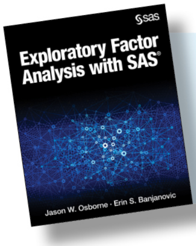


Contents

About This Book vii

About the Author xi

Chapter 1 / Introduction to Exploratory Factor Analysis **1**
- A Tool for Exploration 1
- EFA vs PCA 2
- Steps to Follow When Conducting EFA 5
- Basic Syntax for EFA 7
- Summary 9
- References 9

Chapter 2 / Extraction Method **11**
- What Is Extraction? 11
- Key Concepts 12
- Extraction Techniques 13
- Three Pedagogical Examples 15
- Example Syntax and Output 16
- Does Extraction Method Matter? 20
- Summary 25
- Exercises 26
- Instruments Used in Examples 27
- References 31

Chapter 3 / Factor Extraction Criteria **33**
- How Many Factors? 33
- Extraction Criteria 34
- Example Syntax and Output 36
- How Do the Criteria Compare? 45
- Summary 57
- Exercises 58
- References 58

Chapter 4 / Rotation **61**
- The Magic of Rotation 61
- Types of Rotation 63
- Orthogonal vs Oblique Rotation 64
- Interpretation of Factor Matrices 65

Example Syntax and Output 66
Which Method? 69
Exploring Multiple Solutions 75
Summary 80
Exercises 81
References 82

Chapter 5 / Sample Size Matters **83**
Why Is Sample Size Important? 83
Published Guidelines 84
Sample Size in Practice 85
An Empirical Investigation in Sample Size 86
An Applied Example from Costello & Osborne 87
Application: Impact of Sample Size on Interpretation 89
Summary 93
Exercises 93
References 94

Chapter 6 / Replication Statistics **97**
Importance of Replication 97
Let's Bring Replication to EFA 98
Procedural Aspects of Replicability Analysis 99
Quantifying Replicability in EFA 100
Application: Replication of Marsh SDQ Data 102
Summary 109
Exercises 110
References 110

Chapter 7 / Bootstrap Applications **113**
How Does Bootstrap Resampling Fit into EFA? 113
The Rise of Resampling 114
Bootstrap Resampling Methods 115
What Can Bootstrap Resampling Do, and What Should It Not Be Used For? 116
Example Syntax and Output 117
Application 126
Summary 137
Exercises 138
References 138

Chapter 8 / Data Cleaning **141**
The Importance of Data Cleaning 141

Outliers . . . 142
Application: Random vs Constant Responding . . . 147
Missing Data . . . 149
Application: Nonrandom Missingness and Imputation . . . 152
Summary . . . 155
Exercises . . . 155
References . . . 156

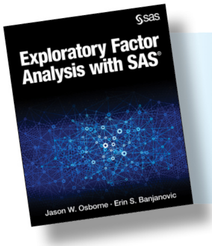
Chapter 9 / Factor Scores . . . **159**
Factor Scores 101 . . . 159
Example Syntax and Output . . . 160
Application: Factor Score Estimation . . . 162
What Are Modern Alternatives? . . . 168
Summary . . . 169
Exercises . . . 169
References . . . 170

Chapter 10 / Higher-Order Factors . . . **171**
What Is a Higher-Order Factor? . . . 171
Did the Initial Solution Get It Right? . . . 172
Mechanics of Performing Second-Order Factor Analysis . . . 173
Application: Replication of Second-Order Factor . . . 177
Summary . . . 179
Exercises . . . 179
References . . . 179

Chapter 11 / After the EFA: Internal Consistency . . . **181**
What Comes Next? . . . 181
What Is Cronbach's Alpha (And What Is It Not)? . . . 182
What Is "Good Enough" for Alpha? . . . 184
Factors That Influence Alpha and Its Use . . . 185
Estimation of Alpha in SAS . . . 187
Would Error-Free Measurement Make a Real Difference? . . . 189
Application . . . 191
Summary . . . 202
Exercises . . . 203
References . . . 204

Chapter 12 / Summary and Conclusions . . . **207**
Best Practices for EFA . . . 207
References . . . 209

Index **211**



1

Introduction to Exploratory Factor Analysis

A Tool for Exploration 1
EFA vs PCA 2
Steps to Follow When Conducting EFA 5
Basic Syntax for EFA 7
Summary 9
References 9

A Tool for Exploration

Exploratory factor analysis (EFA) is a statistical tool used for exploring the underlying structure of data. It was originally developed in the early 1900s during the attempt to determine whether intelligence is a unitary or multidimensional construct (Spearman, 1904). It has since served as a general-purpose dimension reduction tool with many applications. In the modern social sciences it is often used to explore the psychometric properties of an instrument or scale. Exploratory factor analysis examines all the pairwise relationships between individual variables (e.g., items on a scale) and seeks to extract latent factors from the measured variables. During the 110 years since Spearman's seminal work in this area, few statistical techniques have been so widely used (or, unfortunately, misused).

The goal of this book is to explore best practices in applying EFA using SAS. We will review each of the major EFA steps (e.g., extraction, rotation), some associated practices (estimation of factor scores and higher-order factors), and some less common analyses that can inform the generalizability of EFA results (replication analyses and bootstrap analyses). We will review the SAS syntax for each task and highlight best practices according to research and practice. We will also demonstrate the procedures and analyses discussed throughout the book using real data, and we will occasionally survey some poor practices as a learning tool.

