

Fitting the Common Factor Model

PSY544 – Introduction to Factor Analysis

Week 5

The Common Factor Model

- In the past two weeks, we introduced the Common Factor Model in multiple forms.
- First, we introduced the data model:

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

The Common Factor Model

- After making some assumptions, we derived a covariance structure model
- For correlated factors:

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

- For uncorrelated factors:

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

The Common Factor Model

- And we have seen how the covariance structure model can be transformed into a correlation structure model.
- For correlated factors:

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

- For uncorrelated factors:

$$\mathbf{P} = \mathbf{\Lambda}^* \mathbf{\Lambda}^{*'} + \mathbf{D}_{\psi}^*$$

The Common Factor Model

- You learned a LOT already!
- More than a majority of factor analysis practitioners know about the model 😊 (pretty sad, huh?)

Fitting the model

- One important thing to note is that these models are intended for a **population** – they are population models, describing how stuff works in a population.
- Anyway, at the beginning we learned that there are two sides to factor analysis – theory and methodology.
- What we have covered so far was the **theory** (the model itself)
- Now, we will focus on the **methodology** (how to fit the model on data / how to estimate the unknowns in the model)

Fitting the model

- More specifically, we will focus on the **theoretical basis** for fitting the model. Later on in the course, we will cover the actual thing in practice (software and examples).
- A **model** represents some hypothesized structure of data. Different **methods** are available for fitting the model to data and obtaining estimates of model parameters (the elements in model matrices) and providing us with information on how well the model **fits** the data.

Fitting the model

- For the sake of argument, we will consider the hypothetical scenario where the population correlation matrix \mathbf{P} is known, and the model holds exactly in the population (i.e., the model explains \mathbf{P} perfectly)
- This will never ever be the case in practice, but it's a better starting point to begin understanding the principles.
- Later, we will drop these assumptions, no worries.

The population correlation matrix

- Last time we talked about the case where the model holds exactly and the common factors are uncorrelated (orthogonal). In such a case, the population correlation matrix \mathbf{P} has the structure:

$$\mathbf{P} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{D}_{\psi}$$

...or, alternatively:

$$\mathbf{P} - \mathbf{D}_{\psi} = \mathbf{\Lambda}\mathbf{\Lambda}'$$

- Strictly speaking, we should use $\mathbf{\Lambda}^*$ and \mathbf{D}_{ψ}^* , but let's omit the stars now for convenience.

Rotational indeterminacy

- We have a bit of a “problem”.
- If we can find a single $p \times m$ factor loading matrix Λ_1 such that $\mathbf{P} = \Lambda \Lambda' + \mathbf{D}_\psi$, and if $m > 2$ (i.e., we have two or more factors), then there are infinitely many other $p \times m$ factor loading matrices such that $\mathbf{P} = \Lambda_1 \Lambda_1' + \mathbf{D}_\psi = \Lambda_2 \Lambda_2' + \mathbf{D}_\psi = \Lambda_3 \Lambda_3' + \mathbf{D}_\psi = \dots$
- Am I kidding? Nope. Let me show you.

Rotational indeterminacy

- Suppose I'm not wrong and it indeed holds that:

$$\mathbf{P} = \mathbf{\Lambda}_1 \mathbf{\Lambda}'_1 + \mathbf{D}_\psi = \mathbf{\Lambda}_2 \mathbf{\Lambda}'_2 + \mathbf{D}_\psi$$

(we're just considering two solutions now, but there are infinitely many)

- In that case, one solution ($\mathbf{\Lambda}_2$) has to be linked in some way with the other solution ($\mathbf{\Lambda}_1$). To be precise, $\mathbf{\Lambda}_2 = \mathbf{\Lambda}_1 \mathbf{T}$ where \mathbf{T} is a $m \times m$ orthogonal matrix ($\mathbf{T}\mathbf{T}' = \mathbf{I}$)

