Analytic Rotation

PSY544 – Introduction to Factor Analysis

Week 10

Introduction

- A big aspect of the EFA methodology is rotation.
- Rotation is a procedure applied to the factor loading matrix Λ to enhance the interpretability of factor loadings (and, in effect, the factors themselves)

- The process and methods of rotation are often not very well understood by users.
- As a result, the rotation process is often inadequate in practice.

Introduction

- Back in the early days, rotation was done by hand (!) using graphical methods – hence the name *rotation*

- Nowadays, analytical methods are employed using computers

- …however, it's kinda funny to think about early factor analysts turning things around by hand on a big board (well, it was probably not them, but their assistants) [Most computations were done by "computers" – largely women, by the way]

Introduction

- To begin with, recall what the factor loading is. In the common factor model:

$$x = \mu + \Lambda z + u$$

...the factor loadings are regression coefficients, predicting the value of the manifest (dependent) variable from the value of the latent (independent) variable(s).

Example from earlier slides:

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

Introduction

- We know from multiple regression analysis that if the factors are uncorrelated, the λs can be interpreted as simple correlations between the factors and the manifest variables. Therefore, if the factors are orthogonal ($\mathbf{\Phi} = \mathbf{I}$), the loadings are bounded between -1 and +1.

- When the factors are not orthogonal ($\mathbf{\Phi} \neq \mathbf{I}$), the factor loadings can no longer be interpreted as simple correlations, and are no longer bounded between -1 and +1

Introduction

- When the factors are correlated (oblique), the factor loadings can be interpreted as factor weights, representing the linear influence a factor has a particular MV, but no longer as simple correlation.

- This sometimes confuses people when they see a rotated solution with correlated factors that contains factor loadings greater than +1 or smaller than -1.

Introduction

- When we formulate a factor analysis model, one of the things we do is interpreting the factors – giving them meaning, gaining some sense about them.

- This endeavor is greatly simplified if most of the loadings corresponding to a manifest variable j are zero or close to zero. This means that the MV is influenced only by a small number of factors.

- Which is great, because we rely on the structure of loadings to get a sense of what the factor might stand for.

Introduction

- When we want to understand a factor, we look at the column of factor loadings in Λ that corresponds to that particular factor, and we look at a characteristic or a property that is shared by the manifest variables that have high loadings on the factor.

- This allows us to "name" the factor – infer on its nature or meaning through observing which MVs does it strongly affect (...and which MVs it does not)

Introduction

- You can simply change the sign of all elements in any column of Λ. Doing this simply reverses the meaning of the factor. For instance, a factor that we named "spatial reasoning" (because of positive factor loadings on test items supposed to measure spatial reasoning) would become its opposite, "lack of spatial reasoning" (because of negative factor loadings...)

- If factors are orthogonal, no change needs to be done for Φ. If factors are not orthogonal, then if we reverse the sign of factor loadings for this factor, we need to reverse the signs of all Φ elements in the row and column corresponding to the given factor.

Introduction

- You can also re-order the columns of Λ as you like.

- If factors are orthogonal, no change needs to be done for Φ. If factors are not orthogonal, then re-ordering the columns of Λ must be accompanied by re-ordering the rows and columns of Φ.

Orthogonal rotations

- Previously, we have discussed the issue of rotational indeterminacy.

- If $\mathbf{P} = \mathbf{\Lambda_1\Lambda_1'} + \mathbf{D}_\psi$ then we can find other loading matrices:

$$\mathbf{\Lambda_2} = \mathbf{\Lambda_1 T}$$

such that

$$\mathbf{P} = \mathbf{\Lambda_1\Lambda_1'} + \mathbf{D}_\psi = \mathbf{\Lambda_1 TT'\Lambda_1'} + \mathbf{D}_\psi = \mathbf{\Lambda_1 T(\Lambda_1 T)'} + \mathbf{D}_\psi = \mathbf{\Lambda_2\Lambda_2'} + \mathbf{D}_\psi$$

where $\mathbf{T}$ is an orthogonal matrix $(\mathbf{TT'} = \mathbf{I})$

Orthogonal rotations

- This matrix **T** is called a *transformation matrix* and there are infinitely many such transformation matrices available.

- Our goal is to find such **T** so that the resulting Λ_2 is interpretable / is more easily interpretable than the original Λ_1

- However, this is still a mathematical procedure, so we need to find some mathematical definition of "interpretable".

