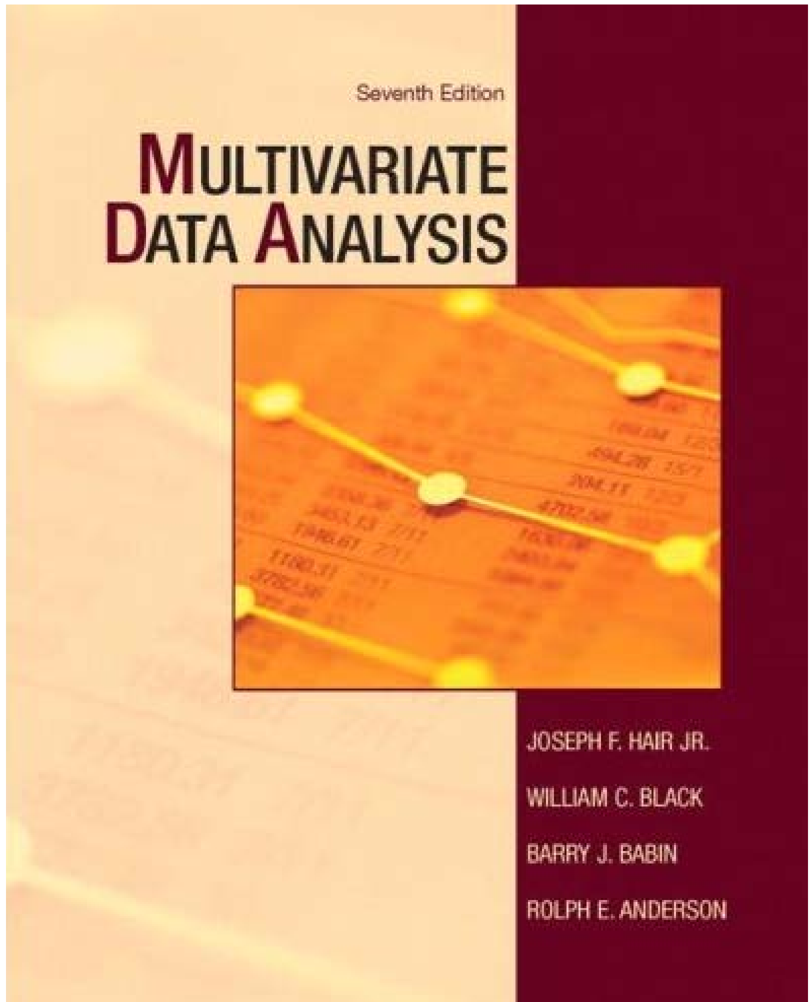
Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Joseph F. Hair, William C. Black, Barry J. Babin, Rolph E. Anderson, Multivariate Data Analysis, 7th Edition, Prentice Hall, 2009



Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Chapter 3
Exploratory Factor Analysis



Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Exploratory Factor Analysis (EFA)

- Definition
  - Exploratory factor analysis (EFA) is an interdependence technique whose primary purpose is to define the underlying structure among the variables in the analysis.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Exploratory Factor Analysis (EFA)

- 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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

