Statistical Methods – Scale reliability analysis with small samples

Research question type: Most

What kind of variables: Ordinal and **interval/scale**

Common applications: Validating a scale in a questionnaire

1. Introduction

Whilst it is common statistical advice not to attempt a reliability analysis with a sample size less than 300 (Kline, 1986) a recent simulation study (Yurdugül, 2008) indicates that this is possible in certain circumstances. The most common statistic used in reliability analysis is **Cronbach's alpha** and an often quoted rule of thumb is a coefficient value above 0.7 is acceptable for psychological constructs (Kline, 1999). However, Cortina (1993) found that the size of a Cronbach's alpha coefficient depends upon the number of items in the scale with scales with more items having higher coefficients.

The advantage of carrying out a reliability analysis is that it can enable a researcher to treat a group of variables on the same subject as a single scale variable, reducing the complexity of further analysis and reducing the risk of Type I errors. However, student researchers often find it hard to obtain sample sizes of 300. The purpose of this worksheet is to advise students about how to go about trying to validate a scale with smaller sample sizes.

2. Scale design

We recommend using **seven point Likert response scales with only the end values anchored to interpretations** for individual scale items, i.e.

Statement

Please circle the number that best reflects your view:

1	2	3	4	5	6	7
Strongly disagree						Strongly agree

This enables the individual items to be treated as **scale data** (allowing more descriptive statistics to be calculated, such as means and standard deviations) with the maximum richness of data taking into account working memory modelling (Miller, 1956). However, such individual item scales are not ideal for statistical analysis so combining them into a multi item scale is preferable. If this turns out not to be possible such individual items are preferable to traditional five point Likert response scales with all values anchored to interpretations (which are ordinal data).

When designing a scale for a psychological construct it is advisable to start with a literature review and not to limit the initial items you include in your scale to your personal interpretation of the literature (Clark and Watson, 1995). This means you would expect to reduce the scale down in the final validated version.

There can be a tendency in student research projects to gloss over this stage and to try to design a final scale first time around. However, Hinkin et al. (1997) recommend that **at least twice as many items** should be generated than those that are finally used.

3. Number of items

Hinkin et al. (1997) recommend that final scales should be **four to six items long**. Short scales also reduce the risk of Cronbach's alpha inflation and misinterpretation. Another problem with long scales is that they may include multiple dimensions (Field, 2013: 709).

4. Scale validation process

Rather than starting with a scale reliability analysis we recommend you start with a Principal Component Analysis. There are two reasons for this:

1. It will show whether the individual items correspond sufficiently to the scale
2. It will show whether the Cronbach's alpha coefficient is stable with a small sample

Provided this process yields an acceptable result (see example below) a reliability analysis should then be carried out. If this also yields an acceptable result then the scale should be constructed as a scale variable either by adding the items together or by saving the regression model.

5. Example

100 members of the public were asked nine questions about their perception of the professionalism of psychologists from which 99 usable responses were obtained. Each question used a traditional five point Likert response scale. One of the items (*Violation_Likelihood*) was reverse worded so the corresponding reversed variable *Violation_Likelihood_Reversed* was computed by subtracting the *Violation_Likelihood* values from 6 and included in the trial scale.

Steps in SPSS

- Analyze > Dimension Reduction > Factor Analysis
- Put all the scale items in the Variables list and click on OK

This returned a single significant component (the default condition is that the returned component eigenvalues must be > 1) accounting for 55.6% of the total variance with an eigenvalue of 5.006. All the component loadings > 0.7 except for *Violation_Likelihood_Reversed* which was 0.476. However, as this had a loading > 0.4, it was also retained in the scale.

Component Matrix[a]

	Component
	1
Psychologists_Competence	.767
Psychologists_Intergrity	.831
Uphold_Prof_Standards	.823
Respect_Rights	.840
Welfare_Concern	.769
Awareness_of_Prof_Responsibility	.711
Positive_Consistency	.711
Contribution_to_Research	.718
Violation_Likelihood_Reversed	.476

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Note: The lower score with the *Violation_Likelihood_Reversed* item may indicate an issue with the validity of all the responses (it is common for responses to be given a slightly positive value as respondents tend to try to please the questionnaire setter; perhaps some of them did not read the reverse worded question as closely and gave it a positive score).

The following simulation studies were then quoted:

- According to Guadagnoli and Velicer (1988) the component pattern is stable for a sample size of 100 provided that the component contains at least four variable loadings > 0.6.
- According to Yurdugül (2008) the Cronbach's alpha coefficient is reliable for a sample size of 100 provided that the first eigenvalue of the Principle Component Analysis matrix > 3.

A reliability analysis was then carried out with the same nine variables:

- Analyze > Scale > Reliability Analysis
- Place all nine items in the Items list

Reliability Statistics

Cronbach's Alpha	N of Items
.894	9

This returned a Cronbach's alpha coefficient of 0.894. It was then concluded that due to the number of items in the scale relative to the size of this coefficient that these items may be considered as a single scale variable. As there is quite a large variation in the component scores it is advisable to use the regression model rather than adding the individual items. This is the default option under Scores… Save as variables.

Note: As there were nine items in this scale a higher threshold should be taken for reliability than the normal cut-off value (0.7) as it is assumed that this applies to scales with the recommended number of items, i.e. between four and six items. However, Hair et al. (1998) recommend a Cronbach's alpha cut-off value of 0.55.

