

Applying the bootstrap to the multivariate case: Bootstrap component/factor analysis

LINDA REICHWEIN ZIENTEK

Sam Houston State University, Huntsville, Texas

AND

BRUCE THOMPSON

*Texas A&M University, College Station, Texas
and Baylor College of Medicine, Houston, Texas*

The bootstrap method, which empirically estimates the sampling distribution for either inferential or descriptive statistical purposes, can be applied to the multivariate case. When conducting bootstrap component, or factor, analysis, resampling results must be located in a common factor space before summary statistics for each estimated parameter can be computed. The present article describes a strategy for applying the bootstrap method to conduct either a bootstrap component or a factor analysis with a program syntax for SPSS. The Holzinger–Swineford data set is employed to make the discussion more concrete.

Historically, researchers have concentrated their efforts on null hypothesis statistical significance testing (NHSST), but in recent years, a debate has ensued over the strengths and limitations associated with NHSST. This debate has resulted in the publication of numerous articles (see, e.g., Anderson, Burnham, & Thompson, 2000) and entire books devoted to the topic (see, e.g., Harlow, Mulaik, & Steiger, 1997; Kline, 2004; Thompson, 2006). In 1999, the American Psychological Association (APA) Task Force on Statistical Inference (TFSI) published recommendations for alternatives to common statistical methods (Wilkinson & TFSI, 1999). These publications have enabled researchers to understand two major limitations of statistical significance: the dependency of statistical significance upon sample size and the fact that statistical significance does not measure the probability that results will replicate (Smith & Henson, 2000; Thompson, 1995). As researchers, we want to know about the population involved in a study, and we want to know whether our results are replicable (Thompson, 1994, 2006). According to Carver (1978), “It is a fantasy to hold that statistical significance reflects the degree of confidence in the replicability or reliability of results” (p. 386).

To empirically estimate result replicability, either external or internal methods may be employed (Thompson, 1993, 1996). External replicability involves collecting a new sample. While this method offers the best evidence of result replicability, most researchers do not have the time or the resources to collect another sample. Internal replication involves investigating the available sample results with one of three methods: cross validation, the jackknife

method, or the bootstrap method (Thompson, 2004). The bootstrap method, developed by Efron and his colleagues, frees researchers from two limiting factors inherent in classical theory: “the assumption that the data conform to a bell-shaped curve and the need to focus on statistical measures whose theoretical properties can be analyzed mathematically” (Diaconis & Efron, 1983, p. 116). Bootstrapping, which became available with advances in technology, allows researchers to move beyond traditional parameter estimates (e.g., mean, standard deviation, and correlation coefficient) to any parameter estimate of interest. With bootstrapping, the only assumption required is that the sample be representative of the population (Guthrie, 2001; Kline, 2005).

The challenge for the social sciences is that generalizing about the idiosyncrasies of individuals is inherently so difficult. The bootstrap method models replication across studies by mixing up the participants in a given study in many different ways to determine whether results are stable across numerous combinations of study participants (Thompson, 1996).

Bootstrap Approaches and Purposes

There are two approaches to bootstrapping: parametric and nonparametric. In the parametric approach, random samples are drawn for a specified probability density function. In the nonparametric approach, hundreds or thousands of resamples are drawn with replacement, and each resample is the same size as the original sample (Kline, 2005; Thompson, 1994). The empirically estimated sampling distribution is comprised of the sample

L. R. Zientek, lrzientek@yahoo.com

statistics from each of the resamples for the parameter estimate of interest.

The ability to create the empirically estimated sampling distribution is exactly what makes the bootstrapping method so powerful (Guthrie, 2001). Standard deviations of the empirically estimated sampling distribution are estimates of the standard errors (*SEs*) of the parameter estimates (Kline, 2005). These *SEs* can be used either for inferential or for descriptive purposes. In inferential applications, the parameters are divided by their empirically estimated *SEs* to create test statistics, from which calculated *p* values can be computed for the purpose of conducting NHSST. As Thompson (1993) explained, "This is only another approach to testing statistical significance" (p. 373).

In descriptive applications, on the other hand, the bootstrap-estimated *SEs* evaluate the replicability of parameter estimates over the numerous resamples. Parameter estimates with relatively small *SEs* are more stable over the resamples, and hence are more apt to replicate. Note that bootstrap resampling can be invoked even when statistical significance is not of interest to the researcher, and instead replicability issues are of primary concern.

The two distinct bootstrap applications can be differentiated by noting which sampling distribution features are of interest in the two applications. In inferential applications, we use the empirically estimated sampling distribution to find critical values in the extreme tails of the sampling distribution, reflecting the fact that we use small p_{critical} values in NHSST. These tails are sparse with respect to density, and so huge numbers of resamples are required in order to accurately estimate these distribution tail densities. In descriptive applications, however, we focus instead on the central tendency and the dispersion of the distribution, which inherently tend to be more stable over resampling.