To get started in our exploration of EFA, we will first discuss the similarities and differences between EFA and principal components analysis (PCA), another technique that is commonly used for the same goal as EFA. We will then briefly summarize the steps to follow when conducting an EFA and conclude with a quick introduction to EFA in SAS.

EFA vs PCA

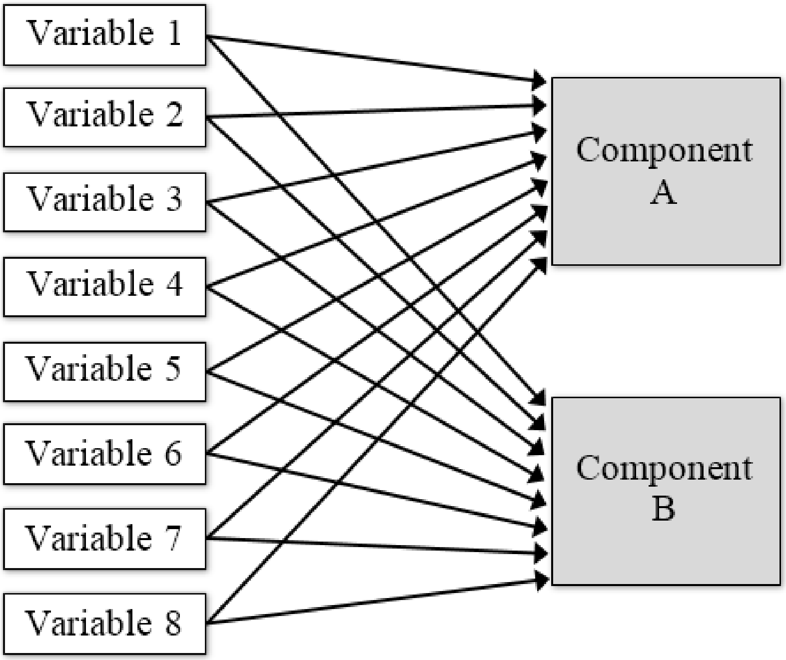
As you will come to learn, EFA is quite different from PCA. Unfortunately, there are many misconceptions about the two analyses, and one of the biggest is that PCA is part of, or synonymous with, EFA. This misconception probably has modern day roots from at least two factors:

1. Statistical software, including SAS, has PCA as the default extraction technique when performing exploratory factor analysis.
2. Many modern researchers use PCA and EFA interchangeably, or use PCA when performing an analysis that is more appropriate for EFA.

Although the two methods generally seem to do the same thing, they are different in some key ways. Principal components analysis is a computationally simplified version of a general class of dimension reduction analyses. EFA was developed before PCA (Hotelling, 1933), thanks to Spearman (1904). EFA was developed prior to the computer age when all statistical calculations were done by hand, often using matrix algebra. As such, these were significant undertakings requiring a great deal of effort. Because of the substantial effort required to perform EFA with hand calculations, significant scholarship and effort went into developing PCA as a legitimate alternative that was less computationally intense but that also provided similar results (Gorsuch, 1990). Computers became available to researchers at universities and industrial research labs later in the 20th century, but remained relatively slow and with limited memory until very late in the 20th century (about the time the first author was in graduate school using mainframes at the university). Our commentary on PCA is not intended to slight these scholars nor to minimize their substantial contributions, but rather to attempt to put PCA and EFA into context for the modern statistician and quantitative researcher. We will therefore focus on EFA, despite the popularity of PCA.

Without getting into the technical details, which are available in other scholarly references on the topic, PCA computes the analysis without regard to the underlying latent structure of the variables, using all the variance in the manifest variables. What this means is that there is a fundamental assumption made when choosing PCA: that the measured variables are themselves of interest, rather than some hypothetical latent construct (as in EFA). This makes PCA similar to multiple regression in some ways, in that it seeks to create optimized weighted linear combinations of variables.

Figure 1.1 *Conceptual overview of principal components analysis*



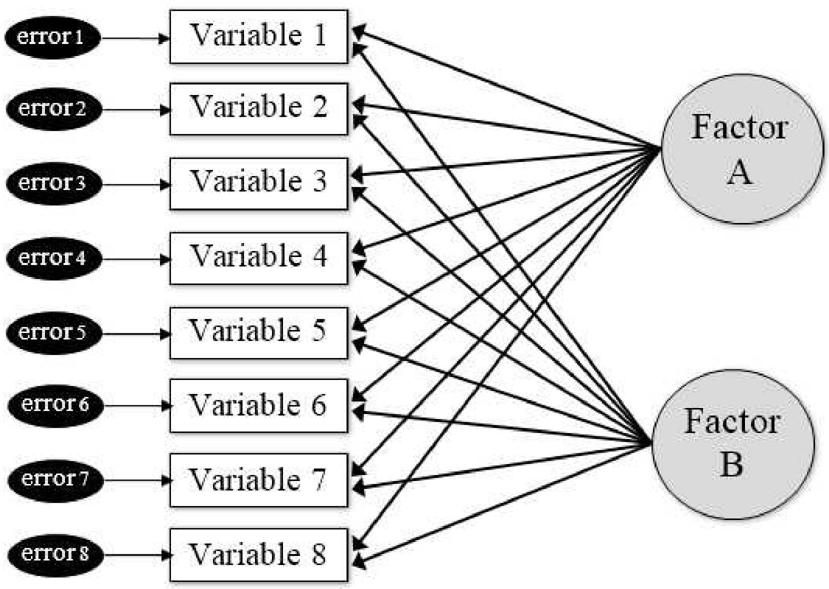
An example of a PCA model to extract two factors is presented in Figure 1.1 on page 3. We have already conducted some initial analyses (to be discussed in Chapter 3) that have convinced us of this two-component structure and led us to this model. Note that all PCA and EFA analyses extract as many components or factors as there are manifest variables, although not all are retained for interpretation; here we simplify for clarity to examine the first two components extracted and to see how they relate to the measured variables. Now the important thing to note in this figure is the direction of the arrows. Notice that they point from the variables to the components. This is because each component is formed as a *weighted linear combination* [1] of the predictor variables. One-hundred percent of what is in those variables ends up becoming the components. As analysts, we can then review the results and identify the primary component each variable loads on to create scales, we can create component scores, or we can do whatever else we would like with the results; but the components themselves are completely defined by the variables. In this way, principal components analysis combines manifest (observed) variables into components.