Orthogonal rotations

- Remember – because the alternative solutions to Λ are equally good solutions, rotation does not affect the fit of the model. Neither does it affect the communalities.

- Back to the definition of „interpretable" – again, a solution that would be easy to interpret would be one with some loadings close to zero, some of high magnitude and an overall detectable pattern of loadings.

- The zero loadings are important – if all the elements in the first column are high and all others are close to zero, that would not be easy to interpret.

Thurstone's "simple structure"

- Thurstone (1947) has defined requirements of a factor loading matrix with so-called *simple structure* to be more easily interpretable

- The requirements are phrased in terms of the number and location of zero (or very small) factor loadings in the Λ matrix

Thurstone's "simple structure"

- For a $p \times m$ factor loading matrix:

1. Each row of Λ should contain at least one zero

2. Each column of Λ should contain at least m zeros

3. Every pair of columns of Λ should have a couple of rows with a zero in one column but not the other

4. If $m \geq 4$, every pair of columns of Λ should have several rows with zeros in both columns

5. Every pair of columns of Λ should have few rows with nonzero loadings in both columns

Thurstone's "simple structure"

- In general, Thurstone's criteria suggest that each factor should be represented by relatively high loadings for a distinct subset of MVs and relatively low loadings for the remaining MVs.

- In addition, these subsets defining different factors should not overlap too much.

- Furthermore, each MV should only be influenced by some subset of the common factors.

- The criteria do **not** imply that each MV should only be influenced by a single factor, which is a common misconception.

Analytic rotation

- In the 1950s, researchers began developing automated procedures for rotation. The guiding principle was to establish an objective, mathematical definition of simple structure.

- Generally, this is achieved by defining a function of the elements in the factor loading matrix (Λ) that expresses the degree of simple structure numerically. We refer to such a function as a *simplicity function* or a *simplicity criterion*. This is a scalar-valued function with a matrix-valued argument, $f(\Lambda)$.

Analytic rotation

- The transformation matrix is then found as the matrix which maximizes the simplicity function (or, rather, yields such a Λ which maximizes the simplicity function).

- This general approach is called *analytic rotation*. Analytic rotation requires no user judgement or subjective input (unlike the earlier methods for manual rotation)

Analytic rotation

- Sometimes, it is convenient not to maximize a function that measures the simplicity of Λ, but to minimize a function that measures the complexity of Λ. Such a function would be called a *complexity function*. The idea is the same.

- Because the signs of the factor loadings depend on arbitrary choices for scoring the latent variables, it is convenient to base the simplicity or complexity functions on squared factor loadings, λ_{jk}^2. Then, the criterion is unaffected by the signs of factor loadings.

Quartimax

- The first suggested simplicity criterion. It's the sum of the fourth powers of factor loadings:

$$Q(\Lambda) = \sum_{j=1}^{p} \sum_{k=1}^{m} (\lambda_{jk}^2)^2$$

- This is equivalent to the overall variance of squared factor loadings.

- *Given an unrotated factor loading matrix Λ, choose a transformation matrix T such that $Q(\Lambda)$ is maximized with the rotated loading matrix ΛT*

Varimax

- It turned out that Quartimax tended to provide a Λ with a single column of large loadings and small loadings in other columns. In most cases, that's not desirable.

- The Varimax criterion was suggested instead. Varimax is the sum of the m within-column variances of squared factor loadings:

$$V(\Lambda) = \sum_{k=1}^{m} \frac{1}{p} \sum_{j=1}^{p} (\lambda_{jk}^2 - \lambda_{.k}^2)^2$$

- where $\lambda_{.k}^2 = \frac{1}{p} \sum_{j=1}^{p} \lambda_{jk}^2$ is the within-column mean squared loading.

Varimax

- As simple structure improves, the squared loadings on factors become more variable (some loadings high, the rest low). Summing the variances of the squared loadings over all m factors provides a measure of simplicity.

- The described criterion is known as *raw Varimax* because it is applied to the raw factor loadings. Kaiser found that it works well, but sometimes, in rows with small communalities, it does not. He therefore standardized rows of the factor matrix by dividing factor loadings by the square roots of communalities before rotation. This is usually called *normal Varimax* or *Varimax with Kaiser normalization*.

Varimax

- Varimax tends to work well as an orthogonal rotation.

