

AI3

*Web content*



Thanks Neverill for the cover photo.

Angelo Riva

**Economic, Business  
and Marketing Research Method**

Advanced Multivariate Statistics and Econometric Research Method  
for Social, Management, Marketing and Economic Sciences





Aracne editrice

[www.aracneeditrice.it](http://www.aracneeditrice.it)  
[info@aracneeditrice.it](mailto:info@aracneeditrice.it)

Copyright © MMXVIII  
Giacchino Onorati editore S.r.l. – unipersonale

[www.giacchinoonoratieditore.it](http://www.giacchinoonoratieditore.it)  
[info@giacchinoonoratieditore.it](mailto:info@giacchinoonoratieditore.it)

via Vittorio Veneto, 20  
00020 Canterano (Rome)  
(06) 45551463

ISBN 978-88-255-1110-9

*No part of this book may be reproduced  
by print, photoprint, microfilm, microfiche, or any other means,  
without publisher's authorization.*

1<sup>st</sup> edition: May 2018

*"Whatever you dream of undertaking, start it.  
Boldness has genius, power, magic."*  
J. W. Von Goethe

*"If I have seen it so far it is because  
I climbed on the shoulders of the giants "*  
I. Newton



**Introduction** 9

**PART I  
RESEACH METHOD IN SOCIAL, MANAGEMENT,  
ECONOMIC SCIENCE USING  
STATISTICS, ECONOMETRICS AND  
MULTIVARIATE ANALYSIS**

**1. Research planning and data analysis** 17

**2. Correlation** 29

**3. Regression** 37

**4. Comparing two means (t-test)** 45

**5. Comparing several means ANOVA (GLM1)** 55

**6. Analysis of covariance ANCOVA (GLM2)** 61

**7. Factorial ANOVA (GLM 3)** 67

**8. Repeated measures design (GLM4)** 71

**9. Mixed design ANOVA (GLM5)** 77

**10. Multivariate analysis of variance** 83  
**(MANOVA)**

**11. Discriminant analysis** 89

**12. Principal components analysis and factor analysis** 95

<b>13. Analysis of the cluster</b>	<b>101</b>
<b>14. Multidimensional Scalling</b>	<b>105</b>
<b>15. Correspondence Analysis, Conjoint Analysis, Canonical Correlation</b>	<b>109</b>
<b>16. Time series</b>	<b>115</b>

## **PART II**

### **APPLICATION OF ADVANCED MULTIVAR- IATE STATISTICS AND ECONOMETRIC: A CASE STUDY**

<b>1. Introduction</b>	<b>129</b>
<b>2.Theoretical background and methodological approach</b>	<b>134</b>
<b>3.Methodology and data</b>	<b>137</b>
<b>4. Result</b>	<b>138</b>
<b>5. Discussion</b>	<b>147</b>

<b>APPENDIX</b>	<b>151</b>
-----------------	------------

<b>REFERENCES</b>	<b>171</b>
-------------------	------------

## Introduction

While there is a broad range of literature on multivariate statistics only few books are written on the process in the topic on social, management<sup>1</sup> and economic sciences.

With this in mind, the primary aim of the present research is explored how to apply the main methodology of statistics, econometric and multivariate techniques in these field of application<sup>2</sup>.

A business research is (Zichmund 2010): “*The application of scientific method in searching for the trust about business phenomena. These activities included defining the business opportunities and problems, generating and evaluating ideas, monitoring performance, and understand the business process*”.

The three main key research questions of this book are:

*Q1: What are the evolution of the main research method in social, management, economic science using statistics, econometrics and multivariate analysis?*

*Q2: What are the best methodology to do good research in business and economy and marketing and social science?*

*Q1: What are some possible applications of this methodology?*

---

<sup>1</sup> See also Saunders, M.N.K., Lewis, P., Thornhill, A. (2012) *Research Methods for Business Students* (6th edition), Harlow, England: Pearson Education, Inc. Ghauri, P., & Grønhaug, K. (2002), *Research Methods in Business Studies : A Practical Guide* (2nd edition), London: FT Prentice Hall Europe; Easterby-Smith, M., Thorpe, R., Jackson, P.R. (2008) *Management Research* (3rd edition). London: Sage; Casse, C., & Lee, B. (eds) (2011) *Challenges and Controversies in Management Research*. New York, NY: Routledge; Zikmund W.G. (2010), *Business research method*, Mason.