Exploratory factor analysis, on the other hand, is a group of extraction and rotation techniques that are all designed to model unobserved or latent constructs. It is referred to as common factor analysis or exploratory factor analysis. EFA assumes and asserts that there are latent variables that give rise to the manifest (observed) variables, and the calculations and results are interpreted very differently in light of this assumption.

1 Weighted linear combinations means that each variable has a different weight: the factor/component loading that determines how much or little each variable contributes to the composite. This is similar to multiple regression where a variable score is composed of different amounts (regression weights) of each variable in the equation.

You can see this very different conceptual vision of the same two-factor model reviewed above in Figure 1.2 on page 4 below. Notice the changed direction of the arrow between the variables and factors as well as the addition of error terms for each variable. Factor analysis recognizes that model variance contains both shared and unique variance across variables. EFA examines only the shared variance from the model each time a factor is created, while allowing the unique variance and error variance to remain in the model. The factors are then created as weighted linear combinations of the shared variance. When the factors are uncorrelated and communalities are moderate, PCA can produce inflated values of variance accounted for by the components (Gorsuch, 1997; McArdle, 1990). Since factor analysis analyzes only shared variance, factor analysis should yield the same general solution (all other things being equal) while also avoiding the illegitimate inflation of variance estimates.

Figure 1.2 *Conceptual overview of exploratory factor analysis*



There are two other issues with PCA that we will briefly note. First, PCA assumes that all variables are measured without error (an untenable assumption in almost any discipline), whereas EFA offers the option of acknowledging less than perfect reliability. Second, PCA parameters are selected in an attempt to reproduce sample, rather than population, characteristics (Thompson, 2004).

Thus, we have many similarities between PCA and some important conceptual and mathematical differences. Most authors agree that there is little compelling reason to choose PCA over other extraction methods, and that PCA can be limited and provide biased parameter estimates. Such a list of authors would include Bentler & Kano, 1990; Floyd & Widaman, 1995; Ford, MacCallum, & Tait, 1986; Gorsuch, 1990; Loehlin, 1990; MacCallum & Tucker, 1991; Mulaik, 1990; Widaman, 1993. If one is to seek best practices, one is hard pressed to conclude PCA is ever a best practice. Widman (1993) puts it very bluntly:

“principal components analysis should not be used if a researcher wishes to obtain parameters reflecting latent constructs or factors.” (p. 263). Unfortunately, it is still the default dimension reduction procedure in much statistical analysis software, even though it is usually not (in our opinion) the conceptually desirable choice, and usually has no clear advantage in modern quantitative methodology that we can detect.

This is a topic that arouses passions among statisticians, and the first author has rarely published a paper or given a talk on this topic without someone getting upset for taking this position so clearly and unapologetically. So let us sidestep this issue for the moment and summarize: PCA is not considered a factor analytic technique, and there is disagreement among statisticians about when it should be used, if at all. More often than not, researchers use PCA when EFA would be appropriate and preferable (for example, see Ford et al., 1986; Gorsuch, 1983; Widaman, 1993).

Steps to Follow When Conducting EFA

Exploratory factor analysis is meant to be exploratory in nature, and thus it is not desirable to prescribe a rigid formula or process for executing an EFA. The steps below are meant to be a loose guide, understanding that a factor analysis often requires returning to previous steps and trying other approaches to ensure the best outcome. The general pattern of performing an EFA falls into six general steps that will guide the discussion through the rest of the book:

1. Data cleaning
2. Deciding on an extraction method to use
3. Deciding how many factors to retain
4. Deciding on a method of rotation (if desired)
5. Interpreting results *(return to #3 if a solution is not ideal)*
6. Replication or evaluation of robustness *(return to the beginning if a solution is not replicable or robust)*

Step 1: Data cleaning. Without clean data, what follows in almost any analysis is moot. This is another point where passions run high among researchers and statisticians because there is considerable controversy about any manipulations of the sample and data (e.g., how to treat outliers, missing data). We have a clear position on the issue—data should be cleaned and issues (e.g., failing to meet assumptions) should be addressed. The first author wrote an entire book on the topic, in which he demonstrated repeatedly how clean data produces results that are better estimates of population parameters and, therefore, more accurate and replicable (Osborne, 2013). Instead of debating the point here, allow me to assert that data that is filled with errors or that fails to meet assumptions of the analysis being performed is likely to lead to poorer outcomes than data that is free of egregious errors and that meets assumptions. We will discuss some other data quality issues later in the book, including the importance of dealing appropriately with missing data.

