Multivariate Behavioral Research ›

Volume 44, 2009 - Issue 2

Original Articles

Exploratory Factor Analysis With Small Sample Sizes

J. C. F. de Winter*, D. Dodou* & P. A. Wieringa

Pages 147-181 | Published online: 14 Apr 2009

Cite this article https://doi.org/10.1080/00273170902794206

Abstract

Exploratory factor analysis (EFA) is generally regarded as a technique for large sample sizes (N), with N = 50 as a reasonable absolute minimum. This study offers a comprehensive overview of the conditions in which EFA can yield good quality results for N below 50. Simulations were carried out to estimate the minimum required N for different levels of loadings (λ), number of factors (f), and number of variables (p) and to examine the extent to which a small N solution can sustain the presence of small distortions such as interfactor correlations, model error, secondary loadings, unequal loadings, and unequal p/f. Factor recovery was assessed in terms of pattern congruence coefficients, factor score correlations, Heywood cases, and the gap size between eigenvalues. A subsampling study was also conducted on a psychological dataset of individuals who filled in a Big Five Inventory via the Internet. Results showed that when

data are well conditioned (i.e., high λ, low f, high p), EFA can yield reliable results for N well below 50, even in the presence of small distortions. Such conditions may be uncommon but should certainly not be ruled out in behavioral research data.

* These authors contributed equally to this work

ACKNOWLEDGMENTS

The research of J. C. F. de Winter has been funded by the Dutch Ministry of Economic Affairs, under its Innovation Program on Man-Machine Interaction, IOP MMI. We are grateful to Samuel D. Gosling for providing the dataset of the Big Five Inventory for the subsampling study.

Notes

* These authors contributed equally to this work

[1] This article defines simple structure as a special case of Thurstonian simple structure, also called independent cluster structure or ideal simple structure.

[2] Lorenzo-Seva and Ten Berge (2006) suggest the .95 threshold for good agreement on the basis of judgments of factor similarity by factor analytic experts. Note that others have used a .92 threshold for good and .98 for excellent agreement (MacCallum et al., 2001).

[3] The ES index in this article was calculated from the eigenvalues of the unreduced correlation matrices (UCM, with 1s in the diagonal). It has been argued that it is more conceptually sensible to use the eigenvalues of the reduced correlation matrix (RCM, with communality estimates in the diagonal) when the goal is to identify the number of common factors (Fabrigar et al., 1999; Preacher & MacCallum, 2003). We have repeated the subsampling study with ES based on the RCM (with communality estimates based on squared multiple correlations). Results showed that the difference in ES based on the UCM and the ES based on the RCM was always smaller than 10% and that overall average ES was higher for the UCM as compared with the RCM.

[a] A different population pattern was produced for each repetition for all conditions of groups 3, 4, and 11.

[b] The numbers refer to the variables per factor with a .8 loading.

[4] A loading of .6 was considered low for the sample size (N = 17) under investigation. This was based on the findings of the first part of the simulations (Table 2): for $\lambda = .6$, $f = 3$, $p = 24$, the required minimum N for good agreement ($K = .95$) was 55.

[5] Cronbach's α was calculated for two conditions of the first simulation series (low loadings: $\lambda = .2$, $f = 2$, $p = 24$, $N = 1{,}438$ and high loadings: $\lambda = .9$, $f = 2$, $p = 24$, $N = 6$), Although factor recovery was identical in those two conditions (see Table 2), average Cronbach's a among variables loading on the factor was .332 for the low loadings and .968 for the high loadings. This demonstrates that high internal consistency is not necessary for good factor recovery. A more detailed discussion of this issue can be found in Boyle (1991).