<sup>2</sup> See also Agresti A. and Finlay B. (1997), *Statistical methods for the social sciences*, third edition, Prentice Hall, Upper Saddle River, NJ: Brambor T., Clark W.R. and Golder M. (2006), “Understanding Interaction Models: Improving Empirical Analyses”, *Political Analysis*, 14: 63-82.; Corbetta P., Gasperoni G. and Pisati M. (2001), *Statistica per la ricerca sociale*, il Mulino, Bologna; Hamilton L.C. (2012), *Statistics with Stata: updated for version 12*, Belmont, CA, Brooks/Cole, Cengage Learning; Miles J. and Shevlin M. (2001), *Applying regression & correlation: a guide for students and researchers*, Los Angeles, Sage Publications; Menard, S. [2002], *Applied logistic regression analysis*, second edition, Thousand Oaks, CA, Sage;

*For answering the first question and second question (in the first part), we discuss and compare in sequence a set of important methodology. The methodologies analyzed follow the structure of the book:*

- 1. Research planning and data analysis (chapter 1)*
- 2. Correlation (chapter 2)*
- 3. Regression (chapter 3)*
- 4. Comparing two means (t-test) (chapter 4)*
- 5. Comparing several means ANOVA (GLM1) (chapter 5)*
- 6. Analysis of covariance ANCOVA (GLM2) (chapter 6)*
- 7. Factorial ANOVA (GLM 3) (chapter 7)*
- 8. Repeated measures design (GLM4) (chapter 8)*
- 9. Mixed design ANOVA (GLM5) (chapter 9)*
- 10. Multivariate analysis of variance (MANOVA) (chapter 10)*
- 11. Discriminant analysis (chapter 11)*
- 12. Principal components analysis and factor analysis (chapter 12)*
- 13. Analysis of the cluster (chapter 13)*
- 14. Multidimensional Scalling (chapter 14)*
- 15. Correspondence Analysis, Conjoint Analysis, Canonical Correlation (chapter 15)*
- 16. Time series (chapter 16)*

*For answering thirds question, we analyze a practical application (second part) of the correct method to develop research in business in special application.*

*The book is designed to develop focus on conducting a good research. It is about research techniques and research methods. It analyses the stages of developing a good research proposal. It will appeal to people who require developing a good understanding of research approaches and skills, and would like to pursue it in their professional life. The book analyses the project of research (from research question to research methods to developing and testing hypothesis).*

*In particular, this book aims for professional researches and it can be used also at the graduate and postgraduate students in Business and Economics Marketing and other Social Science. But these methodologies can be applied in any kind of science (engineering, physics..). For some of these techniques a good knowledge and experience in economics, statistic and econometrics research is required.*



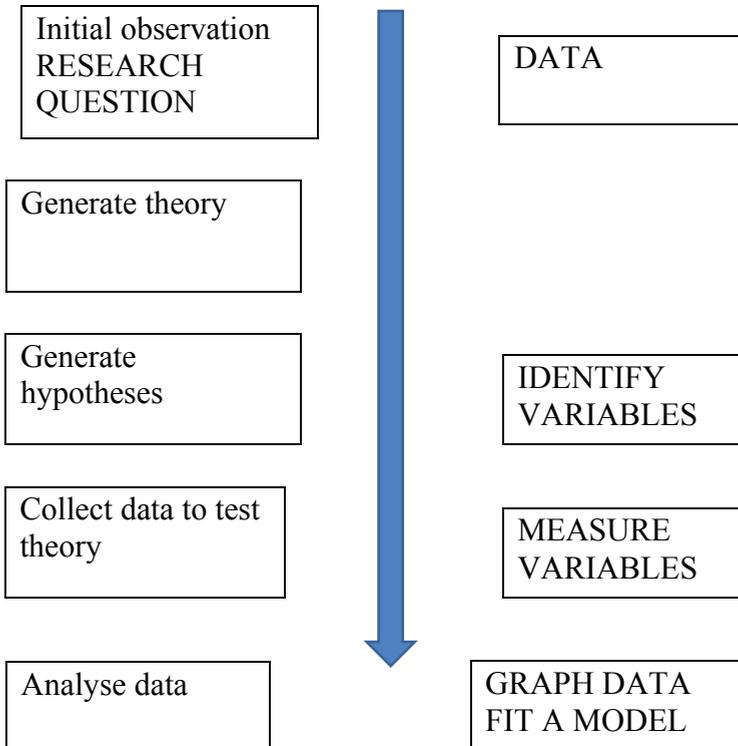
**Figure: Research cycle:** Source: Napier University

For each multivariate technique there is a specific type of research question: some techniques also have strengths and weaknesses. Current statistical packages<sup>3</sup> permit to run a procedure, but the results should be analyzed and understanding the limit of the techniques.