Like any method, the bootstrap has advantages and limitations. Thus, we emphasize that "bootstrap methods do not magically take us beyond the limits of our data" (Thompson, 1993, p. 373). For example, the bootstrap cannot make an unrepresentative sample representative. Nevertheless, given reasonable sample data, advantages of the bootstrap method include freedom from distributional assumptions, the ability to study any statistic of interest (Guthrie, 2001), the ability to model the specific effects of sampling error (Scott, Sexton, & Thompson, 1989), and the ability to inform either inferential or descriptive applications or both. Investigating the impact of sampling error variance is important because sampling error affects result replicability.

When applied to the multivariate case, bootstrap component, or factor, analysis can be useful for determining the number of factors to retain or the replicability of pattern/structure coefficients (Lorenzo-Seva & Ferrando, 2003; Thompson, 1988, 2004) or both. Of course, the bootstrap method cannot compensate for limitations of the study, such as numerous predictor variables, small effect sizes, and small sample sizes (Kline, 2005; Thompson, 2002). And bootstrap results, like the results of any research method, are just one component that facilitates interpretation of results.

Purpose

The purpose of the present article is to describe a strategy for conducting a bootstrap component, or factor, analysis with a program developed in SPSS. The focus will be on descriptive applications of the bootstrap factor analysis. The well-known Holzinger and Swineford (1939, pp. 81–91) data set is employed to make the discussion more concrete.

Here we use a relatively small subset of nine variables only to make the heuristic application more manageable, in terms of both journal pages and ease of use for readers who seek to use our program and reproduce these analyses to confirm that the program is working. The variables we selected were: T6, Paragraph Comprehension Test; T7, Sentence Completion Test; T9, Word Meaning Test; T10, Speeded Addition Test; T12, Speeded Counting of Dots in Shapes Test; T13, Speeded Discrimination of Straight and Curved Caps Test; T14, Memory of Target Words Test; T15, Memory of Target Numbers Test; and T17, Memory of Object–Number Association Targets Test. Note that the program is not limited to having only nine variables, and that such small data sets are not typical in exploratory factor analysis (Thompson, 2004).

Factor analysis enables researchers to investigate empirical relationships between latent and observed variables. With prior knowledge regarding the measured variables, the confirmatory approach is conducted to test the hypothesis regarding links between factors and the observed variables. Absent prior knowledge, the exploratory approach is taken to determine the minimum number of factors underlying the observed variables. Various methods (Gorsuch, 1983; Thompson, 2004) exist for determining the number of factors to extract; these methods include but are not limited to the eigenvalue-greater-than-one rule, the scree plot/test, and parallel analysis. Parallel analysis, which was introduced by Horn (1965), is not an option in SPSS, but the necessary syntax for this analysis can be located on the Web. The eigenvalue-greater-than-one rule and the scree test are both readily available in SPSS.

Problems may arise when applying the bootstrap method to the multivariate case. In the bootstrap resamples, factors may vary in their orders across solutions. As Thompson (1995) explained:

The bootstrap must be applied such that each of the hundreds or thousands of resampling results are all located in a common factor space before the mean, *SD*, skewness and kurtosis are computed . . . If the analyst computed mean structure (or pattern) coefficients for the first variable on the first component across all the repeated samplings, the mean would be a nonsensical mess representing an average of some apples, some oranges, and perhaps some kiwi. The sampled solutions must be rotated to best fit positions with a common target solution, prior to computing means and other statistics across the resamples, so that the results are reasonable. (pp. 88–89)

To correct variations across factors, a Procrustes rotation is conducted on each resample to rotate all results to the best-fit position with a target matrix, which creates a

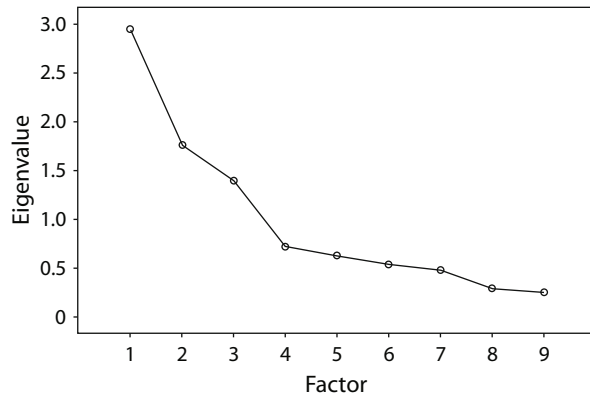


Figure 1. Scree plot results.

single common factor space across resamples. Through the vagaries of sampling error influences, the initial bootstrapped factors may occur in various orders across resamples (e.g., Factor I may appear first in the first resample but second in the next resample).

