Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. (3362 citations as of 4/1/2016)

Online comments by Jack McArdle

There is no best "number of factors" rule, but there is a worst rule — the number of eigenvalues greater than one. Bill Revelle

Psychometric Society, Asheville, NC, July, 2016

The Number of Factors Problem

Henry Kaiser's approach to factor analysis (dating from 1960) was disdainfully labeled "Little Jiffy" by Chet Harris:

"principal components with associated eigenvalues greater than one followed by normal varimax rotation"

Horn (1965) introduced a different strategy for selecting the number of components (aka, "factors"):

select as many components that have eigenvalues greater than the values expected from randomly constructed data

Parallel Analysis

This comparison method in Horn (1965) has become known in the literature as "parallel analysis"

The word "parallel" refers to the construction of random data sets that have the same number of variables and number of subjects as the real correlation matrix being analyzed.

The ordered eigenvalues are obtained for each correlation matrix constructed; these ordered vectors are then averaged to obtain the value expected for a particular (ordered) eigenvalue when generated from random data.

The Guttman Lower Bound

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. (3362 citations as of 4/1/2016)

Online comments by Jack McArdle

The Kaiser rule is based on some earlier work by Guttman on a lower bound for the number of factors:

the number of eigenvalues greater than one for the (observed) correlation matrix

Note the following: if the population correlation matrix is the identity, so all the variables have zero correlations, then there are *no* latent roots of the population correlation matrix that are strictly greater than one.

However, an observed correlation matrix based on data sampled from a population having such an identity correlation matrix, one would expect about one-half of the eigenvalues to be greater than 1.0 and about one-half to be less than 1.0.

This is because the sum of all the eigenvalues (all of which must be nonnegative) is equal to the number of variables.

The Illustration in Horn (1965)

Horn (1965) gives an illustration involving 65 variables and 297 subjects — this is his Table 1 that we show next

Real data number of eigenvalues greater than 1.0: 16

Real data number of eigenvalues greater than that found for random data: 9

Horn also noted something of an inflection at this point: given that Horn was Cattell's student, did he see the proverbial elbow in the scree plot?

Further Work

A lot of work has been stimulated by Horn (1965):

using percentiles of the distribution for the random eigenvalues;

using permutation distributions to generate the parallel random data;

using other criteria, such as Velicer's minimum average partial (MAP) test.