From Data to Information: Using Factor Analysis with Survey Data

Ronald D. Fricker, Jr., Walter W. Kulzy, and Jeffrey A. Appleget, Naval Postgraduate School; rdfricker@nps.edu

Abstract

In irregular warfare, surveys are routinely used to gain insight into population attitudes, perceptions, and beliefs. Understanding these types of population traits can provide insight into the human terrain. However, population attitudes, perceptions, and beliefs often manifest as latent traits that can be only indirectly, incompletely, and sometimes imperfectly measured via single survey questions. Factor analysis is a method for estimating these latent traits from question-level survey data. Because survey analysis in general, and factor analysis in particular, are typically not taught as part of operations research curricula, this paper is intended to provide an introduction to factor analysis for the military operations research analyst.

Introduction

Surveys are routinely used to gain insight into population attitudes, perceptions, and beliefs. In terms of irregular warfare, understanding these types of population traits can provide insight into the human terrain. However, population attitudes, perceptions, and beliefs often manifest as latent traits that can be only indirectly and incompletely measured via single survey questions. Simply said, individual survey questions are often imperfect measures of the population traits of interest and there is frequently a need to distill survey data down into relevant information about the population or populations.

Factor analysis is a method for identifying latent traits from question-level survey data. It is useful in survey analysis whenever the phenomenon of interest is complex and not directly measurable via a single question. In such situations, it is necessary to ask a series of questions about the phenomenon and then appropriately combine the resulting responses into a single measure or "factor." Such factors, then, become the observed measures of the unobservable or latent phenomenon.

Appropriately and correctly applied, factor analysis can be a valuable tool in irregular warfare, allowing analysts to better measure population latent traits, thus turning survey data into useful information. Because survey analysis in general, and factor analysis in particular, are typically not taught as part of operations research curricula, this paper is intended to provide an introduction to factor analysis for the military operations research analyst.

The paper is organized as follows. First, we give a brief overview and some background on factor analysis. Next, we describe the factor analysis model and the necessary steps for fitting the model. We then illustrate the application of factor analysis to actual survey data and present some conclusions about the utility of using factor analysis for survey analysis.

Background

Factor analysis is a hybrid of social and statistical science. First conceived in the early 1900s, the goal was multivariate data reduction, but data reduction of a very specific type. Essentially the idea is to explain the correlation structure observed in p dimensions via a linear combination of r factors, where the number of factors is smaller than the number of observed variables, and where the factors achieve both "statistical simplicity and scientific meaningfulness" (Harman 1976).

Figure 1 illustrates the idea of factor analysis with six observed variables (i.e., survey question responses) that can be effectively summarized in terms of two latent variables (factors). Note that the survey question responses are observed with error (denoted by the ε terms) and the question responses are weighted linear combinations of the factors (where the weights are the λ_{ij} s). What factor analysis does is model the p observed variables as linear combinations of r factors, where the analyst has to prespecify r, such that the model covariance matrix closely matches the sample covariance matrix of the observed variables.

An alternative to factor analysis is principal components, which uses orthogonal transformations to convert a set of possibly correlated variables into a reduced set of uncorrelated variables that capture most of the variation in the original data. The transformation is defined so that the first principal component accounts for as much of the variability in the data as possible, and each succeeding component has the highest variance possible under the constraint that it be orthogonal to the preceding component or components. A principal components analysis, although useful for efficiently summarizing data, does not necessarily result in factors with scientifically meaningful interpretations.

In contrast, factor analysis is specifically designed to look for meaningful commonality in a set of variables (DeCoster 1998). There are two types of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA looks to explore the data to find an acceptable set of factors. In this sense, it is much like exploratory data analysis. The goal is not so much to formally test hypotheses as it is to discover likely factors that will account for at least 50% of the common variation in the observed factors. CFA, on the other hand, begins with a theory or hypothesis about how the factors should be constructed and seeks to test whether the hypothesized structure adequately fits the observed data.

