Contributions to Statistics

For further volumes:
http://www.springer.com/series/2912

Matilde Bini · Paola Monari · Domenico Piccolo ·
Luigi Salmaso
Editors

Statistical Methods for the Evaluation of Educational Services and Quality of Products

Physica-Verlag

Editors
Prof. Matilde Bini
University of Florence
Dept. of Statistics "G. Parenti"
Viale Morgagni, 59
50134 Firenze
Italy
bini@ds.unifi.it

Prof. Paola Monari
University of Bologna
Dept. Statistical Science
Via Belle Arti, 41
40126 Bologna
Italy
paola.monari@unibo.it

Prof. Domenico Piccolo
University of Naples Federico II
Dept. of Statistical Science
Via Leopoldo Rodinò, 22
80138 Napoli
Italy
domenico.piccolo@unina.it

Prof. Luigi Salmaso
University of Padua
Department of Management and Engineering
Stradella San Nicola, 3
36100 Vicenza
Italy
luigi.salmaso@unipd.it

ISBN 978-3-7908-2384-4 e-ISBN 978-3-7908-2385-1
DOI 10.1007/978-3-7908-2385-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009939931

Cover design: Integra Software Services Pvt. Ltd., Pondicherry

Printed on acid-free paper

Physica-Verlag is a brand of Springer-Verlag Berlin Heidelberg
Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The book presents statistical methods and models that can usefully support the evaluation of educational services and quality of products. The contributions collected in this book summarize the work of several researchers from the universities of Bologna, Firenze, Napoli and Padova. The contributions are written with a consistent notation and a unified view, and concern methodological advances developed mostly with reference to specific problems of evaluation using real data sets.

The evaluation of educational services, as well as the analysis of judgements and preferences, poses severe methodological challenges because of the presence of one or more of the following aspects: the observational (non experimental) nature of the context, which is associated with the well-known problems of selection bias and presence of nuisance factors; the hierarchical structure of the data, that entails correlated observations and consideration of effects at different levels of the hierarchy and their interactions (multilevel analysis); the multivariate and qualitative nature of the dependent variable, that requires the use of ad hoc statistical methodologies; the presence of non observable factors, e.g. the satisfaction, calling for the use of latent variables models; the simultaneous presence of components of pleasure and components of uncertainty in the explication of the judgments, that asks for the specification and estimation of mixture models.

The first part of the book deals with latent variable models. In many fields of application most of the variables under investigation are not directly observable, and hence not measurable. In this context latent variable models assume a prominent role. Traditionally, latent variable models were used in psychometrics and have been concerned with measurement error, and latent variable constructs measured with multiple indicators (factor analysis). Nowadays, latent variables are used to represent different phenomena, such as true variables measured with error, hierarchical and longitudinal data, unobserved heterogeneity and missing data. Chapters 2, 3 and 4 illustrate latent variable models with educational behaviour applications. Since the variables under investigation are abilities, initial status, or rate of change in temporal achievement, the models rely on continuous latent variables, but different types of observations can be considered. Latent variable models for hierarchical data, i.e. multilevel models, are considered in Chaps. 5 and 6. In particular, Chapter 5 reviews the use of multilevel models for value-added analysis in education. Chapter 6 describes the specification and estimation of a multilevel mixture factor model with continuous and categorical latent variables.

From a different point of view, Chap. 7 proposes an approach mainly based on individual perceptions about the discrete choices. In this framework, the latent process guiding the preferences and the judgements is represented by a mixture model. Extensions dealing with multi-attribute methods, such as conjoint analysis and choice modelling, are provided in Chap. 8, carrying out a brief and critical review in order to clarify the distinctions between the models as well as to point out their common issues.

A frequently encountered problem in fitting statistical models is the presence of outliers. Chapter 9 deals with a robust diagnostic approach known as Forward Search that detects the presence of outliers and assesses their influence on the estimates of the model parameters. In particular, the use of this approach is investigated in generalized liner models applied in studies on university performance evaluation.

The last chapters are devoted to nonparametric hypotheses testing via permutation methods for complex observational studies and to nonparametric construction of composite indicators. Chapter 10 presents a novel global performance score for the construction of a global performance index when the focus is at evaluating the product performances in connection with more than one aspect (dimension) and/or under several conditions (strata). Chapter 11 considers permutation methods for multivariate testing on ordered categorical variables within the framework of multivariate randomised complete block designs with application to a case study related to food sensorial evaluation. Chapter 12 is devoted to permutation tests for stochastic ordering problems where the main goal is to find out where the treatment peak is located (so called “umbrella alternative”). Chapter 13 deals with a novel method for constructing preference rankings based on the nonparametric combination procedure with application to the evaluation of professional profiles of municipal directors.

The Editors would like to thank all the people who, by their intensive research and aptitude of integration, have contributed to the realization of this book.

We thank Carla Rampichini of University of Florence for her precious collaboration to the editing work.

Matilde Bini - University of Firenze
Paola Monari - University of Bologna
Domenico Piccolo - University of Napoli Federico II
Luigi Salmaso - University of Padova

The Research Units were partially supported by a research grant from the Italian Ministry of University and Research (MIUR): PRIN 2006 “Statistical Methods and Models for the Evaluation of the Educational Processes”, by the University of Padova CPDA088513 and by the CFEPSR, Portici.

