

Critical Eigenvalue Sizes in Standardized Residual Principal Components Analysis

A principal components analysis of Rasch residuals, i.e., of observed responses minus their expectations, is used in Wright (1996) to investigate whether or not there is more than one component explaining the structure of respondent data. Wright postulates that, if the data are unidimensional, then components in the residuals will be at the noise level. Wright uses logit residuals. Linacre (1998) argues in favor of residuals standardized by their model standard deviation. These have the form of random normal deviates and will be adopted here.

The idea of retaining components that are above noise level is common practice in psychometrics. The Cattell (1966) scree test and the Kaiser (1960) rule are the most often used procedures to determine the number of components. They are both based on inspection of the correlation matrix eigenvalues. Cattell's recommendation is to retain only those components above the point of inflection on a plot of eigenvalues ordered by diminishing size. Kaiser (1960) recommends that only eigenvalues at least equal to one are retained. One is the average size of the eigenvalues in a full decomposition.

Smith and Miao (1994, p. 321) observe many components with eigenvalues greater than one in four simulations of unidimensional observational data. In their simulations, the first component corresponds to the Rasch dimension. The eigenvalue of the second component, the largest component in the random noise, never exceeds 1.40, suggesting that 1.40 is a threshold value for randomness.

Humphreys and Montanelli (1975) argue that the Kaiser rule is only true for very large correlation matrices. They propose that criterion eigenvalue thresholds be estimated by simulation studies based on random data formed into matrices of relevant sizes. The number of non-random components is determined by comparing the eigenvalue vector of the empirical data matrix with the vector of mean eigenvalues from the simulations. Only those leading empirical components with eigenvalues greater than their simulated equivalents are retained.

Accordingly, simulations of normal random deviates are performed here. These approximate matrices of Rasch standardized residuals for situations in which the data fit the model. O'Connor's (2000) SAS program was used to efficiently perform multiple simulations.

In Table 1, the average eigenvalues, along with their 5th and 95th percentile values, are presented, obtained from the simulation of different numbers of subjects (N) and items (L). The simulated data are all random noise. The graphs shows the Cattell scree plot for the eigenvalues of the first 20 components.

It is seen that the value of 1.40 is always exceeded by

the first eigenvalue, and usually by the second. Consequently, the recommendation is to decide the criterion eigenvalue directly from relevant simulations.

Gilles Raïche
 Université du Québec à Montréal
 Département d'éducation et pédagogie

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 629-637.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

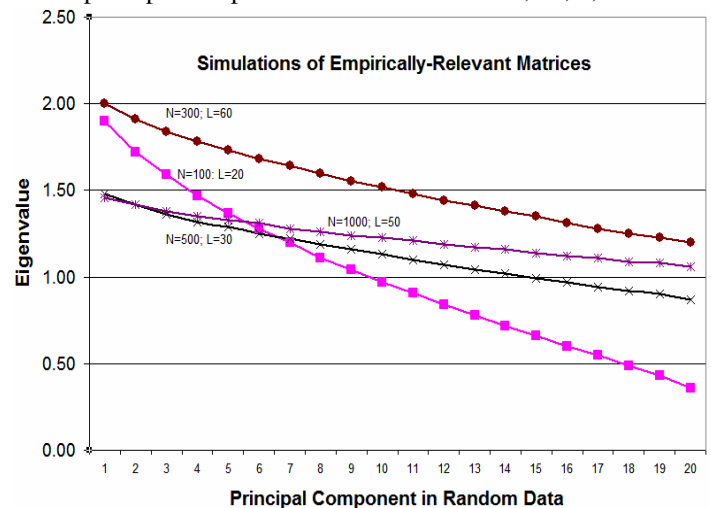
Linacre, J. M. (1998). Detecting multidimensionality: which residual data-type works best? *Journal of Outcome Measurement*, 2, 3, 266-283.

Humphreys, L. G. and Montanelli, R. G. (1975). An examination of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, 10, 193-206.

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

Smith, R. M and Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.): *Objective Measurement: Theory into Practice. Volume 2*. Greenwich: Ablex.

Wright, B.D. (1996) Local dependency, correlations and principal components. *Rasch Meas Trans*, 10, 3, 509-511.



	N=100; L=20			N=500; L=30			N=1000; L=50			N=300; L=60		
	5 th	Mean	95 th	5 th	Mean	95 th	5 th	Mean	95 th	5 th	Mean	95 th
1	1.74	1.90	2.07	1.42	1.48	1.55	1.42	1.46	1.50	1.92	2.00	2.09
2	1.61	1.72	1.85	1.37	1.42	1.47	1.38	1.42	1.45	1.85	1.91	1.98
3	1.50	1.59	1.69	1.33	1.36	1.40	1.35	1.38	1.41	1.79	1.84	1.91
4	1.39	1.47	1.56	1.29	1.32	1.36	1.33	1.35	1.38	1.73	1.78	1.83
5	1.29	1.37	1.46	1.25	1.29	1.32	1.31	1.33	1.35	1.68	1.73	1.78