



Research Methods 3

The Quantitative Researcher

Jesus Canduela

Robert Raeside

This course text is part of the learning content for this Edinburgh Business School course.

Full terms and conditions that apply to students on any of the Edinburgh Business School courses are available on the website www.hw.ac.uk/ebs, and should have been notified to you by Edinburgh Business School, Heriot-Watt University or by the centre or regional partner through whom you purchased your course. If this is not the case, please contact Edinburgh Business School at the address below:

Edinburgh Business School
Heriot-Watt University
Edinburgh
EH14 4AS
United Kingdom

Tel + 44 (0) 131 451 3090

Fax + 44 (0) 131 451 3002

Email enquiries@ebs.hw.ac.uk

Website www.hw.ac.uk/ebs

Research Methods 3:

The Quantitative Researcher

Dr Canduela is the Head of DBA (Doctor of Business Administration) Programme at **Edinburgh Business School, Heriot Watt University**.

Jesus has enjoyed a career in the professional sector as a statistician and consultant, in addition to a career in academia.

He is an expert in online questionnaires having carried out a large number of them in relation to research and consultancy projects. He has years of experience teaching research methods and supervising students at undergraduate and postgraduate levels.

Professor Raeside works in the **Edinburgh Business School of Heriot-Watt University** providing support to Doctorate of Business Administration students. He is a statistician by background and is a Chartered Statistician with the Royal Statistical Society.

Robert's research interests include applications of statistics to business focusing on employment, quality improvement and social networks. In addition to publishing and supervision of research students he has also provided consultancy services to private and public sector businesses.

First Published in Great Britain in 2020.

© Heriot-Watt University 2020

The rights of Jesus Canduela and Robert Raeside to be identified as Authors of this Work have been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved; students who are studying this course may print and download these materials for their own private study only and not for commercial use. Except for this permitted use, no materials may in whole or in part be used, copied (by any electronic, mechanical or other means, including photocopying, printing, downloading, scanning or translating), shared, lent, hired or resold in any way, without the prior consent of Heriot-Watt University. Applications for permission to use the materials should be made by contacting enquiries@ebs.hw.ac.uk.

Contents

Chapter 1	Introduction to Quantitative Research	1/1
	1.1 Introduction	1/1
	1.2 The Methodology of Positivism	1/2
	1.3 What is Statistics?	1/2
	1.4 Statistical Thinking	1/6
	1.5 Book Structure	1/8
Chapter 2	The Basics: SPSS and Descriptive Statistics	2/1
	2.1 Introduction	2/1
	2.2 Data	2/1
	2.3 Descriptive Summaries	2/7
	2.4 Further Commentary on Tables	2/18
	2.5 Graphical Summaries	2/25
	2.6 Simple Rules for Detecting Outliers	2/41
	2.7 Normal Distribution	2/42
	2.8 Conclusions	2/51
Chapter 3	Data Collection	3/1
	3.1 Introduction	3/1
	3.2 Observation and Automated Data Collection	3/1
	3.3 Secondary Data	3/2
	3.4 Content Analysis	3/4
	3.5 Experimentation	3/14
	3.6 Surveys	3/17
	3.7 Sampling	3/18
	3.8 Complex Sample Design	3/27
	3.9 Administration	3/28
	3/10 Conclusions	3/29

Chapter 4	Questionnaires	4/1
	4.1 Introduction	4/1
	4.2 Questionnaire Design	4/1
	4.3 The Questionnaire	4/3
	4.4 Piloting	4/14
	4.5 Improving Response rates and Minimising Problems	4/16
	4.6 Administration	4/17
	4.7 Coding and Data Input	4/18
	4.8 Weighting	4/19
	4.9 Conclusion	4/19
Chapter 5	Data Ethics	5/1
	5.1 Introduction	5/1
	5.2 General Points	5/1
	5.3 Safe Data Processing	5/5
	5.4 Reporting Data	5/6
	5.5 Conclusion	5/12
Chapter 6	Assessing the Statistical Significance of Differences	6/1
	6.1 Introduction	6/1
	6.2 Statistical Significance	6/2
	6.3 Confidence Interval for the Mean	6/4
	6.4 Testing the Statistical Significance of Differences	6/10
	6.5 Non-parametric Statistics	6/34
	6.6 Conclusion	6/53
Chapter 7	Association and Correlations	7/1
	7.1 Introduction	7/1
	7.2 Correlation between Numerical Data	7/3
	7.3 Correlation between Ranked Data	7/15
	7.4 Association between Categorical Variables	7/19
	7.5 Conclusions	7/27

