The Common Factor Model

PSY544 – Introduction to Factor Analysis

Week 4

Homework!

- Homework assignment 1 will be out this week
- I’ll send you an email, along with the deadline

The data model in factor analysis

- Recall the way we formulated the Common Factor Model earlier – we expressed the MVs as a linear function of the **common factors** and the **unique factors:**

$$x_{ij} = \mu_j + \lambda_{j1} z_{i1} + \lambda_{j2} z_{i2} + \cdots + \lambda_{jm} z_{im} + 1u_{ij}$$

Mean + Common factor part + Unique factor part

$$x_{ij} = \mu_j + \sum_{k=1}^{m} \lambda_{jk} z_{ik} + u_{ij}$$

The data model in factor analysis

$$x_{ij} = \mu_j + \sum_{k=1}^{m} \lambda_{jk} z_{ik} + u_{ij}$$

Where:

x_{ij} **is the score of person** ***i*** **on manifest variable** ***j***
μ_j **is the mean of manifest variable** ***j***
z_{ik} **is the common factor score of person** ***i*** **on factor** ***k***
λ_{jk} **is the factor loading of manifest variable** ***j*** **on factor** ***k***
u_{ij} **is the unique factor score of person** ***i*** **on unique factor** ***j;*** **and** $u_{ij} = s_{ij} + e_{ij}$
s_{ij} **is the factor score of person** ***i*** **on specific factor** ***j***
e_{ij} **is the error term for person** ***i*** **on manifest variable** ***j***

The data model in factor analysis

- **We will consider the model as operating in a population, and thus we will consider the data model for a typical individual by omitting the subscript *i*:**

$$x_j = \mu_j + \lambda_{j1} z_1 + \lambda_{j2} z_2 + \cdots + \lambda_{jm} z_m + 1 u_j$$

$$x_j = \mu_j + \sum_{k=1}^{m} \lambda_{jk} z_k + u_j$$

- **Here we actually have *p* equations, one for each manifest variable x_j, but we can express it all as a single equation using matrix notation:**

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda z} + \boldsymbol{u}$$

The data model in factor analysis

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda z} + \boldsymbol{u}$$

Where:

$\boldsymbol{x}$ is a p x 1 vector of typical scores on the p manifest variables

$\boldsymbol{\mu}$ is a p x 1 vector of population means of the p manifest variables

$\boldsymbol{\Lambda}$ is a p x m matrix of factor loadings, where $p > m$ (rectangular matrix)

$\boldsymbol{z}$ is a m x 1 vector of (unobservable) common factor scores

$\boldsymbol{u}$ is a p x 1 vector of (unobservable) unique factor scores

The data model in factor analysis

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda z} + \boldsymbol{u}$$

- **For illustration, let's extract the equation for the third manifest variable. Let's assume that *m* = 3 (there are three common factors):**

$$\begin{bmatrix} \\ \\ x_3 \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \\ \mu_3 \\ \\ \end{bmatrix} + \begin{bmatrix} \\ \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \\ \\ \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} \\ \\ u_3 \\ \\ \end{bmatrix}$$

$$x_3 = \mu_3 + \lambda_{31} z_1 + \lambda_{32} z_2 + \lambda_{33} z_3 + u_3$$

The data model in factor analysis

- The data model represents a typical observation in the population. It is intended to explain the structure of the raw data (i.e., the scores on manifest variables)

- The data model is accompanied by assumptions about the joint distribution of the elements in ***z*** and ***u*** and implies a model for the population covariance matrix. The model for the covariance matrix is known as the **covariance structure** and is intended to explain the variances and covariances of the manifest variables, **not** the raw data.

- Before we proceed to derive the covariance structure model, we'll talk about the important distributional assumptions and lay down some notational rules.

