Factor Analysis

Why Perform Factor Analysis?

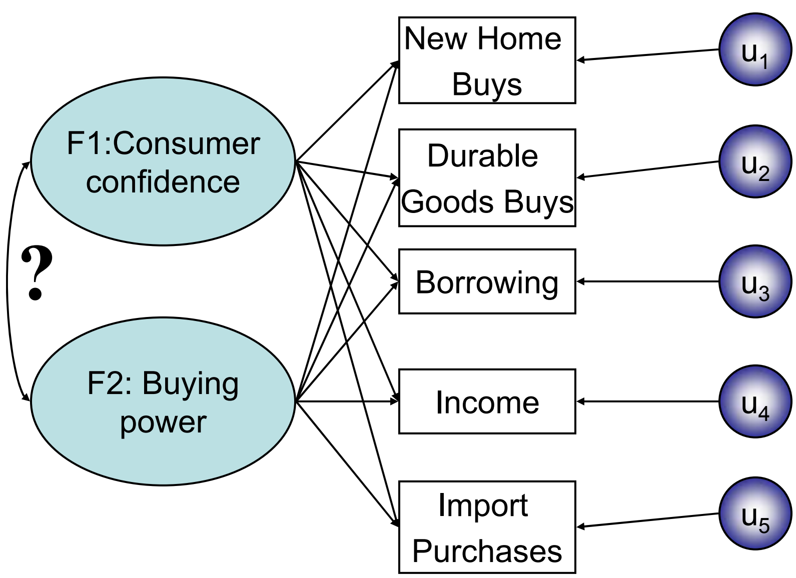
You suspect that the variables you observe (manifest variables) are functions of variables that you cannot observe directly (latent variables).

- 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.

Objectives of Factor Analysis

- To determine whether the p response variables can be partitioned into m subsets, with high correlations within a subset, and low correlations between subsets.
- To determine this new set of m uncorrelated variables (factors) in such a way that these new variables (factors) are interpretable.

Exploratory Factor Analysis



Factor Analysis Model

Let $\mathbf{x} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Let there be m underlying factors

$f_1, f_2, \cdots, f_m$

such that

$\mathbf{x}_{p\times 1} = \boldsymbol{\Lambda}_{p\times m}\mathbf{f}_{m\times 1} + \boldsymbol{\eta}_{p\times 1}$

Factor Analysis Model

$$\mathbf{x}_{p\times 1} = \boldsymbol{\Lambda}_{p\times m}\mathbf{f}_{m\times 1} + \boldsymbol{\eta}_{p\times 1}$$

where

$\mathbf{f} \sim (\mathbf{0}, \mathbf{I}_m)$

$\boldsymbol{\eta} \sim (\mathbf{0}, \boldsymbol{\Psi})$

where $\boldsymbol{\Psi} = \operatorname{diag}\left(\psi_1, \cdots, \psi_p\right)$

Factor Analysis Model

Factor loadings

$$\mathbf{x}_{p\times 1} = \mathbf{\Lambda}_{p\times m}\mathbf{f}_{m\times 1} + \boldsymbol{\eta}_{p\times 1}$$

Common factors

Specific factors

Factor Analysis Model

- $\mathbf{f}$ and η are independent
- $\mathbf{x}$ is either centered or standardized

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:

$$\mathbf{R} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$$

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.

Factor Analysis Equations

Let $\mathbf{x}_{p\times1} = \mathbf{\Lambda}_{p\times m}\mathbf{f}_{m\times1} + \boldsymbol{\eta}_{p\times1}$

Then
$$\begin{aligned}\mathbf{\Sigma} &= \mathrm{Cov}(\mathbf{x}) \\ &= \mathrm{Cov}(\mathbf{\Lambda f} + \boldsymbol{\eta}) \\ &= \mathbf{\Lambda}\ \mathrm{Cov}(\mathbf{f})\ \mathbf{\Lambda}' + \mathrm{Cov}(\boldsymbol{\eta}) \\ &= \mathbf{\Lambda}\ \mathbf{I}_m\ \mathbf{\Lambda}' + \mathbf{\Psi} \\ &= \mathbf{\Lambda}\,\mathbf{\Lambda}' + \mathbf{\Psi}\end{aligned}$$

