Parallel Analysis in SPSS Using O'Connor's Macro

In this document I explain how to use O'Connor's SPSS macro to obtain a parallel analysis (PA) in SPSS. The macro can be obtained from his website at https://people.ok.ubc.ca/brioconn/nfactors/nfactors.html and is referenced in the article:

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32,* 396-402.

The data from this study are based on the Attitudes toward Scientists data from chapter 12 in the text. These data represent scores of 1974 respondents on the nine items shown on page 311 of the text. The data are in the file "Scientist data.sav." Note that only 371 respondents had complete data on all nine variables. The analyses for PA require complete data so are based on these 371 respondents.

To run the macro you will need to open a syntax window. For an explanation of how to do this, see the document "Calculating G-theory Quantities Using SPSS Syntax" in the Chapter 10 materials."

You will need to change some things in the syntax to run a PA for your data. These changes are explained in O'Connor's comments at the beginning of the syntax and I show these below. My comments are in italics.

```
* Enter the name/location of the data file for analyses after "FILE =";
  If you specify "FILE = *", then the program will read the current,
  active SPSS data file; Alternatively, enter the name/location
  of a previously saved SPSS data file instead of "*";
  you can use the "/ VAR =" subcommand after "/ missing=omit"
  subcommand to select variables for the analyses.
```

*The comments above are from O'Connor (note that putting an asterisk " * " before a command causes SPSS to ignore it. This is how comments are included in SPSS syntax files.*

Below I specify how to do this for the Scientist data.

The easiest way to run the macro is to have the data set open in the same location (folder) as the SPSS syntax window. You can then just use the specification ***"FILE = * "*** *for your data file. This is shown in the syntax below.*

*After the specification "****VAR =*** *" I have put in the names of the variables on which the factor analysis will be run. If you are using a different data set, you will need to change these names to reflect the variable names from that data set.*

GET raw / FILE = * / **missing=omit** / **VAR =** ALONE to NOINTRST.

* Enter the desired number of parallel data sets here.

Here, you enter the number of random data sets on which the PA will be based. 100 is a common number, but you can change this if you like by putting in a different number after the equals sign in the command below.

compute ndatsets = 100.

* Enter the desired percentile here.

compute percent = 95.

The "compute" command above specifies whether you want the PA to be based on the 95th percentile of the random data eigenvalues (this is the usual specification) or some other percentile. If you want to change this, just type in another number after the equals sign.

* Enter either
  1 for principal components analysis, or
  2 for principal axis/common factor analysis.

compute kind = 2 .

The compute statement above allows you to specify whether the eigenvalues should be obtained for principal components or principal axis (factor) analysis. This specification should match the analysis you intend to conduct. Type in a "1" or "2" depending on which analysis you would like.

* Enter either
  1 for normally distributed random data generation parallel analysis, or
  2 for permutations of the raw data set.

compute randtype = 2.

Finally, the compute statement above allows you to choose whether the random data eigenvalues should be based on random data from a normal distribution (1) or random permutations of your data set (2). Unless your data are normally distributed, it is best to choose option 2.

This is the end of the changes you need to make.

After you have made these changes, save the syntax file. Then highlight all the syntax and click on the right arrow button at the top of the screen (see the document "Calculating G-theory Quantities Using SPSS Syntax" in the Chapter 10 materials for a screen shot.)

You should get the following output:

```
Run MATRIX procedure:

PARALLEL ANALYSIS:

PAF/Common Factor Analysis & Raw Data Permutation

Specifications for this Run:
Ncases    371
Nvars       9
Ndatsets  100
Percent    95

Raw Data Eigenvalues, & Mean & Percentile Random Data Eigenvalues
          Root     Raw Data        Means     Prcntyle
     1.000000     1.816130      .277286      .364815
     2.000000     1.227235      .193048      .249381
     3.000000      .126715      .123181      .172654
     4.000000     -.046820      .065377      .105490
     5.000000     -.090415      .014626      .052030
     6.000000     -.113539     -.037263     -.006727
     7.000000     -.147892     -.089451     -.058663
     8.000000     -.178573     -.140347     -.111379
     9.000000     -.236608     -.201213     -.158919
```

The "**Specifications for this Run**" indicate the number of cases and of variables on which the PA was based. These should match the number of (complete) cases and number of variables in your data set. Recall that there were 371 respondents with complete data, so the analyses are based on this number. The number of datasets and the percentile that you specified in the macro are also shown.

The first column (labeled "**root**" which is another name for an eigenvalue) lists each factor by number. There will be as many of these as there are variables.

The second column (labeled "**Raw Data**") shows the eigenvalues from your data (in this example, the Scientist data). These eigenvalues should be the same as those produced by running a factor analysis on the data.

The third column shows the mean or average value of the random eigenvalues, and the last column (labeled "**Prcntyle**") shows the 95th percentile of the random eigenvalues.

PA involves comparing the real data eigenvalues in the second column to either the mean (50th percentile) or 95th percentile of the random eigenvalues.

Using the mean of the random data eigenvalues suggests three factors, because the first three values in the raw data column are greater than those in the "means" column.

Using the 95th percentile suggests two factors because the first two values in the raw data column are greater than those in the "Prcntyle" column.
Common practice is to use the 95th percentile. However, if results based on the mean and 95th percentiles disagree, best practice isto examine solutions based on both numbers of factors and to decide based on a) agreement with other criteria for determining the number

of factors, such as the scree plot, and b) the interpretability and utility of the resulting factors.

The text below was included by O'Connor and appears at the end of each PA analysis from the macro.

```
Warning: Parallel analyses of adjusted correlation matrices
eg, with SMCs on the diagonal, tend to indicate more factors
than warranted (Buja, A., & Eyuboglu, N., 1992, Remarks on parallel
analysis. Multivariate Behavioral Research, 27, 509-540.).
The eigenvalues for trivial, negligible factors in the real
data commonly surpass corresponding random data eigenvalues
for the same roots. The eigenvalues from parallel analyses
can be used to determine the real data eigenvalues that are
beyond chance, but additional procedures should then be used
to trim trivial factors.

Principal components eigenvalues are often used to determine
the number of common factors. This is the default in most
statistical software packages, and it is the primary practice
in the literature. It is also the method used by many factor
analysis experts, including Cattell, who often examined
principal components eigenvalues in his scree plots to determine
the number of common factors. But others believe this common
practice is wrong. Principal components eigenvalues are based
on all of the variance in correlation matrices, including both
the variance that is shared among variables and the variances
that are unique to the variables. In contrast, principal
axis eigenvalues are based solely on the shared variance
among the variables. The two procedures are qualitatively
```

different. Some therefore claim that the eigenvalues from one extraction method should not be used to determine the number of factors for the other extraction method. The issue remains neglected and unsettled.