

Statistical Analysis of Management Data

Hubert Gatignon

Statistical Analysis of Management Data

Third Edition



Springer

Hubert Gatignon
INSEAD
Fontainebleau Cedex, France

Statistical Analysis of Management Data. 1st Edition. Kluwer Academic Publishers, 2003
Statistical Analysis of Management Data. 2nd Edition. Springer Science+Business Media,
LLC, 2010

ISBN 978-1-4614-8593-3 ISBN 978-1-4614-8594-0 (eBook)
DOI 10.1007/978-1-4614-8594-0
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013945080

© Springer Science+Business Media New York 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my daughters, Aline and Valérie

Preface

Preface to First Edition

I am very indebted to a number of people without whom I would not have envisioned this book. First, Paul Green helped me tremendously in the preparation of the first doctoral seminar I taught at the Wharton School. The orientations and objectives set for that book reflect those he had for the seminar on data analysis which he used to teach before I did. A second individual, Lee Cooper at UCLA, was determinant in the approach I used for teaching statistics. As my first teacher of multivariate statistics, the exercise of having to program all the methods in APL taught me the benefits of such an approach for the complete understanding of this material. Finally, I owe a debt to all the doctoral students in the various fields of management, both at Wharton and INSEAD, who have, by their questions and feedback, helped me develop this approach. I hope it will benefit future students in learning these statistical tools, which are basic to academic research in the field of management especially. Special thanks go to Bruce Hardie who helped me put together some of the databases and to Frédéric Dalsace who carefully identified sections that needed further explanation and editing. Also, my research assistant at INSEAD, Gueram Sargsyan, was instrumental in preparing the examples used in this manual to illustrate the various methods.

Preface to Second Edition

This second edition reflects a slight evolution in the methods for analysis of data for research in the field of management and in related fields in the social sciences. In particular, it places a greater emphasis on measurement models. This new version includes a separate chapter on confirmatory factor analysis, with new sections on second order factor analytic models and multiple group factor analysis. A new, separate section on analysis of covariance structure discusses multigroup problems

that are particularly useful for testing moderating effects. Some fundamental multivariate methods such as canonical correlation analysis and cluster analysis have also been added. Canonical correlation analysis is useful because it helps better understand other methodologies already covered in the first version of this book. Cluster analysis remains a classic method used across fields and in applied research.

The philosophy of the book remains identical to that of its original version, which I have put in practice continuously in teaching this material in my doctoral classes. The objectives articulated in Chap. 1 have guided the writing of the first edition of this book but also of this new edition.

In addition to all the individuals I am indebted to and who have been identified in the first edition of this book, I would like to express my thanks to the cohorts of students since then. The continuous feedback has helped select the new material covered in this book with the objective to improve the understanding of the material. Finally, I would like to thank my assistant of fifteen years, Georgette Duprat whose commitment to detail never fails.

Preface to Third Edition

The methods for analyzing data are evolving rapidly as are the software packages that are available. On the one hand, this software, combined with more sophisticated hardware, is increasingly user-friendly. On the other hand, the theories that are being empirically tested and the large databases that have become more easily available require more complex statistical methodologies. While preserving the original objective to provide foundations for the analysis of such data, this third edition develops further those methodologies that are particularly well suited to data analysis in the social sciences. This explains the extensive new chapter on the analysis of mediation and moderation effects. For each of these methods, this edition also contains illustrations of analysis using STATA. I have also introduced XLSTAT as an alternative to multidimensional scaling because of its flexibility and ease of use as Excel macros. I would like to thank especially all my students at INSEAD who have provided feedback on the drafts of these chapters. Particular thanks go to Kathy Sheram who has advised me in editing the third edition of this book. Her professionalism and precision allowed me to communicate more clearly. This is particularly important for social scientists who may not have a technical background. Kathy contributed immensely to presenting the complex material of this book with concision, precision, and clarity.