Factor Analysis Equations

Factor Analysis involves finding

$\mathbf{\Lambda}$ and $\mathbf{\Psi}$

such that

$\mathbf{\Sigma} = \mathbf{\Lambda}\,\mathbf{\Lambda}' + \mathbf{\Psi}$

Factor Analysis – Details

If $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ exist, so that

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$$

then, since $\mathbf{\Psi}$ is a diagonal matrix, the common factors $\mathbf{f}$ completely explain the covariances σ_{ij}

Factor Analysis – Details

$$\sigma_j^2 = \sigma_{jj} = \sum_{k=1}^{m} \lambda_{jk}^2 + \psi_j$$

The proportion of variance of $\mathbf{x}_j$ that is explained by the common factors is

$$h_j^2 = \frac{\sum_{k=1}^{m} \lambda_{jk}^2}{\sigma_{jj}} = \text{communality of } x_j$$

Factor Analysis – Details

$$\sigma_{jj} = \sum_{k=1}^{m} \lambda_{ik} \lambda_{jk}$$

$$\lambda_{jk} = \mathrm{Cov}\left(x_j, f_k\right)$$

= loading of the j^{th} response variable on the k^{th} factor

Factor Analysis using $\mathbf{P}$

$$\mathbf{P} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$$

$$\sum_{k=1}^{m} \lambda_{jk}^2 + \psi_j = 1$$

Communality of $x_j = \sum_{k=1}^{m} \lambda_{jk}^2 = h_j^2$

$$\lambda_{jk} = \mathrm{Corr}\left(z_j, f_k\right)$$

where $z_j = \dfrac{x_j - \mu_j}{\sigma_{jj}}$

Nonuniqueness of the Factors

Let $\mathbf{T}$ be an orthogonal matrix, so

$$\mathbf{T}\mathbf{T}' = \mathbf{I}$$

Then
$$\begin{aligned}
\mathbf{P} &= \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \\
&= \boldsymbol{\Lambda}\mathbf{I}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \\
&= \boldsymbol{\Lambda}\mathbf{T}\mathbf{T}'\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \\
&= (\boldsymbol{\Lambda}\mathbf{T})(\boldsymbol{\Lambda}\mathbf{T})' + \boldsymbol{\Psi} \\
&= \left(\boldsymbol{\Lambda}^*\right)\left(\boldsymbol{\Lambda}^*\right)' + \boldsymbol{\Psi}
\end{aligned}$$

Factor Rotation

- Multiplication by an orthogonal matrix describes a rotation in the p-dimensional space.

Solving the Factor Analysis Equations

- In many cases, solutions do not exist at all.
- For example,
  - estimates of λ^2 are negative

    or
  - estimates of ψ are negative

Choosing the Number of Factors m

- Initially, do a PCA, and start with m
  - as the number of PC's that explain most ($\geq$ 70%) of the variance
  - (if PCA on correlation matrix) as the number of components with eigenvalues > 1
  - as the number determined from a scree plot (find the elbow, and select all components occurring in the sharp descent before leveling off)

Choosing the Number of Factors m

- After a FA, drop trivial factors
  (factors that load on only one response variable).

Choosing the Number of Factors m

- Determine the minimum number of factors that account for 100% of the common variance
- 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

Reliability of Factor Analysis

Approx. Sample Size	Reliability
50	Very poor
100	Poor
200	Fair
300	Good
500	Very good
1000	Excellent

Methods for Solving the Factor Analysis Equations

- Principal Factor Method
  - Most commonly used method
  - Computationally efficient

Methods for Solving the Factor Analysis Equations

- Maximum Likelihood Method
  - An iterative procedure that is less efficient computationally
  - Yields better estimates than the Principal Factor method in large samples
  - Produces a statistical test for the number of factors
  - A factor extraction method that produces parameter estimates that are most likely to have produced the observed correlation matrix if the sample is from a multivariate normal distribution. The correlations are weighted by the inverse of the uniqueness of the variables, and an iterative algorithm is employed.