Chapter 8	Regression and Statistical Modelling	8/1
	8.1 Introduction	8/1
	8.2 Simple Linear Regression	8/1
	8.3 Multiple Regression	8/16
	8.4 Regression with Categorical Independent Variables	8/30
	8.5 Statistical Modelling	8/32
	8.6 Binary Logistic Regression	8/37
	8.7 Conclusion	8/51
Chapter 9	Data Reduction	9/1
	9.1 Introduction	9/1
	9.2 Principal Component Analysis	9/2
	9.3 Factor Analysis	9/22
	9.4 Other Approaches to Data Reduction	9/28
	9.5 Conclusions	9/28
Chapter 10	Clustering and Segmentation	10/1
	10.1 Introduction	10/1
	10.2 Hierarchical Methods	10/2
	10.3 Partitioning Methods	10/13
	10.4 Two-Step Clustering	10/22
	10.5 Conclusion	10/34
Chapter 11	Concluding Remarks	11/1
	11.1 Statistical Philosophy	11/1
	11.2 The Importance of Communication	11/2
	11.3 Other Methods	11/3
	11.4 Final Remarks	11/5

Chapter I

Introduction to Quantitative Research

Contents

1.1 Introduction	1/1
1.2 The Methodology of Positivism	1/2
1.3 What is Statistics?	1/2
1.4 Statistical Thinking	1/6
1.5 Book Structure	1/8

1.1 Introduction

This book has been written to support doctoral students and show how quantitative methods can be applied in research. There are a huge range of quantitative approaches that are useful when undertaking research, from a variety of disciplines such as Mathematics, Operational Research, Statistics and the rapidly growing discipline of Data Science. We cannot cover them all in this type of instructional book. Instead, the focus is to give a grounding in applied statistics. We believe that applied statistics contains the methods which will satisfy the needs of most social science researchers. Even with this focus, statistics is a huge field and we take here an applied approach, using mainly classical and frequency-based methods. Following this, we believe, will equip you with a statistical philosophy and a toolkit to address your research objectives and answer your research questions. Consequently, we have not touched on mathematical statistics, probability-based methods and Bayesian approaches. It is not that these methods are not relevant – they have allowed great insights to be made – but space here is limited, so we think it is best to cover the subset of quantitative methods in the kind of depth which will allow you to progress your research. We hope this textbook will help you understand the foundations behind quantitative methods. From that basis of knowledge it is much easier to investigate and understand other techniques that are not covered here. It is likely that when embarking on your own research you will have to consult other books that specialise in the techniques you are going to apply, but we hope that this textbook will provide the foundations to facilitate your understanding of those techniques. The textbook covers a variety of the most commonly used techniques and aims to give you an idea of what is available, so that when you choose your techniques it will be an informed choice and not based on a lack of understanding of the alternatives.

Additionally, the Internet offers a gateway to more knowledge and there are excellent YouTube clips and other videos. Reading through the material in this book will equip you with the language to search for these and understand the material in

the videos. There are also many universities that provide open source material, notably the University of California (UCAL), which can be used to supplement the material in this book. Web sources can certainly be a great help in finding out how to implement techniques in software.