6. Even smaller samples

According to Nunally (1978) there should be less items in the scale than the sample size. Yurdugül (2008) analysed sample sizes of 30 and found that Cronbach's alpha coefficients were reliable provided the first eigenvalue of the Principal Component Analysis was greater than 6. Guadagnoli and Velicer (1988) analysed sample sizes of 50 and found that component patterns were stable provided the component loadings were at least 0.8. As they did not consider cases with less than four variables per component this rule should also be applied. We therefore give the following advice:

- Reliability analysis should not be attempted for sample sizes < 30.
- For sample sizes between 30 and 50, only Yurdugül's article should be cited but we recommend that any items with a component loading < 0.4 are removed from the scale and the Principal Component Analysis is re-run. If the resultant first eigenvalue < 6 then a reliability analysis should not be attempted. If less than four items have a component loading > 0.8 then this should be discussed with the researcher making an informed decision about whether a reliability analysis should be attempted.
- For sample sizes between 50 and 100 both articles can be cited and both conditions should be satisfied. Again, after an initial Principal Component Analysis, any items with a component loading < 0.4 should be removed from the scale and the analysis re-run. If the first eigenvalue is between 3 and 6 an informed decision should be made about a reliability analysis based on the sample size and the eigenvalue size by an interpretation of the graph in Yurdugül's article – see Figure 1. If less than four items have a component loading > 0.8 then the advice in the point above should be followed.

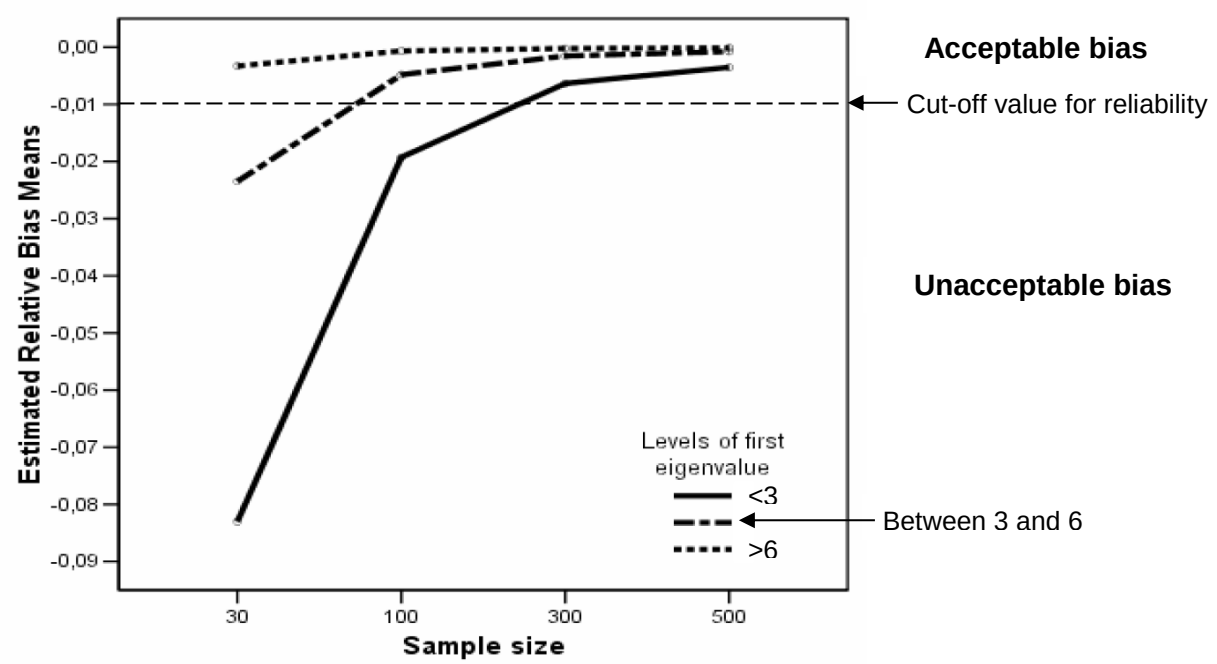

Figure 1: Relationship between Cronbach's alpha bias and sample size (Yurdugül, 2008)

References

Clark, L. A. and Watson, D. (1995) Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), pp. 309-319.

Cortina J. M. (1993) What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, pp. 98-104.

Field, A. (2013) *Discovering Statistics using SPSS*, 4th edn. London: SAGE.

Guadagnoli, E. and Velicer, W. F. (1988) Relation to sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), pp. 265-275.

Hair, J. F., Anderson, R. E., Tatham, R. L. and Black, W. L. (1998) *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall.

Hinkin, T. R., Tracey, J. B. and Enz, C. A. (1997) Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality & Tourism Research*, 21(1), pp. 100-120.

Kline, P. (1986) *A handbook of test construction: Introduction to psychometric design*. New York: Methune.

Kline, P. (1999) *A Handbook of Psychological Testing*, 2nd edn. London: Routledge.

Miller, G. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63, pp. 81-97.

Nunnally, J. C. and Bernstein, I. H. (1994) *Psychometric Theory*, 3rd edn. New York: McGraw-Hill.

Yurdugül, H. (2008) Minimum sample size for Cronbach's coefficient alpha: A Monte-Carlo study. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 35, pp. 397-405.