In this work, the aim is to analyze a set of multivariate analysis techniques about statistics and econometrics. As a matter of fact it is important to understanding the types of research questions and the limits of each methodology in answering those questions. For doing an analysis technique, it is relevant to have a clear understanding of the

<sup>3</sup> For example the package R, Stata, SPSS, S-Plus, SAS and others; Field A. (2013), Discover statistic using ibm SPSS statistics, Sage edition.

form (nonmetric or metric) and quality of the data (how normally distributed the data are).



**Fig. The research process.** Source: Adaption from Field, 2013

Some techniques are sensitive to the linearity, normality distribution, and equal variance assumptions of the data. It is important to test: distribution, skewness, and kurtosis, the magnitude of missing values in observations, outliers. In some field research analyst now has access to a much broader array of sophisticated techniques with which to explore

the data<sup>4</sup>. The challenge becomes knowing which technique to select, and clearly strengths and understanding their weaknesses.



**Figure: Planning the research** Source: National Foundation of Educational Research

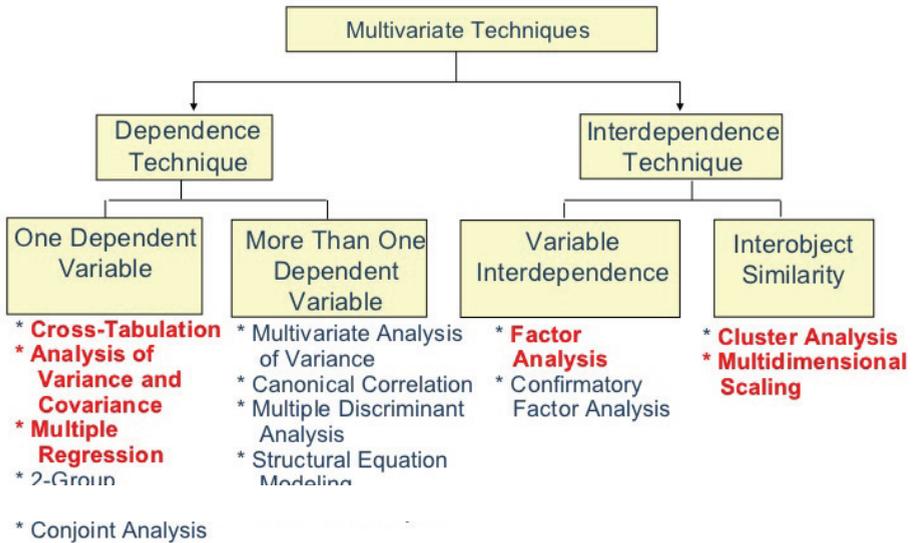
The book gains an appreciation of the types of problems and questions arising with statistics econometrics and multivariate data. In particular the research should be able to apply and interpret methods of dimension reduction including principal component analysis, multidimensional scaling, factor analysis, canonical variates, cluster analysis and discrimination.

The simplest statistic has to do with a single random variable: we can collect a lot of data (we can then summarize these data by calculating mean, variance, etc), we can also compare two different collections

<sup>4</sup> Cooley, W.W. and P. R. Lohnes. (1971). *Multivariate Data Analysis*. John Wiley & Sons, Inc. George H. Dunteman (1984). *Introduction to multivariate analysis*. Thousand Oaks, CA: Sage Publications. Morrison D.F. (1967). *Multivariate Statistical Methods*. McGraw-Hill: New York. Overall, J.E. and C.J. Klett. (1972). *Applied Multivariate Analysis*. McGraw-Hill: New York. Tabachnick, B.G. and L.S. Fidell. (1996). *Using Multivariate Statistics*. Harper Collins College Publishers: New York

of data for the same random variable as is done in the case of control groups.

### A Classification of Multivariate Techniques



**Figure: Multivariate technique** Source: Marketing Research

This book is organized<sup>5</sup> in two part: the first illustrate the main theory of statistics, econometrics and multivariate analysis, the second describe some empirical application.