Rotational indeterminacy

- In that case, one solution (Λ_2) has to be linked in some way with the other solution (Λ_1). To be precise, $\Lambda_2 = \Lambda_1 \mathbf{T}$ where \mathbf{T} is a $m \times m$ orthogonal matrix ($\mathbf{T}\mathbf{T}' = \mathbf{I}$)

$$\Lambda_2 \Lambda_2' = \Lambda_1 \mathbf{T} (\Lambda_1 \mathbf{T})'$$

$$\Lambda_2 \Lambda_2' = \Lambda_1 \mathbf{T} (\mathbf{T}' \Lambda_1')$$

$$\Lambda_2 \Lambda_2' = \Lambda_1 \mathbf{T} \mathbf{T}' \Lambda_1'$$

$$\Lambda_2 \Lambda_2' = \Lambda_1 \mathbf{I} \Lambda_1'$$

$$\Lambda_2 \Lambda_2' = \Lambda_1 \Lambda_1'$$

- See? Λ_1 and Λ_2 are equally fine solutions.

Rotational indeterminacy

- In other words, if we can find one solution, we can find other alternative solutions. We simply choose any matrix \mathbf{T} such that $\mathbf{T}\mathbf{T}' = \mathbf{I}$ and we define $\mathbf{\Lambda}_2 = \mathbf{\Lambda}_1\mathbf{T}$
- We've just seen that $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are equally good solutions, since $\mathbf{\Lambda}_2\mathbf{\Lambda}_2' = \mathbf{\Lambda}_1\mathbf{\Lambda}_1'$
- This is called rotational indeterminacy. Later, when we learn about rotation, we will see that this describes a procedure used to produce alternative (and equally good) solutions to $\mathbf{\Lambda}$.

Rotational indeterminacy

- **However, we must resolve this problem somehow if we want to find a single, unique solution for Λ every time we perform a factor analysis.**
- **In other words, we need to find a criterion for defining this unique solution.**
- **Luckily, we can arrive at a solution with the help of Eigenvalues and Eigenvectors.**

Eigenvalues, eigenvectors, and Λ

- Recall what we have learned about eigenvalues and eigenvectors.
- The eigenstructure of a symmetric matrix \mathbf{S} is the following:

$$\mathbf{S} = \mathbf{U}\mathbf{D}_l\mathbf{U}'$$

...where the columns of \mathbf{U} are **eigenvectors** and the diagonal elements of \mathbf{D}_l are **eigenvalues** (this is a diagonal matrix).

Eigenvalues, eigenvectors, and Λ

- So, a symmetric matrix \mathbf{S} may be expressed as the product of some matrix (\mathbf{U}) which has some vectors for columns (these are the eigenvectors) and some diagonal matrix (\mathbf{D}_λ) which has some values on its diagonal (these are the eigenvalues)
- Thus, we can basically decompose a symmetric matrix \mathbf{S} into two other matrices (one of them is a diagonal matrix) that, when multiplied in some way, give you back your \mathbf{S} matrix.
- We're not trying to understand *why* that is or *what exactly* do the two matrices (\mathbf{U} and \mathbf{D}_λ) represent. We don't care at the moment. All we need to know is that this is possible and (as you will see) useful for FA.

Eigenvalues, eigenvectors, and Λ

- **Alright. For now, we will take the matrix $(\mathbf{P} - \mathbf{D}_\psi)$**
- **Remember, this is the population correlation matrix \mathbf{P} with unique variances (in \mathbf{D}_ψ) subtracted from the diagonal (because \mathbf{D}_ψ is diagonal)**
- **This matrix is also square and symmetric, just like \mathbf{S} from the previous slide.**

Eigenvalues, eigenvectors, and Λ

- The factor model says that $(\mathbf{P} - \mathbf{D}_\psi) = \Lambda\Lambda'$
- So, what we want is to find a unique matrix Λ that satisfies this model.
- Given our rotational indeterminacy problem, we really want to find a **unique** solution in some way, so that we would get this same solution every time we would run a factor analysis.

