

CHAPTER 1

Applied Problems and Methods of Analysis

A Road Map

Neal Schmitt
Fritz Drasgow

Researchers in the organizational sciences and in the social and behavioral sciences in general encounter a host of problems as they seek to describe, predict, and understand behavior. The procedures we use to measure and analyze behavior are constantly improving as a result of the efforts of psychometricians and applied statisticians, among others. Many of the techniques discussed in this book were not available to researchers even a decade or two ago. Moreover, the sheer number of methods has grown tremendously in the past few decades. Maintaining currency in these areas is a constant challenge for substantive researchers in all areas.

Our goal in assembling the chapters in this book was to provide readable and up-to-date discussions of some of the most important advances in measurement, applied statistics, research methods, and data analysis. Each chapter describes a particular research method or data analysis technique, sets out an illustration of its use, and summarizes the advantages and disadvantages of the method. Many of the methods are quite complex, and the authors have made concerted efforts to present key ideas in ways that are as accessible as possible.

In this chapter, we discuss the various types of problems applied researchers face in describing and understanding the data

they collect and what specific chapters and techniques might be used to understand a set of observations best. At the outset, it is important to note several points. First, readers with introductory graduate-level expertise in data analysis should understand the chapters in this book; however, there is variability in the degree to which they can expect to use the techniques immediately. Some require additional study, but we hope that each chapter provides enough detail so that readers can make a judgment as to whether to invest additional learning time so as to use the technique to resolve a specific question. Second, our bias is that more sophisticated tools should never be used when a simple mean, standard deviation, frequency table, correlation, regression analysis, or plot of the data is sufficient to answer a research question. Having made this statement, we are equally certain that these simple techniques must often be supplemented with more complex multivariate, nonlinear procedures and that these procedures contribute in important ways to our understanding of the complexities of human behavior.

Our sense is that in very general terms, the chapters in this book address two issues. The first concern is for the accurate description of some behavioral phenomenon. The second set of questions relates to the interrelationships among different behaviors. Each of these major questions has different components, and some of the chapters address questions related to both of these research objectives. To help in organizing the set of questions addressed in this book and by organizational scientists, we pose a set of questions or research objectives that address these two major objectives and indicate which chapters may help provide techniques or ideas about how to analyze and present research results.

Description and Measurement of Variables

The chapters in Part Two address issues related to variables and how they are measured.

From Observations to Variables

At a very basic level, researchers begin by observing some behavior that is interesting to them for some practical or theoretical reason. After multiple observations, the researcher attempts to

organize these observations in some parsimonious manner or in a way that attempts to provide an explanatory framework. The basic questions at this point are, What are the variables we are studying? and How can we organize our observations and systematize our measurement? We believe that these questions are addressed most directly in Chapter Two by Karen Locke, which describes one form of qualitative research. She begins with comprehensive verbal descriptions of behavior and describes the development of a sophisticated and detailed measurement system. That system can then be used to develop more systematic and easy-to-administer measurement methods that allow for the collection of many more observations. The procedures she describes represent a way to gain insight into data that many organizational scientists have been unable to describe systematically and have ignored or dismissed as situational vagaries.

Issues related to variables are also considered in the penultimate chapter of this book. Charles Hulin, Andrew Miner, and Steven Seitz describe a method of combining empirical research and theoretical notions and considering the full implications of these observations for situations that have not yet been encountered or observed. Their technique also involves the organization of studies and theoretical statements to describe some behavioral phenomenon better.

Both of these chapters are oriented toward the development of additional theoretical questions about which we can and should collect data. If the objective in analyzing a set of observations is the development of new theory and research questions based on a set of seemingly disjointed or not clearly understood observations, these two chapters may be helpful.

Data Collection

Questionnaires, tests, interviews, observational techniques, and ratings of behavior represent some of the more frequent ways in which organizational scientists collect data; less often used in our discipline are the qualitative techniques that Locke describes. The repertoire of data collection procedures has been radically expanded with the use of the computer technology described in Chapter Three by Julie Olson-Buchanan. The author explores useful ways

to standardize data collection, facilitate data coding and processing, and, perhaps most important, measure variables that we have not previously been able to assess efficiently.