- However, Varimax almost monopolized the entire enterprise of orthogonal rotations in applied research (bluntly – everyone uses Varimax all the time)

- Let's take a look at our example data, before and after a Varimax rotation.

Varimax

- Unrotated factor loadings:

	1	2	3
SntComp	0.72	0.38	-0.23
MxdArit	0.90	-0.34	-0.03
MissNum	0.86	-0.13	0.00
Boots	0.48	0.25	0.54

Varimax

- Rotated factor loadings:

(note the deviations from simple structure)

	1	2	3
SntComp	0.16	0.75	0.34
MxdArit	0.18	0.25	0.91
MissNum	0.26	0.36	0.75
Boots	0.72	0.19	0.17

Analytic rotation

- I suggest you perform rotations for various number of extracted factors when exploring the factor structure using EFA. This can also help you in determining the number of factors.

- Under-factoring tends to result in multiple factors collapsed into one, which can manifest as a solution that heavily violates simple structure or that is not easily interpretable.

- Over-factoring can result into a solution which has a column(s) of loadings with only a single non-zero element, or a column(s) of loadings with all elements very small.

Orthogonal rotation?

- As we know, orthogonal rotations require the rotated factors to be orthogonal. In other words, we impose the constraint that the transformation matrix T has to be an orthogonal matrix.

- Is this reasonable, though? With *exploratory* factor analysis, the goal is, after all, to explore the number and nature of the major common factors. How do we know a priori that the factors are uncorrelated?

Orthogonal rotation?

- In reality, this restriction is mostly uncalled for. In the domains we frequently use FA (mental abilities, attitudes, personality, consumer research, public health), we would on the contrary expect the factors to be a priori correlated.

- Orthogonal rotations are, however, still used very often in practice. Why is that?

Orthogonal rotation?

- It's what everyone is doing, so I'll do it, too.
- It's simple.
- It's the default setting in the program I use.
- Lack of understanding of rotation.
- Desire for the factors to be uncorrelated.
- "Varimax" sounds cool.

Orthogonal rotation?

- Does any of that matter? Of course not.
- Orthogonal rotations were made for times when computers were the size of a room and computations were slow.
- We should be using *oblique rotations* instead. Imposing the constraint of uncorrelated factors is, by large, unjustified.
- Moreover – if the best solution (in terms of simple structure) is a one with uncorrelated factors, oblique rotation will find it as such (with oblique rotation, factors are *allowed to* correlate, not *required to*)

Orthogonal rotation?

- With oblique rotations, we can expect the solutions to be more easily interpretable with a simpler structure – just because we have accounted for the potential systematic relationships between the latent variables.

- It's just more realistic. Keep it real, man.

Oblique rotations

- With oblique rotation, we still have a transformation matrix T, we just don't require it to be orthogonal. Instead, we require that:

$$T^{-1}(T^{-1})'$$

...has unit diagonals. The reason is (simply put) the following:

$$T^{-1}(T^{-1})' = \Phi$$

...so we have our full model with a factor correlation matrix Φ:

$$\Sigma = \Lambda\Phi\Lambda' + D_\psi$$

Oblique rotations

- Communalities remain unchanged (because D_ψ remains unchanged)

- Setting aside some of the more historic methods of oblique rotation, two approaches are relevant today – the Oblimin family and the Crawford-Ferguson family.

- The Oblimin family uses a simplicity criterion which includes a "tuning constant", γ, which can be toggled by the user. Different values of the constant lead to different rotation criteria.

Oblique rotations

- Whenever you use a rotation from the Oblimin family, make sure to report the value of γ. The recommended value, however, is 0 (in that case, we're talking about the *Direct Quartimin* criterion).

- The Crawford-Ferguson family is more satisfactory. Let's take a closer look (the CF family is implemented by CEFA, for instance).

- Note: The CF family can produce oblique as well as orthogonal rotations.

Crawford-Ferguson family

- The CF family uses a complexity function of Λ:

$$CF(\Lambda) = (1 - \kappa) \sum_{j=1}^{p} \sum_{k=1}^{m} \sum_{\substack{l=1 \\ k \neq l}}^{m} \lambda_{jk}^2 \lambda_{jl}^2 + \kappa \sum_{k=1}^{m} \sum_{j=1}^{p} \sum_{\substack{h=1 \\ j \neq h}}^{p} \lambda_{jk}^2 \lambda_{hk}^2$$

- κ is a tuning constant between 0 and 1. The function can be understood as follows: $(1 - \kappa)$ times row parsimony plus κ times column parsimony.