1.2 The Methodology of Positivism

In the spectrum of methodological approaches, in terms of ontology and epistemology, positivism lies at the opposite extreme from phenomenology and interpretivism. Positivism considers there to be a single external reality, whereas interpretivism is based on the belief that there is no single external reality. The epistemology of positivism is that it is possible to secure objective knowledge, the research focus is on generalisation and abstraction, and thinking is governed by hypothesis and known theory. In interpretivism understanding is through perceived knowledge, subjectivity is embedded, the focus of research is on the specific, and the approach seeks to understand specific context. In positivism, which some refer to as empiricism, reality is measurable, and theories are founded on testable and replicable facts. There are two affirmations behind positivism: that factual knowledge is based on the 'positive' data of experience, and that beyond factual knowledge is the logic of mathematics. Sir David Hume, the eighteenth-century Scottish philosopher, referred to these notions as the relations of ideas. The French philosopher Auguste Comte refined these ideas and termed them positivism. For a positivist, reality is observable and, through testing, a speculation can be deemed true or false; if it is not testable then the speculation is meaningless.

Positivists have strong faith in science and in their investigations follow the scientific method to form hypotheses; data are collected from measurement, measurements (data) are tested, and theory is accepted, modified or rejected.

In early formulations of positivism some argued that non-measurable influences such as culture, religion, politics and psychology were being ignored and treated as meaningless, and criticised positivism as being too reductive. To overcome this criticism, post-positivism emerged. In post-positivism objective science is still seen as good and useful for understanding but subjective measures are allowed in order to explain influences which are not physically observable, such as feelings and attitude.

1.3 What is Statistics?

A straightforward definition of statistics is: a set of techniques for exploring and interrogating data to convert the raw data, usually numbers, into information (as in Figure 1.1). For example, if I was thinking of opening a coffee shop, wished to maximise profit, and had to choose between one of two towns, I would need information on which town would be best. I would collect data on the number of people in each town, how wealthy they are, their preference for coffee, the amount of competition, and so on, and use these measures to make an estimate of which town I would do best in. However, it is difficult to obtain accurate measures,

especially for preference for coffee, so there will be uncertainty and error in computations made. Statistics helps manage the errors and uncertainties.

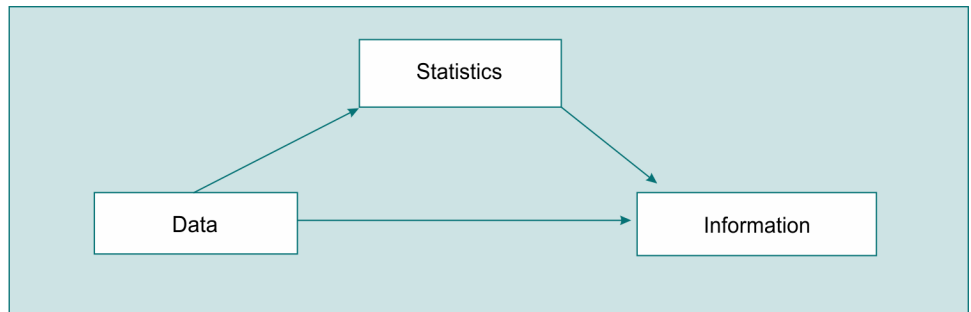


Figure 1.1 The role of statistics

Statistics is more than this, and statistical thinking can help plan a study by formulating the problem and deciding on what data are needed and how to measure and collect them. For example, in the coffee shop example it was assumed I was a profit maximiser, but if I wanted to perform a social good such as establishing a centre where people who are lonely can meet others, the input data to help with the choice of town would be different. Statistics is a process and is associated with a scientific method: formulate questions, collect data, analyse the data, make decisions based on evidence and communicate findings. Because of the connotations of the scientific method, statistics is usually associated with the philosophy of positivism, in that an underlying theory is assumed which explains observations, this theory is then tested, and if the theory is confirmed the theory is validated, otherwise it is rejected. If rejected, work then begins on amending or even proposing a new theory. Observations are made, these are tested against a theory and a process of induction is followed in which hypotheses are generated. From these, deductions are made and confirmed or rejected and the process is repeated. The application of statistics is therefore a circular process, as shown in Figure 1.2. Statistics is not about proving theory, as theory cannot be proven (but only accepted until proven wrong); rather it is about testing and refining theory.

