Psychological Measurement Week 5 Reliability in Jamovi

Last week we looked at the assumption of unidimensionality. This is a primary assumption that must be considered before summing item scores to summarise a scale. This assumption should be checked prior to any further assessment of a scales properties (Fried et al., 2016). It is also noted that the assumption is rather rare in personality psychology (Dunn, Baguely, & Brunsden, 2014; Soĉan, 2000).

If your data are multi-dimensional, there are a few options:

## Treat it as Multidimensional

The first is to treat the scale as multi-dimensional. How does this affect the assignment? If a measure is multi-dimensional this suggests that there are different facets to the scale – for example extraversion might break down into being sociable and talkative. If one of these facets is more appropriate for your question than the other then use that one.

## Model Fit

Second, you could force the items into a single factor and see if this model is appropriate. There are two ways of doing this – is to use EFA but specify the number of factors as 1.

Graphical user interface, application

Description automatically generated



Table

Description automatically generated



The second is to use a method called Confirmatory Factor Analysis which tests the hypothesis that your items reflect a specified structure more than a null model in which there are no relationships between items.

Model fit is a topic in its own right. For the purposes of this assignment I will allow the use of cut-off values. Hu & Bentler (1999) suggest that in some cases a RMSEA of below 0.06 or a TLI (Tucker-Lewis Index) of above 0.95 are indicators of good model fit. There is a lot of nuance with this statement – and just using these cut-off items without justification will cause problems if you are trying to publish. For more information see e.g. McNeish & Wolf (2021) or Flora & Flake (2017)

## Check Factor/Facet Correlations

In the EFA menu there is an option to check factor correlations – this will provide a correlation table with all of the factors identified in the parallel analysis. If these are very high (e.g. above 0.7) then an argument can be made for treating the scores as a single factor (Stochl et al., 2022). In the example below we could not do this because the factors are only moderately correlated.

Graphical user interface, application, table

Description automatically generated

## Ignore?

These checks should be made in every case, and part of the reason that there are failures to replicate and loss of confidence in psychology as a field is that faulty assumptions are made about measures (Flake & Fried, 2020). I do not recommend ignoring the early warnings that you scale is failing assumption checks, but that option is of course open to you! The assignment is a toy example in which you develop your evaluation skills and has no consequences outside of the grades. Bear in mind however that in real world scenarios there will be consequences to using unreliable measures to make inferences about people!

# Tau Equivalence

One of the recommended papers this week is ‘From Alpha to Omega’ by Dunn, Brunsden, & Baguely (2014). This paper discusses issues with using Cronbach’s Alpha by default, and explains the conditions under which alternative measures of item consistency should be used.

As we have seen in the videos for this week, Tau Equivalence is the condition under which all items should have the same mean average. If they do not, then again reliability can be biased.

## Activity 1

Choose one (or more) of the Big Five factors from the fake dataset and answer the following question.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Factor | Factor Loadings approximately equal? | Equal Means\* for all items? (Tau Equivalence) | Item standard deviations approximately equal? |  |
| Neuroticism |  |  |  |  |
| Extraversion |  |  |  |  |
| Openness To Experience |  |  |  |  |
| Conscientiousness |  |  |  |  |
| Agreeableness |  |  |  |  |

\*Absolute so -5 is the same as 5

## Activity 2

Based on the above answers, which model of reliability is most appropriate? Run this analysis

|  |  |  |  |
| --- | --- | --- | --- |
| Scale | Type of Reliability Used | Value | Is there much difference? |
|  |  |  |  |
|  |  |  |  |

## Activity 3

Repeat the above on one of your constructs