Step 2: Deciding on an extraction method. An extraction technique is one of a group of methods that examines the correlation/covariation between all the variables and seeks to “extract” the latent variables from the measured/manifest variables.

There are several factor analysis extraction methods to choose from. SAS has seven *EFA extraction methods*: unweighted least squares (ULS), maximum likelihood (ML), principal axis factoring (PAF), iterated principal axis factoring (iterated PAF), alpha factoring, image factoring, and Harris factoring.[1] Information about the relative strengths and weaknesses of these techniques is not easy to obtain. To complicate matters further, naming conventions for some extraction techniques are not consistent, leaving it difficult to figure out which method a textbook or journal article author is describing, and whether or not it is actually available in the software the researcher is using. This probably explains the popularity of principal components analysis – not only is it the default in much statistical software, but it is one of the more consistent names researchers will see there.

An article by Fabrigar, Wegener, MacCallum and Strahan (1999) argued that if data is relatively normally distributed, maximum likelihood is the best choice because “it allows for the computation of a wide range of indexes of the goodness of fit of the model [and] permits statistical significance testing of factor loadings and correlations among factors and the computation of confidence intervals.” (p. 277). If the assumption of multivariate normality is “severely violated” they recommend iterated PAF or ULS factoring (Fabrigar et al., 1999; Nunnally & Bernstein, 1994). Other authors have argued that in specialized cases, or for particular applications, other extraction techniques (e.g., alpha extraction) are most appropriate, but the evidence of advantage is slim. In general, ML, iterated PAF, or ULS will give you the best results, depending on whether your data is generally normally distributed or significantly non-normal. In Chapter 2, we will compare outcomes between the various factor extraction techniques.

Step 3: Deciding how many factors to retain for analysis. This, too, is an issue that suffers from anachronistic ideas and software defaults that are not always ideal (or even defensible). In this step, you (or the software) decide how many factors you are going to keep for analysis. The statistical software will always initially extract as many factors as there are variables (i.e., if you have 10 items in a scale, your software will extract 10 factors) in order to account for 100% of the variance. However, most of them will be meaningless. Remembering that the goal of EFA is to *explore* your data and *reduce* the number of variables being dealt with. There are several ways of approaching the decision of how many factors to extract and keep for further analysis. Our guide will always focus on the fact that extracted factors should make conceptual and theoretical sense, and be empirically defensible. We will explore guidelines for this later in Chapter 3.

Step 4: Deciding on a rotation method and rotating the factors. Rotation is often a source of some confusion. What exactly is rotation and what is happening when data is rotated? In brief, the goal is to clarify the factor structure and make the results of your EFA most interpretable. There are several rotation methodologies, falling into two general groups: orthogonal rotations and oblique rotations. Orthogonal rotations keep axes at a 90° angle,

1 Please note, there is one other extraction method, principal components analysis. There are also two additional extraction options to recognize and perform analyses on a set of scoring coefficients or a factor pattern matrix.

forcing the factors to be uncorrelated. Oblique rotations allow angles that are not 90°, thus allowing factors to be correlated if that is optimal for the solution. We argue that in most disciplines constructs tend to be at least marginally correlated with each other, and, as such, we should focus on oblique rotations rather than orthogonal. We will discuss these options in more detail in Chapter 4.

Step 5: Interpreting results. Remember that the goal of exploratory factor analysis is to explore whether your data fits a model that makes sense. Ideally, you have a conceptual or theoretical framework for the analysis—a theory or body of literature guiding the development of an instrument, for example. Even if you do not, the results should be sensible in some way. You should be able to construct a simple narrative describing how each factor, and its constituent variables, makes sense and is easily labeled. It is easy to get EFA to produce results. It is much harder to get sensible results.

Note also that EFA is an *exploratory* technique. As such, it should not be used, as many researchers do, in an attempt to *confirm* hypotheses or test competing models. That is what *confirmatory factor analysis* is for. It is a misapplication of EFA to use it in this way, and we need to be careful to avoid confirmatory language when describing the results of an exploratory factor analysis.

If your results do not make sense, it might be useful to return to an earlier step. Perhaps if you extract a different number of factors, the factors or solution will make sense. This is why it is an exploratory technique.

Step 6: Replication of results. One of the hallmarks of science is *replicability*, or the ability for other individuals, using the same materials or methods, to come to the same conclusions. We have not historically placed much emphasis on replication in the social sciences, but we should. As you will see in subsequent chapters, EFA is a slippery technique, and the results are often not clear. Even clear results often do not replicate exactly, even within an extremely similar data set. Thus, in our mind, this step is critical. If the results of your analysis do not replicate (or do not reflect the true nature of the variables in the "real world"), then why should anyone else care about them? Providing evidence that your factor structure is likely to replicate (either through another EFA or through CFA) makes your findings stronger and more relevant. In Chapter 6, we will explore a "traditional" method of replication[1] (similar to cross validation in regression models). In Chapter 7, we will play with the notion of applying a less traditional but perhaps more useful analysis using bootstrap analysis. Confirmatory factor analysis is outside the scope of this book, but is perhaps an even better method of replication.