---

<sup>5</sup> Shevlin M. (2001), *Applying regression & correlation: a guide for students and researchers*, Los Angeles, Sage Publications: .Menard, S. [2002], *Applied logistic regression analysis*, second edition, Thousand Oaks, CA, Sage: Kohler, U. and Kreuter, F. [2012], *Data analysis using stata*, third edition, College Station, Texas: Stata Press: Long, J.S. e Freese, J. [2006], *Regression models for categorical dependent variables using Stata*, second edition, College Station, Texas, Stata Press.; Montgomery (1996), *Introduction to statistical quality control*, 3 ed. John Wiley & Son.

**PART I**

**RESEACH METHOD IN SOCIAL, MANAGEMENT,  
ECONOMIC SCIENCE USING  
STATISTICS, ECONOMETRICS AND  
MULTIVARIATE ANALYSIS**



# Chapter I

## Research planning: data and hypothesis testing

### Research planning

Before starting a research is important to analysis some questions<sup>6</sup>:

1. *what is the aim of the research?*
- 2 *what is the literature review and why and how focus do you need to do one?*
3. *how to develop a research problem?*
4. *how to develop testable hypothesis?*
5. *how to find a research method vs data collection method?*
6. *which strategies choose for social/business research?*
7. *which methods choose for social/business research?*
8. *how to do the analysis?*
9. *how to write up the research?*

In statistic and econometric<sup>7</sup> is important the hypothesis testing, for example understand if two different organization methods result in different firm performances.

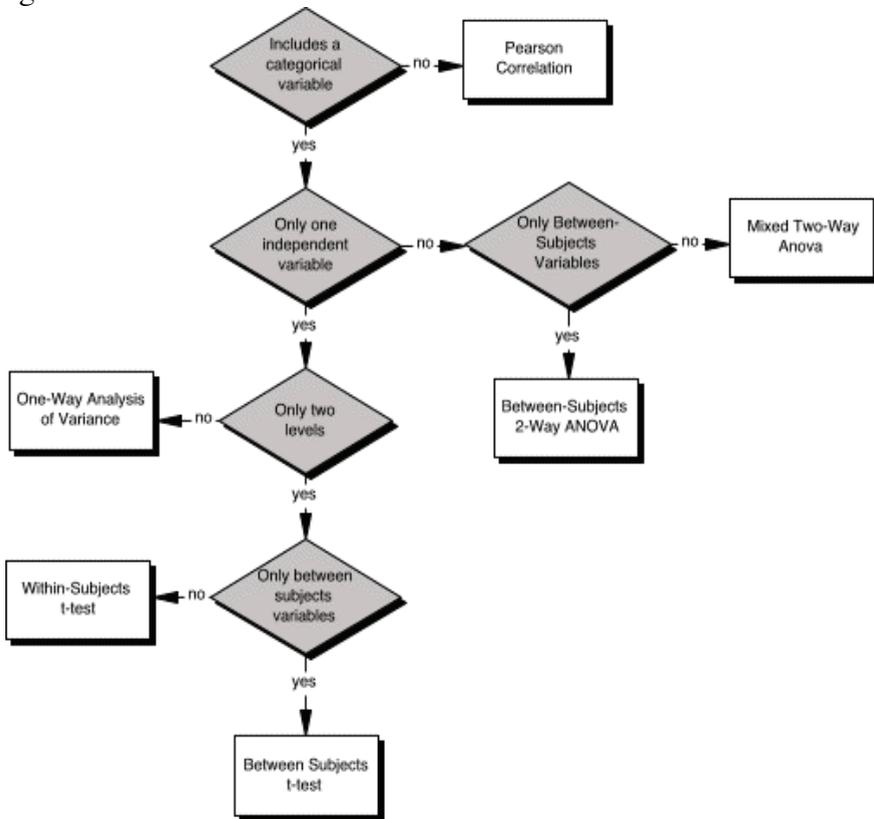
Another example might be that there is no relationship between profit and quality. The alternative hypothesis states the opposite and is usually the hypothesis you are trying to prove (e.g., the two different

---

<sup>6</sup> See also Easterby-Smith, M., Thorpe, R., Jackson, P.R. (2008) *Management Research* (3rd edition). London: Sage; Casse, C., & Lee, B. (eds) (2011) *Challenges and Controversies in Management Research*. New York, NY: Routledge; Zikmund W.G. (2010), *Business research method*, MasonSaunders, M.N.K., Lewis, P., Thornhill, A. (2012) *Research Methods for Business Students* (6th edition), Harlow, England: Pearson Education, Inc. Ghauri, P., & Grønhaug, K. (2002), *Research Methods in Business Studies : A Practical Guide* (2nd edition), London: FT Prentice Hall Europe;.