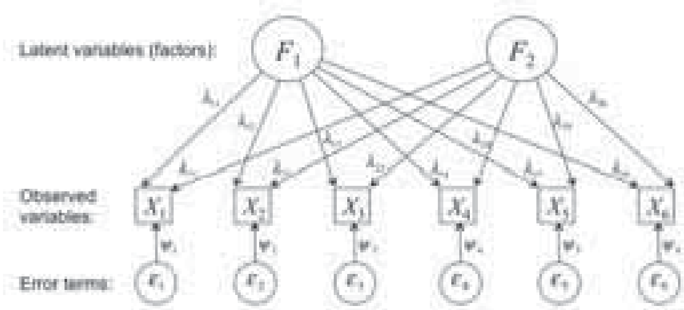

Figure 1. An illustrative example of factor analysis with six observed variables (survey question responses) that can be effectively summarized in terms of two latent variables (factors). Note that the survey question responses are observed with error (the ε terms) and the question responses are weighted linear combinations of the factors (where the weights are the λ_{ij} s).

The literature of factor analysis is burdened by an early history of ad hoc methods compounded by an overemphasis by social scientists on verbal descriptions (vice mathematical derivations) of factor analysis. For those interested in wading through all the historical details, Mulaik gives a thorough verbal and mathematical treatment (Mulaik 2010). For those more interested in a succinct, but statistically rigorous, treatment of the modern application of factor analysis, we recommend Johnson and Wichern (Johnson and Wichern 2002) or, believe it or not, the Wikipedia factor analysis page (http://en.wikipedia.org/wiki/Factor_analysis).

The Factor Analysis Model

Consider a survey consisting of p questions given to n respondents, where respondent i's responses are denoted $\mathbf{y}_i = \{y_{i1}, \ldots, y_{ip}\}$. From the data, a sample covariance matrix $\mathbf{S}$ is calculated in the usual way for the set of centered variables,

$$\mathbf{x}_i \triangleq \{y_{i1} - \bar{y}_1, \ldots, y_{ip} - \bar{y}_p\},$$

where

$$\bar{y}_j = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$$

That is, the (jk)th entry of $\mathbf{S}$ is calculated as

$$s_{jk} = \frac{1}{n-1}\sum_{i=1}^{n} x_{ij}x_{ik}, \quad j \in \{1, 2, \ldots, p\} \text{ and } k \in \{1, 2, \ldots, p\}.$$

The fundamental assumption of factor analysis is that, for some $r < p$, each of the p centered variables ($\mathbf{X} = \{X_1, \ldots, X_p\}$) can be expressed as the sum of r common factors ($\mathbf{F} = \{F_1, \ldots, F_p\}$) multiplied by their loadings ($\lambda_{i1}, \ldots, \lambda_{ir}$) plus a unique factor ($\mathbf{E} = \{\varepsilon_1, \ldots, \varepsilon_p\}$) multiplied by its associated loading ($\psi_1, \ldots, \psi_p$),

$$\begin{aligned}
X_1 &\triangleq Y_1 - \mu_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \cdots + \lambda_{1r}F_r + \psi_1\varepsilon_1 \\
X_2 &\triangleq Y_2 - \mu_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \cdots + \lambda_{2r}F_r + \psi_2\varepsilon_2 \\
&\vdots \\
X_p &\triangleq Y_p - \mu_p = \lambda_{p1}F_1 + \lambda_{p2}F_2 + \ldots + \lambda_{pr}F_r + \psi_p\varepsilon_p
\end{aligned} \quad (1)$$

where $\mu_j = E(Y_j)$. Although the above formulation looks similar in many respects to a series of linear models, note that everything on the right-hand side of the p equations is unobserved. In spite of that, the goal is to estimate the loadings from the data so that the modeled covariance matrix $\mathbf{R}$ is "close to" the observed sample covariance matrix $\mathbf{S}$.