Methods for Solving the Factor Analysis Equations

- Unweighted Least Squares Method
  - A factor extraction method that minimizes the sum of the squared differences between the observed and reproduced correlation matrices ignoring the diagonals.
- Generalized Least Squares Method
  - A factor extraction method that minimizes the sum of the squared differences between the observed and reproduced correlation matrices. Correlations are weighted by the inverse of their uniqueness, so that variables with high uniqueness are given less weight than those with low uniqueness.

Methods for Solving the Factor Analysis Equations

- Alpha
  - A factor extraction method that considers the variables in the analysis to be a sample from the universe of potential variables. It maximizes the alpha reliability of the factors.
- Image
  - A factor extraction method developed by Guttman and based on image theory. The common part of the variable, called the partial image, is defined as its linear regression on remaining variables, rather than a function of hypothetical factors.

Principal Factor Method on $\mathbf{R}$

Model: $\mathbf{R} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$

(1) Estimate $\mathbf{\Psi} = \mathrm{diag}\left(\psi_1, \cdots, \psi_p\right)$

with $\hat{\psi}_j = 1 - R^2_{x_j,\text{all the other } x\text{'s}}$

or equivalently, let the communality

$h_j^2 = R^2_{x_j,\text{all the other } x\text{'s}} = SMC_j$

Principal Factor Method on **R**

Other methods of obtaining initial estimates for the communalities include:

- Estimation using the maximum absolute correlation:

$$h_j^2 = \max_{\substack{j \\ j \neq i}} \left| r_{ij} \right| = r_{\max}^{j}$$

Principal Factor Method on **R**

- Estimation using the adjusted squared multiple correlation:

  It can be shown that

  $\sum_{i=1}^{p} r_{\max}^{i}$ is a better estimator of the total communality $\sum_{i=1}^{p} h_i^2$ than is $\sum_{i=1}^{p} SMC_i$

Principal Factor Method on $\mathbf{R}$

- Estimation using the adjusted squared multiple correlation:

$$h_j^2 = ASMC_j = \frac{\sum_{i=1}^{p} r_{\max}^i}{\sum_{i=1}^{p} SMC_i} SMC_j$$

Principal Factor Method on $\mathbf{R}$

(2) To obtain a unique solution, let

$$\mathbf{\Lambda}\mathbf{\Lambda}' = \mathbf{D} = \mathrm{diag}\left(d_1, \cdots, d_m\right)$$

Also, let $\mathbf{\Lambda} = \left[\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_m\right]$

Principal Factor Method on $\mathbf{R}$

(3)

$$\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = (\mathbf{R} - \boldsymbol{\Psi})\boldsymbol{\Lambda}$$
$$\Rightarrow \boldsymbol{\Lambda}\mathbf{D} = (\mathbf{R} - \boldsymbol{\Psi})\boldsymbol{\Lambda}$$
$$\Rightarrow d_k \lambda_k = (\mathbf{R} - \boldsymbol{\Psi})\lambda_k \quad k = 1, \cdots, m$$

$\Rightarrow d_1, \cdots, d_m$ are the eigenvalues of $\mathbf{R} - \boldsymbol{\Psi}$

$\lambda_1, \cdots, \lambda_m$ are the corresponding eigenvectors

Principal Factor Method on $\mathbf{R}$

- Choose λ's corresponding to the m largest nonnegative eigenvalues
- May have to reduce m

Principal Factor Method with Iteration

- Use communalities obtained in (3) as new estimates to be used in (1).
- Iterate until convergence, or until a nonsensical result is obtained.

Factor Rotation

Common Criteria

- Try to make as many λ_{jk}'s as possible near zero, while maximizing the others.
- Try to develop a solution such that each response variable is not loaded (heavily) on more than one factor.

Methods of Orthogonal Factor Rotation

- Varimax
  - An orthogonal rotation method that minimizes the number of variables that have high loadings on each factor. It simplifies the interpretation of the factors.

Methods of Orthogonal Factor Rotation

- Quartimax
  - A rotation method that minimizes the number of factors needed to explain each variable. It simplifies the interpretation of the observed variables.
- Orthomax
  - A class of rotations that includes quartimax

Methods of Orthogonal Factor Rotation

- Equamax
  - A rotation method that is a combination of the varimax method, which simplifies the factors, and the quartimax method, which simplifies the variables. The number of variables that load highly on a factor and the number of factors needed to explain a variable are minimized.