<sup>7</sup> See also Kohler, U. and Kreuter, F. [2012], *Data analysis using stata*, third edition, College Station, Brambor T., Clark W.R. and Golder M. (2006), "Understanding Interaction Models: Improving Empirical Analyses", *Political Analysis*, 14: 63-82.; Corbetta P., Gasperoni G. and Pisati M. (2001), *Statistica per la ricerca sociale*, il Mulino, Bologna; Hamilton L.C. (2012), *Statistics with Stata: updated for version 12*, Belmont, CA, Brooks/Cole, Cengage Learning;. Miles J. and Shevlin M. (2001), *Applying regression & correlation: a guide for students and researchers*, Los Angeles, Sage Publications;. Menard, S. [2002], *Applied logistic regression analysis*, second edition, Thousand Oaks, CA, Sage: Texas: Stata Press: Long, J.S. e Freese, J. [2006], *Regression models for categorical dependent variables using Stata*, second edition, College Station, Texas, Stata Press.

management methods did result in different performance). The level of statistical significance is based as the so-called  $p$ -value. Let us consider this statement with respect to our example where we are interested in the difference in mean exam performance between two different teaching methods.



**Figure: Flow chart representing choice of analysis depending on design** Source : mst.edu

It is possible to test if there is no difference between the two managerial methods how likely would it be to see a difference in the mean performance between the two management methods. The interest of statistical inference is to draw conclusions on the population and some on the parameters. For this purpose, statistics<sup>8</sup> calculated on the

<sup>8</sup> See also: Morrison D.F. (1967). *Multivariate Statistical Methods*. McGraw-Hill: New York. Overall, J.E. and C.J. Klett. (1972). *Applied Multivariate Analysis*.

basis of the random sample are used to estimate the values of the corresponding parameters of the entire population

The sample mean is a statistic used to estimate the average of a characteristic of interest (eg. the weight of the product unit) referred to the entire population (a given production process).

**Data analysis and matrix**

The starting point of any analysis of multivariate data is a data matrix.

OBSERVATION	VAR. 1 (Y)	VAR. 2	VAR. 3	VAR. 4	VAR. 5	VAR. 6	VAR. 7
1 (X)							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
n							

**Figure: data analysis and matrix**

Each row of X corresponds to an observed unit. Each column of Y corresponds to an observed variable. Starting from the data matrix<sup>9</sup> is a

---

McGraw-Hill: New York. Tabachnick, B.G. and L.S. Fidell. (1996). Using Multivariate Statistics. Harper Collins College Publishers: New York; Cooley, W.W. and P. R. Lohnes. (1971). Multivariate Data Analysis. John Wiley & Sons, Inc. George H. Dunteman (1984). Introduction to multivariate analysis. Thousand Oaks, CA: Sage Publications.

<sup>9</sup> The probability is used to quantify the possibility that a given event will occur. For example, to determine whether there will be the sun in a day. The probability is quantified by assigning a number between 0 and 1 (0% and 100%). The higher the number the more the event is probable: 0 = impossible event 1 = certain event.

series of different matrices can be derived (relationships between variables, leaving the theme of measuring dissimilarities between units). In case of probability the density function  $f(x)$  is used to determine the probability base on tables.

<b>TYOLOGY OF DISCRETE FUNCTION</b>
BINOMIAL PROBABILITY DISTRIBUTION
HYPERGEOMETRIC PROBABILITY DISTRIBUTION
MULTINOMIAL PROBABILITY DISTRIBUTION
NEGATIVE BINOMIAL DISTRIBUTION
POISSON PROBABILITY DISTRIBUTION
<b>TYOLOGY OF PROBABILITY DISTRIBUTIONS</b>
NORMAL PROBABILITY DISTRIBUTION
STUDENT'S DISTRIBUTION
CHI-SQUARE DISTRIBUTION
F DISTRIBUTION

**Figure: probability function**

For example it is possible to get a  $p$ -value such as 0.01 (i.e.,  $p = .01$ ), there is a 1% chance of finding a difference given that the null hypothesis that is true. (1 times in 100 or less) that the difference in the mean management performance between the two management methods (reject the null hypothesis and accept the alternative hypothesis).

---