Exploratory factor analysis and Cronbach's alpha (DrPH)

Dr Wan Nor Arifin
MBBS (IIUM), MSc (Medical Statistics) (USM).
Lecturer, Unit of Biostatistics and Research Methodology, Universiti Sains Malaysia.
E-mail: wnarifin@usm.my

Outlines

Exploratory factor analysis
Cronbach's alpha

Exploratory factor analysis

For the purpose of this hands-on session, extraction method of choice is Principal Axis Factoring. This extraction method is used as it does not assume normally distributed data.

Dataset: Attitude_Statistics v3.sav
R code: efa.R

Preliminary step

1. Clean up the data for wrong entry, missing values. Replace missing values with appropriate imputation method of choice.
2. Descriptive statistics:
   - Check minimum-maximum values per item.
   - n(%) of response to options per item.
3. Normality of data:
   - Univariate normality
     - Maximum-Likelihood extraction requires multivariate normality.
     - Univariate normality → Multivariate normality.
     - If not normal, may use Principal Axis extraction.
   - Multivariate normality
     - Normality of the data at multivariate level.

Step 1

1. Check suitability of data for analysis
   a) Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy

    b) Bartlet's test of sphericity
2. Determine the number of factors by
    a) Eigenvalues
    b) Scree plot
    c) Parallel analysis

<u>Assessment of results for Step 1</u>

<table>
<tr><th>Result</th><th>Cut-off points</th><th>Comments</th></tr>
<tr><td colspan="3" align="center">Suitability of data for analysis</td></tr>
<tr><td>Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy</td><td>> 0.7</td><td>MSA is a relative measure of amount of correlation (Kaiser, 1970). It indicates whether it is worthwhile to analyze a correlation matrix or not. KMO is an overall measure of MSA for a set of items, given as:

$$KMO=\frac{\sum_{i\neq j}^{n}\sum_{i\neq j}^{n} r_{ij}^2}{\sum_{i\neq j}^{n}\sum_{i\neq j}^{n} r_{ij}^2+\sum_{i\neq j}^{n}\sum_{i\neq j}^{n} a_{ij}^2}$$

where
r_{ij} is the correlation between items i and j
a_{ij} is the partial correlation coefficient (or anti-image correlation coefficient) between items i and j

From the formula, we can imply that
KMO→1: Correlation→1 and partial correlation→0.
KMO→1: Correlation→0 and partial correlation→1.

The following is the guideline on interpreting KMO values (Kaiser & Rice, 1974):

< 0.5 – Unacceptable
0.5 – 0.59 – Miserable
0.6 – 0.69 – Mediocre
0.7 – 0.79 – Middling
0.8 – 0.89 – Meritorious
0.9 – 1.00 – Marvelous</td></tr>
<tr><td>Bartlet's test of sphericity</td><td>P-value < 0.05</td><td>Basically it tests whether the correlation matrix is an identity matrix (Bartlett, 1950; Gorsuch, 2014; Revelle, 2015). The determinant of the matrix, R_{vv} is converted to a chi-square statistic and tested for significance:</td></tr>
</table>

		$$\chi^2 = -\left(n-1-\frac{2v+5}{6}\right)\ln\lvert R_{vv}\rvert$$ where n is the sample size v is the number of items while the *df* for the χ^2 is $$df = v\frac{v-1}{2}$$ A significant test indicates that there are worthwhile correlations among the items based on correlation matrix. A non-significant test indicates that the items are not correlated to each other based on the correlation matrix.
Determination of number of factors		
Eigenvalues	> 1	Look at number of factor at Initial Eigenvalues > 1 (Kaiser-Guttman rule). Eigenvalues can be interpreted as how worthwhile a factor in term of item. For an Eigenvalues of 4.5, the extracted factor is worth 4.5 times as much as a single variable. The cut-off value is 1 because if extracted factor is worth less than what a single variable can explain, the factor is not worthwhile to be extracted.
Scree plot	–	Look for last substantial decline or abrupt changes in the plot (elbow). Number of factors is the number of dots (eigenvalues) up to the 'elbow' of the plot. It is also suggested to to fix +/- 1 factor from the decided number of factor.
Parallel analysis	–	Comparison of the scree plot obtained from the data to the scree plot obtained from randomly generated data (Brown, 2006). Number of factors is the number of dots above the intersection between the plots.