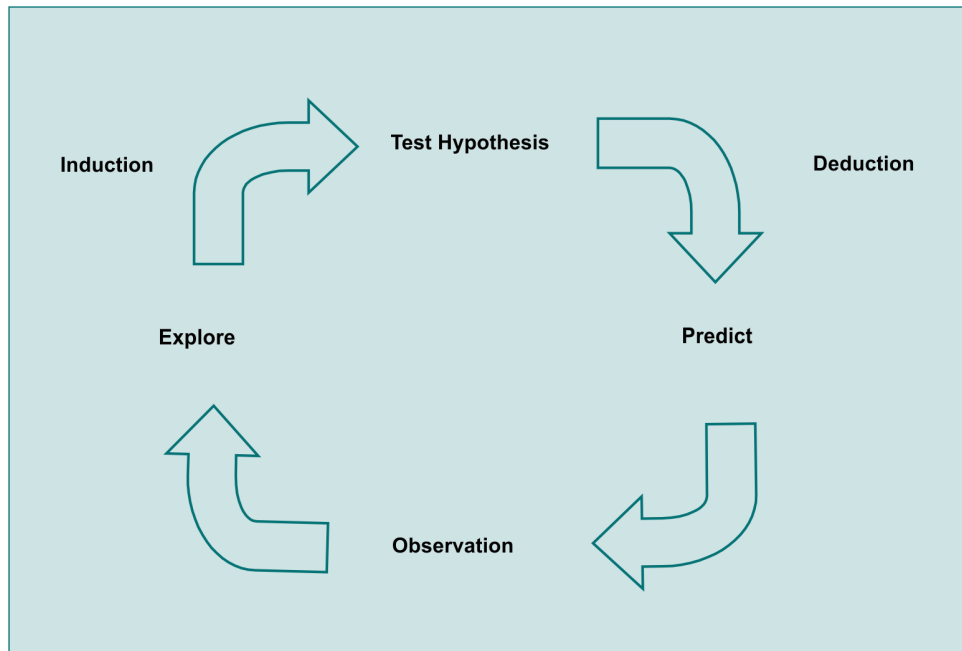


Figure 1.2 The process of statistics

From undertaking tests on observed data, the researcher enters an inductive phase of thinking to determine if findings align with current knowledge. This is formalised by formulating hypotheses, which are a supposition or explanation that the outcome of tests fits with known facts. This now leads to a deductive phase, in which reasoning proceeds from general premises to a specific conclusion. It is here that theory is supported or challenged, and the outcome can be a model of how inputs interplay to give an outcome. This allows predictions to be made, which allow comparison to a future observation, and the cycle begins again. If no model can be formed then a ‘not proven’ verdict results and you skip to gathering more observations, possibly measured in a different way, and the cycle starts again. The process is iterative. In reality the process of statistics is not usually so clearly delineated and experienced researchers often combine inductive and deductive reasoning in a mixed form of reasoning, but the notion remains of measuring to gather data, testing them, and drawing conclusions from which predictions are compared to observations.

Statistics is a scientific approach and is in the frame of the scientific method. The scientific method is a systematic approach in which you pose a research question, collect data, form hypotheses, test these hypotheses and conduct validation efforts (Ackoff, 1962; Popper, 2002; Gauch, 2012). Figure 1.3 presents a schematic of undertaking statistical research in the frame of the scientific method. It shows that there are many backwards iterations, with continual checking, and hence statistical research is a cyclic process. Lindley (2000) gives a comprehensive discussion on the philosophy of statistics.

