

**ASPECTS OF TOURISM**

*Series Editors: Chris Cooper (Oxford Brookes University, UK), C. Michael Hall (University of Canterbury, New Zealand) and Dallen J. Timothy (Arizona State University, USA)*

**Quantitative Methods  
in Tourism  
A Handbook**

Rodolfo Baggio and Jane Klobas

**CHANNEL VIEW PUBLICATIONS**

Bristol • Buffalo • Toronto

**Library of Congress Cataloging in Publication Data**

A catalog record for this book is available from the Library of Congress.

Baggio, Rodolfo.

Quantitative Methods in Tourism: A Handbook/Rodolfo Baggio and Jane Klobas.

Aspects of Tourism: 48

Includes bibliographical references.

1. Tourism. 2. Tourism--Statistics. I. Klobas, Jane E. II. Title.

G155.A1B316 2011

910.01'5195--dc222011000640

**British Library Cataloguing in Publication Data**

A catalogue entry for this book is available from the British Library.

ISBN-13: 978-1-84541-174-9 (hbk)

ISBN-13: 978-1-84541-173-2 (pbk)

**Channel View Publications**

*UK:* St Nicholas House, 31-34 High Street, Bristol BS1 2AW, UK.

*USA:* UTP, 2250 Military Road, Tonawanda, NY 14150, USA.

*Canada:* UTP, 5201 Dufferin Street, North York, Ontario M3H 5T8, Canada.

Copyright © 2011 Rodolfo Baggio and Jane Klobas.

All rights reserved. No part of this work may be reproduced in any form or by any means without permission in writing from the publisher.

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 Datapage International Ltd.

Printed and bound in Great Britain by the MPG Books Group Ltd.

# Contents

List of Examples . . . . .	vii
List of Figures . . . . .	ix
List of Tables . . . . .	xv
Contributors . . . . .	xvii
Foreword . . . . .	xix
Introduction . . . . .	xxi

## Part 1: The Analysis of Data

Introduction to Part 1 . . . . .	3
1 The Nature of Data in Tourism . . . . .	5
Introduction . . . . .	5
Data: A Taxonomy . . . . .	5
Data Harmonization, Standards and Collaboration . . . . .	9
Data Quality . . . . .	12
Concluding Remarks . . . . .	16
Sources of Secondary Tourism Data . . . . .	16
2 Testing Hypotheses . . . . .	21
Introduction . . . . .	21
Parametric and Non-parametric Tests . . . . .	24
Effect Size and Statistical Power . . . . .	26
Sample Size and Significance . . . . .	28
A Summary of Statistical Tests . . . . .	37
3 Data Analysis . . . . .	42
Introduction . . . . .	42
Factor Analysis . . . . .	42
Cluster Analysis . . . . .	73
Multidimensional Scaling and Correspondence Analysis . . . . .	84
4 Model Building . . . . .	88
Simple Regression . . . . .	89
Multiple Regression . . . . .	100
Logistic Regression . . . . .	107
Structural Equation Modeling . . . . .	112
Mediation and Moderation in Model Building . . . . .	130
Multi-Level Modeling . . . . .	133

---

5	Time-dependent Phenomena and Forecasting . . . . .	136
	Introduction . . . . .	136
	Basic Concepts of Time Series . . . . .	138
	Filtering Techniques . . . . .	147
	Comparing Time Series Models. . . . .	149
	Combining Forecasts . . . . .	151
	Stationarity, Stability and System Representations. . . . .	157
<b>Part 2: Numerical Methods</b>		
	Introduction to Part 2 . . . . .	173
6	Maximum Likelihood Estimation . . . . .	175
	Estimating Statistical Parameters. . . . .	175
	Likelihood Ratio Test. . . . .	186
7	Monte Carlo Methods . . . . .	189
	Numerical Experiments. . . . .	189
	Random and Pseudo-Random Numbers . . . . .	193
8	Agent-based Modeling and Simulations . . . . .	199
	Complex Adaptive Systems and Simulations . . . . .	199
	Agent-based Models . . . . .	202
	Issues with Agent-based Models. . . . .	207
	Evaluation of an Agent-based Model . . . . .	209
	Concluding Remarks. . . . .	216
	Appendix: Software Programs . . . . .	220

# **List of Examples**

1	The Nature of Data in Tourism	
	Example: Usage of primary and secondary data . . . . .	8
	Example: Selecting a dataset . . . . .	15
2	Testing Hypotheses	
	Example: Statistical hypothesis testing. . . . .	25
	Example: Statistical power calculation. . . . .	27
	Example: Bootstrap calculations . . . . .	30
	Example: Combining observations. . . . .	34
3	Data Analysis	
	Example: Factor analysis in tourism research . . . . .	44
	Example: Scree plots . . . . .	48
	Example: Rotation . . . . .	53
	Extended example: Evaluating a destination website. . . . .	54
	Example: Dimensions of wine tourism . . . . .	71
	Example: Tourist market clusters. . . . .	74
	Example: Clusters . . . . .	75
	Example: Using reports and plots to guide selection of the number of clusters. . . . .	80
	Example: Preferences of tourists . . . . .	84
4	Model Building	
	Example: A pitfall for the impatient modeler . . . . .	93
	Example: Assessing residuals plots . . . . .	95
	Example: Is it never too hot for tourism? . . . . .	100
	Example: Best subsets method for choosing a parsimonious model . . . . .	106
	Example: Interpreting the results of binary logistic regression analysis . . . . .	110