Assumptions

- **We will make the following assumptions about the common factors *z* and unique factors *u*:**
  1. **The common factors and the unique factors are independently distributed. As such, the common factors are uncorrelated with the unique factors. In other words, $\boldsymbol{\Sigma_{zu} = 0 = \Sigma'_{uz}}$**
  2. **The unique factors are mutually independent. As such, the unique factors for different MVs are uncorrelated with each other. This implies that the covariance matrix $\boldsymbol{\Sigma_{uu}}$ is diagonal.**
  3. **The common factors and the unique factors are standardized to have means of zero.**
  4. **The common factors are also standardized to have unit variances (variances of 1).**

Notation

- **We will use the following notation:**

  The manifest variable covariance matrix: $\mathbf{\Sigma} = \mathbf{\Sigma}_{xx}$
  The common factor covariance matrix: $\mathbf{\Phi} = \mathbf{\Sigma}_{zz}$
  The unique factor covariance matrix: $\boldsymbol{D}_{\psi} = \mathbf{\Sigma}_{uu}$

- **Note that (because of the assumptions we made), the diagonal elements of $\mathbf{\Phi}$ are required to be equal to 1. Thus, $\mathbf{\Phi}$ is a factor correlation matrix.**

Deriving the mean and covariance structures

- The mean and covariance structures are derived from the data model:

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda z} + \boldsymbol{u}$$

- Let's derive the mean structure first. We want an equation that represents the mean vector $\boldsymbol{\mu}$ of the manifest variables. If we take the expectation of both sides of the equation above, we get:

$$\boldsymbol{Ex} = \boldsymbol{\mu} + \boldsymbol{\Lambda Ez} + \boldsymbol{Eu}$$

- Given the assumptions we previously talked about, this follows:

$$\boldsymbol{\mu}_{\mathbf{x}} = \boldsymbol{\mu} + \boldsymbol{\Lambda 0} + \mathbf{0}$$

$$\boldsymbol{\mu}_{\mathbf{x}} = \boldsymbol{\mu}$$

Deriving the mean and covariance structures

- **This implies that the means of the MVs are not restricted by the model.**

- **Alright. Let's consider the derivation of the covariance structure. What we want is to obtain an equation for $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{xx}$, the covariance matrix of the MVs.**

- **Let us subtract the mean vector from both sides of the data model:**

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda z} + \boldsymbol{u}$$

$$\boldsymbol{x} - \boldsymbol{\mu} = \boldsymbol{\Lambda z} + \boldsymbol{u}$$

…this equation expresses deviations of individual scores from population means as functions of factor scores, factor loadings and unique factor scores.

Deriving the mean and covariance structures

- Now, we will post-multiply both sides of the equation by a transpose:

$$(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})' = (\boldsymbol{\Lambda z}+\boldsymbol{u})\ (\boldsymbol{\Lambda z}+\boldsymbol{u})'$$

…because the transpose of a sum equals the sum of the transposes:

$$(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})' = (\boldsymbol{\Lambda z}+\boldsymbol{u})\ ((\boldsymbol{\Lambda z})'+\boldsymbol{u}')$$

…because the transpose of a product is equal to the product of the transposes in reverse order:

$$(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})' = (\boldsymbol{\Lambda z}+\boldsymbol{u})\ (\boldsymbol{z}'\boldsymbol{\Lambda}'+\boldsymbol{u}')$$

…expanding, we get:

$$(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})' = \boldsymbol{\Lambda z z}'\boldsymbol{\Lambda}' + \boldsymbol{\Lambda z u}' + \boldsymbol{u z}'\boldsymbol{\Lambda}' + \boldsymbol{u u}'$$

Deriving the mean and covariance structures

$$(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})' = \boldsymbol{\Lambda zz'\Lambda'} + \boldsymbol{\Lambda zu'} + \boldsymbol{uz'\Lambda'} + \boldsymbol{uu'}$$

