INTERNATIONAL JOURNAL OF TESTING, *5*(2), 159–168

Minimum Sample Size Recommendations for Conducting Factor Analyses

Daniel J. Mundfrom and Dale G. Shaw
Department of Applied Statistics and Research Methods
University of Northern Colorado

Tian Lu Ke
Taipei Physical Education College, Taiwan

There is no shortage of recommendations regarding the appropriate sample size to use when conducting a factor analysis. Suggested minimums for sample size include from 3 to 20 times the number of variables and absolute ranges from 100 to over 1,000. For the most part, there is little empirical evidence to support these recommendations. This simulation study addressed minimum sample size requirements for 180 different population conditions that varied in the number of factors, the number of variables per factor, and the level of communality. Congruence coefficients were calculated to assess the agreement between population solutions and sample solutions generated from the various population conditions. Although absolute minimums are not presented, it was found that, in general, minimum sample sizes appear to be smaller for higher levels of communality; minimum sample sizes appear to be smaller for higher ratios of the number of variables to the number of factors; and when the variables-to-factors ratio exceeds 6, the minimum sample size begins to stabilize regardless of the number of factors or the level of communality.

Key words: factor analysis, sample size

One of the most commonly addressed issues in all of statistical analysis is the issue of sample size. When researchers design a study or plan a project, they must decide

Requests for reprints should be sent to Daniel J. Mundfrom, Department of Applied Statistics and Research Methods, University of Northern Colorado, Greeley, CO 80639. E-mail: daniel.mundfrom@unco.edu

how large a sample to obtain. For many statistical procedures that involve sampling from known distributions, minimum sample size determinations are straightforward and well known. This is not the case with all procedures, however, and, specifically, it is not the case when conducting a factor analysis.

Factor analysis has a variety of applications in research and measurement. One common use is in the development of tests and measures. Here, as with other uses of factor analysis, one of the major decisions facing the researcher is the size of the sample needed to provide an expectation that the factor solution has some foundation in reality. Hence, accurate, reliable guidelines for choosing an appropriate sample size for conducting a factor analysis could be extremely useful in many applications, including international and comparative studies of tests and measures.

The factor analysis literature contains a variety of recommendations regarding the appropriate sample size to use for conducting a factor analysis. For the most part, these recommendations are presented as either a suggested minimum sample size or a suggested minimum ratio of sample size to number of variables. Girshick (1939), Archer and Jennrich (1973), and Cudeck and O'Dell (1994) investigated the minimum sample size issue as it relates to the standard errors of the factor loadings. Other researchers looked at the asymptotic standard errors of unrotated maximum likelihood loadings (Lawley, 1967) and of rotated principal components loadings (Jennrich, 1973). MacCallum, Widaman, Zhang, and Hong (1999), however, pointed out that published guidelines for determining adequately small standard errors of loadings do not exist, and Cudeck and O'Dell concluded that a theoretical answer to the sample size question is too difficult to derive directly.

It is generally accepted that larger samples are better (Comrey & Lee, 1992; Cudeck & O'Dell, 1994; Kline, 1994; MacCallum et al., 1999; Velicer, Peacock, & Jackson, 1982), but existing recommendations are varied, even contradictory (MacCallum et al., 1999). Gorsuch (1983) and Kline recommend a minimum sample size of at least 100, whereas Comrey and Lee said 50 is very poor, 100 is poor, 200 is fair, 300 is good, 500 is very good, and 1,000 is excellent. Kline also recommended that considering the ratio of number of participants to number of variables is a better way to determine a minimum sample size. Kline's recommendation seems reasonable, however, after reviewing various proposed guidelines for this ratio; Arrindell and van der Ende (1985) concluded that these recommendations are inconsistent. For example, Cattell (1978) recommended a ratio of three to six times the number of variables; Everitt (1975) argued for a ratio of at least 10 to 1; and Hair, Anderson, Tatham, and Black (1995) suggested using a ratio of 20 to 1. Other researchers have looked at how different levels of communality (MacCallum et al., 1999; Tucker, Koopman, & Linn, 1969; Velicer et al., 1982) and the ratio of variables to factors affect minimum sample size (Browne, 1968; MacCallum, Widaman, Preacher, & Hong, 2001; MacCallum et al., 1999; Preacher & MacCallum, 2002; Tucker et al., 1969). In sum, these researchers have demonstrated that the minimum necessary sample size in factor analysis is related to the number of variables, the number of factors, the number of variables per factor, and the size of the communalities.