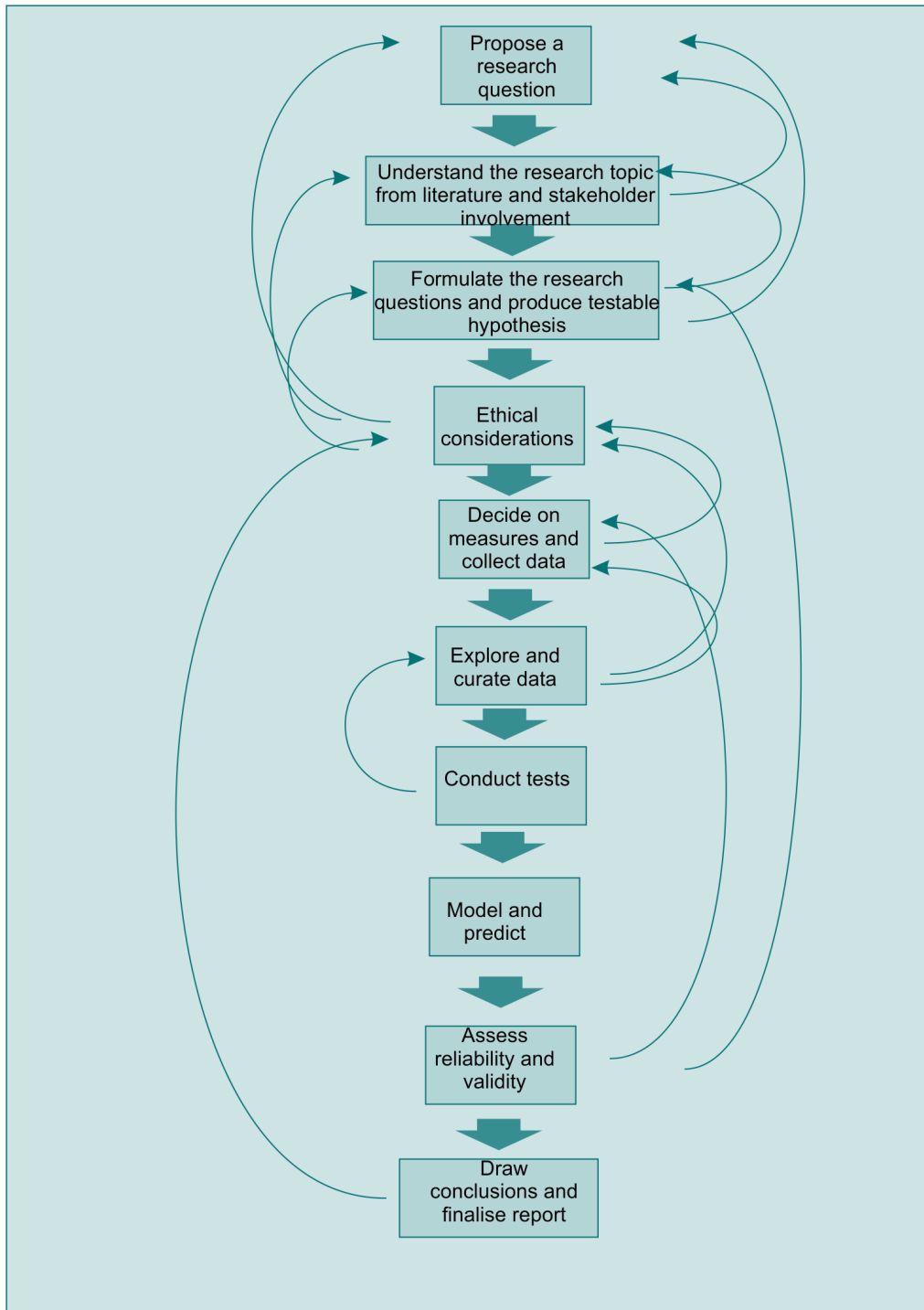


Figure 1.3 The statistical research cycle

In the statistical research cycle a great deal of effort is spent formulating the research: that is, setting the research question, understanding the area through reviewing the literature and involving stakeholders. Central in the cycle are ethical considerations which iterate back to research formulation. Only after ethical concerns are satisfied can data collection begin. Once the data are collected they are explored to gain an initial understanding and judge their suitability for answering the research question. If judged suitable, then modelling occurs in which predictions are made. The predictions are checked for reliability: essentially asking, if the process is repeated will the outcomes be reproduced? Maybe the reproduced outcomes will not be identical to the predictions; some small departures would be expected arising from measurement error. If these deviations are within a tolerable degree, the outcomes are considered reliable. If the deviations are larger and can be attributed to some change in the population, the original results might still be considered reliable. If reliability cannot be shown, the credibility of the research is reduced. Then validity is assessed by checking with current theory or thought (face validity), and by the degree to which the measures are measuring what the researcher intends (construct validity). If validity cannot be assured, the research cannot be considered credible and the researcher has to start again.

