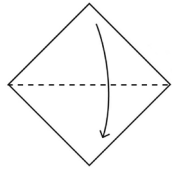# ANALYSING QUANTITATIVE DATA

for BUSINESS *and* MANAGEMENT STUDENTS

1

INTRODUCTION

Quantitative data and analysis have become an integral part of research and business (*Economist*, 2010; Salsburg, 2002; Siegel, 2013). Many scholars and observers have claimed that we are now living in the age of 'big data' and quantitative analysis (e.g. McAfee and Brynjolfsson, 2012) and that the nature of research and business are transforming as a result (e.g. Ayers, 2008; Davenport and Harris, 2007; Siegel, 2013). In fact, it has become impossible to navigate today's academic literature or business media without encountering at least some mention of 'big data' and quantitative analysis. News stories about how quantitative data analysis is being used to fundamentally change some aspect of research, business or life seem to appear daily. Books about the successful use of quantitative analysis, such as Michael Lewis's (2004) *Moneyball*, have become 'required' reading in many areas of business.

The excitement currently being generated by quantitative data analysis is easy to understand, given the impact that quantitative data and analysis have made on a number of fields ranging from professional sports, to medicine, to internet search, to management, to operations and to marketing. The benefit of this renewed interest is that new methods, software, books, websites, conferences and information are proliferating at an astounding rate. For those interested in quantitative data and analysis, these are exciting times.

While exciting, the times are changing for all researchers and practitioners. It appears that we have reached a tipping point where knowledge of quantitative data analysis is no longer optional for most areas of research and many areas of business. We question whether one can be successful in research or industry without some

competence in understanding, using and interpreting quantitative data analysis. Countless reports and surveys point to the conclusion that there is a great demand for individuals with knowledge and skill in quantitative data analysis, but these individuals are in short supply. Some have gone as far as claiming that jobs in quantitative analysis will become one of the most desirable and prestigious careers in the 21st century (Davenport and Patil, 2012).

Even for those that are knowledgeable and skilled in quantitative data analysis, the development and proliferation of tools and methods have made the decision about what one should do an exceedingly challenging one. We find that the most common question asked by experienced researchers and practitioners is, 'How does one determine which analyses are the most appropriate for the data and research method?' Ultimately, the choice of the appropriate quantitative analysis will be based on the purpose of the research (e.g. exploratory or confirmatory), the research questions being asked (and the associated hypotheses to be tested), the research design (e.g. experimental or non-experimental) and the nature of the data (e.g. scale of numerical measurement, shape of the distribution; Scherbaum, 2005). Yet, making these decisions can still be exceedingly difficult because of the unwieldy array of information and options that must be considered in choosing quantitative analyses.

The challenge, especially for students, also stems from the difficulty associated with developing a framework or mental model for understanding quantitative analyses. Quantitative analyses are not simply a set of techniques to be applied to data. Rather, they are a way of systematically thinking about research questions, research methodology and the observed patterns in the data. One could even consider quantitative analysis a language that allows researchers to communicate using an agreed upon vocabulary. It is this mental model of quantitative analysis that serves as an invaluable aid in making choices about appropriate analyses, research methodology and research questions.

The aim of this book is to assist students, researchers and practitioners in navigating the complicated world of quantitative data analysis. Our goal is to provide a foundation for developing one's mental model for quantitative analysis, as well as gaining an understanding of the various methods of quantitative analysis that are currently available. We cover the philosophical and theoretical foundations of quantitative analysis, the interconnections between data collection methods and quantitative analysis, how the nature of the data collected impacts on quantitative analysis and the steps involved with preparing data for quantitative analysis. This book covers many of the most common quantitative analyses available to researchers and practitioners in answering research or business questions. Subsequent books in this series will address many of the less common and advanced methods (e.g. factor analysis, conjoint analysis) that are not covered here.