Item Quality

In constructing measures, we usually recognize that any single item is a highly fallible measure of the underlying construct we wish to assess. Consequently, we use multiple items to increase the reliability of our measures and ensure content coverage. An important research question then relates to the manner in which each item constitutes an acceptable indicator of the construct measured. In classical test theory, item difficulty (the percentage getting an item right or the mean of the item) and the correlation of the item with the remainder of the items measuring a construct are the two major indexes examined. In most test or scale construction exercises today, we use item response theory to calibrate items. Items are described in terms of their relationship with an underlying trait continuum using difficulty, discrimination, and guessing parameters. Once items are calibrated (that is, these three parameters have been estimated), they are selected so as to maximize the information available about respondents at a given trait level. Hence, item quality and person trait level are jointly considered in determining the utility or quality of an item. Item response theory formulations for dichotomously scored items are described in Chapter Four by Steven Reise and Niels Waller; in Chapter Five, Michael Zickar describes similar formulations for items with three or more response alternatives.

Test and Scale Quality

An important conclusion from item response theory is that measures are more or less reliable at different points on the ability continuum. Note that this conclusion contrasts with the usual interpretation of classical test theory's standard error of measurement as a single value. The conclusion based on item response theory results from explicitly modeling responses to individual items, and it is an ineluctable result that the precision of measurement varies as a function of the person's trait level (and, of course, the set of items used

to measure the person). These questions are more fully addressed in Chapter Four for dichotomously scored items and Chapter Five for polytomously scored items.

Bias in Measurement

Organizational scientists have long been interested in the degree to which scores on their instruments are affected by variables that are conceptually unrelated to the construct in which they are interested. The outside contaminants are usually demographic in nature and often include gender, race, age, and so forth, but they can also be psychological, such as social desirability. These issues have been examined at both the item level and the test level. An item is usually considered biased when there are group differences on the item after controlling for an individual's standing on the underlying construct. Similarly, tests are considered biased when there are group differences in the expected total test score (the true score in the terminology of classical test theory) after controlling for the trait presumably assessed by the measure. In Chapter Six, Nambury Raju and Barbara Ellis describe methods to study both item and test bias using methods based on item response theory. In this literature, bias is referred to as differential item functioning (DIF) and differential test functioning (DTF).

Item quality issues can also be examined using generalizability theory (see Chapter Seven by Richard DeShon) and confirmatory factor analysis (see Chapter Eight by Charles Lance and Robert Vandenberg). In Chapters Seven and Eight, researchers can examine or model various influences on item types so the research question often shifts to understanding the nature of bias as well as its presence.

Reliability and Error in Measurement

A fundamental concern of measurement for the past century has been the reliability of the data collected. The question addressed here concerns the replicability of observations. Perhaps the first and most frequently mentioned concern is for replicability of measurement or observation across time. However, researchers are often also concerned with other sources of error in measurement

as well—for example, the items used to measure a construct, the people who record observations, and the situation that is the basis for the observation. Interactions among these different sources of error are also possible. A system to evaluate these multiple sources of error has been available for a relatively long time (Cronbach, Gleser, Nanda, & Rajaratnam, 1971), but we have not frequently used generalizability theory to evaluate the multiple potential sources of error that might be present in a set of observations. The power and uses of generalizability theory to inform us about the quality of our measurement efforts are described in Chapter Seven.

Dimensionality

Finally, we include concerns about the dimensionality of measures as a component of measurement and description. Because of the complexity of human behavior, we are almost never interested in a single behavioral index. Organizational scientists are interested in multiple dimensions of ability and personality. At a lower level, we ask questions about the dimensionality of job satisfaction, organizational commitment, mood, and nearly every other construct of interest in the workplace. Conceptually, we can identify multiple aspects of constructs such as commitment and job satisfaction. However, a major research question involves the degree to which we can empirically distinguish among these aspects. Traditionally, we used exploratory factor analyses to examine the dimensionality of measures. Although there were certainly exceptions, the usual approach was to use some arbitrary quantitative criterion (such as a set of eigenvalues or a scree plot) to determine how many dimensions were represented in a set of item responses. We then interpreted these factors by examining the correlations or loadings of each variable on the derived factors. Using confirmatory factor analysis as described in Chapter Eight, we begin with a priori hypotheses about the number and nature of the dimensionality represented in response to a set of items. These hypotheses are then systematically evaluated using tests of significance and indexes of the degree to which a specific factor model fits the covariances between measures.