Example: A classification table for destination return intentions . . . . .	111
Example: Structural equation modeling of intention to return . . . . .	113
Example: Specifying the measurement model . . . . .	122
Example: A complete structural equation model . . . . .	125
Example: Total, direct and indirect effects . . . . .	128
Example: Options available to the destination intention modeler . . . . .	129
Example: Mediation and moderation. . . . .	131
Example: The mediating role of satisfaction between service quality and intention. . . . .	132
 5 Time-dependent Phenomena and Forecasting	
Example: Simple decomposition of a time series . . . . .	138
Example: Mixed forecasts . . . . .	152
Example: Correlation between time series . . . . .	154
Example: Causal correlations. . . . .	156
Example: Stationarity tests . . . . .	160
Example: Non-linear analysis . . . . .	164
 6 Maximum Likelihood Estimation	
Example: Fitting a probability distribution from empirical data . . . . .	179
Example: Incomplete data . . . . .	182
Example: Fitting a power-law distribution . . . . .	183
Example: Choice of regression models . . . . .	187
 7 Monte Carlo Methods	
Example: Profit of an investment . . . . .	194
Example: Solar panel output power . . . . .	195
 8 Agent-based Modeling and Simulations	
Example: ABM applications . . . . .	204
Example: Simulating the arrivals of tourists . . . . .	210

## **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. Rodolfo has held several courses and lectures at national and international level and has carried out consulting activities for private and public tourism organizations. 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 to the study of tourism destinations.

**Jane Klobas** is Alberto Dondena Research Fellow at the Carlo Dondena Centre for Research on Social Dynamics at Bocconi University in Milan, Italy, and Professor at the University of Western Australia Business School. She teaches quantitative research methods to undergraduate, masters and doctoral degree students in Italy and Australasia, and conducts applied research using both qualitative and quantitative research methods. She is author or co-author of several books and book chapters, and has published research on the psychology and management of technology-mediated learning and knowledge sharing in many journals, including *Internet Research*, *Journal of Information Science*, *Computers and Education*, *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 and a Master in Development Economics from the University of East Anglia. He is currently a PhD candidate at the University of East Anglia where he is also a teaching assistant. He is a member of the Resilience Alliance Young Scientists (RAYS) group. His current PhD research is funded by the Economic and Social Research Council (ESRC) and focuses on the analysis of social-ecological systems with a resilience perspective. He is using agent-based modeling and network analysis methods for these investigations.

## **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, statistics and the quantitative approach are often labeled as 'disagreeable'. Yet, if tourism is to mature effectively as a subject, we cannot hide from the demands of quantitative approaches.

At a stroke, 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' armory.

Chris Cooper  
*Oxford Brookes University*  
*Oxford*



# **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, rationalize 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 favorable conditions and to have a deeper knowledge of the physical world.

Formally, statistics has the objective of collecting, analyzing 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. (p. 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 endeavor 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 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 specialized 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 rigor, 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 to 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 they will find extensive mathematical and conceptual treatment for the different topics to satisfy their curiosities or needs 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 in 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 usage in tourism studies. These techniques are gaining wide attention and a reputation in many disciplines for the study of several types of system, 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 their 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 be applied well only 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 a wide usage 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 organized as follows.

## **The Nature of Data**

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**

A brief review of the basic statistical tests is provided. A discussion of the issues related to sampling techniques is then presented. The chapter extends the description to the family of non-parametric tests, seldom considered by much of the tourism literature, but quite useful in numerous cases. The final section deals with those cases in which a limited sample could create problems in a study, giving the most important methods usable to extract significant information even from a small amount of data, or how to combine different studies.

## **Data Analysis**

This chapter deals with issues related to the processing of data in order to examine relationships and patterns. Data description, reduction

and segmentation for predicting the groups that individual cases fall into are described by using factor analysis, cluster analysis and multidimensional scaling.

### **Model Building**

Several methods are employed for the building of a quantitative relation combining a set of empirical data. Regression models and structural equation modeling are of fundamental importance in all areas of science. Many of these are well known and used in the tourism field. The chapter discusses these methods, highlighting the issues related to the computational techniques used and the reliability of results in different conditions.

### **Time-dependent Phenomena and Forecasting**

This chapter contains a very brief discussion of the main time series analysis methods that are typically used for forecasting purposes. The availability of many other good works is acknowledged and the reader is referred to these for more extensive treatment. Different uses of time series are discussed, such as simple non-linear analysis techniques to provide indicators for some characteristics of a system's structure.

*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 analyzing 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 '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.

They 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 one hundred or one thousand times higher.

The content of this part is organized as follows.

### **Maximum Likelihood Estimation**

The idea behind maximum likelihood estimation is to employ a generalized procedure able to find the value of one or more parameters for a given statistic that makes its likelihood distribution a maximum. These provide efficient methods for quantifying uncertainty and assessing 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.

### **Agent-based Modeling and Simulations**

Agent-based modeling and numerical simulations are means that facilitate exploring the structural and dynamic characteristics of systems that may prove intractable with analytical methods. The chapter examines the issues related to them with particular attention to their use in the study of socio-economic systems. Besides a general description, the possibilities, limitations and their relationship with other more traditional investigation methods are examined. Special focus is put on the assessment of their validation and reliability.

### **References**

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