Simple Structure of the Matrix of Factor Loadings

The matrix of factor loadings is

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{bmatrix}$$

Simple Structure of the Matrix of Factor Loadings

The matrix Λ is said to have simple structure when

1. Each row contains at least one zero.
2. Each column contains at least m zeroes.
3. For each pair of columns, there should be at least m variables whose entries vanish in one column but not in the other.

Factor Rotations and Simple Structure

- The purpose of rotation is to attempt to approximate simple structure.

Varimax Rotation Method

Let $\mathbf{B} = \mathbf{\Lambda T}$ where $\mathbf{T}$ is orthogonal

$$= \left[\mathbf{b}_1, \cdots, \mathbf{b}_m\right]$$

Consider $\mathbf{b}'_q = \left[b_{1q}, \cdots, b_{pq}\right]$

Varimax Rotation Method

Let $v_q^* = \text{variance}\left(b_{1q}^2, \cdots, b_{pq}^2\right)$

Maximizing v_q^* will tend to drive the b'_{jq}s towards 0 or 1.

The <u>raw</u> varimax rotation would maximize

$$\mathrm{V}^* = \sum_{q=1}^{m} v_q^*$$

Varimax Rotation Method

In order to give more weight to variables with higher communalities,

let $v_q^* = \text{variance}\left(\frac{b_{1q}^2}{h_1^2}, \cdots, \frac{b_{pq}^2}{h_p^2}\right)$

$h_j^2 =$ communality of j^{th} response

Varimax Rotation maximizes $\mathrm{V} = \sum_{q=1}^{m} v_q$

Quartimax Rotation

Maximize $\text{Variance}\left(b_{jq}^2\right)$

or equivalently

Maximize $\text{Kurtosis}\left(b_{jq}\right)$

In practice, this method tends to produce solutions in which most of the variables load on the first factor.

Orthomax Rotation Criteria

Maximize

$$\sum_{q=1}^{m}\left\{\sum_{j=1}^{p}\left(\lambda_{jq}^{*}\right)^{4}-\frac{\gamma}{p}\left[\sum_{j=1}^{p}\left(\lambda_{jq}^{*}\right)^{2}\right]^{2}\right\}$$

$\gamma = 0$	quartimax
$\gamma = 1$	raw varimax
$\gamma = m/2$	equamax
$\gamma = p(m-1)/(p+m-2)$	parsimax

Methods of Oblique Factor Rotation

- Direct Oblimin
  - A method for oblique (nonorthogonal) rotation. When delta equals 0, solutions are most oblique. As delta becomes more negative, the factors become less oblique.

Methods of Oblique Factor Rotation

- Promax
  - An oblique rotation, which allows factors to be correlated. It can be calculated more quickly than a direct oblimin rotation, so it is useful for large datasets.

Oblique Rotations

- Promax-Oblique in two steps:
  1. Varimax rotation
  2. Relax orthogonality constraints and rotate further.

Factor Scores

In the model: $\mathbf{x} = \mathbf{\Lambda f} + \eta$
$\mathbf{f}$ is the vector of factor scores,
but Λ is estimated,
and η is unknown,
so $\mathbf{f}$ cannot be determined explicitly.

Methods for Determining Factor Scores

- Bartlett's (WLS) Method
- Thompson's (Regression) Method

Bartlett's (WLS) Method

Let $\mathbf{z}_r$ = standardized data vector.

