

# Statistical Analysis of Management Data

Hubert Gatignon

# Statistical Analysis of Management Data

Second Edition

 Springer

Hubert Gatignon  
Boulevard de Constance  
INSEAD  
77305 Fontainebleau  
France  
hubert.gatignon@insead.edu

ISBN 978-1-4419-1269-5                      e-ISBN 978-1-4419-1270-1  
DOI 10.1007/978-1-4419-1270-1  
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2009937584

© Springer Science+Business Media, LLC 2010

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

*To my daughters, Aline and Valérie*

# 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 the 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 Chapter 1 have guided the writing not only 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 15 years, Georgette Duprat, whose commitment to details never fails.

# Preface

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.

# Contents

<b>1</b>	<b>Introduction</b>	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 Students to Applications and “Hand-On” Use of Various Computer Programs for Carrying Out Statistical Analyses of Data	2
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	5
1.5	Pedagogy	6
	Bibliography	8
<b>2</b>	<b>Multivariate Normal Distribution</b>	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	15
2.4.4	Significance Test: $K$ -Sample Problem	17
2.5	Examples Using SAS	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	27
	Bibliography	28

Basic Technical Readings . . . . .	28
Application Readings . . . . .	28
<b>3 Reliability Alpha, Principle Component Analysis, and Exploratory Factor Analysis . . . . .</b>	<b>29</b>
3.1 Notions of Measurement Theory . . . . .	29
3.1.1 Definition of a Measure . . . . .	29
3.1.2 Parallel Measurements . . . . .	30
3.1.3 Reliability . . . . .	30
3.1.4 Composite Scales . . . . .	31
3.2 Exploratory Factor Analysis . . . . .	34
3.2.1 Axis Rotation . . . . .	34
3.2.2 Variance-Maximizing Rotations (Eigenvalues and Eigenvectors) . . . . .	35
3.2.3 Principal Component Analysis (PCA) . . . . .	39
3.2.4 Exploratory Factor Analysis (EFA) . . . . .	41
3.3 Application Examples Using SAS . . . . .	47
3.4 Assignment . . . . .	53
Bibliography . . . . .	56
Basic Technical Readings . . . . .	56
Application Readings . . . . .	57
<b>4 Confirmatory Factor Analysis . . . . .</b>	<b>59</b>
4.1 Confirmatory Factor Analysis: A Strong Measurement Model . . . . .	59
4.2 Estimation . . . . .	61
4.2.1 Model Fit . . . . .	62
4.2.2 Test of Significance of Model Parameters . . . . .	65
4.3 Summary Procedure for Scale Construction . . . . .	65
4.3.1 Exploratory Factor Analysis . . . . .	65
4.3.2 Confirmatory Factor Analysis . . . . .	66
4.3.3 Reliability Coefficient $\alpha$ . . . . .	66
4.3.4 Discriminant Validity . . . . .	66
4.3.5 Convergent Validity . . . . .	66
4.4 Second-Order Confirmatory Factor Analysis . . . . .	67
4.5 Multi-group Confirmatory Factor Analysis . . . . .	69
4.6 Application Examples Using LISREL . . . . .	72
4.6.1 Example of Confirmatory Factor Analysis . . . . .	72
4.6.2 Example of Model to Test Discriminant Validity Between Two Constructs . . . . .	73
4.6.3 Example of Model to Assess the Convergent Validity of a Construct . . . . .	78
4.6.4 Example of Second-Order Factor Model . . . . .	98
4.6.5 Example of Multi-group Factor Analysis . . . . .	114