The purpose of this study was to determine a minimum necessary sample size in factor analysis that would provide a factor solution in agreement with the population structure from which the sample was taken. To accomplish this purpose, we generated a variety of population correlation matrices under several different conditions, repeatedly sampled from these population structures, and determined coefficients of congruence between the sample solutions and population structures. Minimum necessary sample size recommendations were developed based on obtaining a high percentage of sample solutions that surpassed a preset level of agreement reflected in the congruence coefficients.

METHOD

Population Correlation Matrices

For this simulation study, population correlation matrices were generated, based on continuous data, under a variety of conditions, using a method given by Tucker et al. (1969) for generating correlation matrices that produce solutions having a strong, simple structure. These procedures allowed for the number of common factors to be varied from 1 to 6, the number of variables per factor to be varied from 3 to 12, and three different communality patterns: high—all communalities ranged between .6 and .8, wide—all communalities ranged between .2 and .8, and low—all communalities ranged between .2 and .4. A total of $3 \times 6 \times 10 = 180$ different population conditions were investigated. For each of these 180 conditions, 100 different population correlation matrices were generated.

Sample Correlation Matrices

Sample correlation matrices were generated from each of these 18,000 population matrices using a variety of sample sizes. A "small" sample size was used as a starting point, with the exact size determined by the number of variables. Sample sizes were systematically increased according to the following scheme:

If the sample size was less than 30, then it increased by 1 at each stage.
If the sample size was between 30 and 100, then it increased by 5 at each stage.
If it was between 100 and 300, it increased by 10 at each stage.
If it was between 300 and 500, it increased by 50 at each stage.
If it was greater than 500, it increased by 100 at each stage.

This procedure continued until criteria were satisfied for either good or excellent agreement. These criteria are explained later.

Factor Analyses

All of the factor analyses performed in this study used the maximum likelihood estimation procedure and a varimax rotation. Then, the rotated population solutions were used as target matrices, and each sample solution was subjected to target rotation using a procrustrian rotation method developed by Tucker and Korth (1976). Once the rotated solutions were obtained, a coefficient of congruence between each factor from the sample solution and the corresponding factor from the population solution was calculated using the following formula:

$$\phi_k = \frac{\sum_{j=1}^{p} f_{jk(s)} f_{jk(t)}}{\sqrt{\left(\sum_{j=1}^{p} f_{jk(s)}^2\right)\left(\sum_{j=1}^{p} f_{jk(t)}^2\right)}}$$

where $f_{jk(t)}$ is the population factor loading for variable *j* on factor *k* and $f_{jk(s)}$ is the corresponding sample loading. To assess the degree of congruence across *m* factors, the mean value of Φ_k across the *m* factors is computed. This average is denoted as *K* and is given by

$$K = \frac{\sum_{k=1}^{m} \phi_k}{m}.$$

For any particular *m*-factor solution, however, *m!* different *K*s can be obtained by rearranging the order of the *m* columns. In this study, the maximum value of *K* in this group of *m! K*s was used to represent the measure of agreement between any particular sample's rotated solution and the rotated solution from its parent population. For each of the 18,000 population conditions and for each specific sample size investigated, 100 sample correlation matrices were generated and factor analyzed, resulting in 100 values of *K* for each of the population-sample size scenarios. These 100 *K*s were arranged so that $K_{(1)} \leq K_{(2)} \leq \ldots \leq K_{(100)}$. Then, the value of $\frac{K_{(5)} + K_{(6)}}{2}$ was used to represent the lower boundary of a 95% confidence interval for this population correlation condition with this particular sample size and was denoted as $K_{0.95}$. One hundred $K_{0.95}$s were obtained in this way for each population-sample size scenario.