Basic Syntax for EFA

The `FACTOR` procedure is used to conduct EFA in SAS. You must be careful when conducting an EFA in SAS because the `FACTOR` procedure is also used to conduct PCA—if you do not specify an extraction method or if you just select one of the extraction options,

1 We put that in quotation marks because most researchers reporting results from an EFA fail to do any replication at all.

you could end up with a PCA and not an EFA! Below is a brief summary of the syntax and the basic options available. We will delve into further details and options in the chapters to come.

```
PROC FACTOR   DATA = dataset-name
         NFACTORS = number-of-factors-to-retain
         METHOD = factor-extraction-method
         ROTATE = rotation-method;
   VAR variables-to-include;
               RUN;
```

In the above syntax, you would specify all of the arguments that are highlighted. You would identify the data set to use for the analysis, tell SAS the number of factors to extract (we will discuss the multi-step procedure for this in Chapter 3), specify one of the seven EFA extraction methods (note there are other options, but they are not EFA extraction methods), specify one of the 25 rotation methods, and list the variables to be factored. If you do not specify the **NFACTORS**, **METHOD**, and **ROTATE** options along with the **VAR** statement, then the analysis will still run, but SAS will use its default options of retaining the number of factors identified by the Kaiser Criterion, performing PCA extraction, not conducting rotation, and using all of the variables in the data set. Thus, it is best practice to get into the habit of specifying all of the options, even if you choose to use one of the methods that is a default, so that you do not accidentally overlook a key component of your analysis.

Let's take a minute to quickly review the structure of the syntax above. Note that there is a semicolon on the end of some lines and not others. The placement of the semicolon is of insurmountable importance in SAS—if it is in the wrong place your code will not run and you will get errors galore! The semicolon signals the end of a *statement* and tells SAS how to parse what we are requesting. In the above syntax, we have a **PROC FACTOR** statement, a **VAR** statement, and a **RUN** statement. You will notice that there are some other key terms in the statement that allow users to specify the inputs for our analysis (e.g., **DATA =**, **NFACTORS =**). These are referred to as *options* within the statement. Finally, notice that the syntax starts with **PROC FACTOR**. This tells SAS that we are using the **FACTOR** procedure, and the following statements and options will specify the details of the analysis we would like to conduct. We will use this terminology throughout this book.

We hope most of what was just described was a review. If it was not, we recommend you first review some other excellent introductory texts to using SAS (e.g., Cody, 2007; Delwiche & Slaughter, 2012) before attempting to follow the syntax in this book. It is expected that readers have an understanding of basic data manipulation, the **DATA** step, and at least a few basic procedures (e.g., **MEANS**, **SORT**, **FREQ**). In addition, readers are encouraged to refer to the SAS documentation about **PROC FACTOR** for additional information throughout this book (SAS Institute Inc., 2015).

Summary

The goal of this book is to be a friendly, practical, applied overview of best practices in EFA, with applications to SAS. In the coming chapters, we will explore various aspects of EFA, and the best practices at each step. We will skip the matrix algebra and equations as much as possible. If you are seeking a more mathematical tome on the subject, there are many good ones already available.

Throughout the book, we describe and explain the SAS procedures for each analysis. We also review key syntax for our example analyses. The complete syntax for all of the analyses conducted throughout the book, the data sets that are used, and the solutions to exercises are available on the author's page at http://support.sas.com/publishing/authors.

References

Bentler, P. M, & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research, 25*(1), 67-74.

Cody, R. P. (2007). *Learning SAS by example: A programmer's guide*. Cary, NC: SAS Institute.

Delwiche, L. D., & Slaughter, S. J. (2012). *The little SAS book : A primer : a programming approach* (5th ed.). Cary, NC: SAS Institute.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272 - 299.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment, 7*(3), 286 - 299.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39(2)*, 291-314.

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

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research, 25*(1), 33-39.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of personality assessment, 68*(3), 532-560.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417-441.

Loehlin, J. C. (1990). Component analysis versus common factor analysis: A case of disputed authorship. *Multivariate Behavioral Research, 25*(1), 29-31.

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin, 109*(3), 502 - 511.

McArdle, J. J. (1990). Principles versus principals of structural factor analyses. *Multivariate Behavioral Research, 25*(1), 81-87.

Mulaik, S. A. (1990). Blurring the distinctions between component analysis and common factor analysis. *Multivariate Behavioral Research, 25*(1), 53-59.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.

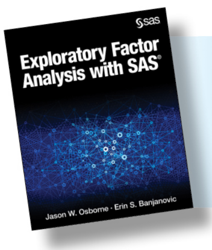
Osborne, J. W. (2013). *Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data*. Thousand Oaks, CA: Sage Publications.

SAS Institute Inc. (2015). *SAS/STAT® 14.1 User's Guide: The FACTOR Procedure*. Cary, NC: SAS Institute Inc.

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology, 15*(2), 201-292.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications: Washington, DC:* American Psychological Association.

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research, 28*(3), 263-311.