- 4.7 Assignment . . . . . 120
- Bibliography . . . . . 121
  - Basic Technical Readings . . . . . 121
  - Application Readings . . . . . 121
- 5 Multiple Regression with a Single Dependent Variable . . . . . 123**
  - 5.1 Statistical Inference: Least Squares and Maximum Likelihood . . . . . 123
    - 5.1.1 The Linear Statistical Model . . . . . 123
    - 5.1.2 Point Estimation . . . . . 125
    - 5.1.3 Maximum Likelihood Estimation . . . . . 127
    - 5.1.4 Properties of Estimator . . . . . 129
    - 5.1.5 R-Squared as a Measure of Fit . . . . . 133
  - 5.2 Pooling Issues . . . . . 135
    - 5.2.1 Linear Restrictions . . . . . 135
    - 5.2.2 Pooling Tests and Dummy Variable Models . . . . . 138
    - 5.2.3 Strategy for Pooling Tests . . . . . 141
  - 5.3 Examples of Linear Model Estimation with SAS . . . . . 141
  - 5.4 Assignment . . . . . 147
  - Bibliography . . . . . 147
    - Basic Technical Readings . . . . . 147
    - Application Readings . . . . . 147
- 6 System of Equations . . . . . 151**
  - 6.1 Seemingly Unrelated Regression (SUR) . . . . . 151
    - 6.1.1 Set of Equations with Contemporaneously Correlated Disturbances . . . . . 151
    - 6.1.2 Estimation . . . . . 153
    - 6.1.3 Special Cases . . . . . 155
  - 6.2 A System of Simultaneous Equations . . . . . 155
    - 6.2.1 The Problem . . . . . 155
    - 6.2.2 Two-Stage Least Squares: 2SLS . . . . . 159
    - 6.2.3 Three-Stage Least Squares: 3SLS . . . . . 160
  - 6.3 Simultaneity and Identification . . . . . 160
    - 6.3.1 The Problem . . . . . 160
    - 6.3.2 Order and Rank Conditions . . . . . 161
  - 6.4 Summary . . . . . 163
    - 6.4.1 Structure of  $\Gamma$  Matrix . . . . . 163
    - 6.4.2 Structure of  $\Sigma$  Matrix . . . . . 163
    - 6.4.3 Test of Covariance Matrix . . . . . 164
    - 6.4.4 3SLS Versus 2SLS . . . . . 165
  - 6.5 Examples Using SAS . . . . . 165
    - 6.5.1 Seemingly Unrelated Regression Example . . . . . 165
    - 6.5.2 Two-Stage Least Squares Example . . . . . 176
    - 6.5.3 Three-Stage Least Squares Example . . . . . 176
  - 6.6 Assignment . . . . . 180

Bibliography . . . . .	184
Basic Technical Readings . . . . .	184
Application Readings . . . . .	184
<b>7 Canonical Correlation Analysis . . . . .</b>	<b>187</b>
7.1 The Method . . . . .	187
7.1.1 Canonical Loadings . . . . .	190
7.1.2 Canonical Redundancy Analysis . . . . .	190
7.2 Testing the Significance of the Canonical Correlations . . . . .	190
7.3 Multiple Regression as a Special Case of Canonical Correlation Analysis . . . . .	192
7.4 Examples Using SAS . . . . .	193
7.5 Assignment . . . . .	198
Bibliography . . . . .	198
Application Readings . . . . .	198
<b>8 Categorical Dependent Variables . . . . .</b>	<b>199</b>
8.1 Discriminant Analysis . . . . .	199
8.1.1 The Discriminant Criterion . . . . .	199
8.1.2 Discriminant Function . . . . .	202
8.1.3 Classification and Fit . . . . .	204
8.2 Quantal Choice Models . . . . .	208
8.2.1 The Difficulties of the Standard Regression Model with Categorical Dependent Variables . . . . .	208
8.2.2 Transformational Logit . . . . .	209
8.2.3 Conditional Logit Model . . . . .	212
8.2.4 Fit Measures . . . . .	215
8.3 Examples . . . . .	217
8.3.1 Example of Discriminant Analysis Using SAS . . . . .	217
8.3.2 Example of Multinomial Logit – Case 1 Analysis Using LIMDEP . . . . .	223
8.3.3 Example of Multinomial Logit – Case 2 Analysis Using LIMDEP . . . . .	225
8.4 Assignment . . . . .	227
Bibliography . . . . .	227
Basic Technical Readings . . . . .	227
Application Readings . . . . .	228
<b>9 Rank-Ordered Data . . . . .</b>	<b>231</b>
9.1 Conjoint Analysis – MONANOVA . . . . .	231
9.1.1 Effect Coding Versus Dummy Variable Coding . . . . .	231
9.1.2 Design Programs . . . . .	238
9.1.3 Estimation of Part-Worth Coefficients . . . . .	238
9.2 Ordered Probit . . . . .	239
9.3 Examples . . . . .	243
9.3.1 Example of MONANOVA Using PC-MDS . . . . .	243