Dimensionality issues are also considered by Reise and Waller in Chapter Four in the context of item response theory. The assumption that a set of items measures a single underlying latent trait is critical for item response theory; failure to satisfy the uni-

dimensionality assumption adequately renders conclusions based on this method invalid.

Examining the Interrelationships Among Variables and Testing Substantive Hypotheses

The remaining chapters in this book are devoted primarily to methods for examining hypotheses about the interrelationships among the variables measured to test substantive hypotheses. It is certainly true that the chapters and analytic methods already described can be and are used to examine substantive questions and that the techniques described in Part Three often yield important information about the quality of measures.

Evaluations of Theoretical Models

Organizational and behavioral scientists have long used regression and analysis of variance to test models of the relationships between a set of independent or predictor variables and one or more outcome variables. In the past several decades, the literature often suggests models that include multiple predictor-outcome relationships and regression equations. The ability to model and simultaneously evaluate several regression equations, as well as complex mediator and reciprocal relationships, is provided by structural equation modeling. Available since the 1960s, the use of structural equation modeling has become widespread in the organizational sciences only in the past fifteen years. In Chapter Nine, Roger Millsap provides a useful descriptive summary and an example of the use of structural equation modeling. This analytic technique can be used to examine simultaneously the relative importance of multiple theoretical explanations of behavioral phenomena. For example, one can compute the size of various direct and indirect effects on some outcome variable given a particular theoretical model.

Modeling Change

When studying time-related phenomena, we are often tempted to compute respondents' change scores. A researcher who has observations across at least three points in time can examine a whole

series of interesting questions about growth or change. A methodology that has evolved from structural equations modeling, latent growth models allow researchers to examine change while avoiding the unreliability of difference scores. As described by David Chan in Chapter Ten, latent growth modeling involves the computation of regression parameters (that is, slopes and intercepts) for the individuals in a study. One then can use latent growth modeling to ask whether there are significant individual differences in slopes and intercepts and whether slopes and intercepts correlate. In addition, one can examine hypotheses about the correlates of the regression parameters and the shape of the change curve, and examine the similarity of the change process and its correlates across multiple groups. Latent growth modeling, which became available only in the 1990s, represents a highly useful way to examine important research questions that cannot be addressed using more familiar data analytic techniques. Since the method involves the analysis of parameters for latent growth constructs and can explicitly model the measurement process, relationships among latent variables are corrected for unreliability in change indexes.

Concern with time and when or if some discrete event occurs are addressed in the set of analytic techniques discussed in Chapter Thirteen by David Harrison.

Differences as Constructs

Change involves a difference between a single variable at two points in time; *fit* is ordinarily assessed as a difference between two variables at a single point in time. Fit (for example, the match between an individual's ability and the requirements of a job) has long been a concern of psychologists interested in individual differences. This concern has been extended to questions of common values or interests and the degree of fit between a person and the organization (Kristof, 1996). The level of congruence is thought to have important implications for a variety of outcome variables, including performance, organizational commitment, satisfaction, and stress.

Psychometricians have long pointed to the fact that, similar to change scores, measures of fit that are computed by subtracting one measure from another often have very low reliability, which makes these measures less than useful in examining substantive

questions. In Chapter Eleven, Jeffrey Edwards examines the assumptions underlying the computation of various measures of fit and reformulates data analyses questions about fit in terms of polynomial regression. He also uses graphical displays and response surfaces methods to explain the nature of fit hypotheses.