Contents

Contributors

Silvia Bacci
Department of Statistics “G. Parenti”, University of Florence, Viale Morgagni 59, 50134, Florence, Italy, s.bacci@ds.unifi.it

Dario Basso
Department of Management and Engineering, University of Padua Stradella S. Nicola 4, 36100, Vicenza, Italy, basso@gest.unipd.it

Rossella Berni
Department of Statistics “G. Parenti”, University of Florence, Viale Morgagni 59, 50134, Florence, Italy, berni@ds.unifi.it

Bruno Bertaccini
Department of Statistics “G. Parenti”, University of Florence, Viale Morgagni 59, 50134, Florence, Italy, brunob@ds.unifi.it

Silvia Bianconcini
Department of Statistical Sciences, University of Bologna, via Belle Arti 41, 40126 Bologna, Italy, silvia.bianconcini@unibo.it

Matilde Bini
Department of Statistics “G. Parenti”, University of Florence, Viale Morgagni 59, 50134, Florence, Italy, bini@ds.unifi.it

Mario Bolzan
Department of Statistics, University of Padua, Via Cesare Battisti 241, 35121- Padua, Italy, mbolzan@stat.unpd.it

Stefano Bonnini
Department of Mathematics, University of Ferrara, Via Machiavelli 35, 44100 Ferrara, Italy, bnnsfn@unife.it

Silvia Cagnone
Department of Statistical Sciences, University of Bologna, via Belle Arti 41, 40126 Bologna, Italy, silvia.cagnone@unibo.it

Livio Corain
Department of Management and Engineering, University of Padua, Stradella S.Nicola 3, 36100 Vicenza, Italy, livio.corain@unipd.it

Antonio Cordellina
Research & Development, Reckitt-Benckiser Ltd, Piazza S. Nicolo' 12/3, 30034 Mira (VE), Italy, Antonio.Cordellina@Reckittbenckiser.com

Marcella Corduas
Department of Statistical Sciences, University of Naples Federico II, Via Leopoldo Rodinò, 22 I-80138, Naples, Italy, marcella.corduas@unina.it

Anna Crestana
Research & Development, Reckitt-Benckiser Ltd, Piazza S. Nicolo' 12/3, 30034 Mira (VE), Italy, Anna.Crestana@Reckittbenckiser.com

Estela Bee Dagum
Department of Statistical Sciences, University of Bologna, via Belle Arti 41, 40126 Bologna, Italy, estela.beedagum@unibo.it

Rosa Arboretti Giancristofaro
Department of Territory and Agro-Forestal Systems, University of Padua, Viale dell'Univeristà 16, 35020 Legnaro (PD), Italy, rosa.arboretti@unipd.it

Caterina Giusti
Department of Statistics and Mathematics Applied to Economics, University of Pisa, Via C. Ridolfi 10, 56124, Pisa, Italy, caterina.giusti@ec.unipi.it

Leonardo Grilli
Department of Statistics "G. Parenti", University of Florence, Viale Morgagni 59, 50134, Florence, Italy, grilli@ds.unifi.it

Maria Iannario
Department of Statistical Sciences, University of Naples Federico II, Via Leopoldo Rodinò, 22 I-80138, Naples, Italy, maria.iannario@unina.it

Mariagiulia Matteucci
Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126 Bologna, Italy, m.matteucci@unibo.it

Stefania Mignani
Department of Statistical Sciences, University of Bologna, via Belle Arti 41, 40126 Bologna, Italy, stefania.mignani@unibo.it

Paola Monari
Department of Statistical Sciences, University of Bologna, via Belle Arti 41, 40126 Bologna, Italy, paola.monari@unibo.it

Irini Moustaki
Department of Statistical Science, London School of Economics, Houghton Street, London,WC2A 2AE UK, i.moustaki@lse.ac.uk

Remigio Musci
Research & Development, Reckitt-Benckiser Ltd, Piazza S. Nicolo’ 12/3, 30034 Mira (VE), Italy, Remigio.Musci@Reckittbenckiser.com

Fortunato Pesarin
Department of Statistics, University of Padua Via Cesare Battisti 241, 35121, Padua, Italy,
pesarin@stat.unipd.it

Domenico Piccolo
Department of Statistical Sciences, University of Naples Federico II, Via Leopoldo Rodinò, 22 I-80138, Naples, Italy, domenico.piccolo@unina.it

Carla Rampichini
Department of Statistics “G. Parenti”, University of Florence, Viale Morgagni 59, 50134, Florence, Italy, rampichini@ds.unifi.it

Riccardo Rivello
Department of Statistics “G. Parenti”, University of Florence, Viale Morgagni 59, 50134, Florence, Italy, rivello@ds.unifi.it

Luigi Salmaso
Department of Management and Engineering, University of Padua, Stradella S.Nicola 3, 36100 Vicenza, Italy, salmaso@gest.unipd.it

Roberta Varriale
Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands; Department of Statistics “G. Parenti”, University of Florence, Florence, Italy, roberta.varriale@ds.unifi.it

Bernard P Veldkamp
Department of Research Methodology, Measurement and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands,
b.p.veldkamp@gw.utwente.nl

Chapter 1
Introduction

Matilde Bini, Paola Monari, Domenico Piccolo and Luigi Salmaso

1.1 Generalized linear latent variable models

In many fields of application most of the variables under investigation are not directly observable and hence not measurable. In these contexts latent variable models assume a prominent role. Their origins can be traced back to the early twentieth century, notably in the study of human abilities. The main ideas lie behind factor analysis and the newer applications of linear structural models. An account of their innovative role in many fields to which statistical methods are applied can be found in Bartholomew (1995) and Bartholomew and Knott (1999). In the recent literature there have been several proposals for generalized latent variable modelling frameworks, integrating specific methodologies in a global theoretical context. One example is the Generalized Linear Latent And Mixed Models (GLLAMM) framework of Skrondal and Rabe-Hesketh (2004). This approach unifies and extends latent variable modelling as multilevel, longitudinal, and structural equation models as well as generalized linear mixed models, random coefficient models, item response models, factor models, and so on. Other two examples are Muthén (2008) and Vermunt (2007), both proposing general frameworks that allow to define models with any