Contents

1	Introduction	1
1.1	Overview	1
1.2	Objectives	2
1.2.1	Develop the Student’s Knowledge of the Technical Details of Various Techniques for Analyzing Data	2
1.2.2	Expose the Student to Applications and Hands-On Use of Various Computer Programs for Carrying Out Statistical Analyses of Data	3
1.3	Types of Scales	3
1.3.1	Definition of Different Types of Scales	4
1.3.2	The Impact of the Type of Scale on Statistical Analysis	4
1.4	Topics Covered	4
1.5	Pedagogy	7
	Bibliography	8
2	Multivariate Normal Distribution	9
2.1	Univariate Normal Distribution	9
2.2	Bivariate Normal Distribution	9
2.3	Generalization to Multivariate Case	11
2.4	Tests About Means	12
2.4.1	Sampling Distribution of Sample Centroids	12
2.4.2	Significance Test: One-Sample Problem	13
2.4.3	Significance Test: Two-Sample Problem	16
2.4.4	Significance Test: K-Sample Problem	17
2.5	Examples	19
2.5.1	Test of the Difference Between Two Mean Vectors: One-Sample Problem	19
2.5.2	Test of the Difference Between Several Mean Vectors: K-Sample Problem	21

2.6	Assignment	26
	Bibliography	29
3	Reliability Alpha, Principal Component Analysis, and Exploratory Factor Analysis	31
3.1	Notions of Measurement Theory	31
3.1.1	Definition of a Measure	31
3.1.2	Parallel Measurements	32
3.1.3	Reliability	32
3.1.4	Composite Scales	33
3.2	Exploratory Factor Analysis	36
3.2.1	Axis Rotation	36
3.2.2	Variance-Maximizing Rotations (Eigenvalues and Eigenvectors)	40
3.2.3	Principal Component Analysis	43
3.2.4	Exploratory Factor Analysis	46
3.3	Application Examples	51
3.3.1	Assignment	66
	Bibliography	75
4	Confirmatory Factor Analysis	77
4.1	Confirmatory Factor Analysis: A Strong Measurement Model	77
4.2	Estimation	79
4.2.1	Model Fit	81
4.2.2	Test of Significance of Model Parameters	84
4.2.3	Factor Scores	84
4.3	Summary Procedures for Scale Construction	84
4.3.1	Exploratory Factor Analysis	84
4.3.2	Confirmatory Factor Analysis	85
4.3.3	Reliability Coefficient Alpha	85
4.3.4	Discriminant Validity	85
4.3.5	Convergent Validity	85
4.4	Second-Order Confirmatory Factor Analysis	86
4.5	Multi-Group Confirmatory Factor Analysis	88
4.6	Application Examples	91
4.6.1	Example of Confirmatory Factor Analysis	91
4.6.2	Example of Model to Test Discriminant Validity Between Two Constructs	98
4.6.3	Example of Model to Assess the Convergent Validity of a Construct	111
4.6.4	Example of Second-Order Factor Model	123
4.6.5	Example of Multi-Group Factor Analysis	126
4.7	Assignment	151
	Bibliography	152

5	Multiple Regression with a Single Dependent Variable	155
5.1	Statistical Inference: Least Squares and Maximum Likelihood	155
5.1.1	The Linear Statistical Model	156
5.1.2	Point Estimation	157
5.1.3	Maximum Likelihood Estimation	159
5.1.4	Properties of Estimator	161
5.1.5	R-Squared as a Measure of Fit	166
5.2	Pooling Issues	169
5.2.1	Linear Restrictions	169
5.2.2	Pooling Tests and Dummy Variable Models	172
5.2.3	Strategy for Pooling Tests	174
5.3	Examples of Linear Model Estimation with SAS and STATA	176
5.4	Assignment	183
	Bibliography	185
6	System of Equations	187
6.1	Seemingly Unrelated Regression	187
6.1.1	Set of Equations with Contemporaneously Correlated Disturbances	187
6.1.2	Estimation	189
6.1.3	Special Cases	191
6.2	A System of Simultaneous Equations	191
6.2.1	The Problem	191
6.2.2	Two-Stage Least Squares (2SLS)	195
6.2.3	Three-Stage Least Squares (3SLS)	196
6.3	Simultaneity and Identification	197
6.3.1	The Problem	197
6.3.2	Order and Rank Conditions	198
6.4	Conclusion	200
6.4.1	Structure of Γ Matrix	200
6.4.2	Structure of Σ Matrix	201
6.4.3	Test of Covariance Matrix	202
6.4.4	Use of 3SLS Versus 2SLS	202
6.5	Examples of Estimation of Systems of Equations Using SAS and STATA	203
6.5.1	Seemingly Unrelated Regression Example	203
6.5.2	Two-Stage Least Squares Example	209
6.5.3	Three-Stage Least Squares Example	213
6.6	Assignment	215
	Bibliography	215
7	Canonical Correlation Analysis	217
7.1	The Method	217
7.1.1	Canonical Loadings	220
7.1.2	Canonical Redundancy Analysis	221