Index

Special Characters

%INCLUDE statement 40

A

absolute sample sizes 84
acquiescence 144
alpha
See Cronbach's coefficient alpha
alpha extraction 14
alpha factoring 6
ALPHA option 187, 188
American Psychological Association (APA) 114
analysis, retaining factors for 6
Armanino, C. 64

B

Baker, F.B. 168
Barrett, P. 100
Berge J.M.F. 100
Bernstein, I. 183
Bianchini, J. 100
Blanchard, M.R. 145
Bond, T.G. 147, 168
bootstrap resampling
aligning procedure for factor loadings 123
application of 126
bootstrapped CI 119
EFA and 113
exercises 138
generalizability 131
Heywood cases 131
methods of 115
rise of 114
traditional CI with ML extraction 117
uses for 116
Briggs, S.R. 99
BY statement 121

C

camps, methods within 72
CFA (confirmatory factor analysis) 7
Cheek, J.M. 99
CI
bootstrapped 119
interpretation of 130
Comfrey, A.L. 84
communalities
about 12
factor loadings and 129
confidence intervals, for alpha 196
confirmatory factor analysis (CFA) 7
constant responding, compared with random responding 147
CONVERGE option 13
convergence 13
CORR procedure 187, 188, 199
Costa, P.T. 99
Costello, A.B. 87, 89, 207
COVARIANCE option 12
criteria, for factor extraction 33
criticality 144
Cronbach, L.J. 182
Cronbach's coefficient alpha
about 182
application of 191

confidence intervals for 196
error-free measurement 189
estimating 187
exercises 203
factors influencing 185
good enough 184
sample size and 191

D

data
incomplete 149
missing 149
response sets in 143
unexpected patterns in 143
data cleaning
about 208
application of 152
as a step in EFA 5
exercises 155
importance of 141
imputation 152
incomplete data 149
missing data 149
nonrandom missingness 152
outliers 142
random vs. constant responding 147
DATA option 102, 119
Davison, A.C. 115
direct oblimin rotation 64
dissimulation 144

E

Edelen, M.O. 146
Educational and Psychological Measurement, Journal of Educational Psychology, Personality and Individual Differences, and Psychological Assessment 86
EFA
See exploratory factor analysis (EFA)
eigenvalues
about 13
variance and 132
EM (expectation-maximization) algorithm 151
Embretson, S.E. 168
engineering data 20, 45, 173
equamax rotation 63
error-free measurement 189
examples
extraction 15
instruments for extraction used in 27
sample size 87
exercises
bootstrap resampling 138
Cronbach's coefficient alpha 203
data cleaning 155
extraction 26
extraction criteria 58
factor scores 169
higher-order factors 179
replication 110
rotation 81
sample size 93
expectation-maximization (EM) algorithm 151
exploratory factor analysis (EFA)
about 1
basic syntax for 7
best practices for 207
bootstrap resampling and 113
compared with principal components analysis (PCA) 2
post 181
quantifying replicability in 100
replication and 98
steps for 5
external replication 101
extraction
about 11
concepts of 12
example syntax and output for 16
exercises 26
instruments used in examples 27
methods of 6, 20
pedagogical examples 15
techniques for 13
extraction criteria

about 34
comparing 45
example syntax and output 36
exercises 58

F

Fabrigar, L.R. 6
factor extraction, criteria for 33
factor loading
aligning procedure for 123
communalities and 129
factor matrices, interpretation of 65
FACTOR procedure 7, 12, 13, 25, 51, 63, 64, 65, 66, 120, 121, 122, 132, 152, 176
factor scores
about 159
application of 162
estimating 162
exercises 169
modern alternatives to 168
proper vs. improper 166
stability of 166
syntax and output 160
FACTOR statement 8, 36, 37, 75, 117, 132, 160
factor weighting, methods of 165
factors
See also higher-order factors
retaining for analysis 6
second-order 177
Fan, X. 196
Fitzpatrick, D.C. 101, 102
Ford, J.K. 85
Forina, M. 64
four-factor replication analysis 107
Fox, C.M. 147, 168
FREQ procedure 142, 149
FUZZ option 75, 89

G

GDS data 23
generalizability 131, 208
geriatric depression scale data 51
GLMMIX procedure 146
Gorusch, R.L. 84
Grice, J.W. 168
Guadagnoli, E. 86, 87, 91

H

Harris extraction 14
Harris factoring 6
Hatcher L. 84
Henson, R.K. 86
Heywood cases 131
HEYWOOD option 132
higher-order factors
about 171
application of 177
exercises 179
initial solution and 172
performing second-order factor analysis 173
replication of second-order factors 177
Hinkley, D.V. 115
Hunka, S. 100

I

image extraction 14
image factoring 6
IML procedure 124
improper factor scores 166
imputation 152
incomplete data 149
inter-item correlation 185
internal consistency 181
internal replication 101
IRT procedure 146
item response theory (IRT) 145, 146, 168

iterated principal axis factoring (iterated PAF) extraction 6, 14
iterations 13

J

Jöreskog, K.G. 84, 93

K

Kaiser Criterion 8, 34, 36, 45, 52, 127, 173
Kaiser, H. 100
Kuder Richardson Formula 20 (KR-20) 183

L

Lanteri, S. 64
Leardi, R. 64
Lee, H.B. 84
legitimate missing data 150
Lord, F.M. 98
Luo, S. 124