Eigenvalues, eigenvectors, and Λ

- Okay, so how do we use the eigen-stuff to find this unique solution?
- Some smart folks have invented the following:
- The eigen-decomposition of $(\mathbf{P} - \mathbf{D}_\psi)$ yields some \mathbf{U} and some \mathbf{D}_l
- If \mathbf{P} and \mathbf{D}_ψ satisfy the common factor model $\mathbf{P} = \Lambda\Lambda' + \mathbf{D}_\psi$ for m factors, then a loading matrix Λ can be constructed from the m largest eigenvalues of \mathbf{D}_l and the corresponding m standardized eigenvectors of \mathbf{U} (remember, each eigenvalue is associated with some eigenvector, hence the word “corresponding”)

Eigenvalues, eigenvectors, and Λ

- The eigen-decomposition of $(\mathbf{P} - \mathbf{D}_\psi)$ yields some \mathbf{U} and some \mathbf{D}_l
- If \mathbf{P} and \mathbf{D}_ψ satisfy the common factor model $\mathbf{P} = \Lambda\Lambda' + \mathbf{D}_\psi$ for m factors, then a loading matrix Λ can be constructed from the m largest eigenvalues of \mathbf{D}_l and the corresponding m standardized eigenvectors of \mathbf{U}
- Actually, the rest of the eigenvalues in \mathbf{D}_l ($p - m$ largest eigenvalues) are all zero.
- Anyway, if we take the eigenvalues in \mathbf{D}_l , take their square roots and put them back again into a matrix we will call $\mathbf{D}_l^{1/2}$, then $\Lambda = \mathbf{U}\mathbf{D}_l^{1/2}$

Eigenvalues, eigenvectors, and Λ

- **Magic.**
- **Remember, we don't care about why this works or how it works. All we care about now is that this procedure allows us to come up with a unique solution to Λ**
- **For the procedure to work, we need to know D_{ψ} , or the unique factor variances.**

Eigenvalues, eigenvectors, and Λ

- Let's look at an example, using the example data we have seen before.
- The matrix \mathbf{P} is given as follows:

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

Eigenvalues, eigenvectors, and Λ

- Assume the unique variances are known:

$$\mathbf{D}_{\psi} = \begin{bmatrix} 0.50 & & & \\ & .51 & & \\ & & .50 & \\ & & & .28 \end{bmatrix}$$

- So the matrix \mathbf{P} with communalities in the diagonal is given by:

$$(\mathbf{P} - \mathbf{D}_{\psi}) = \begin{bmatrix} .50 & & & \\ .49 & .49 & & \\ .14 & .07 & .50 & \\ .48 & .42 & .48 & .72 \end{bmatrix}$$

Eigenvalues, eigenvectors, and Λ

- We can obtain the eigenvalues and eigenvectors of $(P - D_\psi)$
- The non-zero eigenvalues are:

$$D_l = \begin{bmatrix} 1.662 & \\ & .548 \end{bmatrix}$$

- And the corresponding eigenvectors:

$$U = \begin{bmatrix} .502 & -.386 \\ .461 & -.500 \\ .353 & .731 \\ .641 & .259 \end{bmatrix}$$

Eigenvalues, eigenvectors, and Λ

- The factor loading matrix can be obtained: $\Lambda = \mathbf{UD}_t^{1/2}$

$$\Lambda = \begin{bmatrix} .647 & -.285 \\ .594 & -.370 \\ .455 & .541 \\ .826 & .192 \end{bmatrix}$$

- Wait...that's not the loading matrix I have shown you last time for the example data, is it?

Eigenvalues, eigenvectors, and Λ

- It's a transformation of the matrix I have shown you earlier, in the rotational indeterminacy sense, $\Lambda_2 = \Lambda_1 \mathbf{T}$

$$\Lambda = \begin{bmatrix} .647 & -.285 \\ .594 & -.370 \\ .455 & .541 \\ .826 & .192 \end{bmatrix} = \begin{bmatrix} .70 & .10 \\ .70 & .00 \\ .10 & .70 \\ .60 & .60 \end{bmatrix} \begin{bmatrix} .848 & -.529 \\ .529 & .848 \end{bmatrix}$$

- Both Λ matrices provide an exact solution to the model. The procedure involving eigen-stuff allowed us to identify the unique solution, though.