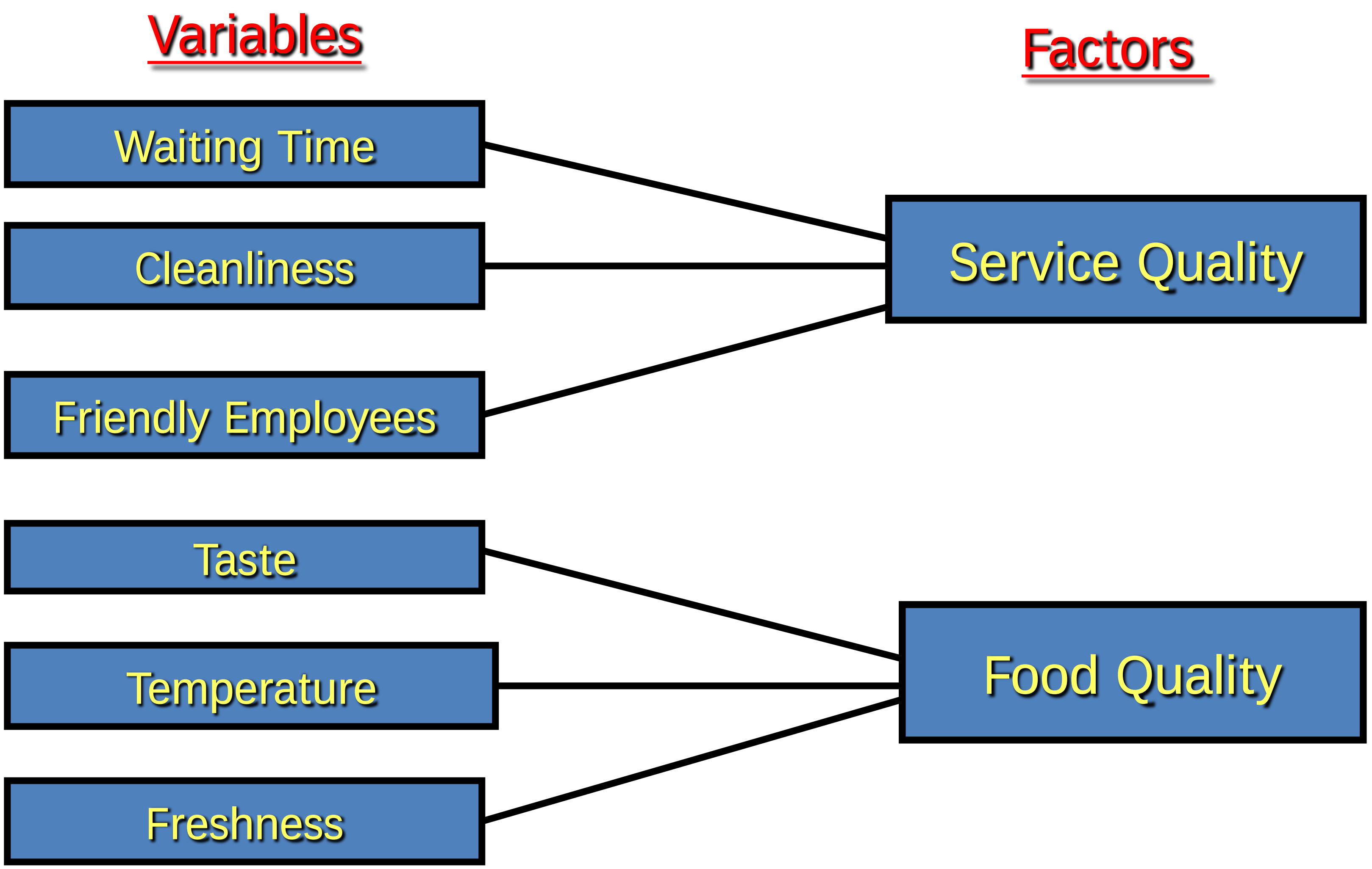
Correlation Matrix of Variables After Grouping Using Factor Analysis

	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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Application of Factor Analysis to a Fast-Food Restaurant



Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Factor Analysis Outcomes

- 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.
- 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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Types of Factor Analysis

- Exploratory Factor Analysis (EFA)
  – is used to discover the factor structure of a construct and examine its reliability.
  It is data driven.
- Confirmatory Factor Analysis (CFA)
  – is used to confirm the fit of the hypothesized factor structure to the observed (sample) data.
  It is theory driven.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

