Graduate Statistics

Factor Analysis

What We Will Cover in This Section

- Introduction.
- Concept.
- Input.
- Component/factor development.
- Number of components/factors.
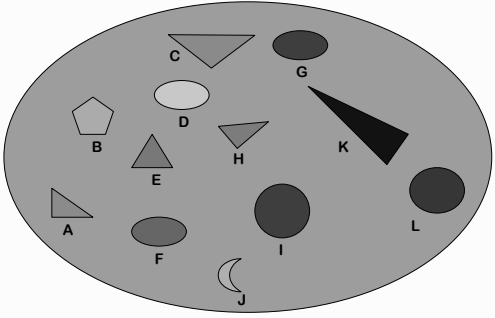- Rotation for interpretation.

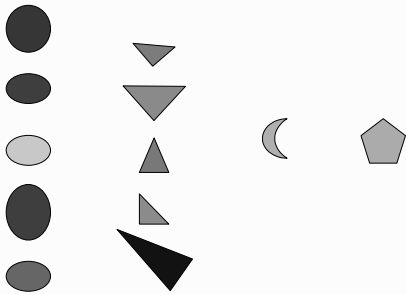## Principal Components and Factor Analysis

Statistical techniques for identifying interrelationships between items with the goal of identifying items that group or cluster together.

How Can These Objects Be Grouped?

Factor 1   Factor 2   Factor 3   Factor 4

Goals of Factor Analysis

1. Summarize patterns of correlations among observed variables.
2. Reduce a large number of observed variables to a smaller number of interpretable groups.
3. Provide an operational definition (regression equation) for an underlying process.
4. Test a hypothesis about the nature of underlying processes.

Input::Output

Input	Intermediate	Output
Correlation Matrix	**Initial Components**	**Rotated Factor Loadings**
What variables. How many variables. How many subjects. Variable relationships.	Type of analysis. Number of factors.	How to rotate. Interpretation of factors. Stability of factors.

◀ ◆ ▶

Fundamental Types of Factor Analysis

1. **Exploratory.**
   Initial examination to describe and summarize a set of variables by grouping them together.

2. **Confirmatory.**
   Does a predicted set of clusters exist demonstrating an <u>a priori</u> set of relationships.

◀ ◆ ▶

Factor Analysis Example

Izzy Wright was interested in examining the qualities that attracted students to a college. Wright designed a survey that he administered to a random sample of high school seniors. The survey assessed four variables: Campus SAFETY, Variety of campus ACTIVITIES, Number of MAJORS, and FACULTY quality.

◀ ◆ ▶

Data

	Safety	Activities	Major	Faculty
S1	32	64	65	67
S2	61	37	62	65
S3	59	40	45	43
S4	36	62	34	35
S5	62	46	43	40

Input Correlation Matrix

	Safety	Activities	Majors	Faculty
Safety	1.00	-.953	-.055	-.130
Activities	-.953	1.00	-.091	-.036
Majors	-.055	-.019	1.00	.990
Faculty	-.130	-.036	.990	1.00

Requirements and Assumptions

1. Correlations should be strong, reliable, and stable.
2. Sample size should be generous.
   - N > 300 is desirable.
   - Depends on the underlying correlations.
3. Missing data have to be dealt with.
4. Correlational assumptions are met.
   - Linearity.
   - Outliers.
   - Multicollinearity.

How to Extract Factors?

Extraction

The mathematical process by which the factors are determined from an larger set of variables.

Factor Extraction Techniques

1. **Principal Components.**
   Goal is to examine all of the variability to extract the most variance with fewest possible factors.

2. **Factor Analysis.**
   Goal is to examine only the variability that the variables have in common. Want remove unique and error variability.

Principal Components Analysis

- Seeks analyze all of the variance in the data set.
- Places 1.00 in the main diagonal.
- Looks at how the variables clump together but does not assume any underlying structure.

Factor Analysis

- Attempts to analyze the structure with random error removed.
- Places shared variance estimates in the main diagonal.
- Assumption is that the Factors cause the variables.
- *N.B.* there is a variety of *Factor Analysis* techniques.

Comments on Extraction Technique

- With strong data, all produce essentially the same factors.
- None of the approaches provides an interpretable solution without rotation.
- 'Data snooping' is common in an attempt to develop interpretable factors.

How Many Factors to Extract?

Alternative 1. Eigenvalue Analysis

Eigenvalue.
The total amount of variance explained by an individual factor.
- The variance has been standardized so S and S^2 =1.
- Thus, total variability in the analysis is equal to *k* or the number of variables in the analysis.
- In Eigenvalue analysis only components with values equal to or greater than 1 (10% of the variance) should be kept.