Eigenvalues, eigenvectors, and Λ

- Okay, so, I have just shown you how to obtain the solution (Λ) if:
 - You know the population correlation matrix, \mathbf{P}
 - You know the contents of \mathbf{D}_ψ , so you know the unique variances or (conversely) the communalities
 - The model holds exactly in the population
- Huh. Putting the “model holds exactly” thing aside, you will never know \mathbf{P} and you will never know \mathbf{D}_ψ , so this is a theoretical scenario.
- That’s true, but this serves as a basis for things to come.

The communality problem

- As I said, the solution obtained by doing the eigen-decomposition of $(\mathbf{P} - \mathbf{D}_\psi)$ requires that you know either the unique variances or the communalities (once you know one, you know the other one, right?)
- But we don't know these, since finding out what they are is a part of the problem we face.
- When factor analysis was young, this was called the "Communality problem"

The communality problem

- Many solutions were suggested to the communality problem.
- The one that “won” (was and is the most widely used) was suggested by Louis Guttman in 1940.
- Guttman suggested *squared multiple correlations* (SMCs) as the initial approximations to communalities.

The communality problem

- Just what is a *squared multiple correlation* (SMC)?
- Imagine you have p manifest variables. You can try to predict the j -th manifest variable from the other $(p - 1)$ manifest variables, linear regression-style.
- This prediction will be imperfect. You can correlate these predicted values of the j -th manifest variable with the actual values of the variable. What you will get is a correlation coefficient, the **multiple correlation coefficient**. Square it and you get the SMC.

The communality problem

- Guttman has shown that if the factor model applies to the population correlation matrix \mathbf{P} , then the squared multiple correlation of the j -th manifest variable on the other $(p - 1)$ manifest variables is the *lower bound* for the communality of the j -th manifest variable.
- So, not knowing the contents of \mathbf{D}_ψ , one might approximate the manifest variable communalities with manifest variable SMCs, computed from \mathbf{P} . These approximations can then be substituted into the diagonal of \mathbf{P} and one can, again, use the eigenvalue-eigenvector approach on this modified \mathbf{P} matrix to obtain $\mathbf{\Lambda}$.

The communality problem

- However, in order to obtain the population SMCs, we need to know \mathbf{P} in the first place. Most often, we don't.
- In practice, we can apply the same procedure to a sample correlation matrix, \mathbf{R} , in order to obtain sample SMCs. Since, in reality, we usually work with sample correlation matrices, let's slowly shift the gear towards thinking more about a sample correlation matrix \mathbf{R} and less about the population correlation matrix, \mathbf{P} .

Working with a sample correlation matrix

- So far, we have studied factor analysis limiting ourselves to the ideal scenario in which we know the population correlation matrix, \mathbf{P} . Moreover, we only considered the case where the model holds exactly in the population.
- Now, let's consider the real world in which we do not have access to \mathbf{P} but we do have access to \mathbf{R} . In this real world scenario, we are not even sure the sample correlation matrix \mathbf{R} is drawn from a population with a correlation matrix \mathbf{P} for which the model holds.
- As before, let's just consider the uncorrelated / orthogonal model for now.

Working with a sample correlation matrix

- First of all, we should tone down the optimism. In our hypothetical scenarios, we could select Λ and D_ψ to reconstruct \mathbf{P} perfectly:

$$\mathbf{P} = \Lambda\Lambda' + D_\psi$$

- In reality, our *estimates* of Λ and D_ψ , $\widehat{\Lambda}$ and \widehat{D}_ψ , will generally not be able to exactly reproduce our sample correlation matrix \mathbf{R} :

$$\mathbf{R} \approx \widehat{\Lambda}\widehat{\Lambda}' + \widehat{D}_\psi$$

Working with a sample correlation matrix

- So, what we want is a **parsimonious** model ($m \ll p$) that provides a relatively good approximation to the data we have observed.
- This degree of approximation (how well the model fits the data) is reflected in the **residual matrix**, defined as $\mathbf{R} - (\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}' + \widehat{\mathbf{D}}_{\psi})$
- The residual matrix tells us how far away the correlation matrix \mathbf{R} we have observed is from the correlation matrix the model predicts. In other words, how far is the observed correlation matrix from the model-implied correlation matrix (which is simply $\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}' + \widehat{\mathbf{D}}_{\psi}$)