The final step, once reliability and validity have been assessed and confirmed, is to draw conclusions and produce a report in which recommendations are made. The report should emphasise the value of the research and in effect 'sell' the findings and recommendations to the stakeholders. To ensure acceptance by stakeholders and to encourage implementation of any recommendations, the report must demonstrate that the work is compliant with good practice, and is reliable, valid and ethical. To do this the report needs to be understandable: good communication often means use of diagrams, clarity of writing and avoidance of over-technical detail (this latter point is dependent on stakeholders' requirements).

In applying statistics, the collecting, organising, analysing, presenting and interpreting of findings involve many decisions which, although based firmly on scientific procedures, lead many to consider statistics as much of an art as a science (see Tippet, 1931; Agresti, Franklin, & Klingenberg, 2017). Through the experience of undertaking applied statistical research, you can develop a 'feel' for data and intuition about how to apply and interpret statistics, and there is some truth in the notion that statistics is both a science and an art. Certainly, statistics should not be considered as mathematics; it is different. As Box (1990, p. 251) commented, 'Statistics is or should be about scientific investigation and how to do it better, but many statisticians believe it is a branch of mathematics.'

I.4 Statistical Thinking

Statistical thinking is more than numbers, charts and tests; it is about systematic and strategic understanding and improvement. As Lord Kelvin (William Thomson) stated in 1883:

When you can measure what you are speaking about and express it in numbers, you know something about it. ... When you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind. (Thomson, 1889: vol. I, p. 73)

Statistics involves more than the definition of measures and the collection of data. It involves the processing of data in such a way as to expose uncertainty. In this way statistics turns data into information which can be used in decision making.

Statistics as a discipline provides a wide range of concepts and methods to understand why things vary. There is variation everywhere and it is important to understand this: for example, to know if a trend is emerging or is just random variation, or if a process is going out of control or is deviation from a target in line with expected variation or 'noise'. Statistics provides the methods to answer these questions and allows you to differentiate between background noise, termed **common cause** variation, and variation that occurs because of a change in some input factor. Where variation increases because of a particular reason it is referred to as **special cause** variation; for example, in a manufacturing process, a change in a critical dimension of a manufactured item might be attributable to wear in a machine, increasing vibration in equipment, a change in the nature of materials supplied, or human error in calibrating machines. Statistics is about identifying and understanding both common and special cause variation and has been called the **science of variation**. In research, understanding special cause variation allows a theory to be confirmed or new theories to be formed, and understanding common cause variation allows prediction and testing. An example of special cause variation might be found in designing an aircraft wing for a new plane: if the design follows an existing model of wing design, vibration might turn out to be too great, so variation is greater than expected and the engineers need to rethink their ideas about turbulence and the application of these ideas to new wings, or perhaps the theory of turbulence needs to be updated. On the other hand, common cause variation is natural variation that arises in replication: for instance, suppose in a manufacturing process metal rods are to be made to a length of 10 cm, but they are not all exactly 10 cm – some are a bit less and others a bit more. This variation might not be apparent to the naked eye and a high-resolution measurement system might be required. In quality improvement the goal is to screen out special cause variation and work to reduce common cause variation. The aim is to produce output that meets the requirements and is consistent. Understanding common cause variation can allow more reliable and repeatable measurements to be made. For a more in-depth discussion of statistical thinking see Deming (2000), Pfannkuch & Wild (2000), John, Whitaker, & Johnson (2005) and Cox & Efron (2017).

Statistical thinking is iterative: you learn, test, learn and observe again and again. Box, Hunter, & Hunter (1978) conceptualise learning from statistical thinking in the diagram, reproduced in Figure 1.4.