Alternative #2. Scree Plot Analysis

- **Scree Plot is a graph of the each component.**
  - The components are plotted in ordinal order.
  - The horizontal axis represents the eigenvalue of the component.
- **Look for the point where the Scree Plot levels off. This is usually the place where additional components have relatively small eigenvalues.**

7

Scree Plot

Eigenvalue vs Component Number (1–18)

Scree Plot

Eigenvalue vs Component Number (1–4)
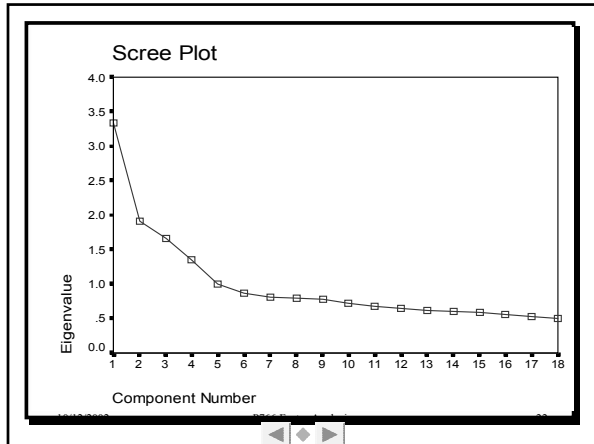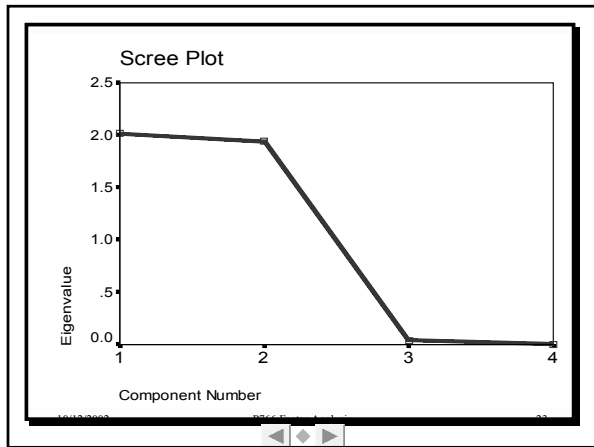
Alternative #3. Total Variance Target

- **Retain the number of factors that will account for a target amount of total variance.**
  - **Problem**
    If there are weak correlations, it may take 20 out of 24 variables to account for your target variability.

Alternative #4: Model Fit

- **When assessing an *a priori* model, then the factors may be compared to the expected amount of variance predicted by the model.**

Factor Rotation

Communality

A. **The proportion of variance that is explained by the factors.**
   or

B. **The squared multiple correlation coefficient R^2 using the factors as predictors and the variable as the criterion.**

Example Communalities

	Initial	Extraction
Safety	1.00	.983
Activity	1.00	.983
Majors	1.00	.995
Faculty	1.00	.997

◄ ◆ ►

Factor Rotation

**Process by which a factor
solution (initial component set)
is made more interpretable
without altering the underlying
mathematical structure.**

◄ ◆ ►

◄ ◆ ►
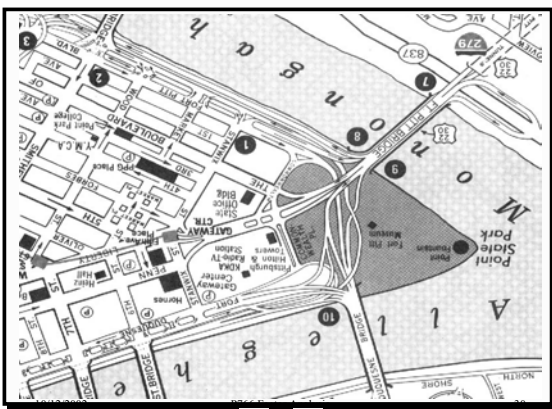
Component Matrix

	Component	
	1	**2**
Safety	-.500	.856
Activity	.357	-.925
Majors	.891	.449
Faculty	.919	.390

◀ ◆ ▶

Orthogonal Rotated Factor Loading Matrix

	Component	
	1	**2**
Safety	-.087	-.988
Activity	-.072	.989
Majors	.997	-.025
Faculty	.998	-.040

Factor Loading

◀ ◆ ▶

Factor Loading

The correlation between a variable and a factor.

- The size of the loading reflects the extent of the relationship between the variable and the factor.
- In orthogonal rotation a variable usually has a strong loading on only one factor.
- In oblique analysis, a variable may have moderate loadings on several factors.

◀ ◆ ▶

11

Factor Score

The estimates of the score a subject would have received on each factor had they been measured on the factors directly.

Potential Issues With Factor Analysis

1. **There is no external criterion against which to verify the solution.**
2. **There is an infinite number ways to rotate the factors all accounting for the same amount of variance.**
3. **Factor analysis is sometimes used to 'resurrect' poorly implemented research.**

The End