Using matrix notation, Equation 1 can be expressed compactly as

$$\mathbf{X} = \Lambda\mathbf{F} + \Psi\mathbf{E}, \quad (2)$$

where Λ is the matrix of the loadings for the common factors of dimension $p \times r$ and Ψ is a matrix of dimension $p \times p$ with $\psi_1 \ldots \psi_p$ on the diagonal and all off diagonal entries zero. Assuming $E(\mathbf{E}) = \mathbf{0}$, we get to the whole point in fitting the factor analysis model, which is that we can use the estimated common factor loadings $\hat{\Lambda}$ to express the factors in terms of their constituent parts:

$$E(\mathbf{F}) = \hat{\Lambda}^{-1}E(\mathbf{X}). \quad (3)$$

One of the most common uses of exploratory factor analysis is to "determine what sets of items hang together in a questionnaire" (DeCoster 1998). Thus, assuming Equation 1 is an appropriate model, via Equation 2 we can determine which of the survey questions are most related and, as desired, use them to estimate the underlying latent factor for any respondent as a linear combination of their responses to the survey questions. Furthermore, if the scientific meaningfulness goal is achieved, the latent variables will have useful and interpretable meanings that provide additional insight into the characteristics of the populations being studied.

Of course, at this point it should be evident that there will be no unique solution to this problem. There are simply too many degrees of freedom in the problem formulation and, even after some assumptions to make the problem solvable, there will still be an infinite set of solutions. This, along with the fact that the choice of solution is subjective, is one of the frequent criticisms of factor analysis. Nonetheless, as we will show, we have found the results to be quite informative and useful in our survey analyses, and there are ways to minimize the number of subjective modeling choices that must be made.

There are three critical steps in fitting a factor analysis model: (1) determining the number of factors, (2) fitting the model in order to estimate the common factor loadings, and (3) rotating the loadings to find the preferred solution. We discuss each of these in turn.

Determining the Number of Factors

To conduct factor analysis, one must prespecify the number of factors r to fit. In so doing, it is crucial not to underestimate or overestimate the number of factors. If too few factors are chosen then the fitted factors become overloaded with irrelevant variables. On the other hand, with an excessive number factors the variables may be spread out too much over the fitted factors. In either case, the result is likely to be that meaningful factors are never properly revealed (Fabrigar et al. 1999 and Hayton et al. 2004).

This seems like a Catch-22: To determine the correct factors, one must first know how many factors there are. However, over the years a number of solutions have been proposed, one of which we found to work quite well.

One early solution is the Kaiser rule, which stipulates that the number of factors used in the model should equal the number of eigenvalues for the original data matrix that are greater than one. Another solution is to use a Scree plot to graph successive eigenvalues versus the number of factors and then setting r to the number of factors where the plotted line visually levels out (indicating that the remaining factors have little explanatory power).

The difficulty with the Kaiser rule and the Scree plot is they are heuristics. The Kaiser rule was designed to help the analyst of the early- to mid-1900s get "into the ballpark" with respect to an acceptable number of factors, but then the analyst was supposed to further refine the acceptable number of factors through trial and error. The Scree plot is also a heuristic because it allows for subjectivity in interpreting the plotted line where the analyst must visually determine when the line in the Scree plot levels out to determine the number of factors.

An alternative to these methods, which we found to work well,

See ***Factor Analysis*** on page *32*

is parallel analysis; although it only became feasible with the widespread availability of significant computing power. As Ledesma and Valero-Mora say, "This method provides a superior alternative to other techniques that are commonly used for the same purpose, such as the Scree test or the Kaiser's eigenvalue-greater-than-one rule" (Ledesma and Valero-Mora 2007).

Parallel analysis involves the construction of multiple correlation matrices from simulated data, where the average eigenvalues from the simulated correlation matrices are then compared to the eigenvalues from the real data correlation matrix. The idea of parallel analysis is that factors derived from the real data should have larger eigenvalues than equivalent factors derived from repeatedly resampled or simulated data of the same sample size and number of variables. Then r is set to the number of factors in the actual data that are greater than the average of the equivalent simulated data factor eigenvalues (Hayton, Allen, and Scarpello 2004).