It has been our observation that books on quantitative analysis rarely focus on the more conceptual aspects of thinking about and planning quantitative analyses. They also do not focus on the link between quantitative analyses and research design and methods. In crafting this book, we attempt to address this gap. We have striven to

provide guidance on the factors that one needs to consider, the steps that one should take and the decisions that one must make when engaging in quantitative analysis. We intentionally avoid providing an inflexible rule-bound system that leads to the 'one and only' quantitative analysis that is appropriate for a given situation. As Abelson (1995) argues, there is rarely only one correct choice or rule. However, taking such an approach does involve a trade-off. The trade-off is that we focus less on providing detailed descriptions of every aspect of every quantitative analysis and the procedures for running them in statistical software programs. Given the wide availability of exceptional resources covering every single detail of quantitative analyses and using the statistical software (e.g. Cohen et al., 2003; Field, 2013; Pedhazur and Pedhazur-Schmelkin, 1991), we felt that it was a trade-off that will benefit those just starting out on the journey of using quantitative analyses.

WHAT IS QUANTITATIVE ANALYSIS?

Quantitative analysis is statistics. It is procedures and rules used to reduce large amounts of data into more manageable forms that allow one to draw conclusions and insights about patterns in the data. Although there are many types of quantitative analyses, this book focuses on common methods that are used to describe quantitative data, to identify differences between groups, to examine associations or relationships between variables, or to make predictions. The most basic form of quantitative analysis is the descriptive quantitative analysis. Descriptive quantitative analyses can be used to condense large amounts of data into a smaller set of numbers representing what is typical in the data and the amount of variability in the data. As described in Part I of Chapter 4, descriptive quantitative analyses include the frequency, mode, median and mean (type of average) as well as the range, variance and standard deviation. Descriptive quantitative analyses can be represented in both numerical and graphical forms (e.g. bar charts and histograms). Regardless of the research question or research method, descriptive quantitative analyses should always be one of the first steps in the process of analysing data.

Related to descriptive analyses are quantitative analyses that examine differences between groups on an outcome or set of outcomes. For example, these analyses might be used to examine differences in customer engagement of those who receive an individualized promotion versus those who receive a generic promotion. As described in Part II of Chapter 4, these analyses typically examine differences in means between groups and include *t*-tests and analysis of variance.

Other quantitative analyses focus on the direction and strength of relationships between variables, and, as an extension of this, aim to predict one variable based on another. These analyses are correlation and regression methods (see Part III of Chapter 4). Correlation is often a first step in understanding whether variables

are related. Regression analyses extend correlations by allowing one to create an equation that can be used to predict or forecast an outcome based on a set of inputs. For example, equations could be developed that allow one to predict future employee sales performance from the current investment in sales training.

HISTORICAL FOUNDATIONS OF QUANTITATIVE ANALYSIS AND CURRENT TRENDS

Although much of the current coverage and discussion of quantitative data analysis treats it as if it is a revolutionary breakthrough that has just occurred, most of the commonly used quantitative analyses were developed over 100 years ago. Although a detailed review of the history of the development of quantitative analysis is beyond the scope of this book (see Lehman, 2011 or Salsburg, 2002 for detailed historical reviews), it is important to consider some challenges in research that led to the initial development of many of the analyses discussed in this book and that still continue to influence quantitative analysis today.

Much of the research in the business and managerial sciences is focused on questions related to large groups. For example, researchers might be interested in the effectiveness of training sales professionals in persuasion techniques for increasing sales productivity. Researchers implement the training at one company with a small group of sales professionals, collect data on the completion of the training and sales productivity, and analyse the data. Ultimately, the researchers are not interested in the specific sales professionals in their study. They are interested in all sales professionals. The question is how one can generalize their findings from the specific group of sales professionals that were part of their study to the much larger group that represents all sales professionals. More specifically, the question is how well a statistic computed from a sample (e.g. an average in the sample) is an adequate estimate of that same statistic in the population (i.e. a population parameter). This question is one that early statisticians developed methods to address. That is, are generalizations about a population justified if analyses are only based on a sample of that population? Populations represent the entire group of interest for a specific research question (e.g. the population of England). Samples represent a subset of the population (e.g. those willing to complete a survey while walking through Trafalgar Square on a given day).