Crawford-Ferguson family

- In CEFA, a couple of criteria within the CF family (such as CF-Quartimax, CF-Varimax, CF-Equamax, etc…) are already implemented without the need for the user to set the value of κ.

- I won't bother you with this much more. I'll provide a couple of practical points at the end of the presentation, and if you're interested, read more about the CF family yourself – you should be able to understand even the more technical papers now.

- Let's use the CF-Quartimax on our example data.

CF-Quartimax

- Rotated factor loadings:

	1	2	3
SntComp	0.14	0.77	-0.03
MxdArit	1.01	-0.05	-0.04
MissNum	0.75	0.13	0.06
Boots	0.03	0.03	0.73

- Factor correlations:

	1	2	3
2	0.59	1	

CF-Quartimax

- As can be seen, the pattern of loadings is much simpler and easier to interpret.

- Factors are substantially correlated.

- Conducting oblique rotation is straightforward in most software. Use it!
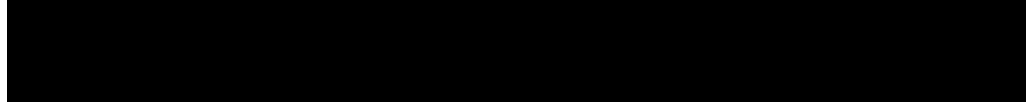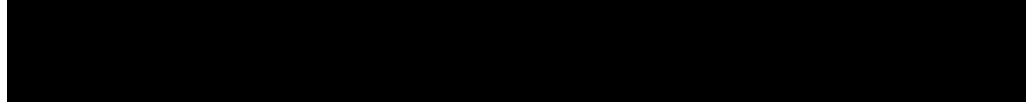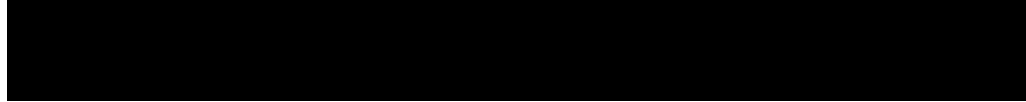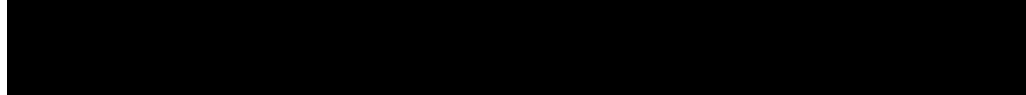
Target rotation
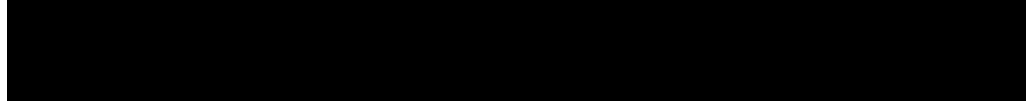
- There is one rotation that can be used in a more confirmatory manner – the target rotation.

- One can think of the target rotation as standing between exploratory and confirmatory factor analysis. It is useful when you already have some prior knowledge about the factor loading pattern, but not enough to warrant a fully confirmatory model.

- Can be oblique or orthogonal.

Target rotation

- The rotation criteria described before are sometimes called *blind rotation* as there is no (or little) room for user input.

- Target rotation, on the other hand, requires input from the user. The analyst sets up a hypothesized pattern of factor loadings, and the software tries to find a loading matrix that is as close as possible (in a least squares sense).

- The *target matrix* has the same dimensions as the loading matrix Λ.

Target rotation

- Target matrix, CEFA-style:

- 0 = loading expected to be small
- ? = unspecified, not small

- The sum of squares of loadings corresponding to the zeros is minimized

	1	2	3
SntComp	?	0	0
MxdArit	0	?	0
MissNum	0	?	0
Boots	0	0	?

Some final points

- Use the CF family, and do oblique rotations. I really don't see a lot of sense in performing orthogonal rotations.

- Try out multiple oblique rotations – CF-Quartimax, CF-Varimax...

- If you have a bit of an idea on what you expect, you might want to try using (oblique) target rotation. CEFA can do it, and this method is pretty under-utilized in applied research.