Chapter 5

Dimension Reduction and Extraction of Meaningful Factors

The FACTOR Procedure

General form of the FACTOR procedure:

```
PROC FACTOR options;
    VAR variables;
RUN;
```


Section 5.2

Exploratory Factor Analysis

Objectives

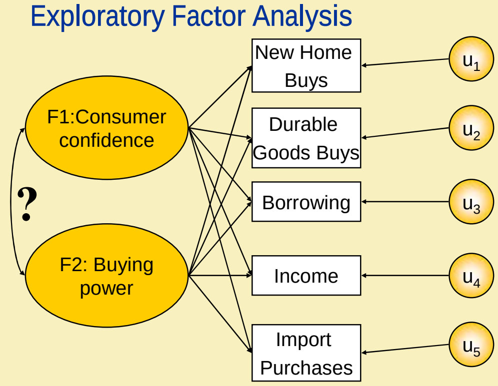
- Explain the distinctions between principal components and common factor analyses.
- Identify several factor extraction methods for factor analysis.
- Differentiate between oblique and orthogonal rotation methods for factor analysis.
- Use the FACTOR procedure to perform exploratory factor analysis.

Why Perform Factor Analysis?

You suspect that the variables you observe (manifest variables) are functions of variables that you cannot observe directly (latent variables).

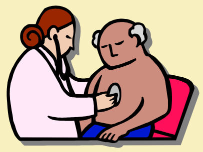
- Identify the latent variables to learn something interesting about the behavior of your population.
- Identify relationships between different latent variables.
- Show that a small number of latent variables underlies the process or behavior you have measured to simplify your theory.
- Explain inter-correlations among observed variables.

Exploratory Factor Analysis



Components versus Factors, Revisited

Principal Components – the symptoms



Latent Factors – the disease

7

The Common Factor Model

$\mathbf{Y} = \mathbf{X}\,\beta + \mathbf{E}$

where

- $\mathbf{Y}$ manifest variables
- $\mathbf{X}$ common factors
- β weights
- $\mathbf{E}$ unique factors + error variation

8

Assumptions of the Common Factor Model

- The unique factors (residuals) are uncorrelated with each other.
- The unique factors (residuals) are uncorrelated with the common (latent) factors.

Under these constraints, you can solve for the correlation matrix:

$$\mathbf{R} = \beta'\beta + \mathbf{U} \text{ or } \mathbf{R} - \mathbf{U} = \beta'\beta$$

9

PCA versus Factor Analysis

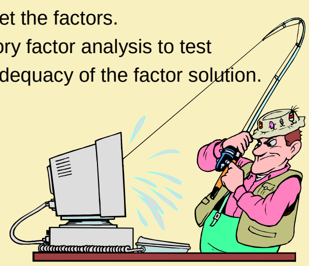
PCA	Factor Analysis
100% of variance accounted for by all components.	Not necessary that 100% of variance be accounted for by the extracted factors.
The components are derived from the variables and explain 100% of the variation in the data.	The variables reflect the common (latent) factors and explain shared variation in the manifest variables.

10

Limitations of Exploratory Factor Analysis

Factor scores are not linear combinations of the variables. They are estimates of latent factors. Try to avoid data fishing problems by:

1. Carefully selecting your manifest variables.
2. Using rotation to interpret the factors.
3. Performing a confirmatory factor analysis to test hypotheses about the adequacy of the factor solution.



11

Factor Extraction Methods: Overview

Principal Factor Analysis

- Computationally efficient
- Most commonly used.

Maximum Likelihood Factor Analysis

- Less efficient computationally; iterative procedure
- Better estimates than principal factor analysis in large samples
- Hypothesis tests for number of factors.

Prior communality estimates are usually the squared multiple correlation of each manifest variable with all the others.

12

How Many Factors?

- Proportion of variance accounted for
  - Minimum # factors to account for 100% of the common variance
- Scree test
  - Find the elbow
- Interpretability criteria
  - At least three items load on each factor
  - Variables within a factor share conceptual meaning
  - Variables between factors measure different constructs
  - Rotated factors demonstrate simple structure.