The process of collecting a sample from the population involves a degree of uncertainty. The uncertainty is that a researcher does not know if his or her sample is an adequate representation of the population. It is no surprise that any given sample will not be a perfect reflection of the population. There is always some degree of error when using a sample statistic as an estimate of a population parameter, aptly named sampling error. This sampling error can lead to fluctuations in results between research studies on the same topic. The major concern with these fluctuations is

that a researcher may collect data from a sample that is not truly reflective of the broad population, and then the inferences that are drawn from this sample about the population can be erroneous. One of the most well-known examples of sampling error gone awry was the 1948 presidential election in the USA. After polling a group of voters that was not representative of the voting population, the *Chicago Tribune* newspaper predicted that Thomas Dewey had defeated Harry Truman and published a headline to that effect. However, Harry Truman had a decisive victory, causing quite an embarrassment for the newspaper. Quantitative analyses were developed to assist in determining whether inferences from a sample to a population are indeed merited, especially when the sample represents a very small percentage of the population.

Practical problems in drawing inferences from a sample of data sparked much of the development of quantitative analysis. However, the current rise in the use of quantitative analysis has been fuelled by both the availability of large amounts of data from different sources that can be integrated (i.e. 'big data') and the development of quantitative analysis software (McAfee and Brynjolfsson, 2012; Mayer-Schönberger and Cukier, 2013). It was once the case that data were difficult to obtain, they were small in quantity and needed to be examined in isolation. With technological advances, large volumes of data are now available from multiple sources and can be integrated. For example, it is now possible to integrate data from point of sale transactions, customer satisfaction surveys, web traffic, customer service training completed, employee engagement, and compensation for every single sales transaction and every single employee in an organization.

Box 1.1   Big data in action

Most large retailers are currently using 'big data' from their point of sales transactions, web traffic, customer surveys and responses to marketing campaigns to tailor coupons and advertisements to customers to help increase purchase behaviours. One North American retailer used its large database of purchase history across all of its customers and its database of customers who signed up for the baby registry to predict which female customers were likely to be starting their second trimester of pregnancy. One can imagine the customers' surprise when brochures for newborn products arrived in the mail before they had even announced the pregnancy to friends and family! Although the use of data in this manner may raise ethical concerns, it is this capacity to integrate and leverage the large amount of data created by normal business operations that has exponentially increased the widespread adoption of quantitative analysis.

The availability of statistical software has substantially impacted on the rapid proliferation of quantitative analysis, particularly in business. Executing quantitative

analyses was once the sole province of highly skilled experts. The software required specialized knowledge to make choices among all of the available analyses and the multitude of options that exist within each analysis. Contemporarily, a considerable amount of new software has been developed for use by non-experts. For example, IBM's SPSS Modeler offers automatic selections among the options for a particular analysis and analyses are conducted using visual schematics in which the user drags a node representing a particular class of analysis (e.g. regression) and connects it to a node representing the data. The software then does the rest. The user does not need to make any other decisions. All that is needed is the software, the data and an idea of the general type of analysis required. As will be argued throughout this book, the blind and uncritical use of quantitative analysis is not recommended and can contribute to the misuse of quantitative analysis, as well as have a negative impact of progress on a field's theories.

KEY CONCEPTS IN QUANTITATIVE ANALYSIS

Regardless of the specific quantitative analysis selected, there are a number of key concepts with which all users of quantitative analyses need to be familiar with. One of the most fundamental concepts in quantitative research and analysis is measurement (Scherbaum and Meade, 2009). Measurement can be conceptualized as the assignment of numbers to properties or attributes of objects, events or people according to a set of rules (Stevens, 1968). It is these properties or attributes of objects, events or people that serve as the variables in quantitative analyses. In much of management research, these variables are used as indirect indicators of unobservable constructs. Constructs are abstractions developed to explain differences, commonalities or patterns in the properties or attributes of objects, events or people. For example, personality is a well-known construct that has been used in the managerial and organizational sciences. Personality is an abstraction developed by psychologists to examine differences between people in their behavioural tendencies and interactions with the environment.