Working with a sample correlation matrix

- Every element in the residual matrix tells us how far is the model-implied (predicted) value of this element from its observed value.
- Alright, so – again, we don't have a population correlation matrix \mathbf{P} which we used for all the computations and methods covered before. What are we going to do?
- Of course, we're going to pretend like the problem isn't there and we'll start by doing things in the exact same way.

Working with a sample correlation matrix

- First, we will obtain some initial communality estimates and plug them into the diagonal of R . We can use the SMCs.
- This way, we will arrive at our $(R - D_\psi)$ matrix just as we did arrive previously at the $(P - D_\psi)$ matrix. [Oh, by the way, did I ever call this a *reduced* correlation matrix? I didn't? Well, now I do.]
- We get the eigenvalues and eigenvectors for this reduced [see?] sample correlation matrix.

Working with a sample correlation matrix

- Again, we will obtain some eigenvalues and some eigenvectors. However, in this case (**not** having a population correlation matrix, **not** being sure the model holds exactly in the population), we will generally not obtain an eigen-solution where the $(p - m)$ smallest eigenvalues are zero.
- Thus, we cannot rely on the number of non-zero eigenvalues to show us the “true” number of factors (m). Thus, we will have to choose m ourselves beforehand, based on our best judgement (more on that later)

Working with a sample correlation matrix

- Thus, having chosen the number m beforehand, we will take the m largest eigenvalues and their corresponding m eigenvectors.
- Just like before, we will take the square root of the eigenvalues, sort them by size and place them in a diagonal matrix $\hat{\mathbf{D}}_{lm}^{1/2}$
- And, just like before, we will create a matrix $\hat{\mathbf{U}}_m$ with the corresponding eigenvectors as columns.

Working with a sample correlation matrix

- Then, we can use the eigenvalues and eigenvector matrices to compute our estimate of factor loadings:

$$\hat{\Lambda} = \hat{U}_m \hat{D}_{lm}^{1/2}$$

- The $\hat{\Lambda}$ obtained in this way minimizes the residual sum of squares (RSS):

$$RSS = \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p [(\mathbf{R} - \hat{\mathbf{D}}_{\psi}) - \hat{\Lambda} \hat{\Lambda}']_{ij}^2$$

- Let me illustrate the formula for RSS now on the board

Working with a sample correlation matrix

- This $\hat{\Lambda}$ results in minimum sum of squared residuals, **conditional** on the given set of prior communality estimates.
- This method is known as the *principal factor method using prior communality estimates* (whoa, that's a LONG name)
- Let's look at an example.

Example

- Previously, we used the population correlation matrix for the four-tests-data for examples. Assume we have a sample correlation matrix **R** from the same data, different from the population matrix **P**:

$$\mathbf{R} = \begin{bmatrix} 1 & & & \\ .51 & 1 & & \\ .11 & .17 & 1 & \\ .41 & .52 & .39 & 1 \end{bmatrix}$$

- We would like to estimate the model, but we do not know the true communalities. Also, we do not know if the model holds in the population.

Example

- We can, however, compute the SMCs as approximations to the true communalities.
- In this case, we would get the following SMCs: .29, .377, .156 and .39 (respectively). We will replace the diagonal elements in R with the communality estimates to get our reduced sample correlation matrix:

$$(R - \hat{D}_\psi) = \begin{bmatrix} .29 & & & \\ .51 & .377 & & \\ .11 & .17 & .156 & \\ .41 & .52 & .39 & .39 \end{bmatrix}$$

Example

- Choosing $m = 2$, we get the following m eigenvalues and m eigenvectors:

$$\hat{\mathbf{D}}_l = \begin{bmatrix} 1.438 & \\ & .184 \end{bmatrix}$$

$$\hat{\mathbf{U}} = \begin{bmatrix} .492 & -.462 \\ .571 & -.331 \\ .296 & .754 \\ .586 & .330 \end{bmatrix}$$

Example

- ...and we compute $\hat{\Lambda}$:

$$\hat{\Lambda} = \hat{U} \hat{D}_l^{1/2} = \begin{bmatrix} .590 & -.198 \\ .685 & -.142 \\ .355 & .323 \\ .703 & .141 \end{bmatrix}$$

- Interesting point: the sums of squares of each column correspond to the eigenvalues.