The naive analyst might think that resampled factors could simply be inspected to determine which should be averaged across solutions. There are two problems with an inspection strategy. First, inspection of 1,000–5,000 resampled structures would be tedious and logistically challenging. Second, any strategy except Procrustes rotation of all solutions to the best-fit position with a target matrix would be incorrect. Best-fit rotation makes the resampled structures as mathematically close in structure as possible given the limits of the resampled data. Any other strategy would yield comparisons that, across resamples, would at best be almost apples to apples, as compared with being the best fit possible to a common factor space.

The target matrix can be (1) the sample varimax-rotated matrix; (2) a matrix composed of ones, negative ones, and zeroes modeling simple structure and computed from the varimax-rotated matrix; or (3) the sample varimax-rotated pattern matrix from a prior study (Thompson, 1988). The Procrustes-rotated pattern/structure coefficients from each resample are then averaged and the standard deviations (*SDs*) of the resampling results computed. The *SDs* of the empirically estimated sampling distribution estimate the *SEs*, which aid in determining replicability.

Exploratory Bootstrap Factor Analysis

For this article, a bootstrap factor analysis (BFA) using the nonparametric approach was conducted with a syntax written in SPSS. Prior to utilizing the program, an exploratory factor analysis with a principal component analysis (PCA) was conducted to determine the number of factors to extract. PCA is the default extraction method in statistical packages and consequently the method most commonly used in journal articles (Chatterjee, 1984; Thompson, 2004). For these reasons, this was the method of choice for the heuristic illustration. However, the syntax can be easily modified to execute other extraction methods.

The appropriateness of PCA as an extraction method has been debated and will not be addressed in this article.

Information on this debate can be found in the various articles arguing both views in a special issue of *Multivariate Behavioral Research* devoted to this topic (Mulaik, 1992). However, it should be noted that differences between PCA and other extraction methods decrease as the number of factored variables increases. This is because iterative estimation of commonalities changes only the diagonal entries of the correlation matrix, and the diagonal of the matrix becomes progressively smaller as more measured variables are analyzed [e.g., 4 diagonal entries with 4 measured variables is 4 out of 16 (25.0%) matrix entries; 8 diagonal entries with 8 measured variables is 8 out of 64 (12.5%) matrix entries]. In practice, we usually factor fairly large numbers of measured variables (Thompson, 2004, p. 56).

To determine the number of factors to extract, a scree plot was investigated. Scree plots have the eigenvalues on the *y*-axis and the factors on the *x*-axis. As Figure 1 shows, in this example, the sample eigenvalues taper off after the third factor, thus indicating that the observed variables are linked to three factors. Note that these plotted eigenvalues are for the sample data and are not the bootstrapped eigenvalues.

Bootstrapped Eigenvalues

One of the first decisions that must be made in BFA is how many factors to extract. As noted previously, a number of methods can be employed to inform this decision, including sophisticated methods such as parallel analysis (Horn, 1965). Unfortunately, the methods most commonly used in applied research (Henson & Roberts, 2007) are exactly the methods shown by Monte Carlo simulations (Zwick & Velicer, 1986) to be least accurate (e.g., the conventional eigenvalue-greater-than-one rule).

Bootstrapping may be used solely for the purpose of determining the correct number of factors to extract. In essence, we seek to consider sampling error as part of using the eigenvalue-greater-than-one rule, which is sometimes incorrectly attributed to Kaiser, but is more accurately attributed to Guttman (1954).

Using the bootstrap, we compute the mean eigenvalues for each factor over repeated resamples (e.g., 1,000 resamples). We also compute the empirically estimated *SEs* of the estimates as the *SDs* of the resampled estimates. In this manner, we can determine whether the mean of resampled

Table 1
Sample Eigenvalues and Mean Bootstrap Results
Across 1,000 Resamples

Sample Eigenvalue	<i>M</i> (BR)	<i>SE</i>	<i>M</i> (BR)/ <i>SE</i>
2.9454	2.9586	0.1808	16.3656
1.7603	1.7782	0.1256	14.1545
1.3963	1.3914	0.0975	14.2729
0.7166	0.7395	0.0518	14.2669
0.6292	0.6265	0.0399	15.7205
0.5351	0.5311	0.0340	15.6128
0.4780	0.4531	0.0343	13.1961
0.2885	0.2866	0.0259	11.0782
0.2506	0.2351	0.0200	11.7313