Constructs form the basis of many theories in business and the management sciences. These theories, in turn, are the source of hypotheses that quantitative analyses are designed to test. Hypotheses are testable statements about the anticipated relationships or differences between variables in the population. Hypotheses are tested using data collected from a sample drawn from the relevant population. The results of a quantitative analysis are used to determine whether the data collected from the sample support inferences about the population. A point that will be made throughout this book is the importance of theory and hypotheses as part of the use of quantitative analyses. Although quantitative analyses can be and often are used without theory or hypotheses, we strongly discourage this practice. There are fields

such as internet search analytics, market research or crime prevention where the only research questions of interest are 'what happens?' or 'do customers who buy one particular product also buy another particular product?' For these descriptive uses of quantitative analysis, theory may not be necessary. However, in much of business and management research, the research questions or hypotheses of interest focus on why some effect or phenomenon happens. These questions cannot be adequately answered without theory guiding the analyses.

As will be elaborated on in Chapter 2, a foundational concept in quantitative analysis is probability. Probability can be described as the likelihood of a particular outcome. Probabilities range from 0.00 (no chance of an outcome occurring) to 1.00 (the outcome is certain to occur). Probabilities serve as the basis for all quantitative analyses that seek to generalize from a sample to a population. Whether or not a hypothesis is considered to be supported rests on the probability associated with the observed result from a quantitative analysis. In other words, probability is one of the primary criteria used to determine whether inferences about the population are supported by the data collected from a sample.

UNDERSTANDING THE INTERPLAY BETWEEN QUANTITATIVE ANALYSIS AND RESEARCH METHODS

As elaborated on in Chapter 3, there are fundamental relationships between research methods, research design and quantitative analysis. The methods and design used in a research study impact on the nature of the conclusions drawn from quantitative analyses, the type of analysis that may be appropriate, the amount of data that is required and the likelihood that the assumptions of a particular quantitative analysis will be met. For example, non-experimental research designs will never support strong conclusions about causality, regardless of the quantitative analysis used. Likewise, if the data collected represent categories (e.g. male/female; yes/no), only a limited number of quantitative analyses is appropriate. Users of quantitative analyses need to be aware of the close relationship between research methods, research design and quantitative analysis. With that said, we do not advocate designing research to support the use of particular quantitative analyses. Theory and the research questions should determine the research design and method which in turn determine the best quantitative analysis.

Research method and design are also related to quantitative analysis in some less obvious ways. Traditions have developed in many scientific fields about the 'right' analysis for certain types of research designs. For example, quantitative data from experimental research study are more frequently analysed using analysis of variance and *t*-tests (see Part II of Chapter 4), while non-experimental data are more frequently analysed using correlation and regression (see Part III of Chapter 4). Although certain

analyses may be more insightful for the types of research questions that are typically studied with a certain research design, there are no explicit rules stating that particular analyses are the most appropriate for particular research designs. We strongly encourage users of quantitative analyses to consider all of the analysis that may be possible in a given context, regardless of whether or not the analysis is what is typical for the selected research design.

SUMMARY

The interest in quantitative analysis has dramatically increased in recent years. Most fields of research and practice have intentionally or unintentionally made understanding, using and interpreting quantitative data analysis a core area of competence. Despite the recent surge in interest, quantitative analysis is by no means new. Many common quantitative analyses were developed long ago to deal with problems such as the uncertainty inherent in drawing inferences from a sample to a population. In this book, we describe the basis of quantitative analysis, many of the basic quantitative analyses available, how to select quantitative analyses and the major considerations in using quantitative analyses. More specifically, this book will help those using quantitative analyses to understand the philosophical underpinnings, the connections to research design and methods, the connections to the properties of the data and the major strengths and limitations of quantitative analyses. It will also provide step-by-step guidance on how to execute the primary quantitative analyses used with experimental and non-experimental research, as well as alternative analyses that can be used when the assumptions of the most common analyses are not met. Each analysis presented will include an example that the reader can use to practise manually computing the analysis using the formulas in the text, as well as computing them in Excel. All formulas for the analyses in Excel will be presented next to the results in the examples and in the Appendix. As we will emphasize throughout this book, quantitative analysis can be a powerful and very useful tool for those conducting research. However, its use requires careful thought and theory or conceptual frameworks of the problems we study. The blind use of quantitative analyses can never lead to true insight into the research questions we ask.