…then, we take the expectations of both sides:

$$E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})'] = \boldsymbol{\Lambda}E[\boldsymbol{zz'}]\boldsymbol{\Lambda'} + \boldsymbol{\Lambda}E[\boldsymbol{zu'}] + E[\boldsymbol{uz'}]\boldsymbol{\Lambda'} + E[\boldsymbol{uu'}]$$

We can simplify, because all the expectations represent a covariance matrix of some sort:

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Lambda\Sigma}_{zz}\boldsymbol{\Lambda'} + \boldsymbol{\Lambda\Sigma}_{zu} + \boldsymbol{\Sigma}_{uz}\boldsymbol{\Lambda'} + \boldsymbol{\Sigma}_{uu}$$

Deriving the mean and covariance structures

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{zu} + \boldsymbol{\Sigma}_{uz}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_{uu}$$

- However, we assumed that both $\boldsymbol{\Sigma}_{zu}$ and $\boldsymbol{\Sigma}_{uz}$ are zero – the common factors are not correlated with the unique factors:

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\mathbf{0} + \mathbf{0}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_{uu}$$

- And using the notation we defined previously:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{D}_{\psi}$$

Deriving the mean and covariance structures

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda \Phi \Lambda'} + \boldsymbol{D_\psi}$$

- Ta-daaaaah! The equation above is the **factor analysis covariance structure**. It represents the factor structure of the population covariance matrix of the manifest variables. The variances and covariances in $\boldsymbol{\Sigma}$ are functions of the common factor loadings ($\boldsymbol{\Lambda}$), common factor correlations ($\boldsymbol{\Phi}$) and unique factor variances ($\boldsymbol{D_\psi}$). This equation is **super-important.**
- We have just derived a model that explains the variances and covariances of the MVs. Note that this model does not contain **any** factor scores, common or unique. We don't need them – the variances and covariances of manifest variables do not depend on them.

Deriving the mean and covariance structures

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{D}_{\psi}$$

…yes, the equation is so important that I have included it again in a separate slide, just so you can admire it.

Deriving the mean and covariance structures

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{D}_{\psi}$$

- As you can see, the model equation assumes the common factors can be potentially correlated ($\boldsymbol{\Phi}$). In unrestricted (exploratory) factor analysis it is sometimes (at least initially) assumed that they are uncorrelated, so that $\boldsymbol{\Phi} = \boldsymbol{I}$. In that case, the covariance structure becomes:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{D}_{\psi}$$

Deriving the mean and covariance structures

- **The *j*-th diagonal element ψ_{jj} of $\boldsymbol{D}_{\psi}$ is the *j*-th unique variance. The *j*-th communality (proportion of variance of MV *j* due to common factors) can be written as:**

$$h_{jj} = \frac{[\boldsymbol{\Lambda\Phi\Lambda'}]_{jj}}{\sigma_{jj}} = 1 - \frac{\psi_{jj}}{\sigma_{jj}}$$

- **If the factors are uncorrelated, then:**

$$h_{jj} = \frac{[\boldsymbol{\Lambda\Lambda'}]_{jj}}{\sigma_{jj}} = 1 - \frac{\psi_{jj}}{\sigma_{jj}}$$

…that is, the sum of squares of row *j* of $\boldsymbol{\Lambda}$ divided by the variance of the *j*-th MV.

Correlation structure

- In the **covariance structure**, factor loadings are regression coefficients (weights) that represent the linear effect of an LV on a particular MV. The latent factor acts like an independent variable, and the MV acts like a dependent variable.

- We have seen that the common factor covariance matrix (Φ) is standardized to have units on the diagonals (= the variance of the LVs is set to 1). As such, the common factor covariances (off-diagonal elements) are, in fact, correlations.

- Sometimes (well, a LOT of times) it is useful to standardize the manifest variables as well – so that the factor loadings are **standardized** regression coefficients, and that we are working with a **correlation matrix** of MVs rather than a covariance matrix.