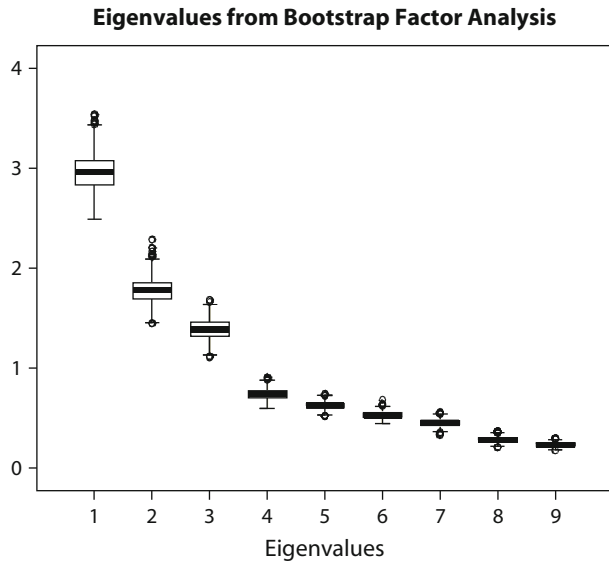


Figure 2. Empirically estimated sampling distribution of eigenvalues.

eigenvalues is greater than 1 and whether *SEs* of mean estimates near 1.0 are large or small.

Table 1 presents these results across our 1,000 resamples. Plots of the empirically estimated sampling distributions for the eigenvalues are presented in Figure 2. For example, across the 1,000 resamples, the first eigenvalue ranged from 2.49 to 3.54. The bootstrap results suggest three factors for these results. For instructional purposes, the mean eigenvalues divided by their *SEs* are presented in Table 1. We want these values to be large, but note that these test the null hypothesis that eigenvalues equal 0.

We instead want to test whether eigenvalues are greater than 1. Therefore, in deciding how many components or factors to extract, using the bootstrap to create confidence intervals (CIs) for the eigenvalues is of greater interest. Once the CIs are created, bootstrap can then determine whether these CIs subsume 1.0, as will be illustrated shortly. Alternatively, we could subtract 1.0 from each eigenvalue and then divide this difference by the *SE* of a given eigenvalue and in that manner conduct a non-nil null hypothesis test (Cohen, 1994) of whether a given eigenvalue equaled 1.0.

Bootstrapped Pattern/Structure Coefficients

After determining the number of factors to extract, a BFA was conducted. The BFA syntax consists of two files. Program 1 first creates a target matrix from the varimax-rotated matrix of the sample. The target matrix is created from the sample varimax-rotated matrix by placing a 1 or a -1 in the corresponding entry of the target matrix if a variable is linked to both a factor and a 0 on the remaining entries in the corresponding row. Results from the Holzinger-Swineford data set are presented in Table 2. After the pattern/structure coefficients are computed from the sample and the target matrix is created, Program 1 calls Program 2.

In Program 2, resamples are drawn with replacement, and a Procrustes rotation into the single common factor

space is conducted on each resampling result. In each resample, cases are randomly selected, and each resample is exactly the same size as the original sample. When sampling with replacement, a case may be drawn once, more than once, or not at all. For example, one of the resamples from this BFA drew Person 12 five times, Person 301 one time, and Person 1 not at all, as reported in Table 3. The procedure is similar to the method employed by Raykov and Little (1999), who used the bootstrap method to determine the success of the Procrustes rotation.

The resampling is conducted hundreds or thousands of times. The resulting Procrustes-rotated pattern/structure coefficients are saved, and the resampling results are concatenated in Program 1. The resulting file is the empirically estimated sampling distribution for both the eigenvalues and structures estimates. In this program, the output file TEIGENVORIG contains the 1,000 eigenvalues, and the output file TBROTORIG contains the 1,000 Procrustes-rotated pattern/structure coefficients. Researchers can now conduct inferential and descriptive analyses of the eigenvalues and the pattern/structure coefficients that would not have been available in classical test theory, without the need to investigate the assumption of multivariate normality.

To illustrate the necessity of the common-factor-space rotation, a BFA was conducted without a Procrustes rotation. On the first 333 resamples, T14, T15, and T17 were aligned on Factor III, but on the 334th resample, these variables were aligned on Factor II, as reported in Table 4.

Table 2
Sample Pattern/Structure Coefficients From the Varimax Rotated Matrix

Variable	Varimax-Rotated Factor			Target Factor		
	I	II	III	I	II	III
T6	.89	.09	.12	1	0	0
T7	.91	.10	.00	1	0	0
T9	.88	.11	.07	1	0	0
T10	.03	.77	.15	0	1	0
T12	.05	.84	.01	0	1	0
T13	.20	.70	.06	0	1	0
T14	.19	-.02	.79	0	0	1
T15	-.04	.02	.79	0	0	1
T17	.05	.36	.64	0	0	1

Table 3
Resample With Replacement

Case (Person)	Frequency	Valid Percentage	Cumulative Percentage
2	2	.7	0.7
4	2	.7	1.3
5	3	1.0	2.3
8	1	.3	2.7
9	2	.7	3.3
10	2	.7	4.0
12	5	.7	5.6
⋮	⋮	⋮	⋮
301	1	.3	100.00

Note—The table partially illustrates one resample of 301 rows of data. Person 1 was not sampled at all; Person 2 was sampled twice, and so forth, until the 301st person was sampled once.

