# Guide to Decisionmaking in Factor Analysis

Since the late Ron Nuttall introduced me to factor analysis in 2001, I have been an avid practitioner. Along the way I have gradually assembled guidelines to help me negotiate the sometimes bewildering array of choices involved in this method (really, *these* methods, as many texts will emphasize). I'm planning to create a more complete document that includes citations. I'm sharing the current draft in the hopes of helping others and getting feedback. If you have comments about this guide, you're very welcome to send them to me at Roland@IntegrativeStatistics.com.

**Overview**

A. Decide how to treat missing data
B. Assess suitability of data for factor analysis
C. Extraction method
D. Number of factors in a principal axis factor solution
E. Rotation method

***

A. <u>Decide how to treat missing data</u>
1. Listwise deletion may shrink the sample size too much.
2. Pairwise deletion may make the analysis impossible depending on missing data patterns.
3. Either listwise or pairwise deletion has the potential to bias results; consider imputation.
4. Be prepared to run multiple analyses employing different missing data strategies.

B. <u>Assess suitability of data for factor analysis</u>
1. Distributions should not be excessively far from normal; may need to transform.
2. Relationships need to be essentially linear; may need to transform.
3. Sampling adequacy (often optional). Desirable to have high *r*s relative to partial *r*s.
   a. As a global measure, Kaiser-Meyer-Olkin statistic should be at least 0.5 and probably should be > 0.6.
   b. Anti-image correlations: for individual variables, use the same standard as for KMO, and this can help decide inclusion/exclusion.
4. See additional criteria in section D.

C. <u>Extraction method</u>
1. Use principal components only if looking to use all information in pure data reduction (i.e., if it is reasonable to assume ~0% of the info is measurement error).
2. Can use maximum likelihood extraction, and its associated goodness of fit test for number of factors, if
   a. variable distributions are fairly close to normal, i.e., skewness <2 and kurtosis <7, or if willing to transform or drop nonnormal variables.
   b. sample size is moderate: goodness of fit test is sensitive to small or large *n*.
3. In most social science cases, principal axis factoring works well.

D. Number of factors in a principal axis factor solution
1. Plan on needing to try multiple solutions.
2. Begin by extracting factors from all initial eigenvalues of at least 0.7. (Much research has shown that including only initial eigenvalues >=1 can under- or over-estimate the number of factors.) Once you see how solutions turn out after extraction and rotation, you can specify a minimum-eigenvalue or number-of-factors criterion based on something more informed than just the Kaiser Guttman rule of "eigenvalues >=1."
3. Scree test, visual or numerical, is best performed on extracted eigenvalues, rather than initial (although initial is what is displayed in SPSS).
4. Factor correlations should not be too high. If they are, and if largest initial eigenvalue is more than 10 times as large as the next, consider a one-factor solution (Harman; Kriebeck; Tracey).
   a. Assuming correlations are not too high, the variance explained by each factor *after rotation* should be at least 5 or 10%.
   b. Consider Terence (T. J. G.) Tracey's method of partialing out the first *unrotated* factor as a "general factor" attributable to bias, e.g., to social desirability or common method bias.
5. Each factor should have at the very least 2 and probably 3 variables with high loadings.
6. Only retain those factors that are interpretable. (In connection with this, see 4.b.)
7. (Parallel Analysis method: too many different versions; no one version has been sufficiently validated or met with sufficient consensus.)
8. (Minimum Average Partial correlation test: promising but may not be worth the effort required.)

E. Rotation method
1. Use oblique rotation as the default. There is no nothing to lose: if correlations among factors are very low, you can always switch to varimax (in which case results should hardly change anyway). But by starting with varimax you risk ignoring meaningful correlations.
2. To display two factors' relationship after finalizing a solution, can use varimax to plot the two clusters of variables and show the factors' correlation (which will be represented by the angle between them, with *r* equal to the cosine).