9.3.2	Example of Conjoint Analysis Using SAS . . . . .	244
9.3.3	Example of Ordered Probit Analysis Using LIMDEP . . . . .	246
9.4	Assignment . . . . .	248
Bibliography	. . . . .	250
Basic Technical Readings	. . . . .	250
Application Readings	. . . . .	250
<b>10</b>	<b>Error in Variables – Analysis of Covariance Structure . . . . .</b>	<b>253</b>
10.1	The Impact of Imperfect Measures . . . . .	253
10.1.1	Effect of Errors-in-Variables . . . . .	253
10.1.2	Reversed Regression . . . . .	255
10.1.3	Case with Multiple Independent Variables . . . . .	256
10.2	Analysis of Covariance Structures . . . . .	257
10.2.1	Description of Model . . . . .	257
10.2.2	Estimation . . . . .	259
10.2.3	Model Fit . . . . .	262
10.2.4	Test of Significance of Model Parameters . . . . .	263
10.2.5	Simultaneous Estimation of Measurement Model Parameters with Structural Relationship Parameters Versus Sequential Estimation . . . . .	263
10.2.6	Identification . . . . .	263
10.2.7	Special Cases of Analysis of Covariance Structure . . . . .	264
10.3	Analysis of Covariance Structure with Means . . . . .	266
10.4	Examples of Structural Model with Measurement Models Using LISREL . . . . .	267
10.5	Assignment . . . . .	268
Bibliography	. . . . .	291
Basic Technical Readings	. . . . .	291
Application Readings	. . . . .	291
<b>11</b>	<b>Cluster Analysis . . . . .</b>	<b>295</b>
11.1	The Clustering Methods . . . . .	295
11.1.1	Similarity Measures . . . . .	296
11.1.2	The Centroid Method . . . . .	296
11.1.3	Ward’s Method . . . . .	300
11.1.4	Nonhierarchical Clustering: <i>K</i> -Means Method (FASTCLUS) . . . . .	305
11.2	Examples Using SAS . . . . .	306
11.2.1	Example of Clustering with the Centroid Method . . . . .	306
11.2.2	Example of Clustering with Ward’s Method . . . . .	310
11.2.3	Example of FASTCLUS . . . . .	310
11.3	Evaluation and Interpretation of Clustering Results . . . . .	312
11.3.1	Determining the Number of Clusters . . . . .	312
11.3.2	Size, Density, and Separation of Clusters . . . . .	320
11.3.3	Tests of Significance on Other Variables than Those Used to Create Clusters . . . . .	320

11.3.4	Stability of Results . . . . .	320
11.4	Assignment . . . . .	321
Bibliography	. . . . .	321
Basic Technical Readings	. . . . .	321
Application Readings	. . . . .	321
<b>12</b>	<b>Analysis of Similarity and Preference Data</b> . . . . .	<b>323</b>
12.1	Proximity Matrices . . . . .	323
12.1.1	Metric Versus Nonmetric Data . . . . .	323
12.1.2	Unconditional Versus Conditional Data . . . . .	324
12.1.3	Derived Measures of Proximity . . . . .	324
12.1.4	Alternative Proximity Matrices . . . . .	324
12.2	Problem Definition . . . . .	325
12.2.1	Objective Function . . . . .	326
12.2.2	Stress as an Index of Fit . . . . .	326
12.2.3	Metric . . . . .	327
12.2.4	Minimum Number of Stimuli . . . . .	328
12.2.5	Dimensionality . . . . .	328
12.2.6	Interpretation of MDS Solution . . . . .	328
12.2.7	The KYST Algorithm . . . . .	329
12.3	Individual Differences in Similarity Judgments . . . . .	330
12.4	Analysis of Preference Data . . . . .	331
12.4.1	Vector Model of Preferences . . . . .	331
12.4.2	Ideal Point Model of Preferences . . . . .	331
12.5	Examples Using PC-MDS . . . . .	332
12.5.1	Example of KYST . . . . .	332
12.5.2	Example of INDSCAL . . . . .	335
12.5.3	Example of PROFIT (Property Fitting) Analysis . . . . .	341
12.5.4	Example of MDPREF . . . . .	350
12.5.5	Example of PREFMAP . . . . .	356
12.6	Assignment . . . . .	358
Bibliography	. . . . .	368
Basic Technical Readings	. . . . .	368
Application Readings	. . . . .	368
<b>13</b>	<b>Appendices</b> . . . . .	<b>369</b>
Appendix A: Rules in Matrix Algebra . . . . .		369
Vector and Matrix Differentiation . . . . .		369
Kronecker Products . . . . .		369
Determinants . . . . .		369
Trace . . . . .		369
Appendix B: Statistical Tables . . . . .		370
Cumulative Normal Distribution . . . . .		370
Chi-Squared Distribution . . . . .		370
F Distribution . . . . .		371
Appendix C: Description of Data Sets . . . . .		372

The MARKSTRAT® Environment . . . . .	373
Marketing Mix Decisions . . . . .	375
Survey . . . . .	376
Indup . . . . .	381
Panel . . . . .	381
Scan . . . . .	382
Bibliography . . . . .	384
<b>About the Author</b> . . . . .	385
<b>Index</b> . . . . .	387