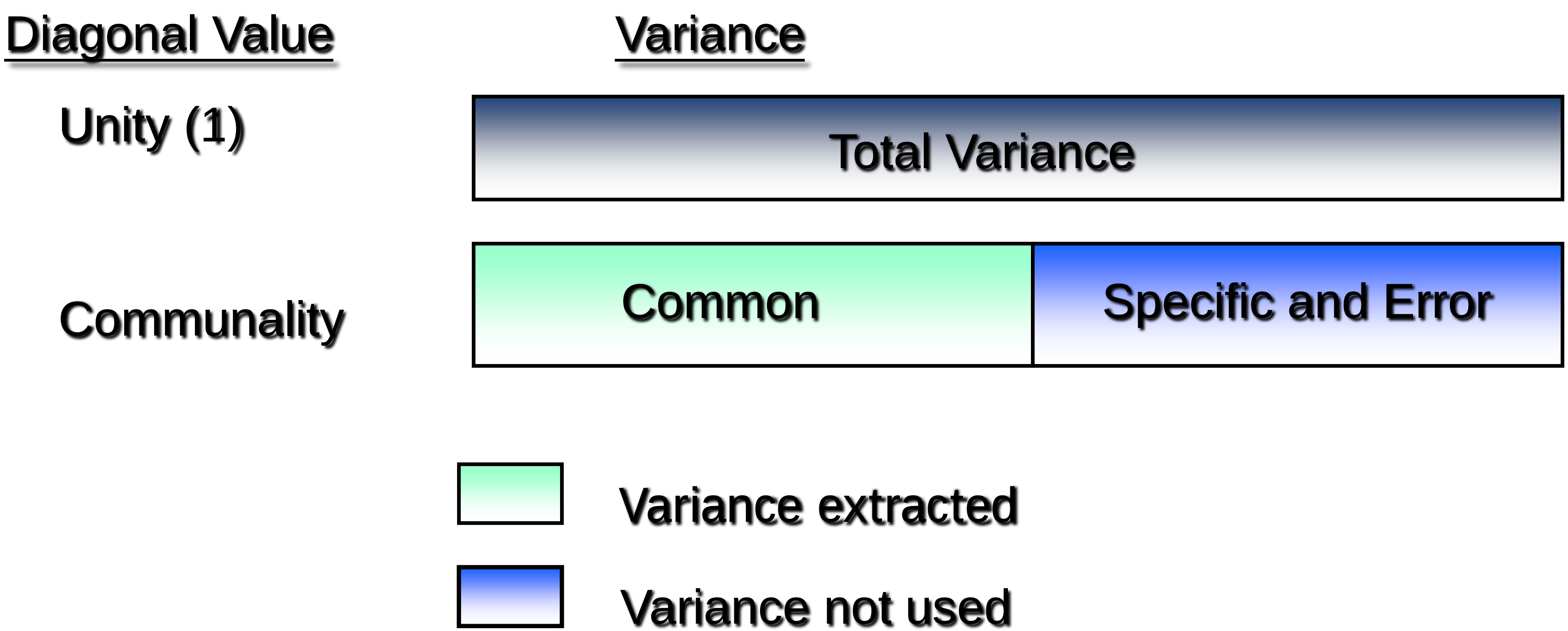
Extraction Decisions

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

- How to rotate?
  - Orthogonal or Oblique rotation

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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



Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Principal Components vs. Common?

- Two Criteria
  - Objectives of the factor analysis.
  - Amount of prior knowledge about the variance in the variables.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Number of Factors?

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

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Eigenvalue Plot for Scree Test Criterion