Criteria for Good and Excellent Agreement

MacCallum et al. (1999) followed the guidelines set forth in Tucker et al. (1969) for the interpretation of the coefficient of congruence, *K*. A value of *K* between 0.98 and 1.00 is considered excellent agreement, between 0.92 and 0.98 is considered good agreement, between 0.82 and 0.92 is considered borderline, between 0.68 and 0.82 is considered poor, and below 0.68 is considered terrible. In this study, because the focus is on minimum necessary sample size to achieve acceptable agreement, values of *K* set at 0.98 and 0.92 were used as thresholds to determine sample size recommendations. In particular, $P_{0.92}$ is defined as the percentage of the number of $K_{0.95}$s that are larger than 0.92 in the 100 $K_{0.95}$s from a particular population-sample size scenario, and $P_{0.98}$ is defined as the percentage of the number of $K_{0.95}$s that are larger than 0.98 in the 100 $K_{0.95}$s from a particular population-sample size scenario.

To determine the minimum necessary sample size for any population correlation matrix condition, the following criteria were used as a threshold for "good" agreement between the population solution and the sample solutions:

1. The $P_{0.92}$ values from three successive sample sizes are greater than or equal to 0.95, or
2. The $P_{0.92}$ values from two successive sample sizes are greater than or equal to 0.95, the $P_{0.92}$ value from the next sample size is less than 0.95, and the $P_{0.92}$ values from the next two successive sample sizes are greater than or equal to 0.95.

The following criteria were used as a threshold for "excellent" agreement:

3. The $P_{0.98}$ values from three successive sample sizes are greater than or equal to 0.95, or
4. The $P_{0.98}$ values from two successive sample sizes are greater than or equal to 0.95, the $P_{0.98}$ value from the next sample size is less than 0.95, and the $P_{0.98}$ values from the next two successive sample sizes are greater than or equal to 0.95.

RESULTS

In this study, two minimum necessary sample sizes are determined for 180 different population conditions: one that allows for "good" agreement between sample and population solutions (i.e., *K* values at least as high as 0.92) and one that allows for "excellent" agreement between sample and population solutions (i.e., *K* values at least as high as 0.98). To make these determinations, 371,600 population correlation matrices were generated, and 37,160,000 sample correlation matrices were generated. The results are presented in Table 1.

TABLE 1
The Minimum Necessary Sample Sizes for Each Condition Under Two Criteria

	Excellent-Level Criterion (0.98)						*Good-Level Criterion (0.92)*					
p/f	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*
Level of communality: High												
3	32	320	600	800	1,000	1,200	13	90	170	260	300	350
4	27	150	260	350	450	500	13	75	120	170	220	170
5	21	75	130	260	260	300	11	45	65	90	130	110
6	19	55	95	160	200	160	12	40	50	55	70	70
7	18	45	75	110	130	110	11	40	40	55	55	55
8	18	45	75	90	75	70	11	40	30	40	45	55
9	17	40	60	65	80	80	12	35	30	40	50	60
10	15	35	60	70	65	65	13	35	35	45	55	65
11	16	35	55	60	60	75	14	35	40	55	60	70
12	15	35	55	55	65	75	15	35	40	55	65	75
Level of communality: Wide												
3	110	710	1,300	1,400	1,400	1,600	35	160	450	500	700	600
4	65	220	350	700	900	900	25	90	130	240	320	300
5	50	130	200	300	300	350	30	60	80	110	140	130
6	50	95	140	180	200	180	20	55	65	75	70	100
7	40	75	105	160	150	130	20	50	55	75	65	60
8	36	65	90	90	130	110	15	45	45	50	55	60
9	33	55	70	85	90	100	15	40	40	50	50	60
10	32	55	75	80	85	95	14	35	35	45	55	65
11	36	50	65	75	85	95	14	35	40	50	60	70
12	30	50	70	75	85	95	15	35	40	50	65	75
Level of communality: Low												
3	150	900	1,700	2,600	3,000	3,800	45	600	1,200	1,200	1,300	1,200
4	95	270	450	800	1,000	1,400	35	120	230	250	400	400
5	75	150	220	370	430	400	35	75	85	170	180	160
6	70	120	160	190	200	260	30	60	85	130	120	120
7	60	80	100	180	170	140	30	60	65	75	85	80
8	55	75	100	100	130	130	23	60	60	75	80	75
9	50	70	85	110	100	120	22	50	60	60	65	70
10	50	70	85	90	110	110	20	45	40	60	60	70
11	50	60	75	95	95	105	20	45	45	50	60	70
12	50	60	75	85	100	110	20	40	40	55	65	75