Matilde Bini
Department of Statistics "G. Parenti", University of Florence, Viale Morgagni 59, 50134, Florence, Italy, e-mail: bini@ds.unifi.it

Paola Monari
Department of Statistical Sciences, University of Bologna, via Belle Arti 41, 40126 Bologna, Italy, e-mail: paola.monari@unibo.it

Domenico Piccolo
Department of Statistical Sciences, University of Naples Federico II, Via Leopoldo Rodinó, 22 I-80138, Napoli, Italy, e-mail: domenico.piccolo@unina.it

Luigi Salmaso
Department of Management and Engineering, University of Padua, Stradella S. Nicola 3, 36100 Vicenza, Italy, e-mail: salmaso@gest.unipd.it

M. Bini et al. (eds.), *Statistical Methods for the Evaluation of Educational Services and Quality of Products*, Contributions to Statistics,
DOI 10.1007/978-3-7908-2385-1_1,

combination of categorical and continuous latent variables at each level of the hierarchy.

Latent variable models specify the joint distribution of a set of observed and latent variables. Variables that are directly observed, also known as manifest variables, will be denoted by Y. A collection of K manifest variables will be distinguished by subscripts and written as column vector $\boldsymbol{y} = (Y_1, ..., Y_K)'$. Latent variables will be denoted by X, and Q latent variables will form the column vector $\boldsymbol{\eta}$. In practice, Q will be much smaller than K. Both latent and manifest variables can be metrical and/or categorical and vary from one individual to another. The relationships between them must be expressed in terms of probability distributions, so that, after the Y's have been observed, the information we have about $\boldsymbol{\eta}$ is given by its conditional distribution given $\boldsymbol{y}$

$$h(\boldsymbol{\eta}|\boldsymbol{y}) = \frac{h(\boldsymbol{\eta})g(\boldsymbol{y}|\boldsymbol{\eta})}{f(\boldsymbol{y})} \tag{1.1}$$

where $h(\boldsymbol{\eta})$ is the prior distribution of $\boldsymbol{\eta}$, and $g(\boldsymbol{y}|\boldsymbol{\eta})$ is the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{\eta}$. As only $\boldsymbol{y}$ can be observed, any inference must be based on the joint distribution whose density may be expressed as

$$f(\boldsymbol{y}) = \int_{\mathbf{R}_\eta} h(\boldsymbol{\eta})g(\boldsymbol{y}|\boldsymbol{\eta})\mathrm{d}\boldsymbol{\eta}, \tag{1.2}$$

where $\mathbf{R}_\eta$ is the range space of $\boldsymbol{\eta}$.

The main assumption in this framework is the *conditional (or local) independence* of the observations $\boldsymbol{y}$ given the latent variables $\boldsymbol{\eta}$. Hence, Q must be chosen so that

$$g(\boldsymbol{y}|\boldsymbol{\eta}) = \prod_{i=1}^{K} g_i(y_i|\boldsymbol{\eta}) \tag{1.3}$$

A latent variable model consists of two parts. The first part is given by the prior distribution of the latent variables $h(\boldsymbol{\eta})$. This accounts for the nature of $\boldsymbol{\eta}$, but it was seen to be essentially arbitrary and its choice is largely a matter of convention. The second element in the model is the set of conditional distributions of the manifest variables given the latent variables $g_i(y_i|\boldsymbol{\eta})$. A convenient family of distributions which allows to account for both discrete and continuous observations is the *exponential family*

$$g_i(y_i|\boldsymbol{\eta}) = \exp\left\{\frac{\boldsymbol{y}\theta - b(\theta_i)}{\phi_i} + \mathrm{d}(y, \phi_i)\right\} \tag{1.4}$$

where θ_i is some function of $\boldsymbol{\eta}$. The simplest assumption about the form of this function is to suppose that it is a linear function, in which case we have

$$\theta_i = \alpha_{i0} + \alpha_{i1}\eta_{i1} + ... + \alpha_{iQ}\eta_{iQ} \qquad i = 1, 2, ..., K \tag{1.5}$$

This is the General Linear Latent Variable Model (GLLVM). The term “linear” refers to its linearity in the αs.

Several statistical methodologies based on observed and latent variables of different nature are encompassed in the GLLVM described above and they are formalized by Bartholomew and Knott (1999). It provides a generalization of both the classical Generalized Linear Models (GLMs) by including latent dependent variables, and the classical factor model by allowing observations of different nature as well as linear relationships among the factors. From this point of view, GLLVM also generalizes the LISREL model by describing the relationship between dependent and independent latent variables in terms of probability distributions.

Chapters 2, 3 and 4 illustrate GLLVM with application in the educational evaluation. Since the variables under investigation are abilities, initial status and rate of change in temporal achievement, we deal with continuous latent variables, but different types of observations are considered. From a different point of view, Chapter 7 proposes an approach mainly based on individual perceptions about the discrete choice; thus, latent variables are a fundamental issues but they are quantified by explicit parameters in the model and by subjects covariates when it is convenient.

Chapter 2 deals with the problem of ordinal observations. In the literature (Jöreskog and Moustaki, 2001) there are two main approaches for conducting latent variable analysis with categorical observed data. The most popular is the Underlying Variable Approach (UVA) which assumes that each manifest variable is an indirect observation of a standardized normal variable. This approach is used in the general framework of structural equation modelling (LISREL). The other main approach is the Item Response Function (IRF) approach by which the manifest variables are treated as they are. The unit of analysis is the entire response pattern of a subject, so no loss of information occurs. The models for ordinal data within the IRF has been recently developed by Moustaki (2000). After a review of basic concepts of the two approaches, some methodological developments are introduced. This methodological extension requires an improvement of the computational algorithms for parameter estimation. Furthermore some theoretical results on the goodness of fit problems due to the severe sparseness, typical of variables with many categories, are presented.

