**How Many Components or Factors to Retain**

 **Parallel Analysis.** In this analysis one determines how many components account for more variance than do components derived from random data. One creates many random sets of data, with the constraint that each set has the same number of variables and the same number of cases as the actual set of data. One then compares the eigenvalues from an analysis of the actual data with the distribution of eigenvalues from the random sets of data. Starting with the first component, the eigenvalue from the original data is compared to the 95th percentile of the eigenvalues of the first components from the random data. If the original data produced an eigenvalue greater than that 95th percentile, then that component is retained and you move on to the second component. This process is continued until for the *nth* component, the eigenvalue from the actual data is not greater than the 95th percentile from the random data. At that point, one stops and decides to retain *n – 1* components.

 SAS, SPSS, and Matlab scripts for conducting parallel analysis can be found at <https://people.ok.ubc.ca/brioconn/nfactors/nfactors.html> . Here I illustrate use of the SPSS script. First, be sure that the file with the original data contain only the data to be used for the components analysis – any other variables need be deleted. With the data file open in SPSS, edit the script to indicate how many variables there are, how many cases there are, and how many random data sets to create. Then simply run the script. Here is output:

|  |  |
| --- | --- |
| Parallel Analysis | Actual Data |
|  **PARALLEL ANALYSIS:****Principal Components****Specifications for this Run:****Ncases 231****Nvars 7****Ndatsets 1000****Percent 95****Random Data Eigenvalues** **Root Means Prcntyle** **1.000000 1.251612 1.344920** **2.000000 1.146058 1.207526** **3.000000 1.064757 1.118462** **4.000000 .992964 1.038794** **5.000000 .926129 .973311** **6.000000 .852369 .907173** **7.000000 .766109 .830506** |  |

 Notice that for only the first two components, the eigenvalues from the actual data have values greater than those of the 95th percentile of the random data. Accordingly, a two component solution is indicated.

 The script also includes the option to replace the main diagonal of the input correlation matrix with estimates of the communalities of the variables. While this is controversial, it is my preference when the researcher intends to conduct a factor analysis rather than a components analysis.

**Velicer’s MAP Test**. With this procedure one considers how much common (shared by variables) variance remains in the data after extracting *n* components. In the first step, the first component is removed from the original correlation matrix, resulting in a matrix of partial correlations. The off-diagonal elements, when squared, represent variance shared by the variables in potential components other than the first. The mean squared off-diagonal partial correlation coefficient is computed for this step. On the second step the first two components are removed from the original correlation matrix. On the *nth* step, the first *n* components are removed from the original correlation matrix. One retains the component which has the smallest mean squared off-diagonal partial correlation (and all earlier components).

 The correlation matrix is the input for the SPSS script. It can be brought in from an external file or simply pasted into the program. Here I have done the latter:

matrix.

compute cr = {

 1.00, .83, .77, -.41, .02, -.05, -.06;

.83, 1.00, .90, -.39, .18, .10, .03;

.77, .90, 1.00, -.46, .07, .04, .01;

-.41, -.39, -.46, 1.00, -.37, -.44, -.44;

.02, .18, .07, -.37, 1.00, .91, .90;

-.05, .10, .04, -.44, .91, 1.00, .87;

-.06, .03, .01, -.44, .90, .87, 1.00 }. <snip, snip>

The output:

**Velicer's Minimum Average Partial (MAP) Test:**

**Velicer's Average Squared Correlations**

 **.000000 .266624**

 **1.000000 .440869**

 **2.000000 .129252**

 **3.000000 .170272**

 **4.000000 .331686**

 **5.000000 .486046**

 **6.000000 1.000000**

**The smallest average squared correlation is**

 **.129252**

**The number of components is**

 **2**

 Which Test to Use? With luck, both parallel analysis and the MAP test will indicate the same decision. Parallel analysis has a slight tendency to extract too many components, the MAP test too few. If the two disagree, try increasing the number of random data sets in the parallel analysis and look carefully at the two smallest squared partial correlation coefficients from the MAP test. You may end up having to use the interpretability criterion.

 This material was extracted from the document at <http://core.ecu.edu/psyc/wuenschk/MV/FA/PCA-SPSS.docx>