13

The FACTOR Procedure, revisited

General form of the FACTOR procedure:

```
PROC FACTOR options;
     VAR variables;
RUN;
```

14

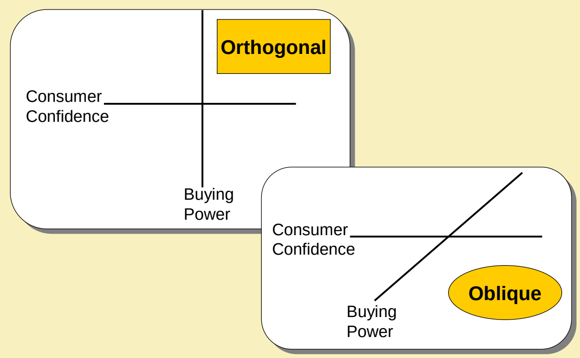
Using the FACTOR Procedure

ch5s2d1.sas

This demonstration illustrates exploratory factor analysis using the FACTOR procedure

15

Are Factors Correlated? Rotation Methods



16

Rotation Methods

- Varimax-Orthogonal: Maximizes the variance of columns of the factor pattern matrix.
- Promax-Oblique in two steps:
  1. Varimax rotation
  2. Relax orthogonality constraints and rotate further.

PROC FACTOR can also perform many other rotation methods. See the *SAS/STAT User's Guide*.

17

Displayed Factor Analysis Output

Eigenvalues

In factor analysis, the eigenvalues displayed are related to the reduced correlation matrix (**R**-**U**).

- In PCA, eigenvalues are of **R**.
- Rule of eigenvalue > 1 is less meaningful in determining the number of factors to retain for factor analysis.
- Scree plot of eigenvalues is often useful in factor analysis.

18

Displayed Factor Analysis Output

Factor Pattern Matrix

- The matrix of standardized regression coefficients for **Y** = **XB** + **E**
- Equal to the matrix of correlations between the variables and the extracted (orthogonal) common factors.

Displayed Factor Analysis Output

Rotated Factor Pattern Matrix

- The matrix of standardized regression coefficients for rotated factors
- Equal to the correlations between the variables and the rotated common factors for orthogonal rotations.

Displayed Factor Analysis Output

Structure Matrix

- Generated for oblique rotations only
- The matrix of the correlations between variables and rotated common factors.

Displayed Factor Analysis Output

Reference Structure Matrix

- Generated for oblique rotations only
- The matrix of semipartial correlations between variables and common factors, removing from each common factor the effects of other common factors.

Displayed Factor Analysis Output

Correlation between Factors

- Generated for oblique rotations only

Factor plots

Final communality estimates

- R^2 for predicting variables from factors
- Called squared canonical correlations when ML method is used

Variance explained by each factor

Rotated Factor Analysis

ch5s2d2.sas

This demonstration illustrates the FACTOR procedure for exploratory factor analysis with rotation.

Exercises

This exercise reinforces the concepts discussed previously.

25

Section 5.3

Cronbach's Coefficient Alpha for Scale Reliability

Objectives

- Perform reliability analysis of scale data using Cronbach's coefficient alpha.
- Interpret output from the CORR procedure for Cronbach's alpha.

27

Internal Consistency of a Scale

When measuring a latent variable, you need a way to quantify the latent variable.

Items that load on a factor for the latent variable are often summed to create a scale score.

But how reliable is the scale?

The "true" reliability is the squared correlation between the scale score (Y) and the true value of the latent variable (T).

28

Cronbach's Alpha

One way of estimating the reliability of a scale is to compute Cronbach's alpha

$$\alpha = \left(\frac{p}{p-1}\right)\left(\frac{\sum_{i\neq j}\operatorname{cov}(Y_iY_j)}{\operatorname{var}(Y_0)}\right)$$

where **Y** are the variables that make up the scale, *p is* the number of variables, and Y_0 is the sum of all the variables in **Y**.

Be sure to reverse-code items as necessary!

29

Other Statistics

- Standardized alpha
  - Useful when variables have different variances
- Correlation with total for each item (standardized and raw)
  - Find the items that best characterize the latent variable
- Alpha if item deleted for each item (standardized and raw)
  - Identify well-fitting and poorly fitting items.

30

The CORR Procedure

General form of the CORR procedure:

```
PROC CORR options;
      VAR variables;
RUN;
```

31

PROC CORR for Reliability Analysis

ch5s3d1.sas

This demonstration illustrates PROC CORR for reliability analysis.

32

Exercises

This exercise reinforces the concepts discussed previously.

33