Chapter 3 deals with Item Response Theory (IRT), or latent trait models for the study of individual responses to a set of items designed to measure latent abilities. IRT is a measurement theory that was first formalized in the Sixties with the fundamental work of Lord and Novick (1968) and it has a predominant role in educational testing. An IRT model describes the relationship between the observable examinee performance in the test, typically in the form of responses to categorical items, and the unobservable latent ability. Therefore, IRT models can be included in the GLLVM framework. IRT is used in all phases of test administration, from the test calibration to the estimation of individual abilities, in which the estimated item parameters are used to characterize the examinees. After a brief presentation of the main assumptions of IRT models, several aspects related to specific problems in the context of test administration are treated. This decision has been motivated by the many advances introduced over the last few years, that allow both to support more complex models and to improve the estimation algorithms. In particular, issues on multidimensionality (Wang et al., 2004), incomplete design (Béguin and Glas, 2001) and the inclusion of prior information (van der Linden, 1999) are discussed, referring both to current literature and to some contributions of the authors. A particular attention is given to the use of the Gibbs sampler, in the Markov Chain Monte Carlo (MCMC) methods, for the estimation of IRT models (Albert, 1992; Fox and Glas, 2001). Finally, applications related to these topics are presented in the context of educational assessment.

Chapter 4 describes the application of GLLVM for the analysis of individual data repeated over a period of time, that allows dynamical studies of social processes, rather than static cross-sectional analyses. The analysis of repeated measures has been considered from different points of view, such as individual growth techniques (Singer and Willett, 2004), time series and econometric analysis (Diggle et al., 1994), and multilevel modelling (Skrondal and Rabe-Hesketh, 2004). They can be encompassed into the general class of random coefficient models, in which random effects are incorporated into the model in view of reflecting unobserved heterogeneity in the individual behavior. More generally, the random coefficients can be incorporated into GLLVM by considering them as latent variables. Borrowing from Meredith and Tisak (1990), we refer to these models as Latent Curve Models (LCMs), since random coefficients permit each case in the sample to have a different trajectory over time. Growth curve models are studied to compare University student careers over time. We focus on continuous response variables, using conventional normal-theory estimators, such as maximum likelihood, into the framework of GLLVM.

In Chapter 7 we assume that evaluation is a psychological process where a rater/judge expresses the agreement within a prefixed scale. This process is generated by the perception of value/quality/performance and is governed by latent variables. In order to model the empirical results of a survey and to infer on the stochastic mechanism that generated ordinal data, we suppose that the final choice is determined by personal feeling/attractiveness towards the item and intrinsic uncertainty always present in human decisions. These aspects are combined in an effective way by introducing a mixture random variable where both components are expressed and weighted, as in D'Elia and Piccolo (2005). Thus, we will introduce CUB models by considering the observed ordinal response y as a realization of a discrete random variable Y defined on the support $\{y = 1, 2, \dots, m\}$, for a given integer $m > 3$, as a mixture of Uniform and Shifted Binomial random variables. Formally, its probability mass function is defined by:

$$\Pr(Y = y) = \pi \binom{m-1}{y-1} (1-\xi)^{y-1} \xi^{m-y} + (1-\pi)\frac{1}{m}, \quad y = 1, 2, \dots, m,$$

where $\pi \in (0,1]$ and $\xi \in [0,1]$. By examining the π parameter we quantify the *propensity* of the respondent to adhere to a completely random choice whereas $1-\xi$ parameter is related to the strength of feeling. Recently, Iannario (2009c) proved that these models are fully identifiable. This probability structure adheres to most of observed shapes for real ordinal data and it has been generalized to take into account the effect of significant covariates (Piccolo and D'Elia, 2008) or atypical situations (Iannario, 2009b). Then, asymptotic maximum likelihood inference has been developed (Piccolo, 2006) by using EM algorithm and a software in ***R*** is currently available (Iannario and Piccolo 2009) for the estimation of CUB models, without and with covariates. In this regard, a few application to real data set concerning university evaluation of teaching and services will be discussed. A special topic is a model-based clustering procedure, firstly performed by Corduas (2008a,c), where a Kullback-Liebler divergence criterion is applied for selecting subgroups of expressed ratings by university students. The characteristic of the proposal is the possibility to assess classical classification methods by an inferential approach within the unique framework of ordinal modelling. Although CUB models are focused on the marginal distribution of the respondents, their use seems effective for investigating sound relationships among ordinal responses and covariates and for enhancing unobserved traits in the data. Thus, differences and integrations with IRT are worth of interest.

1.2 Multilevel models

The class of multilevel models is suitable for the analysis of hierarchical data, where level 1 units are nested in level 2 units, which are possibly nested in level 3 units and so on. For example, students nested in classrooms, classrooms nested in schools, schools nested in districts. Longitudinal and repeated measures data can be seen as special cases of hierarchical data, with occasions nested in subjects.

The basic two-level model is the linear random intercept model:

$$y_{ij} = \alpha + \boldsymbol{\beta}\mathbf{x}_{ij} + \boldsymbol{\gamma}\mathbf{w}_j + u_j + e_{ij} \tag{1.6}$$

where j indexes the level 2 units (clusters) and i indexes the level 1 units (subjects). For example, in the evaluation of schools the clusters are the schools and the subjects are the students. The variables in the model are:

- y_{ij} the outcome of subject i of cluster j;
- $\mathbf{x}_{ij}$ a vector with the features of subject i of cluster j;
- $\mathbf{w}_j$ a vector with the features of cluster j.

