ASPECTS OF TOURISM: 79

Quantitative Methods in Tourism

A Handbook

2nd edition

Rodolfo Baggio and Jane Klobas

CHANNEL VIEW PUBLICATIONS
Bristol • Blue Ridge Summit

DOI 10.21832/BAGGIO6195

Library of Congress Cataloging in Publication Data
A catalog record for this book is available from the Library of Congress.
Names: Baggio, Rodolfo, author. | Klobas, Jane E., author.
Title: Quantitative Methods in Tourism: A Handbook/Rodolfo Baggio and Jane Klobas.
Description: Second edition. | Bristol, UK; Blue Ridge Summit, PA, USA:
Channel View Publications, [2017] | Series: Aspects of Tourism: 79 |
Includes bibliographical references and index.
Identifiers: LCCN 2017006686| ISBN 9781845416195 (hbk: alk. paper) | ISBN
9781845416188 (pbk: alk. paper) | ISBN 9781845416225 (kindle)
Subjects: LCSH: Tourism. | Tourism–Statistics.
Classification: LCC G155.A1 B316 2017 | DDC 910.01/5195–dc23 LC record available at
https://lccn.loc.gov/2017006686

British Library Cataloguing in Publication Data
A catalogue entry for this book is available from the British Library.

ISBN-13: 978-1-84541-619-5 (hbk)
ISBN-13: 978-1-84541-618-8 (pbk)

Channel View Publications
UK: St Nicholas House, 31-34 High Street, Bristol BS1 2AW, UK.
USA: NBN, Blue Ridge Summit, PA, USA.

Website: www.channelviewpublications.com
Twitter: Channel_View
Facebook: https://www.facebook.com/channelviewpublications
Blog: www.channelviewpublications.wordpress.com

The policy of Multilingual Matters/Channel View Publications is to use papers that are natural, renewable and recyclable products, made from wood grown in sustainable forests. In the manufacturing process of our books, and to further support our policy, preference is given to printers that have FSC and PEFC Chain of Custody certification. The FSC and/or PEFC logos will appear on those books where full certification has been granted to the printer concerned.

Typeset by Deanta Global Publishing Services Limited.
Printed and bound in the UK by Short Run Press Ltd.
Printed and bound in the US by Edwards Brothers Malloy, Inc.

Contents

Contributors

Rodolfo Baggio holds a 'Laurea' degree in Physics (MPhys) from the University of Milan, Italy, and a PhD from the School of Tourism at the University of Queensland, Australia. After working for leading information technology firms for over 20 years, he is currently a professor at the Bocconi University where he teaches courses in Computer Science and coordinates the Information and Communication Technologies area at the Master in Economics and Tourism. He is also Research Fellow at the 'Carlo F. Dondena' Centre for Research on Social Dynamics and Public Policy. Rodolfo has held several courses and lectures at national and international level and has carried out consulting activities for private and public tourism organisations. He has managed several international research projects and actively researches and publishes in the field of information technology and tourism. His current research combines complexity theory and network analysis methods with the study of tourism destinations.

Jane Klobas is an Education and Research Consultant, based in Australia and Italy. She is an Adjunct Professor at Murdoch University, Western Australia and a Visiting Professor to the University of Bergamo and other universities in Europe and Asia. She was previously at Bocconi University in Milan and the University of Western Australia. She supervises doctoral students for the University of Liverpool Online and teaches research methods to doctoral students and faculty at several universities. She is author or co-author of several books and book chapters, and has published widely across disciplines in journals including *The Internet and Higher Education*, *Computers in Human Behavior*, *Library and Information Science Research*, *Demographic Research*, *Journal of Organizational Behavior* and *Decision Support Systems*.

Jacopo A. Baggio holds a degree in Economic and Social Sciences from the University of Milan Bicocca, a Master in Development Economics and a PhD in International Development from the University of East Anglia, which was funded by the UK Economic and Social Research Council (ESRC). He subsequently worked as a postdoctoral research associate with the Center for Behavior, Institutions and the Environment (CBIE) at

Arizona State University, and is now Assistant Professor at the Department of Environment and Society, Utah State University. His research focuses on the analysis and modelling of social-ecological systems. His main interests can be divided into two macro areas. One focuses on the conditions under which collective action succeeds in human societies, analysing what drives collective action and how it is influenced by uncertainty. The other centres upon social-ecological networks, characterising interdependencies between biodiversity, food, water, energy and decision-making.

