Workshop Construction and Analysis of Tests and Questionnaires: Lab Session 2

The goal of an explorative factor analysis (EFA) is to study the internal structure of a set of questionnaire items. More specifically, using EFA you want to investigate whether the association between items can be explained by a limited number of factors. These factors represent hypothetical constructs, also referred as latent variables (i.e., characteristics on which people differ, but which cannot be directly observed; examples include personality, attitude, values, etc.).

In practice, EFA amounts to answering two main questions:

- Do the items measure one single factor, or multiple factors?
- What is the internal structure of the questionnaire/test: Which items "load" on which factor?

Rotation of factors

To ensure that we have an interpretable solution we use so called rotation. By using rotation we may see if the items satisfy a simple structure, which means that for each factor we have items that load high on that factor, and low on the other factors.

A distinction is made between orthogonal and non-orthogonal rotation. If we use orthogonal rotation we look for an interpretable solution assuming the factor are independent (i.e., uncorrelated).

Lab session

The goal of this lab session is to see the foundations of EFA, how to apply it in SPSS, and how to interpret the basic SPSS output resulting from the factor analysis.

The data you will be using consists of the individual answers of $N = 295$ second-year bachelor students on a personality questionnaire that consists of a total of $K = 30$ items. You will perform an EFA to see if we can answer the questions above.

Exercises

Part I: Validity (Internal structure)

Use SPSS to apply the following steps & answer the questions:

1. Test if the relationship between the items is linear
2. Perform the first EFA:
   Does it make sense to perform a EFA, are all assumptions met?
   How many factors do you choose, and why?
3. Perform a second EFA with **varimax** rotation & fixed number of factors:
   How many items violate the rule of a simple structure?
4. Perform a third EFA with **oblimin** rotation & fixed number of factors:
   How man items violate the rule of a simple structure?
   How strongly are factors correlated with each other?
5. Select a solution from the results obtained in steps 3-4 (choose the form of rotation)

The first difficulty you might encounter in step 2 is the determination of the number of factors. More often than not, it is not immediately clear how many factors you should choose. The larger-than-one rule is a clear rule but often indicates too much factors. The Scree plot is better in this sense but does not always give an unequivocal answer to the following question: How many dots clearly fall above the line you draw (the "elbow", the point where the slope of the curve is clearly leveling off)? If it is unclear how many factors you have to choose, then perform several EFAs, each with a different number of factors. In the end you choose from all these possible results, for example based on interpretation (always go with the solution that is easier to interpret). Another way to make sense of the data (and to choose the number of factors) is to look at the theory that the questionnaire was based on.

Once you have chosen the number of factors and the form of rotation, you can start interpreting the results & evaluating your final EFA solution/model:

1. How much % of the variance can be explained by the EFA solution you chose?
2. Look at the residual correlations: does the factor model fit the data well?
3. Consult the item loadings for every item & write down which items belong to what factor/dimension
4. Based on the item content, provide an appropriate label for every factor/dimension
5. For every factor: How should we interpret low scores and high scores on the scale that measure this factor? (e.g., how would you characterize a low scoring person and a high scoring person?)
6. The items are taken from a well-known personality questionnaire that is used a lot in the social sciences and in particular in psychology: Do you recognize it? (hint: think of the bigger construct that the questionnaire is supposed measure)

Part II: Reliability

Now that you have performed an EFA, you know the internal structure of the questionnaire and you can calculate scale scores to perform a reliability analysis for every dimension of the scale.

Apply the knowledge you gained on the first day of this workshop and compute the sum scores for all individuals and then perform a reliability analysis for each scale separately. Don't forget to recode items if needed!

In SPSS, make sure that you ask for Descriptives for Scale and Scale if item Deleted. In addition, ask for the Guttman's Lambda statistics.

For every scale separately, answer the following questions:

1. What is the reliability of the scale?
2. Look at the contribution of each item to its scale. Would you consider removing an item from one of the scales? Motivate your choice!
3. Which item contributes the most to the reliability of the scale?
4. Make a histogram of the sum scores (via graphs -> legacy dialogs -> histogram). Comment on the shape of the distribution. What does it convey about the "difficulty" of the scale?

Appendix

Guidelines

The following rules and questions will help you during the analyses:

- Consult the table with factor loadings (in case of an oblimin rotation, this is the pattern matrix). This matrix gives the factor loadings (correlation between item and a factor). The larger the value, the larger the effect of the factor. If the value is 0, it means that the factor does not explain any of the variance in item responses. In practice, usually we want the loadings to be at least .3. So when inspecting the matrix, we determine for every item, on which factor the item loads the strongest (e.g., has the highest correlation with) to see if we have a pattern/can identify dimensions in the data.
- In general, we consider the model as good if the residual correlations are less than .05, and acceptable if most of the residual correlations are less .05 and the others less than .15.
- After rotation, does each factor have at least three items with a high loading on the factor? If the answer is no, then you have to select a solution with fewer factors because a scale with less than 3 items is rarely a good and appropriate scale.
- If there is an item that belongs to no scale (all loadings of the item, in absolute value, <0.3), remove this item and then do the analysis again. This is an item not belonging to any of the scales.
- If a set of items is too small (less than 3 items) or this set yields a scale with an insufficient reliability (alpha clearly smaller than 0.6, or multiple items with low contribution to reliability), then do not make a scale with these items.
- What if the rule for a simple structure is clearly violated and an item is associated with multiple components? There are a number of options. The best procedure is the following one: If the item clearly has a much higher loading on one factor, then add the item to that factor. If the item loads similarly on multiple components, then you have to check whether it contributes to a scale (using reliability analysis). If it does not fit any scale, remove it. If it fits only one of the scales, the problem is solved. If it fits multiple scales, add it to the scale with the least number of items. In the end, you may attribute the item to only one scale!
- **Do not forget to reverse items before you perform a reliability analysis!**

SPSS commands

To run a simple EFA (no fixed number of factors, no rotation) 1in SPSS, proceed as follows:

- Choose: Analyze → dimension reduction →> factor analysis. Choose all variables (all items) included in the dataset.
- *For an easier interpretation of tables*: Click on options. Choose suppress small coefficients and choose the value .3. SPSS will now only display loadings (and other coefficients) if the absolute value is larger than .3.
- **SPSS' default extraction method is a principal component analysis (PCA)**, but we want to perform Principal Axis Factoring (see also lecture slides). Choose extraction and then select "Principal Axis Factoring". Also ask for the scree plot (display).
- Choose descriptives and under the header correlation matrix, ask for the Coefficients for KMO and Bartlett's sphericity, and Reproduced. Click Continue.
- Via Extraction ask for the scree plot under display

Further:

- For an EFA with **fixed number of factors**: Choose factors to extract, and fill in the number you of numbers.
- For a **rotated EFA**: choose the kind of rotation in the tab extraction (e.g., Direct Oblimin or Varimax)