Then, u_j is the random effect of cluster j, i.e. an unobservable quantity characterizing such a cluster and shared by all its subjects. The term u_j is a residual component that captures all the relevant factors at the cluster level not accounted for by the covariates and thus its meaning depends on which covariates enter the model. The effect u_j is called random because it is a random variable, assuming independence among the clusters. For consistency of the estimates, the crucial assumption on u_j is that its expectation conditionally on the covariates is null (exogeneity). Less crucial, but standard assumptions are the homoscedasticity, i.e. u_j has constant variance, and the normality of the distribution. Finally, the level 1 errors e_{ij} are residual components taking into account all the unobserved factors at the subject level making the outcome different from what predicted by the covariates and the random effect. The e_{ij} are assumed independent among subjects and independent of u_j. The other standard assumptions are similar to those on u_j, i.e. exogeneity, homoscedasticity and normality. Model (1.6) is named random intercept since each cluster has its own intercept that has both fixed and random components. However, the slopes are assumed to be constant across clusters, so the regression lines are parallel.

The simple random intercept model (1.6) can be extended in several ways. For example, it is often found that the relationship between the outcome y_{ij} and a level 1 covariate x_{ij} varies from cluster to cluster, so the regression lines are no longer parallel. This leads to the so called random coefficient model that can be written as

$$y_{ij} = \alpha + u_{0j} + (\beta + u_{1j})x_{ij} + e_{ij} \tag{1.7}$$

where it is usually assumed that (u_{0j}, u_{1j}) is bivariate normal. The random coefficient also implies that the between-cluster variance is a quadratic function of the covariate.

Now there are plenty of textbooks on multilevel modelling. Snijders and Bosker (1999) is an excellent introduction. Hox (2002) has fewer details, but it covers a wider range of topics. Raudenbush and Bryk (2002) present the models in a careful way along with thoroughly discussed applications. Goldstein (2003) is a classical, though not easy, reference with wide coverage and many educational applications.

Chapter 5 deals with the use of multilevel models for value-added analysis in education. The chapter reviews the concept of effectiveness in the educational setting and outlines the value-added approach. Multilevel models are presented as a tool for measuring effectiveness, with a discussion of several issues in model specification, such as the choice of the set of the covariates and the modelling of the achievement progress. The chapter ends with some remarks on the use of the model results for ranking the schools and for predicting the outcome of an hypothetical student.

1.2.1 Multilevel mixture factor models

Factor analysis is a well-known statistical method used to describe the correlations among some manifest variables, indicators, in terms of fewer latent variables, factors. In its standard formulation, factor analysis assumes that the variables are measured on a set of independent units; this assumption may be inadequate when units are nested in clusters assuming what is called a hierarchical structure (Goldstein, 2003; Snijders and Bosker, 1999). These differences can be modeled including group dummies in the model, as in the multigroup approach, or can be modeled with a multilevel factor model with continuous latent variables at all level of the analysis. Besides the difference in the nature, fixed or random, of the group effects, these two models differ in their perspective: the multilevel factor model usually aims at exploring the latent structure underlying the observed phenomenon at different levels of the analysis (see, as some examples, Goldstein and McDonald (1988), Longford and Muthén (1992) and Grilli and Rampichini (2007a)) while the multigroup factor model has a confirmatory approach and aims at comparing the observed groups of units with respect to the different parameters of the factor model (Bollen, 1989; Meredith, 1993; Muthén, 1989). In a confirmatory perspective, another model useful to compare the observed groups of units is the multilevel mixture factor model with a categorical latent variable at the higher level of the analysis. This model evaluates the existence of unobserved subpopulation (classes) of groups with similar features with respect to the factor model parameters and overcomes the creation of over-detailed information of the multigroup factor model, which estimates as many group coefficients as the groups. Mixture factor analyses have been developed and largely used in the one-level context (McLachlan and Peel, 2000; Magidson and Vermunt, 2001; Lubke and Muthén, 2005). More recently, the specification of mixture factor models in the multilevel context has received a growing interest. As en example, Palardy and Vermunt (2009) and Muthén (2008) use a two-level mixture model in the context of growth analysis and Vermunt (2007) use a mixture model in the context of IRT analysis; Muthén and Asparouhov (2008) also describe a more general two-level mixture model with different types of latent variables.

Chapter 6 deals with the use of multilevel models in the context of factor analysis and, more precisely, in the context of mixture factor models. This chapter describes the specification and estimation of a multilevel mixture factor model with continuous latent variables at the lower level of the analysis and a categorical latent variable at the higher level focusing, on one hand, on the illustration of some theoretical issues of the model and, on the other hand, on the applied results that can be achieved Varriale (2007). Then, a multilevel mixture factor model is used in order to evaluate the external effectiveness of the Italian university using, as indicator of the phenomenon, the information on the job satisfaction expressed by the graduates. In particular, the model is used to analyse the underlying structure of the job satisfaction at the individual level and, at the same time, to cluster higher level units represented by the programs that the individuals attended during the university in classes with some typical characteristics.

1.3 Choices and conjoint analysis: critical aspects and recent developments

Standard conjoint analysis (CA) is a multi-attribute quantitative method useful to study the evaluation of a consumer/user about a new product/service. In the literature many authors (see for example Alvarez-Farizo and Hanley (2002)), have studied and applied this method; the main theoretical problems are faced by Green and Srinivasan (1990) about statistical models and by Moore (1980), related to the insertion of baseline variables related to the respondent.

In Chapter 8 a joint study including a modified conjoint analysis and the Response Surface Methodology, in order to improve the analysis of multi-attribute valuation methods, is presented.

Our proposal is based on the conjoint analysis jointly with the status-quo evaluation, Hartman et al. (1991), which is the alternative related to the current situation.

The statistical analysis is carried out through the Response Surface Methodology (RSM) (for more details see Khuri and Cornell (1987) and Myers and Montgomery (1995)) by considering the quantitative judgement of each respondent for each profile with respect to the assessed score about the status-quo and taking the individual information into account. The final result is achieved carrying out an optimization procedure on the estimated statistical models, by defining an objective function in order to reach the optimal solution for the revised (or new) service/product. In this context, it is relevant to point out the modified structured data, through a new questionnaire, in order to collect information about the baseline variables of the respondent, the quantitative data about the current situation (status-quo) of the product/service and the proper CA analysis by means of the planning of an experimental design.