Foreword

The tourism subject continues to mature, evidenced by debates on research approaches and the ever-increasing sophistication of the techniques used to investigate the activity that is tourism. These debates are often focused around the quantitative versus qualitative debate, and as Rodolfo and Jane say in their introduction to the first edition of this book, statistics and the quantitative approach are often labelled as 'disagreeable'. Yet, if tourism is to mature effectively as a subject, we cannot hide from the demands of quantitative approaches.

At a stroke, the second edition of this book progresses the maturity of tourism while also removing the mystique surrounding numbers and tourism. This is not a quantitative methods textbook; rather, it is a manual to guide tourism researchers through the minefield of advanced quantitative methods and how to apply them to tourism research. The book is unusual because it is written by experts in mathematics and quantitative methods; experts who have since moved into the tourism subject area. As such, this is a 'grown up' book that makes a number of demands and assumptions of its readers, providing researchers with the practical tools necessary for the analysis of complex tourism data sets, without shying away from the word 'complex'. This book will considerably enhance the standing of tourism as a subject and I know that it will be a valuable addition to the researchers' armoury.

Chris Cooper
Oxford Brookes University, Oxford

Introduction to the Second Edition

Five years ago, when we wrote the first edition of this book, we thought of it as a one-time-only project. We were inspired by our experiences, as advisers to researchers, analysts and students, to provide an accessible, sensible and rigorous guide to useful methods for statistical inquiry into tourism matters of all but the most econometrically complex kind. We were delighted by the response to our book and happy to bask in the pleasure of a job well done. But, of course, quantitative methods, technological tools and sources of data continue to develop, and the expectations of supervisors, examiners and peer reviewers of research papers evolve. What was 'enough to know' five years ago, is not enough to know now. Thus, this second edition.

This edition retains the overall approach taken in the first edition. The first part of the book concerns common issues in the statistical analysis of data and the most widely used techniques. The second part describes and discusses several newer and less common approaches to data analysis that we believe are useful for tourism researchers and analysts, and which we encourage readers to consider. We have added material to both sections.

The first part of the book now includes sections on issues that, while always important, have become more transparent as software evolves and makes it easier to adopt and present the results of analyses undertaken using both older and newer techniques. We focus on techniques that, having become more accessible are, in the reports and papers we read, often applied without a great deal of thought, in a textbook sequence that does not necessarily fit the data and context of the project being described. We have added consideration of data screening and cleaning to Chapter 1 and methods for measuring similarity and dissimilarity to Chapter 2. Chapter 4 has been extended to include observations about the partial least squares (PLS) approach to path modelling (sometimes equated with structural equation modelling [SEM]). Chapter 4 also includes new sections on multilevel modelling and accounting for common method variance in SEM.

A new chapter on 'Big Data' has been added to Part 2. This chapter aims not only to inform readers about the many aspects that come together to make Big Data more than a data-based revolution, but also to consider controversies about whether Big Data means the end of statistics. The chapter guides users through decisions to be made about when and how to use Big Data and how to interpret and evaluate the findings of Big Data projects. The final chapter, on agent-based modelling and simulations, has been updated and revised.

Once again, many people have provided encouragement and support for this edition. We thank you all.

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.

Part 1

The Analysis of Data

Introduction to Part 1

The first part of this book contains a discussion of standard methods in statistical data analysis: hypothesis tests, regressions, cluster and factor analysis and time series analysis. They have been chosen for their importance in the field of tourism studies, even though they are scarcely treated in general tourism textbooks.

We have avoided highly sophisticated methods that, usually, can only be applied well in special circumstances, but we have included some extensions to the standard techniques. These, although well diffused in other disciplines (e.g. non-linear analysis techniques for time series), have not had wide use in tourism studies. Their effectiveness has been demonstrated many times in other fields and we think they will prove useful in this area too.

The content of this part is organised as follows.

The Nature of Data in Tourism

Data are the main ingredient of all the methods discussed in this book and are examined from a general perspective. The various types are described and examined. The quality of data is then discussed and practical suggestions for assessing and evaluating the suitability of data in relation to the objective of an investigation are given. Finally, a list of electronic sources of tourism data is provided.

Testing Hypotheses and Comparing Samples

This chapter contains a review of the main concepts and techniques connected with statistical hypotheses testing. Issues regarding the power of tests and the effects of sample size are discussed. Also, bootstrap and meta-analysis as methods to improve the reliability of the outcomes are presented. A summary of the most commonly used statistical tests is included. The chapter closes with a description of different methods to assess similarity (or diversity) within and between samples.