Note. p/f is the ratio of variables to factors. F1 denotes a one-factor solution, F2 a two-factor solution, and so forth.

The following relations are evident from the table: (a) for a fixed variables-to-factors ratio (*p*/*f*), as the number of factors increases, the minimum sample size increases; (b) for a fixed number of factors, as the level of communality decreases, the minimum sample size increases; (c) for a fixed number of factors, as the number of variables increases, the minimum sample size decreases (although in the lower portion of each section of the table for some factor numbers, the minimum sample size increases slightly because the sample size cannot be smaller than the total number of variables); and (d) to have "good" agreement between sample and population, once the variables-to-factors ratio reaches at least 7, the minimum sample size begins to stabilize.

Tables 2 and 3 summarize the ranges for the minimum necessary sample sizes with the variables-to-factors ratio set at 7 and the number of factors varying. For example, in Table 3, with high levels of communality and the "excellent" agreement criterion, the minimum sample size varies from 75 to 100 depending on the number of factors that are present in the data. And if one is satisfied with "good" agreement, then the sample size need not ever exceed 85, regardless of the level of communality or the number of factors.

Tables 1, 2, and 3 are presented in terms of the number of variables to number of factors ratio (*p*/*f*). If one considers the number of variables involved rather than the variables-to-factors ratio, it is also evident in Table 1 that as the total number of variables increases into the 40 to 50 range, the minimum sample size required to

TABLE 2
The Ranges of Minimum Necessary Sample Sizes in 12 Different Conditions for Variables-to-Factors Ratio of 7

Number of Factors	*Excellent-Level Criterion 0.98*	*Good-Level Criterion 0.92*
1	18–60	11–30
2	45–80	40–60
3	75–100	40–65
4	110–180	55–75
5	130–170	55–85
6	110–140	55–80

TABLE 3
The Range of the Maximum Necessary Sample Sizes With Number of Factors From 2 to 6, Three Levels of Communality, and Two Criteria When the Variables-to-Factors Ratio is 7

Criterion	*High*	*Wide*	*Low*
0.98	75–130	105–160	100–180
0.92	40–55	50–75	60–85

meet the “good” level of agreement criterion begins to stabilize, in just about every case, at a point below 100.

CONCLUSION

The aim of this study was to investigate the minimum necessary sample size for conducting factor analyses under a variety of conditions. The intent was that recommendations could be made that could serve as more or less universal guidelines for use when planning studies involving factor analyses. With that in mind, the following conclusions are drawn.

First, the number of variables may not be an appropriate index to use to determine the size of the sample. Cattell (1978), Nunnally (1967), Everitt (1975), and others all proposed various ratios of participants to variables ranging from 3 to 1, 10 to 1, and even higher for such an index. In this study, with the number of factors fixed, the ratio of sample size to number of variables exhibits an inverse relation. Using the coefficient of congruence criteria described earlier and a fixed number of factors, a small number of variables requires a larger minimum sample size than does a large number of variables. The relation between the minimum necessary sample size and the number of variables, for a fixed number of factors, appears to be compensatory, not proportional. In addition, the relation between the minimum necessary sample size and the ratio of the number of variables to the number of factors also appears to be compensatory. This result is not inconsistent with Marsh, Hau, Balla, and Grayson (1998), in which the same conclusion was drawn in regard to confirmatory factor analysis.