Example

- Then, we can compute the model-implied **reduced** correlation matrix:

$$\widehat{\Lambda}\widehat{\Lambda}' = \begin{bmatrix} .387 & & & \\ .432 & .489 & & \\ .146 & .198 & .231 & \\ .387 & .461 & .295 & .514 \end{bmatrix}$$

- On the diagonal of this matrix are the communalities for this solution. These values are the sums of squares of factor loading in each row of $\widehat{\Lambda}$
- Off the diagonal are the correlations between the MVs as reconstructed by this particular model solution.

Example

- Compare the model-implied reduced correlation matrix to the observed reduced correlation matrix:

$$\widehat{\Lambda}\widehat{\Lambda}' = \begin{bmatrix} .387 & & & \\ .432 & .489 & & \\ .146 & .198 & .231 & \\ .387 & .461 & .295 & .514 \end{bmatrix}$$

$$(R - \widehat{D}_\psi) = \begin{bmatrix} .29 & & & \\ .51 & .377 & & \\ .11 & .17 & .156 & \\ .41 & .52 & .39 & .39 \end{bmatrix}$$

Example

- They don't match – the model does not fit the data perfectly.

$$\widehat{\Lambda}\widehat{\Lambda}' = \begin{bmatrix} .387 & & & \\ .432 & .489 & & \\ .146 & .198 & .231 & \\ .387 & .461 & .295 & .514 \end{bmatrix}$$

$$(R - \widehat{D}_\psi) = \begin{bmatrix} .29 & & & \\ .51 & .377 & & \\ .11 & .17 & .156 & \\ .41 & .52 & .39 & .39 \end{bmatrix}$$

Example

- This lack of fit is represented in the residual matrix:

$$(R - \hat{D}_\psi) - \hat{\Lambda}\hat{\Lambda}' = \begin{bmatrix} -.097 & & & \\ .078 & -.113 & & \\ -.036 & -.028 & -.075 & \\ .023 & .059 & .095 & -.125 \end{bmatrix}$$

Example

- The solution produced a residual matrix with minimum sum of squares, conditional on the prior communality estimates. If the prior communality estimates would be different, a different residual matrix would satisfy the RSS criterion.

Short review

- So, what was the principle behind the *principal factor method using prior communality estimates*? Let's do a short recap:
- 1) First, we obtain some communality estimates (like SMCs) and plug them into the diagonal of R . Thus, we get our estimate of $(R - D_\psi)$
- 2) Then, we obtain the eigen-solution of $(R - D_\psi)$
- 3) We use the eigen-solution to obtain $\hat{\Lambda}$
- 4) What we just got is a solution that minimizes the Residual Sum of Squares (RSS) given our initial \hat{D}_ψ

Iterative procedure

- Alternatively, we could try to minimize the RSS criterion by estimating \mathbf{D}_ψ alongside $\mathbf{\Lambda}$ and not sticking only with the initial $\hat{\mathbf{D}}_\psi$
- This can be done iteratively, by minimizing RSS with respect to both $\hat{\mathbf{D}}_\psi$ and $\hat{\mathbf{\Lambda}}$
- This technique is called *iterative principal factors* or *ordinary least squares (OLS)* or *unweighted least squares (ULS)* or *minres*