Levels of Analyses

In the past decade or so, psychologists who usually focus on individual determinants of behavior have come to realize that individual behavior is embedded in teams or groups, which are parts of an organization. These organizations, in turn, are units in a larger society or culture. This embeddedness makes multiple theoretical and analytical demands on the organizational scientist (Klein & Kozlowski, 2000). One popular analytic method by which these multilevel hypotheses or concerns can be addressed is hierarchical linear modeling. In Chapter Twelve, Paul Bliese provides a very readable exposition of the reasons that levels of analysis issues are important analytically. He then proceeds to use examples to show how to evaluate hypotheses about relationships among variables at a single level (including the appropriateness of doing so) as well as cross levels of analysis.

Discrete Outcome Variables

Occasionally, organizational researchers are interested in the occurrence or nonoccurrence of some variable, such as employees' leaving an organization or accidents or job acceptance. Occasionally, also, several potential outcomes that are not ordered are considered a variable, such as job choices. A whole family of techniques has been developed to handle these special data analytic problems, and some organizational researchers are beginning to use these methods to help them understand data. Harrison in Chapter Thirteen begins by showing why and when these alternative methods should be used. He also provides a variety of examples of their use in addressing various questions. Some of these situations involve individuals whose data are censored, that is, observations of these cases begins and ends at different times. In these situations, the question that is addressed is when or if some

event occurs. Harrison describes nonlinear methods (for example, Cox regression) that are applicable in these instances.

Computational Modeling

When empirical data are lacking or difficult to obtain, researchers may test various theoretical propositions using computational modeling. In Chapter Fourteen, Hulin, Miner, and Seitz describe situations in which traditional correlational and experimental research designs cannot be implemented or will not be helpful. In those instances, researchers who are willing to use existing data and various theoretical propositions can go well beyond currently available data to develop new knowledge about the potential interrelationships of variables. Interestingly, Hulin, Miner, and Seitz note that most theories are incomplete and underspecified; attempts to write computer programs to simulate the behavior described by the theories reveal many ambiguities and uncertainties about what should take place. Researchers can also use computational modeling to develop new hypotheses about these interrelationships, which can then be tested in subsequent empirical studies. Chapter Fourteen contains several examples in which the use of this novel technique provided information unavailable using any more traditional method of research.

Cumulating Data Across Studies

On some organizational questions, researchers have conducted dozens or even hundreds of studies, and very often, if not always, the results of these studies appear to differ. In the past twenty-five years, researchers have developed several different methods of meta-analysis that involve the analysis of data collected in different studies by different researchers. These methods are designed to estimate an overall effect size across studies to quantify the magnitude of a relationship between two or more variables. In addition to the effect size itself, many meta-analysis methods estimate the variance of the effect size in order to quantify the variability of the relationship due to moderators. The most common and least interesting reason that a measure of effect size may vary across studies is sampling error; often other methodological artifacts, such as variabil-

ity in range restriction, also affect the magnitude of relation. When variance in effect size remains after sampling error and methodological artifacts have been accounted for, researchers turn to the consideration of substantively interesting moderator variables.

In the final chapter of this volume, Hannah Rothstein, Michael McDaniel, and Michael Borenstein provide a guide to different methods of meta-analyses and examples of the use of meta-analyses to represent the cumulative results of large bodies of data. The burgeoning use of meta-analysis in literature reviews demonstrates that it provides a significant advance over the more qualitative summaries previously used, in which reviewers often mistakenly concluded that a body of literature produced conflicting results, when in fact the variability in results could be easily explained on the basis of limitations of individual studies.

Conclusion

We believe that the methods described in this book help address questions that either cannot be addressed using simpler and more familiar techniques or that represent significant advances over the more traditional methods. We hope that this book will help researchers recognize the value of the methods described, educate themselves concerning the applicability and use of these techniques in their own research, and begin to use the methods to address important research questions. When this happens, we believe that we will have advanced the frontiers of organizational science.

References

- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley & Sons.
- Klein, K. J., & Kozlowski, S.W.J. (Eds.). (2000). *Multilevel theory, research, and methods in organizations*. San Francisco: Jossey-Bass.
- Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology, 49*, 1–49.