Table 4
Results From One Resample With Procrustes Rotation

Variable	Factor		
	I	II	III
T6	.87	.11	.05
T7	.90	-.08	.07
T9	.89	.04	.08
T10	-.03	.28	.73
T12	.02	.07	.81
T13	.18	-.04	.71
T14	.12	.81	-.03
T15	-.05	.80	.05
T17	.01	.70	.29

Note—Results associated with pattern/structure coefficients deemed salient (i.e., > |0.5|) are presented in italics.

Table 5
Variability in Standard Errors From Bootstrap for Selected Variables

Variable	Factor	Procrustes Rotation	SD	Mean Bootstrap Results/SE
T6	I	0.9824	0.0073	133.9587
T7	I	0.9913	0.0055	181.5362
T9	I	0.9863	0.0078	126.3891
T13	II	0.9502	0.0271	35.0915
T17	III	0.8806	0.0469	18.7631

Therefore, the average of the pattern/structure coefficients for Factor II would be averaging the same thing most but not all of the time, and the results would be meaningless. The Procrustes rotation with the target matrix aligns each resample's factors to the single best-fit position with the common factor space solution so that the summary statistics will be correct.

The empirically estimated sampling distribution, which consists of the summary statistics across all 1,000 resamples for the eigenvalues and the Procrustes-rotated pattern/structure coefficients, is the primary focus of the bootstrap method. The average bootstrap results, the estimated SEs, and the ratios of the average bootstrap results to the estimated SEs can be used for either inferential or descriptive purposes or both and are reported as output from the program.

Small SEs indicate parameter stability across resampling. Larger effect sizes tend to have smaller SEs. Therefore, if pattern/structure coefficients are large in magnitude, the SEs tend to be smaller and the parameter estimates more stable. The BFA results for this example are consistent with this theory. The sample statistics and parameter estimates for variables T6, T7, and T9 are close to 1.0 on Factor I with relatively small SEs, whereas the sample statistics and parameter estimates for T13 on Factor II and for T17 on Factor III are further from 1.0 with comparatively larger SEs, as reported in Table 5.

Over each resample, factor analysis results may diverge from the sample statistics. The more the resampling results deviate from the mean bootstrap results, the larger the estimated SEs and the less stable the parameter estimates will be. Box plots of the empirically estimated sampling distribution (i.e., the 1,000 resampling statistics) give a visual representation of how the variables

deviated from the mean bootstrap result on each factor and are presented in Figures 3, 4, and 5. On Factors I, II, and III, the SEs of the variables linked to the prospective factors were small in comparison with the mean bootstrap parameter estimates. Variable T7 was the most stable, and the 1,000 Procrustes-rotated pattern/structure coefficients for T7 ranged from .97 to 1.00 on Factor I. T17 is the least stable across the resamples, with an SE of approximately .05 on Factor III. The Procrustes-rotated pattern/structure coefficient results for T17 deviated more from the mean bootstrap results, and the 1,000 resampling results ranged from .62 to .98. In comparison with T7, this range is much larger, as reported in Table 6.

When testing stability across resamples, the hope is for the mean bootstrap results and sample statistics to be relatively close and the empirically estimated SEs to

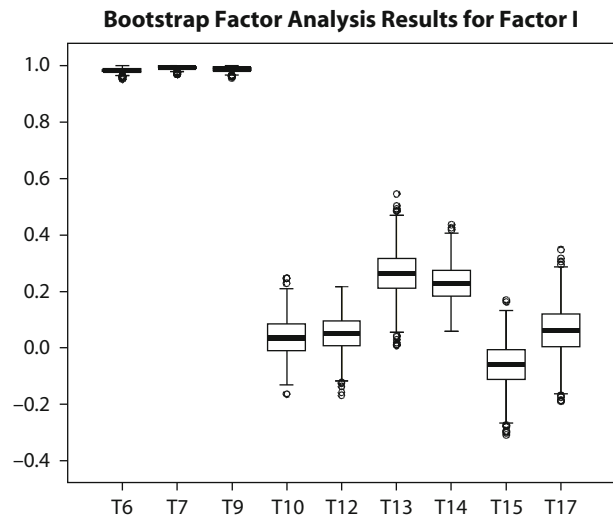


Figure 3. Empirically estimated sampling distribution for Factor I.