Iterative procedure

- We will start by doing things the same way we did previously, using the *principal factors method*:
- 1) First, we obtain some communality estimates (like SMCs) and plug them into the diagonal of \mathbf{R} . Thus, we get our estimate of $(\mathbf{R} - \mathbf{D}_\psi)$
- 2) Then, we obtain the eigen-solution of $(\mathbf{R} - \mathbf{D}_\psi)$
- 3) We use the eigen-solution to obtain $\hat{\Lambda}$
- ...but we won't end here. We will use the computed $\hat{\Lambda}$ to obtain new communality estimates by summing the squared elements in each row of $\hat{\Lambda}$ (diagonal elements of $\hat{\Lambda}\hat{\Lambda}'$)

Iterative procedure

- We shall take the new communality estimates and plug them into the diagonal of \mathbf{R} . Thus, we get a new $(\mathbf{R} - \mathbf{D}_\psi)$
- Again, we obtain the eigen-solution of this new $(\mathbf{R} - \mathbf{D}_\psi)$ and use it to compute a new $\hat{\Lambda}$
- ...and repeat (use the newly computed $\hat{\Lambda}$ to again obtain new communality estimates). We continue this process until the communalities obtained in successive iterations do not significantly differ by some pre-set criterion (convergence criterion).

Iterative procedure

- That's really all there is (in principle) about OLS.
- By the way, the RSS function (the formula we have seen before) is a *discrepancy function* – it quantifies the distance between the observed and model-implied correlation matrices. In other words, it expresses the degree of lack of model fit.
- Being a discrepancy function, it is always greater than or equal to zero and is zero **only** when the observed and model-implied correlation matrices are the same.

Heywood cases

- One nasty thing can happen when using OLS estimation
- That is, some communalities can, in the course of the iterations, be greater than one. Conversely, the unique variances can become less than zero (because in a standardized solution, the communality and the unique variance of an MV add up to one)
- But there's no such thing as negative variance. Thus, such a solution would be nonsensical and unacceptable. We call these occurrences *Heywood cases*

Heywood cases

- If you're using smart software, you should be notified whenever a *Heywood case* occurs
- If you're using smart software, it can help you circumvent the problem by placing a constraint on the associated unique variance such that it can only be greater than or equal to zero.

Summary

- We considered multiple scenarios of fitting the model to data. Let's do a quick review.
 - 1) You know \mathbf{P} and you know \mathbf{D}_ψ . You can obtain the eigen-solution of $(\mathbf{P} - \mathbf{D}_\psi)$ to compute $\mathbf{\Lambda}$.
-however, this will never be the case in practice.

Summary

- We considered multiple scenarios of fitting the model to data. Let's do a quick review.
- 2) You know \mathbf{P} but you do not know \mathbf{D}_ψ . You can estimate communalities using SMCs and plug them into the diagonal of \mathbf{P} to obtain $(\mathbf{P} - \mathbf{D}_\psi)$. Afterwards, you obtain the eigen-solution of $(\mathbf{P} - \mathbf{D}_\psi)$ to obtain $\mathbf{\Lambda}$.

....however, this will **also** never be the case in practice.

Summary

- We considered multiple scenarios of fitting the model to data. Let's do a quick review.
- 3) You do not know \mathbf{P} and you do not know \mathbf{D}_ψ . All you have is \mathbf{R} . You can estimate communalities using SMCs and plug them into the diagonal of \mathbf{R} to obtain $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$. Obtain the eigen-solution of $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ to get $\hat{\mathbf{\Lambda}}$.

...the solution minimizes RSS given your original $\hat{\mathbf{D}}_\psi$. This can happen very often in practice, although we would normally use a better option coming up next.

Summary

- We considered multiple scenarios of fitting the model to data. Let's do a quick review.
- 4) You do not know \mathbf{P} and you do not know \mathbf{D}_ψ . All you have is \mathbf{R} . You can estimate communalities using SMCs and plug them into the diagonal of \mathbf{R} to obtain $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$. Obtain the eigen-solution of $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ to get $\hat{\mathbf{\Lambda}}$. Use the computed $\hat{\mathbf{\Lambda}}$ to obtain new communality estimates from the diagonal of $\hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}'$. Return to the beginning with fresh new communality estimates, repeat until convergence.