

# Introduction

*Data is like garbage. You had better know what you are going to do with it before you collect it*

Mark Twain

Many people consider statistics a disagreeable discipline. Probably because for centuries it has been used to allow power (whether public or private) to achieve its objectives. Did a king want to declare war? His mathematicians counted people fit for military service, their available means and their equipment. Were funds for building a palace or a castle insufficient? Incomes were calculated, and taxes were increased just enough, if the regency was astute, to collect the amount of money required to satisfy all the wishes without squeezing the taxpayers too much. Was a firm in need of increasing production or profit levels? Statisticians were employed to count, measure, highlight weak areas, rationalise costs, remove or add workers and suggest possible solutions. Yet, with its methods, medicine, technology, economics and many other disciplines have reached levels that have allowed us to live longer and better, to work in more favourable conditions and to have a deeper knowledge of the physical world.

Formally, statistics has the objective of collecting, analysing and interpreting data collected in various ways and assessing methods and procedures for performing these activities. The objective of a statistician is to derive universally valid conclusions from a collection of partial observations. With a very practical approach, knowing that measuring all the aspects of a phenomenon can be impossible for many reasons, we employ well studied and discussed scientific methods to do the work, and, more importantly, to give some measure of the reliability of the conclusions drawn. In his book, *The Rise of Statistical Thinking 1820–1900*, Theodore Porter states:

Statistics has become known in the twentieth century as the mathematical tool for analysing experimental and observational data. Enshrined by public policy as the only reliable basis for judgements as to the efficacy of medical procedures or the safety of chemicals, and adopted by business for such uses as industrial quality control, it is evidently among the products of science whose influence on public and private life has been most pervasive. Statistical analysis has also come to be seen in many scientific disciplines as indispensable for drawing reliable conclusions from empirical results. For some modern fields, such as quantitative genetics, statistical mechanics, and the psychological

field of intelligence testing, statistical mathematics is inseparable from actual theory. Not since the invention of calculus, if ever, has a new field of mathematics found so extensive a domain of applications. (Porter, 1986: 3)

Tourism, like many other human activities, relies heavily on data of all sorts and the quantitative treatment of data and information collected in a wide variety of ways is a crucial endeavour for both academics and practitioners. Yet, numbers and formulas are not the most widely diffused objects in the tourism field and our experience in this area tells us that the application of mathematical and statistical concepts and procedures is far from common practice.

In its long history, statistics has implemented a large number of techniques for dealing with different situations and giving answers in different conditions. Very sophisticated, and sometimes complicated, procedures enable us to derive justified outcomes that, in many cases, prove to be crucial for decision-making, or for the implementation of development plans or policies, or simply for understanding how tourism activities unfold.

Many of these techniques, however, can only be found in scholarly journal papers or in advanced specialised books. There is, generally, little practical information on a variety of methods and, mainly, on the way they can be applied to tourism cases. Advanced quantitative methods are rarely described in tourism textbooks, and the treatment given in more standard statistical textbooks is, at times, too theoretical and gives little operational information. On the other hand, a quick survey of the tourism literature shows a certain limitation in the number of methods and techniques.

This book aims to fill this information gap by providing practical tools for the quantitative analysis of data in the tourism field. The main objective is to make available a usable reference book rather than a theoretical text discussing the methods. For a full treatment of the different methods described, the reader will be supplied with relevant references on the different topics. Most of the methods presented have been chosen after a survey of the tourism literature. We have also taken into account many current techniques used in journals and scientific publications as well as our experience in teaching these topics and the efforts spent in trying to find instructional materials with the right mix of arguments and the right balance between scientific rigour, practical usefulness and simplicity of language. This work has highlighted a number of approaches that have been shown to provide interesting outcomes. To these, a number of more recent topics have been added. They are well consolidated in other disciplines and their effectiveness allows us to see a promising future for their application in tourism studies.

Different from a standard statistics textbook, this work gives little space to the theoretical discussion of the methods presented. Rather, it aims at providing practical hints on their applicability and, where appropriate, a discussion on their advantages and disadvantages. Many examples are presented and references to similar studies are illustrated; they are an integral part of the text and, in many cases, replace the theoretical exposition of the methods discussed.

This book has been designed for graduate students at master and PhD level, researchers in both tourism and the social sciences and practitioners or industry consultants. It is assumed that the reader has at least a basic understanding and some (good) familiarity with elementary statistics (descriptive and inferential) and with concepts and terms such as confidence limits, significance levels, degrees of freedom, probability and probability distributions and so on. In any case, numerous references in the book will point the reader to noteworthy works in which he/she will find extensive mathematical and conceptual treatment for the different topics to satisfy his/her curiosity or need to explore all the nuances of the methods discussed here. Many of the techniques described definitely require the use of some software program, and in many cases, the standard statistical analysis programs do not contain dedicated functions for them. Nevertheless, these can be found without much effort on the internet as small executable programs or scripts for some widely used application development environments, such as Matlab or GAUSS. References have been given with the text and an appendix contains a list of these programs with their internet addresses. Needless to say, some familiarity with the use of a computer is an unavoidable skill today.

Many authors report, as diffuse wisdom, the fact that every equation included in a book would halve the sales. Caring much for the economic health of our publisher, we have tried to reduce mathematical expressions to a minimum. However, as the reader will understand, some of them are unavoidable when speaking the language of numbers.

Finally, it is important to remark here that, although it is commonly considered to be a scientific discipline, statistics might be more accurately thought of as a craft or an art, where experience plays a central role and numerous different interpretations of even basic concepts and procedures exist. What is presented in this book is the interpretation (grounded) of the authors. We have taken care to present the most widely accepted readings, but in some cases our views might be questioned and different versions may be found in the literature.

The book is divided into two parts. The first part deals with data analysis methods that are widely used by the tourism research community, but not described much in standard tourism books. The second part describes some numerical methods that, to date, have seen limited use in tourism studies.

These techniques are gaining wide attention and a reputation in many disciplines for the study of several types of systems, especially when the issues investigated are difficult or not tractable with analytical methods. They have been made practically usable through the operation of modern computer systems. Although, in some cases, highly computationally intensive, they have proved to be able to provide useful insights that can complement the conclusions attained by more traditional methods and may give, in the future, different perspectives to the field of tourism. An appendix describing some of the more used software tools closes the book.

All the chapters have been written to be independent of one another, and for this reason the references have been listed separately at the end of each chapter. In this way, the reader is not forced to go through the book with a predetermined sequence, but is free to hop here and there, following his/her own curiosity or needs.

As a final note, the authors wish to advise the reader that all the internet addresses contained in the book have been checked before releasing the final version of the text. However, nothing can guarantee that they will not change or disappear. Should this happen, an online search will surely enable the reader to find moved pages or similar contents.

The authors would like to thank a number of people who have helped and supported us in our work, but the list risks being quite long and tedious for the reader. All who have helped us are aware of the importance of their contributions, and to them our sincere thanks.