Step 2

1. Re-run the analysis by fixing number of factors as decided from previous step.
2. Decide on rotation method. Choose an oblique rotation, **Promax**.

<u>Assessment of results for Step 2</u>

Result	Cut-off points	Comments
Judge quality of items by looking at the following results. Remove poor quality items.		
Communalities (Extraction)	Ideally > 0.5 Practically > 0.25	It is the % of item variance explained by the extracted factors. A cut-off of 0.5 is practical (Hair Jr. et. at., 2009), which means that 50% of item variance is explained by all extracted factors. The cut-off value depends on researcher as to what amount of explained variance is acceptable to him/her. However, for practical purpose I consider 0.25 cut-off point, considering factor loading > 0.5 is accepted, thus variance = square of factor loading = 0.5^2 = 0.25
Pattern matrix (pattern coefficients) OR Rotated factor matrix (when orthogonal rotation is used)	Ideally > 0.5	Only available when oblique rotation is used. Usually the pattern coefficients in the matrix are interpreted similarly to factor loadings. The coefficients are partials correlation coefficients of factors to the item. The factor loadings are ideally > 0.5 as recommended by Hair Jr. et. al. (2009), although 0.3 cut-off value is commonly used in EFA. Factor loadings of > 0.5 for all items in a construct indicates correlation between items and construct. Also check for cross-loading problem of an item across factors. This problem is indicated by having almost comparable factor loadings in two or more factors. It indicates that the item is not specific for a construct and to general, thus should be removed.
Factor correlations	< 0.85	Only available when oblique rotation is used. If > 0.85, the is a multicollinearity between the factors, thus the factors are not distinct from each other, thus can be combined (change number of fixed factors) (Brown, 2006).

Step 3

1. Re-run the analysis similar to **Step 2** every time an item is removed. Make judgment based on the results.
2. The analysis is finished when:
    - Satisfying number of factors.
    - Satisfactory quality of items.

Cronbach's alpha

- The reliability is checked for each factor as extracted from EFA.
- Selected good items per factor.

Step

1. Determine the reliability for each factor separately by including the selected items only.

Assessment of results

Result	Cut-off points	Comments
Cronbach's alpha	OK > 0.7 Caution > 0.9	Satisfactory=0.7-0.8; Clinical use=0.9 (Bland & Altman, 1997). Indicates the internal consistency reliability. Although a higher value indicates a higher reliability, a value of > 0.90 indicates that some items are redundant and should be removed (Streiner, 2003).
Corrected Item-Total Correlation	> 0.5	Ideally > 0.5 (Hair Jr. et. al., 2009) It is the correlation between value of an item to total value of others in a construct. A negative CITC indicates that an item is negatively correlated to the total, so reverse coding the item might be indicated.
Squared Multiple Correlation	-	% of contribution of an item to internal consistency. Low SMC shows that the item contribute minimally to the reliability of a construct, thus might indicate removal of the item.
Cronbach's alpha if item deleted	-	If the value is a marked improvement of Cronbach's alpha, it might justify removing the item. Retain the item if the value is less than reported Cronbach's alpha or the improvement is very minimal.

References

Bartlett, M. S., (1951), The effect of standardization on a chi square approximation in factor analysis. *Biometrika*, 38, 337-344.

Bland, J. M., Altman, D. G. (1997). Cronbach's alpha. *BMJ*, 314, 572.

Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York: The Guilford Press.

Gorsuch, R. L. (2014). Factor analysis: Classic edition. New York: Rouledge.

Hair Jr., J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Pearson Prentice-Hall.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35(4), 401-415.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, Mark IV. *Educational and Psychological Measurement*, 34, 111-117.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Publications.

Revelle, W. (2015). *psych: Procedures for Personality and Psychological Research*. Evanston, Illinois, USA: Northwestern University. http://CRAN.R-project.org/package=psych Version = 1.5.8.