Fitting the Model

Given that by definition $E(\mathbf{X}) = \mathbf{0}$, and assuming that the common factors are independent of the unique factors, it is straightforward to show that the covariance matrix for $\mathbf{X}$ from Equation 2 is

$$\mathbf{R} = \Lambda \mathbf{R}_F \Lambda' + \Psi^2, \tag{4}$$

where $\mathbf{R}_F$ is the covariance matrix of the factors (Mulaik 2010, p. 136; Johnson and Wichern 2002, p. 479). Further, assuming that $E(\mathbf{F}) = 0$ and $\mathrm{cov}(\mathbf{X}) = \mathbf{I}$, where the former condition follows because the factors can always be rescaled and the latter because we assume the factors are independent, Equation 4 simples to

$$\mathbf{R} = \Lambda\Lambda' + \Psi^2 \tag{5}$$

Then, from Equation 5, Λ and Ψ are estimated via maximum likelihood.

Note that the maximum likelihood estimators (MLEs) are not analytically derivable and must be solved for numerically using an iterative approach. Under the assumption that $\mathbf{F}$ and $\mathbf{E}$ are jointly normally distributed, the calculations essentially follow the usual estimation methods with an additional uniqueness condition added because of the indeterminacy of the factor analysis model (Johnson and Wichern 2002, pp. 492-498 and supplement 9A). Mulaik also mentions an alternative approach due to Howe that maximizes the determinant $|\Psi^{-1}(\mathbf{S} - \Lambda\Lambda')\Psi^{-1}|$ for a model that makes no distributional assumptions (Howe 1955).

Choosing the Preferred Rotation

Maximum likelihood estimation results in a nonunique solution for how the variables load onto the factors. That is, for any estimated common factor loading matrix $\hat{\Lambda}$ there are infinitely many other matrices that will fit the observed sample covariance matrix $\mathbf{S}$ equally well because

$$\hat{\Lambda}\mathbf{F} = \hat{\Lambda}\mathbf{T}\mathbf{T}^{-1}\mathbf{F} = \Lambda^*\mathbf{F}^* \tag{6}$$

where $\Lambda^* = \hat{\Lambda}\mathbf{T}$ and $\mathbf{F}^* = \mathbf{T}^{-1}\mathbf{F}$ for some transformation matrix $\mathbf{T}$.

Thus, after an initial solution is found, the final step in factor analysis is to rotate the variables to simplify their factor loadings. The rotation process is critical to factor analysis because it allows the analyst to identify the desired factor constructs, usually in terms of a simple structure of substantively interesting variables. However, this procedure is susceptible to criticism because all rotations are mathematically equivalent and thus the final choice is subjective.

There are two main types of rotation: (1) oblique, and (2) orthogonal. Orthogonal rotation is most commonly associated with what is called the "varimax" method, and oblique rotations are most commonly associated with what is called the "promax" method. The distinction between the two rotations is whether the factors are assumed to be correlated or not; orthogonal rotations are uncorrelated while oblique rotations may be correlated.

Kline says the most accepted method for creating factors with simple structure is varimax (Kline 1994). On the other hand, the oblique method is recommended by Costello and Osborne because it can account for both correlated and uncorrelated factors (Costello and Osborne 2005). Habing recommends an orthogonal rotation when using maximum likelihood estimation (Habing 2003).

We used the varimax rotation on our survey data and found it to work well. As defined in Johnson and Wichern (2002, p. 504), the varimax procedure finds an orthogonal transformation matrix $\mathbf{T}$ that maximizes

$$V = \sum_{j=1}^{r}\left[\sum_{i=1}^{p}\tilde{\lambda}_{ij}^4 - \frac{1}{p}\left(\sum_{i=1}^{p}\tilde{\lambda}_{ij}^2\right)^2\right] \tag{7}$$

where

$$\tilde{\lambda}_{ij} = \lambda_{ij} \Big/ \sqrt{\textstyle\sum_{j=1}^{r}\lambda_{ij}^2}$$

Equation 7 is akin to calculating the sum of the variances of the factor loadings across the r factors. Varimax finds the rotation that makes the high loadings as high as possible while simultaneously making the low loadings as low as possible on each factor.