In general, we may define the set of experimental variables, which influence the measurement process: $\mathbf{x} = [\mathbf{x}_1,..,\mathbf{x}_k,..,\mathbf{x}_K]$; and the set of noise variables: $\mathbf{z} = [\mathbf{z}_1,..,\mathbf{z}_s,..,\mathbf{z}_S]$.

The general RSM model can be written as:

$$\mathbf{Y}_{ij}(\mathbf{x},\mathbf{z}) = \beta_0 + \mathbf{x}^{'}\beta + \mathbf{x}^{'}\mathbf{B}\mathbf{x} + \mathbf{z}^{'}\delta + \mathbf{z}^{'}\boldsymbol{\Delta}\mathbf{z} + \mathbf{x}^{'}\boldsymbol{\Lambda}\mathbf{z} + \mathbf{e}_{ij} \qquad i = 1,..,I; j = 1,..,J \quad (1.8)$$

where $\mathbf{x}$ and $\mathbf{z}$ are the vectors of variables as described above; β, $\mathbf{B}$, δ, $\boldsymbol{\Delta}$, and $\boldsymbol{\Lambda}$ are vectors and matrices of the model parameters, $\mathbf{e}_{ij}$ is the random error which is assumed Normally distributed with zero mean and variance equal to σ. $\boldsymbol{\Lambda}$ is a $K \times S$ matrix which plays an important role since it contains the parameters of the interaction effects between the $\mathbf{x}$ and $\mathbf{z}$ sets.

Note that, in general, if J are the profiles and I the respondents, the observations are IxJ. In this context, the set $\mathbf{x}$ are the judgements, expressed through votes in a metric scale [0,100], on the attributes involved in the experimental planning; while the set $\mathbf{z}$ is related to the baseline individual variables, which are relevant for the service or product studied and that may change according to the specific situation. The response variable $\mathbf{Y}$ is defined as a quantitative variable of the process; in this case, the judgements expressed, on each full profile of the plan, by the respondents in the same metric scale.

The final aim is to find the best preference, by evaluating both the quantitative judgements about the full profiles and the judgements about the current situation, which is the most insensitive to the heterogeneity of the respondent.

1.4 Robust diagnostic analysis with forward search

A frequently encountered difficulty in statistical inference problems is the presence of outliers in the data. Outliers can be defined as observations which appear to be inconsistent or somewhat different from the rest of the data. They can arise from models different from the one we intend to estimate (contaminants) or can be atypical observations generated by the assumed model. Their identification is of extreme importance since they can have strong negative effects on classical estimator efficiency, and should hence be eliminated or down-weighted in the estimation of the model. Furthermore, their pattern should be thoroughly examined since they could provide valuable new information on the problem being analyzed. Unfortunately, their identification is often very difficult, particularly when multivariate distributions are being dealt with.

The Forward Search, introduced by Atkinson and Riani (2000), is a general diagnostic approach for detecting the presence of outliers and assessing their influence on the estimates of the model parameters. The method was applied to regression analysis, but it could as well be applied to almost any model and multivariate method (Atkinson et al., 2004).

This algorithm is based on the following steps: the start is a robust fit to very few observations and then a successive fit is done with larger subsets. More specifically, it starts by finding a presumably outlier-free subset of observations, for example the set proposed by Rousseeuw (1984) to find the least median of squares estimators (LMS), i.e., the value of the parameters that yields the smallest median squared residual. The surface to be minimized has many local minima and the minimum value can only be obtained by approximation. Rousseeuw proposed restricting the search to all the estimates obtained by using only subsets of size p. The starting subset of the Forward Search, is given by the p observations which yield the smallest median squared residual. This is an approximation of the real LMS estimate and unfortunately still requires the evaluation of all possible subsets of size p (Bertaccini and Bini, 2007).

Formally, let $Z = (X, y)$ a data matrix of dimension $nx(p+1)$. If n is moderate and $p << n$, the choice of the initial subset can be performed by exhaustive enumeration of all $\binom{n}{p}$ distinct $p$tuple $S^{(p)}_{i_1,\ldots,i_p} \equiv \{z_1, \ldots, z_p\}$, where $Z^T_{i_j}$ is the i_jth row of Z, for and $1 < i_j \neq i_{j^*} < n$. Specifically, let $\iota^T = [i_1, \ldots, i_p]$ and let $e_{\iota, S^{(p)}_\iota}$ be the least squares residual for the unit i given the model has been fitted with the observations in $S^{(p)}_\iota$. The initial subset is which satisfies $e^2_{[\text{med}], S^{(p)}_*} = \min_\iota \left[e^2_{[\text{med}], S^{(p)}_\iota} \right]$ where $e^2_{[k], S^{(p)}_*}$ is the kth ordered squared residual among $e^2_{[i], S^{(p)}_*}$, with $i = 1, \ldots, n$ and med = integer part of $(n + p + 1)/2$. If $\binom{n}{p}$ is too large, the choice is made using 3,000 $p$tuples sampled from Z matrix.

The subset size is increased by one and the model refitted to the observations with the smallest residuals for the increased subset size.

The initial subset $S^{(m)}_*$ of dimension $m \geq p$ is increased by one and the new subset $S^{(*)}_{m+1}$ consists of $m * 1$ units with the smallest ordered residuals $e^2_{[k], S^{(m)}_*}$. The model is refitted to the new subset and the procedure continues with increasing subset sizes until all the data are fitted, i.e. when $S^{(m)}_* = S^{(n)}$. The result is an ordering of the observations by their closeness to the assumed model. Usually one observation enters the subset used for fitting, but sometimes two or more observations enter the subset as one or more leave.