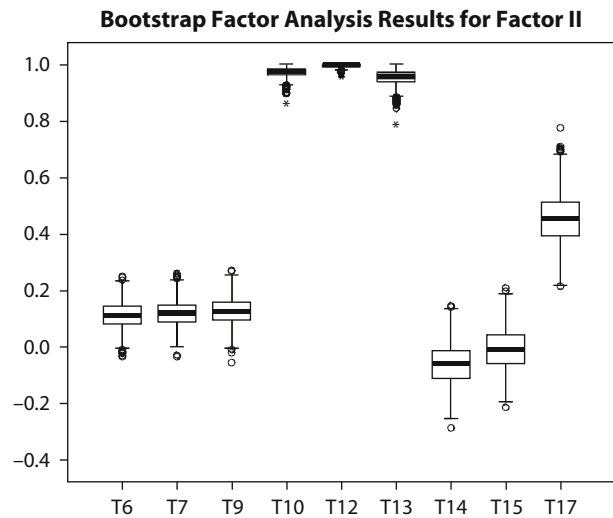


Figure 4. Empirically estimated sampling distribution for Factor II.

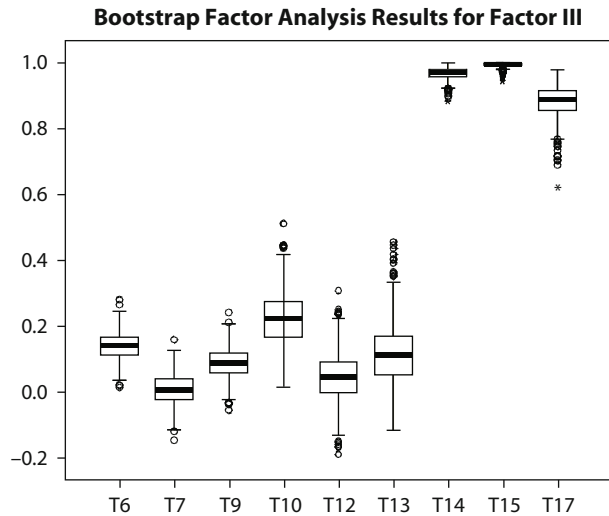


Figure 5. Empirically estimated sampling distribution for Factor III.

be comparatively small. As Guthrie (2001) noted, “If the sample statistic is relatively equal to the mean bootstrap estimate and *SE* is small in relation to the mean bootstrap estimate, then the sample statistic can be thought of as stable” (p. 12). On the other hand, if the sample statistics and parameter estimates are not close or if the *SE* is large in relation to the mean bootstrap estimate, then bias may be reflected and caution is warranted when interpreting the sample data.

Results of the heuristic bootstrap component analysis indicate that the nine variables from this study are relatively stable over the 1,000 resamples. Variables T6, T7, and T9 have mean bootstrap results that are close to the sample Procrustes-rotated pattern/structure statistics and have *SEs* that are small in relation to the mean bootstrap estimates, as reported in Table 7. This indicates that the sample statistics are stable across the resamples and more likely to replicate. In this data set, all of the variables are stable and no bias is reflected. Variable T17 has the largest *SE* in comparison with the mean bootstrap results. The ratio of the mean bootstrap results and the *SE*, which behaves like a *t* statistic, is greater

than or equal to 2.0 for each variable deemed salient to given factors, indicating that these coefficients are not 0. Therefore, the researcher can be more confident that the results replicate over the resamples and that the variables are correctly linked to the prospective factors.

Confidence Intervals

The output of the empirically estimated eigenvalues saved in the file TEIGENVORIG.SAV and the Procrustes-rotated pattern/structure coefficients saved in the file TBROTORIG.SAV can be utilized for further investigation of the data (Ichikawa & Konishi, 1995). CIs of the Procrustes-rotated pattern/structure coefficients and eigenvalues could be computed and could serve as an interpretation aid that takes sampling variability into account (Lambert, Wildt, & Durand, 1991). In 1999, the APA Task Force on Statistical Inference (Wilkinson & TFSI, 1999) strongly recommended the reporting of CIs. CIs take into account sampling error and can be established for any parameter of interest, including but not limited to means, regression coefficients, and effect sizes (Steiger, 2004; Thompson, 2006; Wilkinson & TFSI, 1999).

CIs have advantages over NHSST tests because they provide “more information than does a statement about rejecting (or failing to reject) a null hypothesis, which is almost false anyway” (Pedhazur, 1997, p. 30), and they can be used to test all possible null hypotheses simultaneously, to aid in compiling studies, and to display results graphically (Capraro, 2005; Pedhazur, 1997). When computing CIs with the bootstrap results, adjustments need to be made for biases inherent in the sample (Lambert et al., 1991).