Correlation structure

- **How do we convert a covariance matrix (say, $\boldsymbol{\Sigma}$) into a correlation matrix (say, P)?**

- **Let's create a diagonal matrix $\boldsymbol{D}_\sigma$ which contains the diagonal elements of $\boldsymbol{\Sigma}$:**

$$\boldsymbol{D}_\sigma = \text{Diag}[\boldsymbol{\Sigma}]$$

…the diagonal elements of $\boldsymbol{D}_\sigma$ are the variances of the manifest variables.

Correlation structure

- **If we would want to transform these into standard deviations of the MVs, we would have to take the square root of each element. We'll go a little further and define a new diagonal matrix $D_\sigma^{-1/2}$ which contains the reciprocals of standard deviations of the elements in D_σ**

- **The manifest variable covariance matrix Σ can be transformed into a manifest variable correlation matrix P in the following way:**

$$\mathbf{P} = \mathbf{D}_\sigma^{-1/2} \boldsymbol{\Sigma} \mathbf{D}_\sigma^{-1/2}$$

Correlation structure

- **For example:**

$$\mathbf{P} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\sigma_{33}}} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\sigma_{33}}} \end{bmatrix}$$

Correlation structure

$$= \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{13}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{33}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \frac{\sigma_{23}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{33}}} \\ \frac{\sigma_{31}}{\sqrt{\sigma_{33}}\sqrt{\sigma_{11}}} & \frac{\sigma_{32}}{\sqrt{\sigma_{33}}\sqrt{\sigma_{22}}} & \frac{\sigma_{33}}{\sqrt{\sigma_{33}}\sqrt{\sigma_{33}}} \end{bmatrix}$$

- Remember - pre-multiplication by a diagonal matrix scales rows, post-multiplication scales columns. Can you see it in there?

Correlation structure

$$\begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{13}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{33}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \frac{\sigma_{23}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{33}}} \\ \frac{\sigma_{31}}{\sqrt{\sigma_{33}}\sqrt{\sigma_{11}}} & \frac{\sigma_{32}}{\sqrt{\sigma_{33}}\sqrt{\sigma_{22}}} & \frac{\sigma_{33}}{\sqrt{\sigma_{33}}\sqrt{\sigma_{33}}} \end{bmatrix}$$

- **In this matrix, the diagonal entries will be 1 and the off-diagonal entries will be correlation coefficients (covariance divided by the standard deviations of both variables)**

Correlation structure

- **So, what we have now is the covariance structure:**

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \boldsymbol{D}_{\psi}$$

- **…and a way of transforming the covariance matrix $\mathbf{\Sigma}$ into a correlation matrix P:**

$$\mathbf{P} = \mathbf{D}_{\sigma}^{-1/2}\mathbf{\Sigma}\mathbf{D}_{\sigma}^{-1/2}$$

- **We can substitute:**

$$\mathbf{P} = \mathbf{D}_{\sigma}^{-1/2}(\mathbf{\Lambda\Phi\Lambda}' + \boldsymbol{D}_{\psi})\mathbf{D}_{\sigma}^{-1/2}$$
$$\mathbf{P} = \mathbf{\Lambda}^{*}\mathbf{\Phi}\mathbf{\Lambda}^{*\prime} + \boldsymbol{D}_{\psi}^{*}$$

Correlation structure

$$\mathbf{P} = \mathbf{D}_{\boldsymbol{\sigma}}^{-1/2}(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{D}_{\psi})\mathbf{D}_{\boldsymbol{\sigma}}^{-1/2}$$

$$\mathbf{P} = \boldsymbol{\Lambda}^{*}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{*\prime} + \boldsymbol{D}_{\psi}^{*}$$

- **Where:**

$$\boldsymbol{\Lambda}^{*} = \mathbf{D}_{\boldsymbol{\sigma}}^{-1/2}\boldsymbol{\Lambda}$$

$$\boldsymbol{D}_{\psi}^{*} = \mathbf{D}_{\boldsymbol{\sigma}}^{-1/2}\,\boldsymbol{D}_{\psi}\,\mathbf{D}_{\boldsymbol{\sigma}}^{-1/2} = \mathbf{D}_{\boldsymbol{\sigma}}^{-1}\,\boldsymbol{D}_{\psi}$$