Chapter 9 proposes to validate this robust diagnostic approach when university performance analyses are carried out. In particular, the algorithm is investigated in generalized liner models. The analysis also reviews some robust studies recently performed about effectiveness and efficiency of Italian universities (Bini et al., 2002; Biggeri and Bini, 2003); Bini et al., 2003; Bini and Bertaccini, 2004; Bini, 2004a, 2004b; Bertaccini and Polverini, 2006).

1.5 Nonparametric combination of dependent permutation tests and rankings

Chapters 10, 11, 12 and 13 of the book deal with the Nonparametric Combination approach of dependent permutation Tests (NPC Test) and Rankings (NPC Ranking) to face a variety of univariate and multivariate problems for the evaluation of educational services and quality of products. After a short abstract of each chapter, in this section we provide an introduction on notation and basic theory of nonparametric combination methodology of permutation tests or rankings.

Chapter 10 presents a novel Global Performance Score (GPS) for the construction of a global performance index when we are facing a complex problem of product quality evaluation, that is when the focus is on evaluating the product performances in connection with more than one aspect (dimension) and/or under several conditions (strata). The methodological solution we propose to cope with this problem is described and applied, considering different possible data transformation and an application problem related to the performance evaluation of new detergents.

Chapter 11 considers permutation methods for testing on ordered categorical variables within the framework of randomised complete block designs. The proposed approach is studied and validated via a Monte Carlo simulation study and it has been applied to a food sensorial evaluation study.

Chapter 12 is devoted to permutation tests for stochastic ordering problems where the main goal is to find out where the treatment peak is located (so called "umbrella alternative"). The proposed solution involves testing for stochastic ordering of continuous variables and the nonparametric combination methodology. Since the location of the peak is generally unknown, it can be detected by sequential tests on possible picks and combining together those tests.

Chapter 13 deals with a novel method for constructing preference rankings based on the nonparametric combination procedure and the proposed method is compared with that based on the arithmetic mean. Subsequently, in order to verify to what extent two rankings concord, a new permutation test for the evaluation of concordance between dependent rankings is developed. Finally, the method is applied to the evaluation of professional profiles of municipal directors.

1.5.1 Introduction to permutation tests

The importance of the permutation approach in resolving a large number of inferential problems is well-documented in the literature, where the relevant theoretical aspects emerge, as well as the extreme effectiveness and flexibility from an applicatory point of view (Manly, 1997; Pesarin, 2001; Edgington and Onghena, 2007; Basso et al., 2009).

The great majority of univariate problems may be usefully and effectively solved within standard parametric or nonparametric methods as well, although in relatively mild conditions their permutation counterparts are generally asymptotically as good as the best parametric ones. Moreover, it should be noted that permutation methods are essentially of a nonparametrically exact nature in a conditional context. In addition, there are a number of parametric tests the distributional behavior of which is only known asymptotically. Thus, for most sample sizes of practical interest, the relative lack of efficiency of permutation solutions may sometimes be compensated by the lack of approximation of parametric asymptotic counterparts. In addition, assumptions regarding the validity of parametric methods (such as normality and random sampling) are rarely satisfied in practice, so that consequent inferences, when not improper, are necessarily approximated, and their approximations are often difficult to assess.

For any general testing problem, in the null hypothesis (H_0), which usually assumes that data come from only one (with respect to groups) unknown population distribution P, the whole set of observed data $\mathbf{x}$ is considered to be a random sample, taking values on sample space $\mathscr{X}^n$, where $\mathbf{x}$ is one observation of the n-dimensional sampling variable $\mathbf{X}^{(n)}$ and where this random sample does not necessarily have independent and identically distributed (i.i.d.) components. We note that the observed data set $\mathbf{x}$ is always a set of sufficient statistics in H_0 for any underlying distribution.

Given a sample point $\mathbf{x}$, if $\mathbf{x}^* \in \mathscr{X}^n$ is such that the likelihood ratio $f_P^{(n)}(\mathbf{x})/f_P^{(n)}(\mathbf{x}^*) = \rho(\mathbf{x},\mathbf{x}^*)$ is not dependent on f_P for whatever $P \in \mathscr{P}$, then $\mathbf{x}$ and $\mathbf{x}^*$ are said to *contain essentially the same amount of information with respect to P*, so that they are equivalent for inferential purposes. The set of points that are equivalent to $\mathbf{x}$, with respect to the information contained, is called the *orbit associated with* $\mathbf{x}$, and is denoted by $\mathscr{X}^n_{/\mathbf{x}}$, so that $\mathscr{X}^n_{/\mathbf{x}} = \{\mathbf{x}^* : \rho(\mathbf{x},\mathbf{x}^*) \textit{ is } f_P\textit{-independent}\}$.

The same conclusion is obtained if $f_P^{(n)}(\mathbf{x})$ is assumed to be invariant with respect to permutations of the arguments of $\mathbf{x}$; i.e., the elements $(x_1,\ldots,x_n)$. This happens when the assumption of independence for observable data is replaced by that of *exchangeability*, $f_P^{(n)}(x_1,\ldots,x_n) = f_P^{(n)}\left(x_{u_1^*},\ldots,x_{u_n^*}\right)$, where $(u_1^*,\ldots,u_n^*)$ is any permutation of $(1,\ldots,n)$. Note that, in the context of permutation tests, this concept of exchangeability is often referred to as the *exchangeability of the observed data with respect to groups*. Orbits $\mathscr{X}^n_{/\mathbf{x}}$ are also called *permutation sample spaces*. It is important to note that orbits $\mathscr{X}^n_{/\mathbf{x}}$ associated with data sets $\mathbf{x} \in \mathscr{X}^n$ always contain a finite number of points, as n is finite.