- 7.2 Testing the Significance of the Canonical Correlations 221
- 7.3 Multiple Regression as a Special Case of Canonical
Correlation Analysis 223
- 7.4 Examples 224
- 7.5 Assignment 230
- Bibliography 230
- 8 Categorical Dependent Variables 231**
 - 8.1 Discriminant Analysis 231
 - 8.1.1 The Discriminant Criterion 232
 - 8.1.2 Discriminant Function 235
 - 8.1.3 Classification and Fit 237
 - 8.2 Quantal Choice Models 240
 - 8.2.1 The Difficulties of the Standard Regression Model
with Categorical Dependent Variables 240
 - 8.2.2 Transformational Logit 241
 - 8.2.3 Conditional Logit Model 245
 - 8.2.4 Fit Measures 249
 - 8.3 Examples 251
 - 8.3.1 Example of Discriminant Analysis 251
 - 8.3.2 Example of Multinomial Logit: Case 1 Analysis
Using LIMDEP 259
 - 8.3.3 Example of Conditional Logit: Case 2 Analysis
Using LIMDEP and STATA 261
 - 8.4 Assignment 263
 - Bibliography 267
- 9 Rank-Ordered Data 269**
 - 9.1 Conjoint Analysis: MONANOVA 269
 - 9.1.1 Effect Coding Versus Dummy Variable Coding 269
 - 9.1.2 Design Programs 276
 - 9.1.3 Estimation of Part-Worth Coefficients 276
 - 9.2 Ordered Probit 278
 - 9.3 Examples 281
 - 9.3.1 Example of MONANOVA Using PC-MDS
and XLSTAT 281
 - 9.3.2 Example of Conjoint Analysis with Interval
Scale Rating Data 284
 - 9.3.3 Example of Ordered Probit Analysis Using LIMDEP 289
 - 9.4 Assignment 294
 - Bibliography 295
- 10 Error in Variables: Analysis of Covariance
Structure – Structural Equation Models 297**
 - 10.1 Impact of Imperfect Measures 297
 - 10.1.1 Effect of Errors-in-Variables 297
 - 10.1.2 Reverse Regression 299
 - 10.1.3 Case with Multiple Independent Variables 300