Correlation structure

$$\boldsymbol{\Lambda}^* = \mathbf{D}_\sigma^{-1/2}\boldsymbol{\Lambda}$$

$$\boldsymbol{D}_\psi^* = \mathbf{D}_\sigma^{-1/2}\,\boldsymbol{D}_\psi\,\mathbf{D}_\sigma^{-1/2} = \mathbf{D}_\sigma^{-1}\,\boldsymbol{D}_\psi$$

- The factor loadings for the correlation structure are equal to the factor loadings for the covariance structure divided by the standard deviation of the given MV
- The unique variances for the correlation structure are equal to the unique variances for the covariance structure divided by the variance of the given MV. This means that the unique variances for the correlation structure are really the proportions of variance of the particular MV that is not explained by the common factors

Correlation structure

- The **communalities** (proportion of variance of MV *j* due to common factors), then, can be written as:

$$h_{jj} = [\mathbf{\Lambda}^* \mathbf{\Phi} \mathbf{\Lambda}^{*\prime}]_{jj} = 1 - \psi_{jj}^*$$

- If the factors are uncorrelated, then:

$$h_{jj} = [\mathbf{\Lambda}^* \mathbf{\Lambda}^{*\prime}]_{jj} = \sum_{k=1}^{m} \lambda_{jk}^{*2}$$

Correlation structure

- Consider the correlation structure for uncorrelated (orthogonal) factors to better understand the relationship between elements of $\mathbf{P}$ and the elements of $\boldsymbol{\Lambda}^*$ and $\boldsymbol{D}_{\psi}^*$. An example:

$$\mathbf{P} = \begin{bmatrix} 1 & & & \\ \rho_{21} & 1 & & \\ \rho_{31} & \rho_{32} & 1 & \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^* & \lambda_{12}^* \\ \lambda_{21}^* & \lambda_{22}^* \\ \lambda_{31}^* & \lambda_{32}^* \\ \lambda_{41}^* & \lambda_{42}^* \end{bmatrix} \begin{bmatrix} \lambda_{11}^* & \lambda_{21}^* & \lambda_{31}^* & \lambda_{41}^* \\ \lambda_{12}^* & \lambda_{22}^* & \lambda_{32}^* & \lambda_{42}^* \end{bmatrix} + \begin{bmatrix} \psi_{11}^* & & & \\ & \psi_{22}^* & & \\ & & \psi_{33}^* & \\ & & & \psi_{44}^* \end{bmatrix}$$

- This shows us that $\rho_{11} = 1 = \lambda_{11}^{*2} + \lambda_{12}^{*2} + \psi_{11}^*$
- Also, $\rho_{21} = \lambda_{21}^* \lambda_{11}^* + \lambda_{22}^* \lambda_{12}^*$

- So here, the correlation between two MVs is the sum of the products of their loadings on the common factors

An example

- Remember the example correlation matrix I have shown earlier?

(4 performance measures: paragraph comprehension, vocabulary, arithmetic skills, and mathematical problem solving)

	PC	VO	AR	MPS
PC	1			
VO	.49	1		
AR	.14	.07	1	
MPS	.48	.42	.48	1

An example

- **The factor loading matrix is:**

	Factor 1	Factor 2
PC	.70	.10
VO	.70	.00
AR	.10	.70
MPS	.60	.60

- **Let's compute the communality and the unique variance of PC by hand**
- **The correlation between PC and VO:**

$$\rho_{21} = \lambda^*_{21}\lambda^*_{11} + \lambda^*_{22}\lambda^*_{12} = 0.7 * 0.7 + 0.0 * 0.1 = 0.49$$