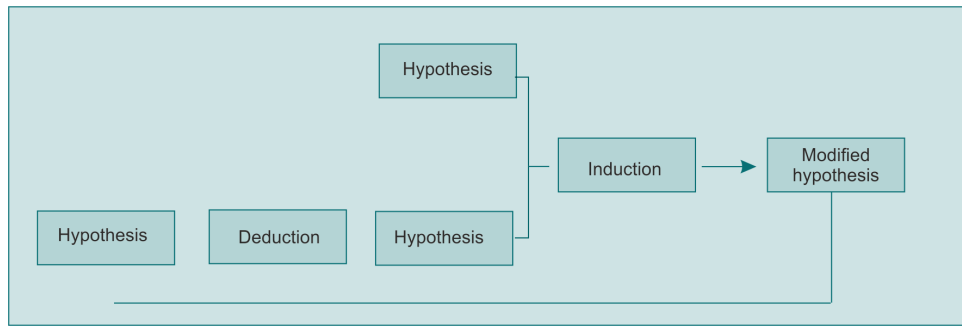


Figure 1.4 Learning using statistical thinking (Source: Box et al., 1978, p. 2)

Considering the value of statistics, it is not surprising that Wilks (1951) in his opening presidential address to the American Statistical Association paraphrased H.G. Wells in *The Making of Mankind* thus:

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.

1.5 Book Structure

The primary function of the book is to equip you with a methodology and a set of methods to undertake quantitative research. We have attempted to cover the main facets of applied statistics, but there are more which are beyond the scope of this book, and when designing your research it is likely that you will need to consult additional sources. We hope that from this book you will develop a vocabulary to search and understand how different methods are used, be able to apply these yourself, and be able to investigate references and other sources to successfully complete your research.

At this point you might be wondering why a doctoral student needs to be versed in both qualitative and quantitative methods. For instance, you might wonder why you would bother with numbers and complicated statistics if your research project involves only qualitative data. Well, there are two important reasons. First, choosing from a limited pool of data and techniques might make your research biased or limited, and this might happen if you are not aware of the qualitative and quantitative options available. Secondly, as you have learned from previous books in this series, the literature review is one of the most important elements of your thesis (if not the most important), and in order to produce an effective and comprehensive literature review you will have to consult previous research. It is more than likely that there will be a mixture of qualitative and quantitative studies in your field and therefore you need to have a reasonable understanding of both sets of techniques in order to make sense of them. Reading only literature using qualitative techniques because you do not understand the quantitative studies, or vice versa, will give you a biased view of previous research and affect the quality of your literature review.

Statistical methods are presented chapter by chapter, avoiding theoretical development, but rather demonstrating through examples how the methods might be

used. Chapter 2 introduces the concept of data and its types, and then the use of statistical software packages is introduced and applied to cover the basics of descriptive statistics. Finally, this chapter introduces the concept of normal distribution. Although many people manage to undertake quantitative work by using spreadsheets such as Microsoft Excel, SPSS is easier to use and offers a greater range of techniques and is the recommended software. There are alternatives, notably SAS, Stata, Minitab and R, and if you have access to these and are familiar with them then by all means use them.

The next two chapters (Chapters 3 and 4) address issues of data collection. Chapter 3 starts with a discussion about automated data collection and the idea of secondary data, and goes on to talk about quantitative content analysis. Experimental design as a method is briefly covered, with an example. Finally, the very important element of sampling is introduced and treated, with a consideration of sample size calculations and the concept of complex sample design. The chapter finishes with a reflection about the administration of the data collection process. Chapter 4 focuses on questionnaires. It starts by focusing on the design of questionnaires in terms of format and type of questions, giving some general advice as to what works best. We then talk about piloting, what is it and how it should be done. The next sections cover ways to improve response rates, minimise problems and administrative arrangements. The last part of this chapter focuses on coding of the data and data entry in preparation for analysis.

At this point in this book, we have covered the different data types and data collection procedures so once you have an understanding of that, you should be in a good position to understand data ethics and that is the topic of Chapter 5. This chapter will consider ethical issues which arise in working with data. We cover the main points you should consider in order to assure your research complies with ethical standards. General points are covered first before moving on to discuss safe data processing and reporting. This chapter brings the ethical discussion beyond the data collection element into the analysis and reporting of data.