Second, the difference in minimum necessary sample size between two different levels of communality will decrease as the ratio of the number of variables to the number of factors increases—that is, as the total number of variables increases. Hence, with more variables, the influence of the level of communality diminishes. In practice, however, this result may not be of much importance. It is relatively easy to generate correlation matrices with varying levels of communality in a simulation study. But with real data, it is much more difficult to ascertain the communality levels beforehand. Consequently, in practice, using a higher variables-to-factors ratio, which diminishes the effect of the level of communality, appears to be a reasonable recommendation, with this ratio being at least 7, if possible.

Third, attempts to provide absolute minimum necessary sample sizes are probably unrealistic. Gorsuch (1983) and Kline (1994) offered 100 as an absolute minimum, and Comrey and Lee (1992) offered sample sizes of 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1,000 or more as excellent. In this study, however, with a variables-to-factors ratio of at least 7, and even with low communality, the minimum necessary sample size for excellent agreement is never greater than 180 and, in most cases, less than 150. On the other hand, with a variables-to-factors

TABLE 4
Recommendation of Minimum Necessary Sample Sizes for Various p/f Ratios for Three Levels of Communality and Two Criteria

	High Communality		Wide Communality		Low Communality	
Criterion	*p/f*	*n*	*p/f*	*n*	*p/f*	*n*
Excellent (0.98)	4	500	4	900	4	1,400
	6	250	6	200	6	260
	8	100	8	130	8	130
Good (0.92)	5	130	5	140	5	200
	7	55	8	60	8	80

Note. p/f = variables-to-factors ratio.

ratio of 3, the number of factors between 3 and 6, and low communalities, the minimum necessary sample size is at least 1,200. It appears that the variety of conditions investigated here are too diverse to provide absolute minimum sample sizes.

Fourth, although the results are quite varied, it appears to be possible to make some general recommendations regarding minimum necessary sample sizes. These recommendations are presented in Table 4 for both “excellent” and “good” criteria, three levels of communalities, and selected variables-to-factors ratios. Again, it is clear that the more variables that are measured per factor and the greater the level of communality, the smaller the sample sizes need to be.

And finally, a word about the data used here. In this study, the data were simulated correlation matrices based on continuous data. In practice, factor analyses are frequently performed on ordinal and even binary data. Although correlation coefficients can be, and often are, calculated with noncontinuous data so that these results may apply in those situations as well, we did not address that situation. The reader may wish to use caution in applying these recommendations with ordinal or binary data.

ACKNOWLEDGMENT

This article is based on Tian Lu Ke’s doctoral dissertation.

REFERENCES

Archer, C. O., & Jennrich, R. I. (1973). Standard errors for rotated factor loadings. *Psychometrika*, *38*, 581–605.

Arrindell, W. A., & van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, *9*, 165–178.

Browne, M. W. (1968). A comparison of factor analytic techniques. *Psychometrika*, *33*, 267–334.

Cattell, R. B. (1978). *The scientific use of factor analysis*. New York: Plenum.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cudeck, R., & O'Dell, L. L. (1994). Applications of standard error estimation in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin*, *115*, 475–487.

Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry*, *126*, 237–240.

Girshick, M. A. (1939). On the sampling theory of roots of determinantal equations. *Annals of Mathematical Statistics*, *10*, 203–224.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (4th ed.). Saddle River, NJ: Prentice Hall.

Jennrich, R. I. (1973). Standard errors for obliquely rotated factor loadings. *Psychometrika*, *38*, 593–604.

Kline, P. (1994). *An easy guide to factor analysis*. New York: Routledge.

Lawley, D. N. (1967). Some new results in maximum likelihood factor analysis. *Proceedings of the Royal Statistical Society of Edinburgh*, *67A*, 256–264.

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*, 611–637.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84–99.

Marsh, H. W., Hau, K., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, *33*, 181–220.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Preacher, K. J., & MacCallum, R. C. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes. *Behavior Genetics*, *32,* 153–161.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*, 421–459.

Tucker, L. R., & Korth, B. (1976). Procrustes matching by congruence coefficients. *Psychometrika*, *41*, 531–535.

Velicer, W. F., Peacock, A. C., & Jackson, D. N. (1982). A comparison of component and factor pattern: A Monte Carlo approach. *Multivariate Behavioral Research*, *17*, 371–388.