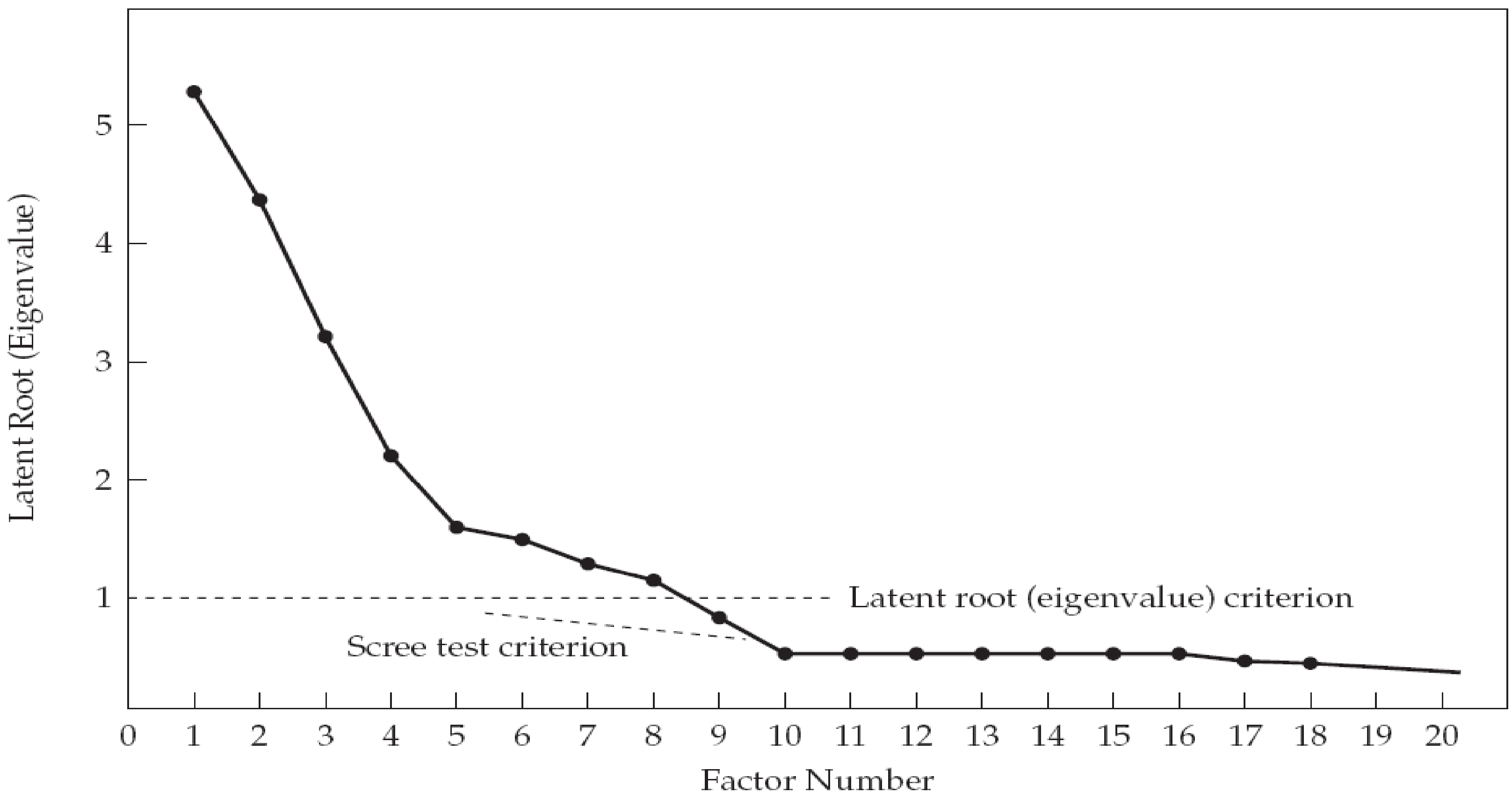


FIGURE 3-6 Eigenvalue Plot for Scree Test Criterion

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

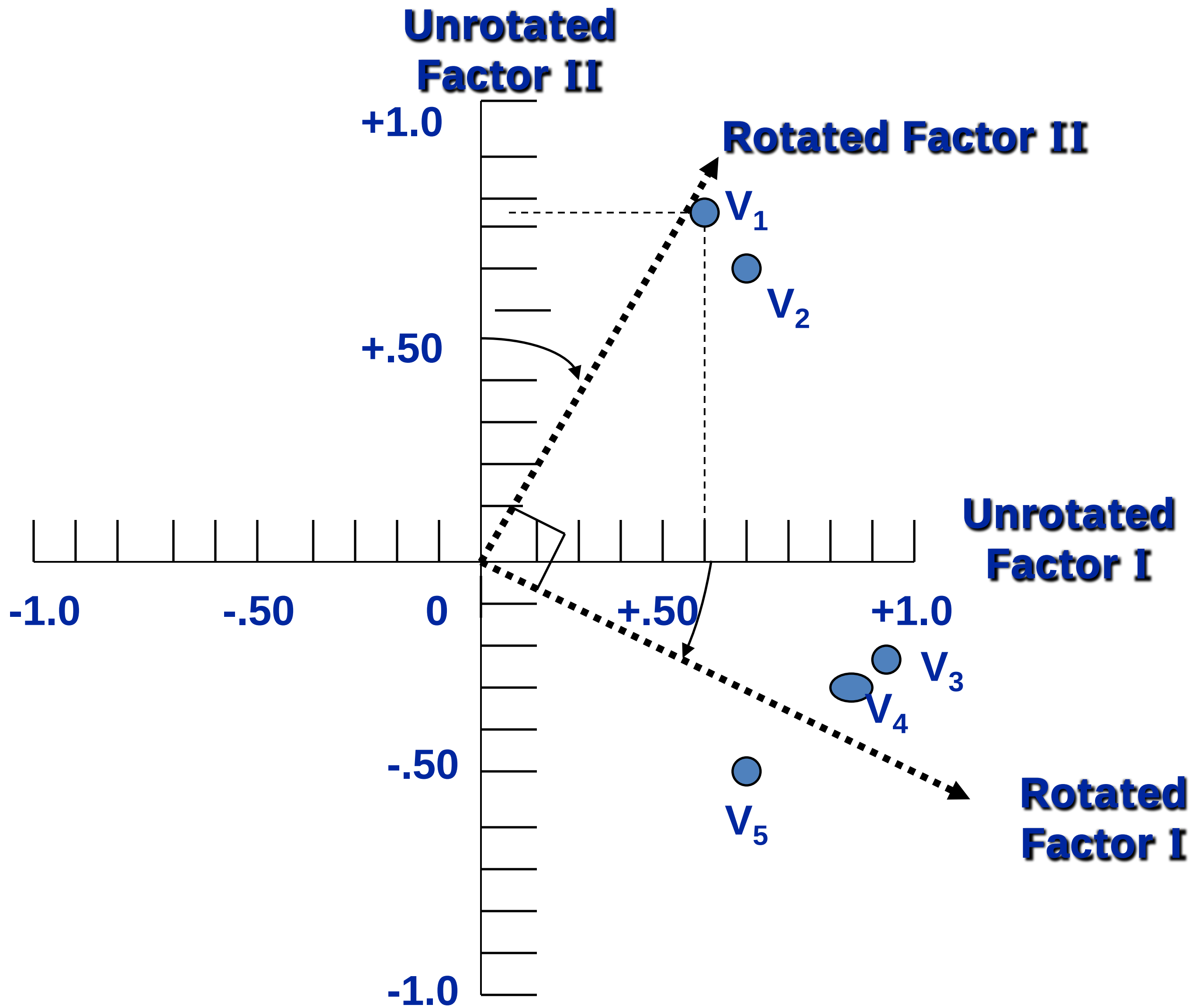
Two Rotational Approaches

1. Orthogonal
   – axes are maintained at 90 degrees.

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

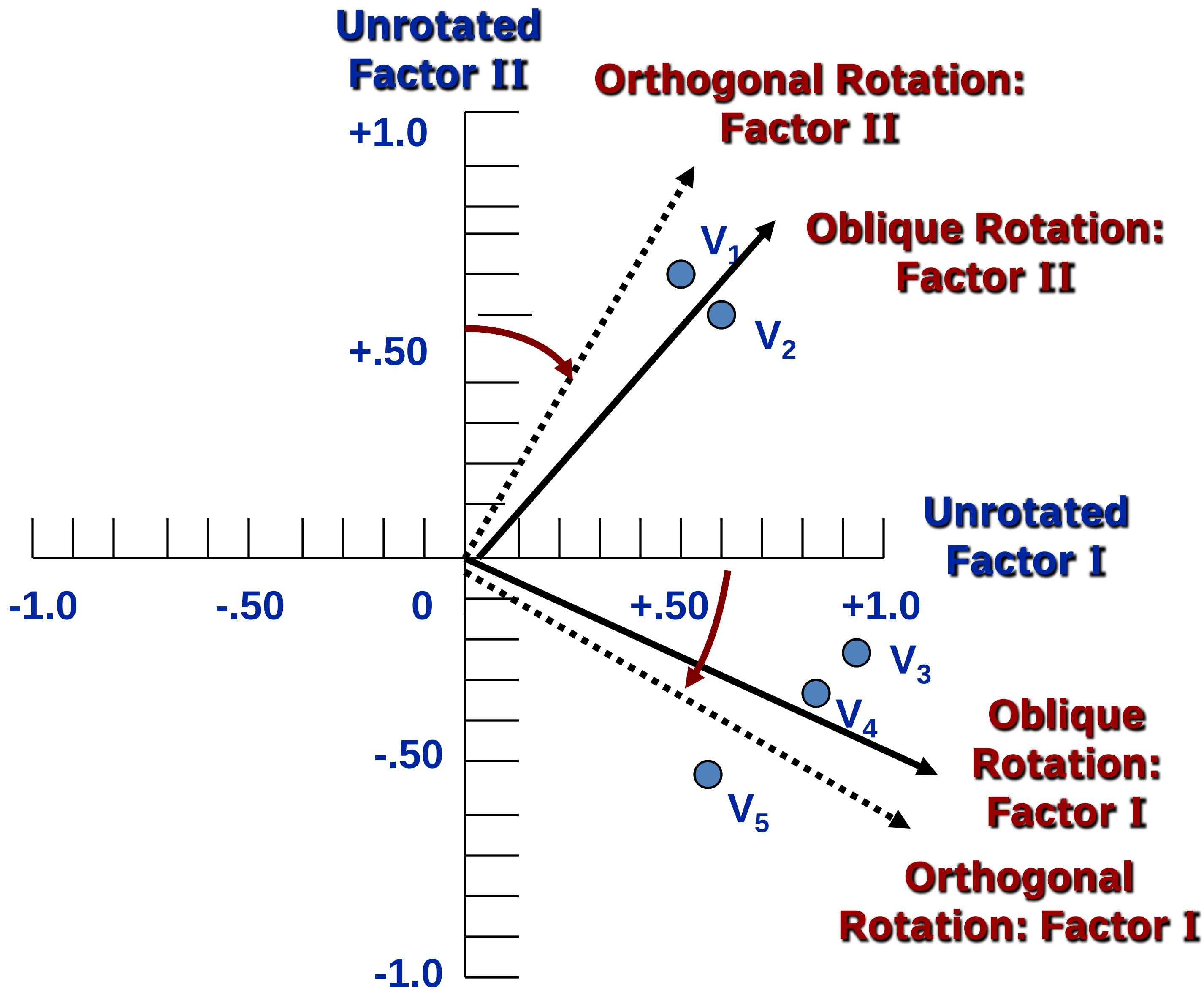
Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Orthogonal Factor Rotation



Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Oblique Factor Rotation



Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Orthogonal Rotation Methods

- Quartimax (simplify rows)
- Varimax (simplify columns)
- Equimax (combination)

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Which Factor Loadings Are Significant?

- Customary Criteria = Practical Significance.
- Sample Size & Statistical Significance.
- Number of Factors (⇧ = >) and/or Variables (⇧ = <) .

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Stage 5: Interpreting the Factors

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

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Stage 6: Validation of Factor Analysis

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

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Stage 7: Additional Uses of Factor Analysis Results

- Selecting Surrogate Variables
- Creating Summated Scales
- Computing Factor Scores

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Rules of Thumb 3–8

Representing Factor Analysis In Other Analyses

- The single surrogate variable:
  - Advantages: simple to administer and interpret.
  - Disadvantages:
    - does not represent all “facets” of a factor
    - prone to measurement error.
- Factor scores:
  - Advantages:
    - represents all variables loading on the factor,
    - best method for complete data reduction.
    - Are by default orthogonal and can avoid complications caused by multicollinearity.
  - Disadvantages:
    - interpretation more difficult since all variables contribute through loadings
    - Difficult to replicate across studies.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Rules of Thumb 3–8 (cont.)

