

Parallel Analysis Using O'Connor's SAS Macro

In this document I explain how to use O'Connor's SAS macro to obtain a parallel analysis (PA). 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 on the nine items shown on page 311 of the text. Note that the macro cannot handle missing data, so a data set based on complete data only must be created. I have done this for the Scientist data; the resulting SAS data file is "Scientist_nomiss.sas7bdat." The data set contains responses from 371 respondents on the nine variables.

Parallel Analysis

The syntax for PA is shown below.

You will need to change some things in the syntax to run a PA for your data. I explain these below with comments in italics.

```
proc iml;  
reset noname;  
options nocenter nodate nonumber linesize=100 pagesize=500; title;  
  
use Scientist_nomiss;  
read all var {alone better boring nofun good help odd norealign nointrst} into raw;
```

The **use** and **read** commands above should specify a data set available in the "work" library. In the **read** command, list the variables to be used in the factor analysis within the braces

The data set into which the data is read must be called "raw" as shown at the end of the read statement.

The typical number of data sets for PA is 100, but you can change this by typing in another number after the equals sign below.

```
/* Enter the desired number of parallel data sets here */  
ndatsets = 100;
```

The 95th percentile of the random data eigenvalues is typically used as the value to which the real data eigenvalues are compared. This can be changed in the **percent =** specification below.

```
/* Enter the desired percentile here */
```

percent = 95;

In the commands below you can specify whether your analysis will be a principal component (1) or factor analysis (2). Type in the appropriate number after the **kind =** specification.

/* Specify the desired kind of parallel analysis, where:

1 = principal components analysis

2 = principal axis/common factor analysis */

kind = 2 ;

The **randtype** statement below 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.

/* Enter either

1 for normally distributed random data generation parallel analysis, or

2 for permutations of the raw data set */

randtype = 2;

Finally, the **seed** statement can be changed in the statement below. Using a seed value of 0 will result in randomly different results each time the program is run. If you anticipate a need to rerun the program and want to obtain identical results, you can type in a different number. Typing in the same number in subsequent runs will ensure that you obtain the same values each time.

/* When seed = 0, the clock is used as the seed for the random number generations. This produces different random numbers on different runs of the program. To use the same random numbers on different runs of the program, set seed to a value other than 0 */

seed = 0;

Leave everything below the line

“/*** End of user specifications *****/”**

as is.

Running all the syntax will yield the following output:

Ncases	371
Nvars	9
Ndatsets	100
Percent	95

This table provides the numbers of cases and variables on which the PA was based.

These should match the number of (complete) cases and number of variables in your data set. Remember that there were 371 respondents with complete data, so the analyses are based on this number.

The number of datasets and the percent (percentile) that you specified in the macro are also shown.

Raw Data Eigenvalues, & Mean & Percentile Random Data Eigenvalues

Root	Raw Data	Means	Prcentyle
1.000000	1.816130	0.275005	0.332724
2.000000	1.227235	0.181194	0.232440
3.000000	0.126715	0.119565	0.160184
4.000000	-0.046820	0.059612	0.100665
5.000000	-0.090415	0.009003	0.037241
6.000000	-0.113539	-0.038157	-0.011246
7.000000	-0.147892	-0.085952	-0.052660
8.000000	-0.178573	-0.134266	-0.093732
9.000000	-0.236608	-0.194430	-0.152461

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 “Prcentyle”) 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 “Prcentyle” column.

Common practice is to use the 95th percentile. However, if results based on the mean and 95th percentiles disagree, best practice would be to 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.