M

MacCallum, R.C. 6, 85
malingering 144
MAP (minimum average partial) analysis 35, 40
Markov chain Monte Carlo (MCMC) simulation 151
Marsh SDQ data, replication of 102
Marsh's Self-Description Questionnaire 87
matrix of associations 12
maximum likelihood (ML) extraction 6, 14, 117
MAXIT ER option 13
McCrae, R.R. 99
MCMC (Markov chain Monte Carlo) simulation 151
Meier, R.R. 144
METHOD option 8, 15, 20, 102, 119
methods
  of bootstrap resampling 115
  of factor weighting 165
  rotation 69
  within camps 72
MI procedure 152
MIANALYZE procedure 151
minimum average partial (MAP) analysis 35, 40
minimum eigenvalue, as an extraction criteria 34, 36, 45
Minnesota Multiphasic Personality Inventory (MMPI-2) 145
missing data 149, 151, 208
ML (maximum likelihood) extraction 6, 14
MMPI-2 (Minnesota Multiphasic Personality Inventory) 145
Monte Carlo methods 86
multidimensionality, influence on alpha of 186

N

negative item-total correlations 186
Neuhaus, J.O. 100
NFACTORS option 8
NHST (null statistical hypothesis testing) 114
NLMIXED procedure 146
NODUPKEY option 122
nonrandom missingness 152
Novick, M.R. 98
null statistical hypothesis testing (NHST) 114
Nunnally, J.C. 183

O

O'Connor, B.P. 40, 41, 208
oblique rotation, compared with
orthogonal rotation 64, 69
ODS (Output Delivery System) 36
orthogonal rotations 63, 64, 69
Osborne, J.W. 85, 87, 88, 89, 101, 102,
116, 145, 207
OUT option 102, 122, 162, 166
OUTHITS option 119
outliers
influence on alpha of 186
types of 142
output
examples for extraction criteria 36
extraction examples 16
factor scores 160
rotation examples 66
Output Delivery System (ODS) 36
OUTPUT statement 122
OUTSTAT option 162

P

PAF (principal axis factoring) extraction 6,
13
parallel analysis, as an extraction criteria
35, 40, 52
pattern matrix 65, 66
patterns in data 143
PCA
See principal components analysis
(PCA)
PCTLPTS option 122
PLOTS option 45
POWER option 64
Preacher, K.J. 124
PREROTATE option 63
principal axis factoring (PAF) extraction 6,
13
principal components analysis (PCA)
about 2, 208
compared with exploratory factor
analysis (EFA) 2
Prior Communality Estimates 12
PRIOR S = SMC option 15
PRIORS option 20
promax rotation 63, 72
proper factor scores 166
proportion of variance, as an extraction
criteria 35, 36, 45, 52
PROPORTION= option 35

Q

quartimax rotation 63

R

random responding
about 144
commonality of 145
compared with constant responding
147
detection of 145
influence on alpha of 186
Rao-Guttman method 14
Rasch measurement 146, 168
ratios, case for 84
Reeve, B.B. 146
reference structure matrix coefficients 66
Reise, S.P. 168
REORDER option 75, 89, 90
REP option 119
replicability analysis, procedural aspects
of 99
replication
about 209
EFA and 98
exercises 110
importance of 97
of Marsh SDQ data 102
of second-order factors 177
procedural aspects of replicability
analysis 99

quantifying replicability in EFA 100
resampling, rise of 114
response sets
in data 143
influence on alpha of 186
response styles, peculiar to educational testing 144
results
interpreting 7
replicating 7
reverse coded items 186
Roberts, J.K. 86
ROTATE option 8, 64
rotation
about 61
example syntax and output 66
exercises 81
interpretation of factor matrices 65
methods of 69
multiple solutions 75
oblique 64, 69
orthogonal 63, 64, 69
types of 63
rotation method 6
RUN statement 8

S

sample size
about 208
alpha and 191
applied example 87
empirical investigation in 86
exercises 93
impact on interpretation of 89
importance of 83
in practice 85
published guidelines for 84
SAMPRATE option 119
SCORE option 162
SCREE option 36, 38
scree plot, as an extraction criteria 35, 36, 52
second-order factors
analyzing 173, 176
replicating 177
SEED option 102, 119
self-description questionnaire data 21, 46
SEM (structural equation modeling) 168, 190, 203
SGPLOT procedure 194
simple correlation matrix 12
Smith, E.V. 147
Smith, R.M. 147
social desirability 144
Sörbom, D. 84, 93
SORT procedure 122
Spearman, C. 1, 2
Stevens, J.P. 84
Strahan, E.J. 6
structural equation modeling (SEM) 168, 190, 203
structure matrix coefficients 66
SURVEYSELECT procedure 89, 102, 119, 120, 122, 194, 197
syntax
examples for extraction criteria 36
extraction examples 16
factor scores 160
for EFA 7
rotation examples 66

T

TABLE statement 149
Tait, M. 85
Task Force on Statistical Inference 114
techniques, for extraction 13
theory, as an extraction criteria 34, 45, 52
Thompson, B. 62, 63, 66, 98, 179, 196
three-factor replication analysis 103
traditional CI, with ML extraction 117
Tucker, L.R. 100
two-factor replication analysis 105

U

ULS (unweighted least squares)
extraction 6, 14
ULTRAHEYWOOD option 132, 134
UNIVARIATE procedure 122, 126, 127, 199
unrotated factor matrix coefficients 65
unweighted least squares (ULS)
extraction 6, 14

V

VAR statement 8, 122, 152, 187
variable 142
variance, eigenvalues and 132
varimax rotation 63, 72
Velicer, W.F. 35, 86, 87, 91

W

Wegener, D.T. 6
weighted linear combination 3
Widaman, K.F. 4
Wrigley, C. 100

Z

Zhang, G. 124

About This Book

Purpose

This book is designed to present a modern overview of best practices in exploratory factor analysis. We present these concepts from an applied, reader-friendly perspective, guiding you through many different aspects of factor analysis and presenting evidence-based recommendations about how to get the most from your analysis (and how to present it to others clearly).