Representing Factor Analysis In Other Analyses

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

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

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

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

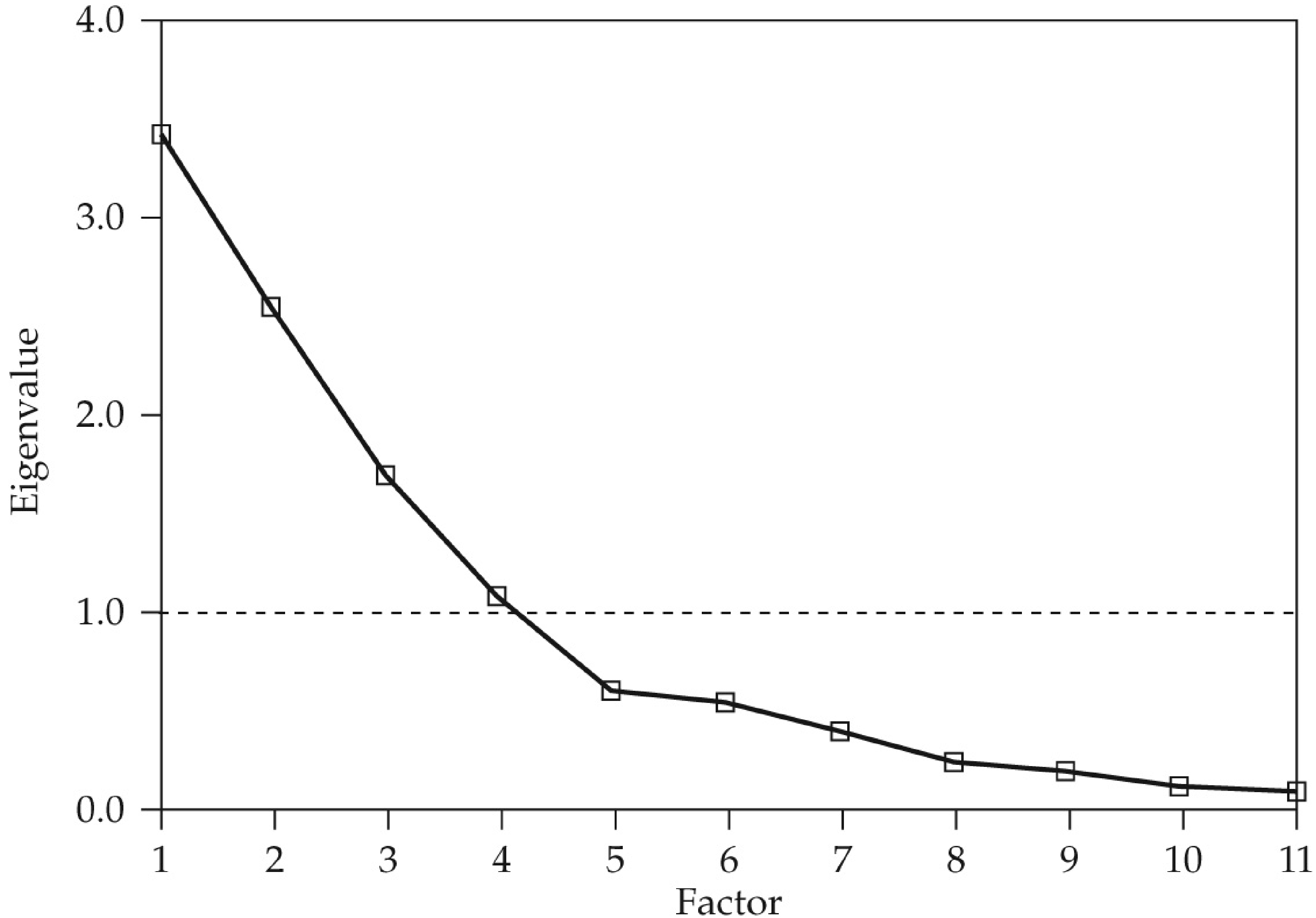
Rotated Component Matrix
"Reduced Set" of HBAT Perceptions Variables

	Component				Communality
	1	2	3	4	
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.