Applying Factor Analysis to Actual Survey Data: An Example

In this section, we illustrate the utility of factor analysis using some actual survey data from three national-level surveys in countries we'll just refer to as "A," "B," and "C." With 140 questions common across all three countries, the survey asked about quality of life, governance, politics, security, social tolerance, international relations, and respondent demographics. Because of question commonality across the three countries, we were able to compare and contrast the factors derived for each of the three countries.

The surveys were fielded in 2010 to 3,770 respondents in Country A, 1,661 respondents in Country B, and 1,481 respondents in Country C. A sample of sufficient size is an important consideration because the sample covariance matrix $\mathbf{S}$ is an estimate of some underlying true covariance matrix Σ. That is, because factor analysis focuses only on the sample covariance matrix, it is important that $\mathbf{S}$ is in fact a good estimate of Σ to ensure the resulting factors represent underlying features of the population and not the noise or other artifacts of the sample.

We fit the factor analysis models using the R statistical package (CRAN 2012), which is available at http://cran.r-project.org. In particular, we used the "factanal" function in the base package to fit the factor analysis model and rotate the loadings to get the final solution. We also used the "fa.parallel" of the R psych package to do the parallel analyses (Revelle 2011).

Prior to fitting the factor analysis models, we first cleaned and coded the data, and then we imputed a small number of missing values to prepare the data. Each of these steps was nontrivial, but the discussion is not included in this article. The most important point to make is that factor analysis can only be done with complete data and thus imputation can be a critical step to complete prior to doing factor analysis. For our data, approximately 6% of the data was missing (due, for example, to respondents refusing or failing to answer one or more questions), but they were spread throughout the data set. Thus, if we had only used complete records, we would have eliminated 60% of the respondents. Imputation allowed us to use all the data and subsequent sensitivity analyses demonstrated that our imputation assumptions had no practical effect on the factor analysis results.

Returning to factor analysis, as discussed earlier, we first used parallel analysis to determine *r*, the number of factors. Figure 2 shows the results from "fa.parallel" for Country A, where the eigenvalues for 27 factors were greater than those from the simulated data (the solid line is greater than the dashed line), so we set *r* = 27. Sensitivity analysis using other values of *r* subsequently confirmed that *r* = 27 was appropriate for Country A.

For Country B we found *r* = 28 to be the appropriate number of factors, and for Country C it was *r* = 25. Thus, factor analysis dramatically reduced the dimensionality of the data from 140 separate questions to 25–28 factors plus some individual questions that did not load onto any of the factors. In terms of the questions that did not load on any of the factors, note that the failure to load does not mean those observed variables are not important; it only means they do not share some underlying commonality with any of the other observed variables.

For the purposes of illustrating our factor analytic results, we focus on four governance-related factors derived from our data: Trust in Policy Makers, Trust in Agencies, Democracy, and Trust in Government and Democracy. These four factors were derived from 12 survey questions and, as shown in Figure 3, they vary in terms of how they are defined across the three countries. In Figure 3, the numbers in the table are the loadings that correspond to each of the questions on the left and the boxes show how the factors are defined in terms of the questions and associated loadings.

In Figure 3 we see that, for these questions, Country A has two factors: Trust in Policy Makers and Trust in Agencies. The former are formed from questions about the country's president and prime minister while the latter is comprised of six questions asking about the respondent's trust in a series of government organizations and agencies. Contrasting Country A with Country B, we see a number of similarities and differences. In particular, Country B's Trust in Policy Makers factor also includes trust in the National Assembly while its Trust in Agencies does not include this question nor the question on trust in political parties. In addition, we see that Country B has a Democracy factor comprised of two questions about democracy while for Country A the same two questions never coalesced into a factor. Finally, for Country C we see that all of these questions formed into one overarching factor about both trust in government (both policymakers and agencies) and democracy.