Since, in the null hypothesis and assuming exchangeability, the conditional probability distribution of a generic point $\mathbf{x}' \in \mathscr{X}^n_{/\mathbf{x}}$, for any underlying population distribution $P \in \mathscr{P}$, is P-independent, permutation inferences are invariant with respect to P in H_0. Some authors, emphasizing this invariance property , prefer to give them the name of *invariant tests*. However, due to this invariance property, permutation tests are distribution-free and nonparametric.

Formally, let $\mathscr{X}^n/\mathbf{x}$ be the orbit associated with the observed vector of data $\mathbf{x}$. The points of $\mathscr{X}^n/\mathbf{x}$ can also be defined as $\mathbf{x}^* : \mathbf{x}^* = \pi\mathbf{x}$ where π is a random permutation of indexes $1, 2, \ldots, n$. Define a suitable test statistic T on $\mathscr{X}^n/\mathbf{x}$ for which large values are significant for a right-handed one-sided alternative: The support of $\mathscr{X}^n/\mathbf{x}$ through T is the set $\mathscr{T}$ that consists of C elements (if there are no ties in the given data). Let

$$T^*_{(1)} \leq T^*_{(2)} \leq \ldots \leq T^*_{(C)}$$

be the ordered values of $\mathscr{T}$. Let T^o be the observed value of the test statistic, $T^o = T(\mathbf{x})$. For a chosen attainable significance level $\alpha \in \{1/C, 2/C, \ldots, (C-1)/C\}$, let $k = C(1-\alpha)$. Define a permutation test, the function $\phi^* = \phi(T^*)$ for a one-sided alternative

$$\phi^*(T) = \begin{cases} 1 & \text{if} \quad T^o \geq T^*_{(k)} \\ 0 & \text{if} \quad T^o < T^*_{(k)} \end{cases} .$$

Permutation tests have general good properties such as exactness, unbiasedness and consistency (see Pesarin, 2001; Hoeffding, 1952).

1.5.2 Multivariate permutation tests and nonparametric combination methodology

In this section, we provide details on the construction of multivariate permutation tests via nonparametric combination approach. Consider, for instance, a multivariate problem where q (possibly dependent) variables are considered. The main difficulties arise because of the underlying dependence structure among variables (or aspects), which is generally unknown. Moreover, a global answer involving several dependent variables (aspects) is often required, so the question is how to combine the information related to the q variables (aspects) into one global test.

In a multivariate problem, when the aim is to compare two o more groups, the matrix of data is generally partitioned into n q-dimensional arrays; that is,

$$\mathbf{X_{n\times q}} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1q} \\ x_{21} & x_{22} & \ldots & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nq} \end{bmatrix} .$$

Each row of $\mathbf{X}$ is a determination of the multivariate variable $[X_1, X_2, \ldots, X_q]$, which has distribution P with unknown dependence structure.

In this framework the null hypothesis H_0, which states the equality in distribution of the multivariate distribution of the q variables in all groups, is supposed to be properly decomposed into q sub-hypotheses H_{0j} each appropriate for partial (univariate) aspects,

$$H_0 : \bigcap_{j=1}^{q} H_{0j}.$$

Hence, the *global* null hypothesis H_0 can be viewed as an intersection of *partial* null hypotheses H_{0j}. Under the global null hypothesis, the rows of $\mathbf{X}$ are exchangeable. We can thus define q *partial* test statistics. Let T_j, $j = 1, \ldots, q$, be a *partial* test statistic for the univariate hypothesis H_{0j} involving each of the q variables.

A desirable property of a multivariate test is that the global null hypothesis should be rejected whenever one of the partial null hypothesis is rejected. To this end, let us consider the rule *large is significant*, which means that the global test statistic should assume large values whenever at least one of its arguments leads to the rejection of at least one partial null hypothesis H_{0j}. Accordingly, the global test ψ^* should be based on a suitable combining function ψ that satisfies the following requirements:

1. A combining function ψ must be non-increasing in each argument:

$$\psi(\lambda_1, \ldots, \lambda_j, \ldots, \lambda_q) \geq \psi\left(\lambda_1, \ldots, \lambda_j', \ldots, \lambda_q\right) \quad \text{if} \quad \lambda_j < \lambda_j',\ j \in \{1, \ldots, q\}.$$

1. Every combining function ψ must attain its supremum value $\bar{\psi}$, possibly not finite, even when only one argument attains zero:

$$\psi(.., \lambda_j, ..) \rightarrow \bar{\psi} \quad \text{if} \quad \lambda_j \rightarrow 0,\ j \in \{1, \ldots, q\}.$$

2. $\forall \alpha > 0$, the critical value of every ψ is assumed to be finite and strictly smaller than the supremum value: $T_\alpha'' < \bar{\psi}$.

The λ's in the definition of the combining function are p-values: $\alpha_i = \Pr\{T_i^* \geq T_i^o | H_{0i}\}$. It is possible, of course, to express ψ also in terms of partial statistics. For instance, if the λ's are test statistics that are significant for large values (as in the bivariate example), some suitable combining functions are the following:

- the direct combining function: $\psi = \sum_{j=1}^{q} \lambda_j$;
- the $\max_T$ combining function: $\psi = \max_j \lambda_j$.

Instead, if the combining function is based on the partial p-values (i.e., $\lambda_j = p_j = \Pr\left[T_j^* \geq T_j | \mathbf{Y}\right]$, which are significant against H_{0j} for small values), the following combining functions are of interest:

- Fisher's: $\psi = -2\sum_{j=1}^{q} \log(p_j)$, $0 \leq \psi \leq +\infty$;
- Tippett's: $\psi = 1 - \min_j p_j$, $0 \leq \psi \leq 1$;