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Scree Test for HBAT Component Analysis



Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

Summary

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?

Source: Hair et al. (2009), Multivariate Data Analysis, 7th Edition, Prentice Hall

蕭文龍,
多變量分析最佳入門實用書--SPSS+LISREL, 第二版,
基峰資訊, 2009

Source: http://shopping.pchome.com.tw/?mod=item&func=exhibit&IT_NO=DJAA1L-A41336032&c=A05

Source: http://shopping.pchome.com.tw/?mod=item&func=exhibit&IT_NO=DJAA1L-A41336032&c=A05

蕭文龍,
多變量分析最佳入門實用書--SPSS+LISREL, 第二版, 碁峰資訊, 2009

- 本書通過Scientific Software International (SSI) LISREL原廠審核通過，成為LISREL原廠推薦的第四本華文書，
  相關網址：http://www.ssicentral.com/cn/books.html#sem
- 本書可作為Hair（2006）Multivariate Data Analysis一書的最佳輔助參考書籍
- 從實用的角度出發，完整介紹社會科學概念、統計軟體的運用以及統計分析，協助學習者完成量化的研究及其相關專題或論文。
- 內容整合了社會科學概念、量化研究、量表發展與統計分析。
- 文中納入第二代統計技術，包括結構方程模式(SEM)、LISREL基本操作SEM結構方程模式範例與SEM結構方程模式實例。
- 特別介紹研究流程、論文結構與研究範例、EndNote書目管理軟體使用說明、LISREL和Nvivo軟體的取得與說明。
- 隨書光碟附贈LISREL For Windows學生版

Source: http://shopping.pchome.com.tw/?mod=item&func=exhibit&IT_NO=DJAA1L-A41336032&c=A05

蕭文龍, 多變量分析最佳入門實用書--SPSS+LISREL, 第二版, 基峰資訊, 2009

- Ch01 社會科學的研究與數量方法的基礎
- Ch02 SPSS的基本操作
- Ch03 檢視資料與敘述性統計
- Ch04 相關分析
- Ch05 卡方檢定
- Ch06 平均數比較
- Ch07 因素分析
- Ch08 迴歸分析
- Ch09 區別分析與邏輯迴歸
- Ch10 單變量變異數分析
- Ch11 多變量變異數分析
- Ch12 典型相關

Source: http://shopping.pchome.com.tw/?mod=item&func=exhibit&IT_NO=DJAA1L-A41336032&c=A05

蕭文龍,
多變量分析最佳入門實用書--SPSS+LISREL, 第二版,
碁峰資訊, 2009

- Ch13 量表的發展、信度和效度
- Ch14 SEM結構方程模式
- Ch15 LISREL的基本操作
- Ch16 結構方程模式的學習範例
- Ch17 結構方程模式的學習範例進階
- Ch18 SEM結構方程模式實例
- Ch19 聯合分析、多元尺度方法和集群分析
- Ch20 交互作用、中介和調節(干擾)效果之驗證
- Ch21 研究流程、論文結構與研究範例
- 附錄A 統計分配表
- 附錄B ENDNOTE書目管理軟體使用說明
- 附錄C 軟體的取得與說明LISREL

Source: http://shopping.pchome.com.tw/?mod=item&func=exhibit&IT_NO=DJAA1L-A41336032&c=A05

References

- Joseph F. Hair, William C. Black, Barry J. Babin, Rolph E. Anderson (2009),
  Multivariate Data Analysis, 7th Edition, Prentice Hall
- 蕭文龍 (2009), 多變量分析最佳入門實用書--SPSS+LISREL, 第二版, 碁峰資訊
- 吳明隆 (2006), SPSS 統計應用學習實務：問卷分析與應用統計, 三版, 知城數位科技