Note that in Figure 3 the blank cells correspond to questions with small loadings that were subsequently set to zero. In spite of the definition of the factor analysis model in Equation 1, it is common practice when using surveys to have a break point in which small factor loadings are simply set to zero so that the associated variables are completely removed from the factor (Neill 2012). For this research, variables with loadings between 0.4 and –0.4 were removed. The reason is that we subsequently fit regression models where we sought to explain one factor as a function of the others. Zeroing the loadings between 0.4 and –0.4 both resulted in clearly defined factors and questions only being associated with a single factor. The latter point ensured the factors used in the regression (particularly those we used as dependent variables) were not only orthogonal but truly independent of the other factors, at least in terms of question composition.

Summary and Conclusions

One of the major challenges with large surveys is reducing the mass of data into useful information. Another challenge with surveys aimed at understanding the human terrain, particularly when applied to irregular warfare, is that the population characteristics of interest may not be directly measured via single questions. Factor analysis helps address both of these issues.

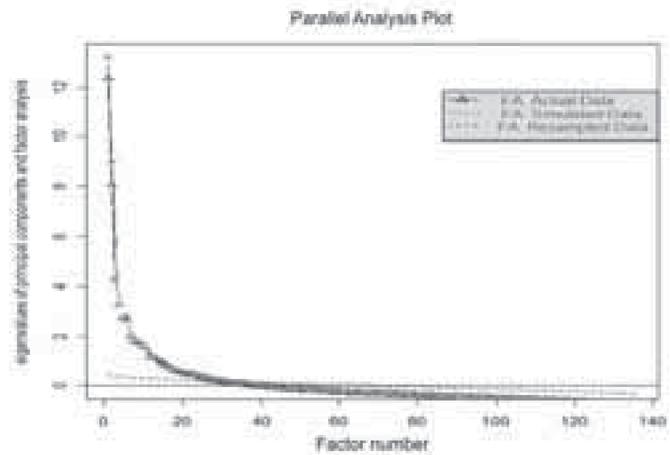

Figure 2. Parallel analysis plot for Country A, where the eigenvalues for 27 factors were greater than those from the simulated data (the solid line is greater than the dashed line), so we set r = 27. Sensitivity analysis subsequently confirmed that r = 27 was appropriate for Country A.

Survey Question		Country A Loadings		Country B Loadings		Country C Loadings
How much trust do you have in President [NAME]?	Trust in Policy Makers	0.80	Trust in Policy Makers	0.91	Trust in Government and Democracy	0.82
To what extent do you approve of [President NAME]'s overall performance?		0.52		0.75		0.66
How much trust do you have in Prime Minister [NAME]?		0.56		0.83		0.80
How much trust do you have in your National Assembly?	Trust in Agencies	0.83		0.64		0.82
How much trust do you have in your political parties?		0.67				0.71
How much trust do you have in your armed forces?		0.60	Trust in Agencies	0.54		0.72
How much trust do you have in your courts and judges?		0.57		0.48		0.77
How much trust do you have in your local police?		0.56		0.64		0.76
How much trust do you have in your local government?		0.63		0.45		0.81
How much trust do you have in your religious leaders?						0.66
How satisfied are you with the way democracy works in your country?			Democracy	0.80		0.50
In your opinion, how stable is democracy in your country?				0.83		0.54

Figure 3. Four factors defined from 12 questions for the three countries. The numbers in the table are the loadings that correspond to each of the questions on the left. The boxes show how the factors are defined in terms of the questions and the associated loadings.

See *Factor Analysis* on page *34*

... **Factor Analysis** *from previous page*

Critics of factor analysis argue that its inherent subjectivity and flexibility allows analysts to manipulate the output. The nonunique solution of the factor loadings is often a particular focus of this criticism. However, all mathematical and statistical models can be manipulated, and most involve making numerous subjective choices (choice of variables, model parameterization, etc.). In this sense, factor analysis is no different. As with those methods, and research in general, it is incumbent on the researcher to ensure his or her results are not sensitive to, or dependent on, modeling choices. That said, remember that the goal of factor analysis is to create factors that are both statistically and substantively meaningful, and the latter implies—perhaps requires—a degree of subjectivity.