Data Reduction

An analysis of multivariate data is presented here. Factor analysis and cluster analysis as well as multidimensional scaling techniques are also described and discussed along with the main issue, advantages, disadvantages and applicability.

Model Building

The chapter discusses regression models and structural equation modelling. Focusing on the tourism field, the chapter highlights the issues related to computational techniques and the reliability of the results in different conditions.

Time-Dependent Phenomena and Forecasting

This chapter contains a quick overview of time series analysis methods and their use for forecasting purposes. In addition, different uses of time series are discussed, such as simple non-linear analysis techniques to provide different ways of studying the basic characteristics of the structure and the behaviour of a tourism system.

Part 2

Numerical Methods

Introduction to Part 2

The methods and techniques presented and discussed in the previous chapters can be considered standard means for analysing and describing the data most commonly collected by academics and practitioners in tourism and hospitality studies. Part 2 presents a few methods that, although used by some, can be seen as more advanced additions to the tourism researcher's toolbox.

The methods are quite common in other disciplines and have a solid foundation both from a theoretical and a 'practical' point of view. Maximum likelihood estimates, Monte Carlo methods and agent-based models are described and discussed with examples of their application to tourism-related problems.

In addition, a chapter is devoted to a brief account of the recent and rapidly evolving world of Big Data, one of the most important by-products of the extremely diffused use of the modern online environments.

These chapters are grouped together as they share an important common feature: the necessity of a computer. For the methods discussed so far, a computer is a useful tool and today no one would venture starting a factor analysis or the computation of regression coefficients without some calculating machinery. However, the techniques discussed in this part are absolutely not usable without the power and ease of use of modern personal computers. The alternative would be to replicate efforts such as the one described by W.S. Gosset (better known by his pen name Student) in his milestone paper 'The probable error of a mean':

> Before I had succeeded in solving my problem analytically, I had endeavoured to do so empirically. The material used was a correlation table containing the height and left middle finger measurements of 3000 criminals, from a paper by W. R. Macdonell (Biometrika, Vol. I, p. 219). The measurements were written out on 3000 pieces of cardboard, which were then very thoroughly shuffled and drawn at random. As each card was drawn its numbers were written down in a book, which thus contains the measurements of 3000 criminals in a random order.

> Finally, each consecutive set of 4 was taken as a sample – 750 in all – and the mean, standard deviation, and correlation of each sample determined. The difference between the mean of each sample and the mean of the population was then divided by the standard deviation of the sample, giving us the z of Section III. This provides us with two sets of 750 standard deviations and two sets of 750 z's on which to test the theoretical results arrived at. (Student, 1908: 13)

Definitely not practical, also when considering that the typical number of replications in a Monte Carlo simulation, for example, is 100 or 1000 times higher, or that we can collect millions of records from online sources.

The content of this part is organised as follows.

Maximum Likelihood Estimation

The idea behind maximum likelihood estimation is to employ a generalised procedure able to find the value of one or more parameters for a given statistic which makes its likelihood distribution a maximum. These provide efficient methods for quantifying uncertainty and to assess confidence limits. As for the other methods described in this part, the estimation methodology is simple, but the calculations involved can be quite intense. Modern computer power is therefore the only practical way of using these methods.

Monte Carlo Methods

Monte Carlo methods are a class of computational algorithms that provide approximate solutions to a variety of mathematical problems. They depend on repeated random sampling to perform statistical experiments and can be loosely defined as statistical simulation methods. Monte Carlo simulation methods are especially useful in studying systems with a large number of interdependent degrees of freedom.

Big Data

This chapter contains a brief description of the much discussed topic of Big Data. A characterisation and definition is followed by a brief description of the technological architecture used and consideration of some of the statistical issues surrounding Big Data. A discussion of machine learning techniques, along with an example, closes the chapter.

Simulations and Agent-Based Modelling

Agent-based modelling and numerical simulations are means that facilitate exploring the structural and dynamic characteristics of systems that may prove intractable with analytical methods. This chapter examines what are complex adaptive systems, when are agent-based models a useful methodology to analyse systems, what are issues that relate to them and where they have been applied successfully. As an application example, a simple model is built to analyse the movements of tourists and the relationship between these and the attractiveness of a tourism destination.

References

Student (1908) The probable error of a mean. *Biometrika* 6 (1), 1–25.