- 10.2 Analysis of Covariance Structures 301
 - 10.2.1 Description of Model 301
 - 10.2.2 Estimation 304
 - 10.2.3 Model Fit 307
 - 10.2.4 Test of Significance of Model Parameters 307
 - 10.2.5 Simultaneous Estimation of Measurement
Model Parameters with Structural Relationship
Parameters Versus Sequential Estimation 307
 - 10.2.6 Identification 308
 - 10.2.7 Special Cases of Analysis of Covariance Structure 308
- 10.3 Analysis of Covariance Structure with Means 310
- 10.4 Examples 312
 - 10.4.1 Example of Structural Model with Measurement
Models 312
- 10.5 Assignment 346
- Bibliography 346
- 11 Testing Mediation and Moderation Effects 349**
 - 11.1 Mediation vs. Moderation Effects 349
 - 11.1.1 Mediation Effects 349
 - 11.1.2 Moderation Effects 350
 - 11.1.3 Mediated Moderation and Moderated Mediation
Effects 352
 - 11.2 Testing Mediation Effects 354
 - 11.2.1 Baron and Kenny’s Procedure 354
 - 11.2.2 Best Practice 356
 - 11.2.3 Sequential Multiple Mediation Effects 370
 - 11.2.4 Testing Mediation When Constituent Paths
Are Nonlinear 373
 - 11.2.5 Experimental vs. Non-experimental Data 382
 - 11.2.6 Regression vs. Structural Equation Modeling 383
 - 11.2.7 Other Issues 401
 - 11.3 Testing Moderation Effects 404
 - 11.3.1 Moderated Regression 405
 - 11.3.2 Incorporating Moderating Effects in Analysis
of Covariance Structure 412
 - 11.4 Testing Moderated Mediation Effects 443
 - 11.5 Stating Mediation and Moderation Effect Hypotheses 446
 - 11.5.1 Stating Hypotheses About Mediation 446
 - 11.5.2 Stating Hypotheses About Moderation 446
 - 11.6 Assignment 447
 - Bibliography 447
- 12 Cluster Analysis 453**
 - 12.1 The Clustering Methods 453
 - 12.1.1 Similarity Measures 454
 - 12.1.2 The Centroid Method 454

12.1.3	Ward's Method	457
12.1.4	Nonhierarchical Clustering: K-Means Method	462
12.2	Examples	463
12.2.1	Example of Clustering with the Centroid Method	463
12.2.2	Example of Clustering with Ward's Method	471
12.2.3	Examples of K-Means Analysis	472
12.3	Evaluation and Interpretation of Clustering Results	472
12.3.1	Determining the Number of Clusters	476
12.3.2	Size, Density, and Separation of Clusters	478
12.3.3	Tests of Significance on Variables Other than Those Used to Create Clusters	483
12.3.4	Stability of Results	484
12.4	Assignment	484
	Bibliography	484
13	Analysis of Similarity and Preference Data	487
13.1	Proximity Matrices	487
13.1.1	Metric Versus Nonmetric Data	487
13.1.2	Unconditional Versus Conditional Data	488
13.1.3	Derived Measures of Proximity	488
13.1.4	Alternative Proximity Matrices	489
13.2	Problem Definition	489
13.2.1	Objective Function	490
13.2.2	Stress as an Index of Fit	491
13.2.3	Metric	491
13.2.4	Minimum Number of Stimuli	492
13.2.5	Dimensionality	492
13.2.6	Interpretation of MDS Solution	493
13.2.7	The KYST Algorithm	493
13.3	Individual Differences in Similarity Judgments	494
13.4	Analysis of Preference Data	495
13.4.1	Vector Model of Preferences	495
13.4.2	Ideal Point Model of Preferences	496
13.5	Examples	496
13.5.1	Example of KYST	496
13.5.2	Example of INDSCAL	501
13.5.3	Example of PROFIT (Property Fitting) Analysis	508
13.5.4	Example of MDPREF	517
13.5.5	Example of PREFMAP	524
13.6	Assignment	541
	Bibliography	542
14	Appendices	543
14.1	Appendix A: Rules in Matrix Algebra	543
14.1.1	Vector and Matrix Differentiation	543
14.1.2	Kronecker Products	543

- 14.1.3 Determinants 543
- 14.1.4 Trace 544
- 14.2 Appendix B: Statistical Tables 544
 - 14.2.1 Cumulative Normal Distribution 544
 - 14.2.2 Chi-Square Distribution 545
 - 14.2.3 F Distribution 546
- 14.3 Appendix C: Description of Data Sets 547
 - 14.3.1 The MARKSTRAT® Environment 547
 - 14.3.2 Marketing Mix Decisions 549
 - 14.3.3 Survey 551
 - 14.3.4 Indup 552
 - 14.3.5 Panel 552
 - 14.3.6 Scan 552
- Index 561**