Find $\mathbf{f}$ such that

$$\left(\mathbf{z}_r - \hat{\mathbf{\Lambda}}\mathbf{f}\right)' \hat{\mathbf{\Psi}}^{-1}\left(\mathbf{z}_r - \hat{\mathbf{\Lambda}}\mathbf{f}\right)$$

is minimized
(i.e., minimized Euclidean distance)

Bartlett's (WLS) Method

$$\Rightarrow \mathbf{f}_r = \left(\hat{\mathbf{\Lambda}}' \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{\Lambda}}\right)^{-1} \hat{\mathbf{\Lambda}}' \hat{\mathbf{\Psi}}^{-1} \mathbf{z}_r$$

$$r = 1, \cdots, N$$

Thompson's (Regression) Method

$$\begin{bmatrix} \mathbf{z} \\ \mathbf{f} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{P} & \mathbf{\Lambda} \\ \mathbf{\Lambda}' & \mathbf{I} \end{bmatrix} \right)$$

$$\Rightarrow E\left[\mathbf{f} \mid \mathbf{z}\right] = \mathbf{\Lambda}' \mathbf{P}^{-1} \mathbf{z}_r$$

So, the vector of factor scores for the r^{th} individual is

$$\mathbf{f}_r = \hat{\mathbf{\Lambda}}' \mathbf{R}^{-1} \mathbf{z}_r$$

Thompson's (Regression) Method

Alternatively, some software computes

$$\mathbf{f}_r = \hat{\mathbf{\Lambda}}' \left(\hat{\mathbf{\Lambda}} \hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}} \right)^{-1} \mathbf{z}_r$$

Factor Scores in SAS

In SAS PROC FACTOR, the OUT= option produces a data set containing the factor scores on each factor for each case.

Maximum Likelihood Method for Solving the Factor Analysis Equations

Assume $\mathbf{x}_1, \cdots, \mathbf{x}_p \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Maximum Likelihood Method for Solving the Factor Analysis Equations

Then the maximum likelihood estimates

$\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Psi}}$

satisfy the equation

$$\mathbf{R}\,\hat{\mathbf{\Psi}}\,\hat{\mathbf{\Lambda}} = \hat{\mathbf{\Lambda}}\left(\mathbf{I} + \hat{\mathbf{\Lambda}}'\,\hat{\mathbf{\Psi}}^{-1}\,\hat{\mathbf{\Lambda}}\right)$$

where $\hat{\mathbf{\Psi}} = \mathrm{diag}\left(\mathbf{R} - \hat{\mathbf{\Lambda}}\,\hat{\mathbf{\Lambda}}'\right)$

and $\hat{\mathbf{\Lambda}}'\,\hat{\mathbf{\Psi}}^{-1}\,\hat{\mathbf{\Lambda}}$ is diagonal

Large Sample Test for the Number of Factors

$$H_0 : \mathbf{\Sigma} = \mathbf{\Lambda}\,\mathbf{\Lambda}' + \mathbf{\Psi}$$

$$H_1 : \mathbf{\Sigma} \neq \mathbf{\Lambda}\,\mathbf{\Lambda}' + \mathbf{\Psi}$$

Test statistic:

$$\left(N - \frac{2p + 4m + 11}{6}\right)\ln\left(\frac{\left|\hat{\mathbf{\Lambda}}\,\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}}\right|}{\left|\mathbf{S}\right|}\right)$$

Large Sample Test for the Number of Factors

Reference Distribution: χ^2_ν

where $\nu = \frac{1}{2}\left[\left(p-m\right)^2 - p - m\right]$

$$\Rightarrow m < \frac{1}{2}\left(2p+1-\sqrt{8p+1}\right)$$

Displayed Factor Analysis Output

Eigenvalues

In factor analysis, the eigenvalues displayed are related to the reduced correlation matrix $\mathbf{R} - \hat{\mathbf{\Psi}}$

- In PCA, eigenvalues are of $\mathbf{R}$.
- 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.

Displayed Factor Analysis Output

Factor Pattern Matrix

– Equal to the matrix of correlations between the variables and the extracted (orthogonal) common factors.

Displayed Factor Analysis Output

Rotated Factor Pattern Matrix

– 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

Anomalies

- Heywood cases
  - Since communalities are squared correlations, they should lie between 0 & 1.
  - However, the mathematics of the common factor model may yield final communality estimates that are greater than 1.
  - If a communality equals 1, this situation is referred to as a Heywood case.
  - If a communality exceeds 1, it is an ultra-Heywood case.

Ultra-Heywood case

- Implies that some unique factor has negative variance.
- Possible causes:
  - Bad prior communality estimates
  - Too many common factors
  - Too few common factors
  - Not enough data
  - Common factor model not appropriate