Program Availability

To aid in usability, the program illustrated here is saved in Microsoft Word, and specific sample items requiring modification are highlighted in color. The program can be downloaded from www.coe.tamu.edu/~bthompson/data sets. Because some items, such as variable names, need to be changed in multiple places, the most efficient method for doing so is to use the Find and Replace function in Word or SPSS. Also, factors may need to be added or deleted. Executing Find in Program 2 for “FAC” will expedite the process of adding and deleting the corre-

Table 6
Range and Standard Deviations of Bootstrap Results for Factors I, II, and III

Variable	Factor I			Factor II			Factor III		
	Min	Max	<i>SD</i>	Min	Max	<i>SD</i>	Min	Max	<i>SD</i>
T6	.95	1.00	.00733	-.04	.25	.04539	.01	.28	.03881
T7	.97	1.00	.00546	-.04	.26	.04508	-.15	.16	.04557
T9	.96	1.00	.00780	-.06	.27	.04902	-.06	.24	.04427
T10	-.16	.25	.06875	.86	1.00	.01894	.01	.51	.07679
T12	-.17	.22	.06135	.95	1.00	.00611	-.19	.31	.06997
T13	.01	.55	.08111	.79	1.00	.02708	-.12	.45	.09082
T14	.06	.44	.06474	-.29	.14	.07190	.88	1.00	.01742
T15	-.31	.17	.07757	-.22	.21	.07180	.94	1.00	.00742
T17	-.19	.35	.08286	.21	.77	.08816	.62	.98	.04694

Note—Results associated with pattern/structure coefficients deemed salient (i.e., > |0.5|) are presented in italics.

Table 7
Sample and Bootstrap Results Across 1,000 Resamples

Variable	Factor I				Factor II				Factor III			
	Sample	Bootstrap	SE	M(BR)/SE	Sample	Bootstrap	SE	M(BR)/SE	Sample	Bootstrap	SE	M(BR)/SE
T6	<i>.9845</i>	.9824	.0073	133.96	<i>.1077</i>	.1089	.0454	2.40	<i>.1381</i>	.1391	.0388	3.59
T7	<i>.9932</i>	.9913	.0055	181.54	<i>.1160</i>	.1149	.0451	2.55	<i>.0032</i>	.0055	.0456	0.12
T9	<i>.9883</i>	.9863	.0078	126.39	<i>.1238</i>	.1239	.0490	2.53	<i>.0890</i>	.0867	.0443	1.96
T10	<i>.0365</i>	.0369	.0687	0.54	<i>.9749</i>	.9689	.0189	51.15	<i>.2198</i>	.2210	.0768	2.88
T12	<i>.0468</i>	.0486	.0614	0.80	<i>.9978</i>	.9935	.0061	162.58	<i>.0470</i>	.0433	.0700	0.62
T13	<i>.2643</i>	.2622	.0811	3.23	<i>.9579</i>	.9502	.0271	35.09	<i>.1122</i>	.1134	.0908	1.25
T14	<i>.2283</i>	.2296	.0647	3.55	<i>-.0602</i>	-.0633	.0719	-0.88	<i>.9717</i>	.9662	.0174	55.46
T15	<i>-.0580</i>	-.0583	.0776	-0.75	<i>-.0174</i>	-.0107	.0718	-0.15	<i>.9982</i>	.9926	.0074	133.77
T17	<i>.0600</i>	.0600	.0829	0.72	<i>.4566</i>	.4517	.0882	5.12	<i>.8876</i>	.8806	.0469	18.76

Note—*M*(BR) stands for mean bootstrap results. The sample results are italicized to distinguish the sample results from the bootstrap results.

sponding factors. A limitation of the program is the length of time it requires to run large numbers (e.g., 1,000) of resamples, which can take several hours depending on hardware. This is a limitation due to workspace complications in SPSS. However, Version 13.0 of SPSS runs in less than half the time of prior SPSS versions.

Discussion

Bootstrap methods allow researchers to focus on statistics other than those whose sampling distributions have been theoretically derived (Diaconis & Efron, 1983). The bootstrap method is one method that can be applied to any statistic of interest, even in the multivariate case. However, in the multivariate case, results must be rotated into the best-fit position in a single common factor space prior to averaging resampled parameter estimates. One advantage of the bootstrap method is its ability to create an empirically estimated sampling distribution, which can then be used for either descriptive (i.e., replicability analyses) or inferential purposes or both.