Is This Book for You?

Factor analysis is used in many disciplines, for many purposes. If you use SAS and also use factor analysis, this book will help you choose options that are defensible and help you meet your goals. If you have never used factor analysis and want to learn, this book is also written for you! We guide you through concepts and their application gently, using minimal technical jargon, and we suggest ways to think about the results that reflect modern practices.

Prerequisites

We wrote this book expecting readers to have a basic working knowledge of SAS and basic knowledge of statistics fundamentals. We do *not* assume you can do matrix algebra in your sleep, that you already know what an eigenvalue is, or that you know the binary language of moisture vaporators. We do, however, assume that most readers understand basic data manipulation in SAS, the DATA step, and at least a few procedures (e.g., MEANS, SORT, FREQ).

Scope of This Book

This book is designed to cover everything you need to know about exploratory factor analysis, and some closely related constructs such as reliability and replicability. Our goal was to stay focused on this one topic, as it is complex and confusing, with many options and steps. We chose, therefore, not to cover related topics like confirmatory factor analysis, construction and validation of scales, latent variable modeling such as structural equation modeling, and other measurement methodologies (for example, item response theory or Rasch measurement).

About the Examples

Software Used to Develop the Book's Content

SAS 9.4 was used for this book.

Example Code and Data

To access the book's example code and data, visit the authors' page at http://support.sas.com/publishing/authors. Select the name of one of the authors. Then, look for the book cover, and select Example Code and Data.

If you are unable to access the code through the website, send email to saspress@sas.com.

This book is compatible with SAS® University Edition.

If you are using SAS® University Edition you can use the code and data sets provided with this book. This helpful link will get you started: http://support.sas.com/publishing/import_ue.data.html.

Output and Graphics Used in This Book

All output in this book was generated using SAS.

Exercise Solutions

Exercises are presented at the end of each chapter. The data sets that are used for these exercises can be found on the SAS Press book website, along with syntax from the chapter that can be helpful in solving the exercises. All data, syntax, and promised reader-oriented materials can also be found on Jason Osborne's website at http://jwosborne.com

Additional Resources

Although this book illustrates many analyses regularly performed in businesses across industries, questions specific to your aims and issues might arise. To fully support you, SAS Institute and SAS Press offer you the following help resources:

- For questions about topics covered in this book, contact the author through SAS Press:
  - Send questions by email to saspress@sas.com; include the book title in your correspondence.
  - Submit feedback on the author's page at http://support.sas.com/author_feedback.
- For questions about topics in or beyond the scope of this book, post queries to the relevant SAS Support Communities at https://communities.sas.com/welcome.
- SAS Institute maintains a comprehensive website with up-to-date information. One page that is particularly useful to both the novice and the seasoned SAS user is the Knowledge Base. Search for relevant notes in the "Samples and SAS Notes" section of the Knowledge Base at http://support.sas.com/resources.
- Registered SAS users or their organizations can access SAS Customer Support at http://support.sas.com. Here you can pose specific questions to SAS Customer Support; under Support, click Submit a Problem. You will need to provide an email address to which replies can be sent, identify your organization, and provide a customer site number or license information. This information can be found in your SAS logs.

Keep in Touch

We look forward to hearing from you. We invite questions, comments, and concerns. If you want to contact us about a specific book, please include the book title in your correspondence.

To Contact the Author through SAS Press

By e-mail: saspress@sas.com

Via the Web: http://support.sas.com/author_feedback

SAS Books

For a complete list of books available through SAS, visit http://support.sas.com/bookstore.

Phone: 1-800-727-3228

Fax: 1-919-677-8166

E-mail: sasbook@sas.com

SAS Book Report

Receive up-to-date information about all new SAS publications via e-mail by subscribing to the SAS Book Report monthly eNewsletter. Visit http://support.sas.com/sbr.

Publish with SAS

SAS is recruiting authors! Are you interested in writing a book? Visit http://support.sas.com/saspress for more information.

About the Authors

Jason W. Osborne, PhD, is an Educational Psychologist and Statistician and is currently Dean of the Graduate School at Clemson University. He is the author of several books and over 70 peer-reviewed articles; he is an Accredited Professional Statistician (American Statistical Association), and his work has been cited over 8,000 times. His passion is dissemination of best practices in statistics and research methods. He is a 3rd degree black belt in Songahm Taekwondo, the father of three, and the rest, as Anne McCaffrey often said, changes without notice.

Erin S. Banjanovic is a Research Scientist at the Human Resources Research Organization (HumRRO), a nonprofit research organization. She works in the Validity Investigations for Education and the Workplace (VIEW) group and specializes in educational evaluation and measurement. She thoroughly enjoys using research, statistics, and programming to address real-world problems. She has served as a lead psychometric analyst on several state and national-level projects for state departments of education and national assessment consortia. She has also assisted with state-wide program evaluations and served as a database manager for a number of multisource databases. She holds a BS in human development, an MA in educational psychology, and she is currently working on her PhD in educational psychology, measurement, and evaluation.