Acknowledgements

We wish to express our appreciation to the Office of Naval Research, particularly the Human Social, Cultural and Behavioral Modeling Program, for supporting this research.

References

Costello, A., and Osborne, J. 2005. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment Research and Evaluation*, Vol. 10. No. 7, 1-9.

CRAN. 2012. *The Comprehensive R Archive Network*.

DeCoster, J. 1998. *Overview of Factor Analysis*. Accessed May 5, 2012 at www.stat-help.com/notes.html.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., and Strahan, E.J. 1999. Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods*, Vol. 3, 272-299.

Habing, B. 2003. *Exploratory Factor Analysis*. Tech. report. University of South Carolina. Accessed at www.stat.sc.edu/habing/courses/530efa.pdf on June 10, 2011.

Harman, H. 1976. *Modern Factor Analysis*. 3rd ed. University of Chicago Press.

Hayton, J.C., Allen, D.G., and Scarpello, V. 2004. Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organizational Research Methods*, Vol. 7, 191-205.

Howe, W.G. 1955. *Some Contributions to Factor Analysis*. Tech. report ONRL-1919. Oak Ridge National Laboratory. Accessed at www.ornl.gov/info/reports/1955/3445603609833.pdf on May 14, 2012.

Johnson, R.A., and Wichern, D.W. 2002. *Applied Multivariate Statistical Analysis*. 5th ed. Prentice-Hall.

Kline, P. 1994. *An Easy Guide to Factor Analysis*. Routledge.

Ledesma, R.D., and Valero-Mora, P. 2007. Determining the Number of Factors to Retain in EFA: An Easy-to-Use Computer Program for Carrying Out Parallel Analysis. *Practical Assessment Research & Evaluation*, Vol. 12 (February), 1-11.

Mulaik, S. 2010. *Foundations of Factor Analysis*. 2nd ed. CRC Press.

Neill, J. 2012. "Exploratory Factor Analysis," Survey Research & Design in Psychology, Lecture 5, slide 82. Accessed at http://www.slideshare.net/jtneill/exploratory-factor-analysis on May 19, 2012.

Revelle, W. 2011. *psych: Procedures for Psychological, Psychometric, and Personality Research*. R package version 1.0-98.

Wikipedia. 2012. Factor Analysis. http://en.wikipedia.org/wiki/Factor_analysis.

About the Authors

Ronald D. Fricker, Jr. *is an associate professor of operations research at the Naval Postgraduate School. He holds a PhD and an MS in statistics from Yale University, an MS in operations research from the George Washington University, and a bachelor's degree from the US Naval Academy. Professor Fricker is a fellow of the American Statistical Association and a former chair of the Section on Statistics in Defense and National Security and the Committee on Statisticians in Defense and National Security, both of the American Statistical Association. He is a contributing editor to* Interfaces *and is on the editorial boards of* Statistics, Politics, and Policy *and* International Journal of Quality Engineering and Technology.

Walter W. Kulzy, III *is an active duty Lieutenant Commander in the United States Navy Supply Corps. He holds an MS in operations research from the Naval Postgraduate School, and a bachelor's degree in economics from Virginia Tech. LCDR Kulzy is a certified Lean Six Sigma black-belt and he is a certified DAWIA Level III contracting officer. He is also the recipient of the 2012 MORS/Tisdale Outstanding Thesis Award for his analytical work of survey data.*

Dr. Jeff Appleget *is a senior lecturer of operations research at the Naval Postgraduate School. He holds a PhD in operations research from the Naval Postgraduate School, an MS in operations research and statistics from Rensselaer Polytechnic Institute and a BS from the US Military Academy. He retired from the US Army as an operations research/systems analyst in 2009 after a 30-year career. He teaches the Wargaming Applications, Combat Modeling, and Modeling and Simulation of Societies in Conflict courses at NPS. His research interests include modeling combat, irregular warfare, counterinsurgency, and peacekeeping operations; survey data modeling and analysis, verification, validation, and accreditation (VV&A) of models, data quality control, wargaming, and integer programming.*■