In this study, a small heuristic data set was used to illustrate how the bootstrap method can be utilized for descriptive purposes and applied to the multivariate case. The results show how the BFA can be useful for “resolving the number of factors, for estimating structure coefficients, and for estimating invariance of results across samples” (Thompson, 1988, p. 681). Stability of parameter estimates was determined by examining *SEs*, comparing the sample statistics to mean bootstrap results, and investigating the ratio of the mean or median bootstrap results to *SEs*. The descriptive bootstrap can be used to evaluate the likely replicability of results, which is exactly the sine qua non of science—finding relationships that recur under stated conditions.

AUTHOR NOTE

Correspondence concerning this article should be addressed to L. R. Zientek, Department of Mathematics, Sam Houston State University, Huntsville, Texas 77341 (e-mail: lrzientek@yahoo.com).

REFERENCES

- ANDERSON, D. R., BURNHAM, K. P., & THOMPSON, W. (2000). Null hypothesis testing: Problems, prevalence, and an alternative. *Journal of Wildlife Management*, *64*, 912-923.
- CAPRARO, M. M. (2005). An introduction to confidence intervals for

- both statistical estimates and effect sizes. *Research in the Schools*, *12*, 22-32.
- CARVER, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review*, *48*, 378-399.
- CHATTERJECC, S. (1984). Variance estimation in factor analysis: An application of the bootstrap. *British Journal of Mathematical & Statistical Psychology*, *37*, 252-262.
- COHEN, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997-1003.
- DIACONIS, P., & EFRON, B. (1983). Computer-intensive methods in statistics. *Scientific American*, *248*, 116-130.
- GORSUCH, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- GUTHRIE, A. C. (2001, February). *Using bootstrap methods with popular statistical programs*. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans.
- GUTTMAN, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*, 149-161.
- HARLOW, L. L., MULAİK, S. A., & STEIGER, J. H. (EDS.), (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.
- HENSON, R. K., & ROBERTS, J. K. (2007). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational & Psychological Measurement*, *66*, 393-416.
- HOLZINGER, K. J., & SWINEFORD, F. (1939). *A study in factor analysis: The stability of a bi-factor solution*. Chicago: University of Chicago Press.
- HORN, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.
- ICHIKAWA, M., & KONISHI, S. (1995). Application of the bootstrap method in factor analysis. *Psychometrika*, *60*, 77-93.
- KLINE, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.
- KLINE, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.
- LAMBERT, Z. V., WILDT, A. R., & DURAND, R. M. (1991). Approximating confidence intervals for factor loadings. *Multivariate Behavioral Research*, *26*, 421-434.
- LORENZO-SEVA, U., & FERRANDO, P. J. (2003). IMINCE: An unrestricted factor-analysis-based program for assessing measurement invariance. *Behavior Research Methods, Instruments, & Computers*, *35*, 318-321.
- MULAİK, S. A. (ED.) (1992). Theme issue on principal components analysis. *Multivariate Behavioral Research*, *27* (3).
- PEDHAZUR, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Stamford, CT: Thomson Learning.
- RAYKOV, T., & LITTLE, T. D. (1999). A note on Procrustes rotation in exploratory factor analysis: A computer intensive approach to goodness-of-fit evaluation. *Educational & Psychological Measurement*, *59*, 47-57.
- SCOTT, R. L., SEXTON, D., & THOMPSON, B. (1989). Structure of a short form of the questionnaire on resources and stress: A bootstrap factor analysis. *Educational & Psychological Measurement*, *49*, 409-419.

- SMITH, A. D., & HENSON, R. K. (2000, January). *State of the art in statistical significance testing: A review of the APA Task Force on Statistical Inference*. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX.
- STEIGER, J. H. (2004). Beyond the *F* test: Effect size, confidence intervals, and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, **9**, 164-182.
- THOMPSON, B. (1988). Program FACSTRAP: A program that computes bootstrap estimates of factor structure. *Educational & Psychological Measurement*, **48**, 681-686.
- THOMPSON, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. *Journal of Experimental Education*, **61**, 361-377.
- THOMPSON, B. (1994). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. *Journal of Personality*, **62**, 157-176.
- THOMPSON, B. (1995). Exploring the replicability of a study's results: Bootstrap statistics for the multivariate case. *Educational & Psychological Measurement*, **55**, 84-94.
- THOMPSON, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, **25**(2), 26-30.
- THOMPSON, B. (2002). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling & Development*, **80**, 64-71.
- THOMPSON, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.
- THOMPSON, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York: Guilford.
- WILKINSON, L., & TASK FORCE ON STATISTICAL INFERENCE (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, **54**, 594-604.
- ZWICK, W. R., & VELICER, W. F. (1986). Factors influencing five rules for determining the number of components to retain. *Psychological Bulletin*, **99**, 432-442.

(Manuscript received May 7, 2005;
revision accepted for publication February 28, 2006.)