In Chapter 6 we move into the realm of statistical inference which is where statistical procedures start getting interesting and very useful. The basics of statistical inference are that by applying some techniques in the forms of statistical tests we will be able to infer the characteristics of the population based on our observed data sample. We start this journey with a chapter dedicated to differences where we present a number of techniques we can apply in order to assess the statistical significance of a difference. These tests vary depending on the data type and characteristics of the data distribution, we cover the two main families of parametric and non-parametric tests by giving examples.

Chapter 7 lays the foundations for statistical modelling by covering the concept of association between variables with special focus on correlations. First, we explore the concept of correlations between numerical data with the use of the Pearson product-moment correlation coefficient. We then we cover associations between ranked data with the use of the Spearman rank correlation coefficient and we finish this chapter investigating association between categorical variables using the chi-square statistic.

Chapter 8 covers the concepts of regression and statistical modelling. The techniques and information from the previous chapter has given you the foundations to now explore more statistically advanced procedures like statistical modelling. This will allow you to produce equations that can be used to either make predictions or explain the impact of some variables on others. We start this chapter with an example of simple linear regression in order to explore the basic elements of statistical modelling with its procedures and requirements. We produce an equation model, we assess how well that model fits the data, we evaluate the normality and randomness of the residuals and we apply statistical inference techniques to assess the model estimation. From here we move into the more complex case of multiple regression. This chapter ends covering the concept of binary logistic regression where the dependent variable has two categories such as yes or no, or failed or succeeded.

Chapter 9 covers the concept of data reduction where we combine a number of variables to produce a single measure in a more efficient way than by purely using averages of the standardised variables. The approaches covered in this chapter use weighted averages based on the correlation between variables. The techniques covered here are principal components and factor analysis; they can both be used in a general sense to reduce the dimensionality of a data set where groups of variables are correlated with one another.

The final chapter (Chapter 10) covers the concepts of clustering and segmentation. A company might be interested in grouping its customers into homogeneous groups in order to apply targeted marketing techniques. These groupings can be generated with clustering and segmentation statistical techniques. There are a variety of procedures for data clustering and segmentation; in this chapter we will cover the most common ones: hierarchical methods for small samples and partitioning methods for larger samples.

References

- Ackoff, R.L. (1962). *The scientific method* (99th ed.). Chichester: John Wiley & Sons.
- Agresti, A., Franklin, C.A., & Klingenberg, B. (2017). *Statistics: The art and science of learning from data* (4th ed.). New York: Pearson.
- Box, G.E.P. (1990). Commentary. *Technometrics*, 32(3), 251–252.
- Box, G.E.P., Hunter, W.G., & Hunter, J.S. (1978). *Statistics for experimenters*. Chichester: John Wiley & Sons.
- Cox, D.R. & Efron, B. (2017). Statistical thinking for 21st century science, *Science Advances*, 3(6): 3: e1700768.
- Deming, E.W. (2000). *Out of the crisis*. Boston: MIT Press.
- Gauch, H.G. (2012). *The scientific method in brief*. Cambridge: Cambridge University Press.
- John, J.A., Whitaker, D., & Johnson, D.G. (2005). *Statistical thinking in business*. London: Chapman & Hall/CRC.
- Lindley, D.V. (2000). The philosophy of statistics. *Journal of the Royal Statistical Society Series D (The Statistician)*, 49(3), 293–337.

- Pfannkuch, M. & Wild, C.J. (2000). Statistical thinking and statistical practice: themes gleaned from professional statisticians, *Statistical Science*, 15(2), 132–152.
- Popper, K. (2002). *The logic of scientific discovery*. London: Routledge Classics.
- Thomson, W. (1889). Lecture to the Institute of Civil Engineers (1883). In: Thomson, W. *Popular lectures and addresses*, vol. 1. London: Macmillan & Co.
- Tippet, L.H.C. (1931). *The method of statistics*. London: Williams & Norgate.
- Wilks, S.S. (1951). Undergraduate statistical education. *Journal of the American Statistical Association*, 46(253), 1–18.