

# Chapter 5

## A Study on University Students' Opinions about Teaching Quality: a Model Based Approach for Clustering Ordinal Data

Marcella Corduas

### 5.1 Introduction

In complex surveys aimed at measuring the satisfaction level of final users of a given product or service, several items are generally investigated. Also, respondents often belong to different categories since they are stratified according to relevant features, such as geographic location or gender. In such situations, the comparison among the distributions of ratings given to a selection of items by interviewees or to a single item by different groups of interviewees can provide a meaningful summary of observed data.

Sometimes, decision makers, who are interested in information arising from statistical analysis, are not very familiar with statistical theory. For this reason, graphical techniques or simple statistical indices (such as the average or mode) have been widely used in empirical analysis. However, this approach wastes relevant information about the distribution of ratings since other aspects concerning shape are not considered. In this respect, for instance, [7] introduced a statistical index based on a distance measure between ordinal data distributions; [2] discussed the fundamental principles for constructing composite indicators and examined some specific measures for assessing university teaching quality. In this chapter, a mixture distribution is used for modeling ratings and a procedure for detecting significant similarities and differences in the distribution of judgements expressed by raters is proposed. Specifically, the case study refers to the yearly survey done at the University of Naples Federico II in order to assess teaching quality.

The chapter is organized as follows. In Sect. 5.2, some results concerning a mixture distribution for ordinal data are briefly illustrated. Then, in Sects. 5.3 and 5.4, a testing procedure based on Kullback-Liebler (KL) divergence and a clustering technique are discussed, and finally, in Sect. 5.5, a case study is presented.

---

M. Corduas (✉)

Dipartimento di Scienze Statistiche, Università di Napoli Federico II, 80138 Napoli, Italy  
e-mail: corduas@unina.it

## 5.2 A Mixture Distribution for Ordinal Data

A statistical model for ordinal data has been recently proposed by [12]. The model provides the probability distribution of the random variable generating ordinal data describing judgments or evaluations that respondents express on a given item.

Specifically, the preference or score is represented by the random variable  $R$  such that:

$$P(R = r) = \pi \binom{m-1}{r-1} (1-\xi)^{r-1} \xi^{m-r} + (1-\pi) \frac{1}{m}, \quad r = 1, 2, \dots, m \quad (1)$$

where  $\pi \in (0, 1]$ ,  $\xi \in [0, 1]$ , and  $m$  is the number of grades for evaluating an item.<sup>1</sup> For  $m > 3$ , (1) is a mixture of a *Uniform* and a (shifted) *Binomial* distribution. The parameter  $\pi$  determines the role of *uncertainty* in the final judgment: the lower the weight  $(1-\pi)$  the smaller the contribution of the Uniform distribution in the mixture and then the smaller is the subject uncertainty in selecting the final rating. Moreover, the parameter  $\xi$  characterizes the shifted Binomial distribution and, therefore, it is related to the strength of the intimate belief of respondents concerning the object of evaluation. In other words,  $(1-\xi)$  is the strength of the positive *feeling* expressed by raters about the item (see, [10, 21] for a discussion). Then, the closer  $\xi$  is to 1 the less the item has been rated positively.

The model is very flexible and is capable of describing distributions having very different shapes. The formulation of the asymmetry and kurtosis coefficients of  $R$  as function of the  $\pi$  and  $\xi$  parameters have been derived by [20]. Specifically, it can be shown that  $Asim(\pi, \xi) = 0$  for  $\xi = 0.5$  and, in addition,  $Asim(\pi, \xi) = -Asim(\pi, 1-\xi)$ , for a given  $\pi \in (0, 1]$ . When  $\xi < 0.5$ , the distribution of  $R$  is skewed negatively and the probability that raters express positive opinions about the given item increases as  $\xi$  moves towards 0. The opposite consideration applies when  $\xi > 0.5$ , the distribution of  $R$  is skewed positively and the probability that raters express negative opinions increases as  $\xi$  moves towards 1. Also, for a given  $\pi \in (0, 1]$ , the kurtosis increases as  $\xi$  approaches the borders of the parameter space, and  $Kurt(\pi, \xi) = Kurt(\pi, 1-\xi)$ .

The influence of external factors in the final judgement may be introduced by adding two relations which connect the model parameters to significant *covariates* by means of a logistic link function (see [22]). This fact originated the acronym CUB which the authors used to identify the model. Note that, in the rest of this chapter, that acronym will simply denote the model (1).

Finally, given the observed ratings  $\mathbf{r} = (r_1, r_2, \dots, r_N)'$  expressed by  $N$  judges towards a certain item, the log-likelihood function for the model (1) is:

$$\log L(\boldsymbol{\theta}) = \sum_{r=1}^m n_r \log(p_r(\boldsymbol{\theta})), \quad (2)$$

---

<sup>1</sup> In order to facilitate the reading, according to the special case study which will be examined in the final section, in the following discussion, we assume that 1 refers to the worst judgement and  $m$  to the best one.

being  $\theta = (\pi, \xi)'$  the parameter vector,  $n_r$  the observed frequency of  $R = r$ , ( $r = 1, \dots, m$ ) and  $p_r(\theta) = P(R = r|\theta)$ . Maximum likelihood estimation of  $\theta$  can be performed by E-M algorithm; an efficient procedure is discussed by [21].

### 5.3 The Kullback-Liebler Divergence

CUB models provide a meaningful and parsimonious parametric representation of rating distributions which can be used for clustering purposes. In this respect, a measure of dissimilarity is needed. Specifically, we introduce the KL divergence measure and, in this section, we briefly recall some useful results.

In general, KL divergence measures the dissimilarity between two probability distributions  $f_1(x, \theta_1)$  and  $f_2(x, \theta_2)$  characterizing a random variable  $X$  under two different hypotheses, respectively [16].

Specifically, the KL divergence is defined as:

$$J(f_1, f_2) = I(f_1, f_2) + I(f_2, f_1), \quad (3)$$

where, assuming the case of a continuous random variable:

$$I(f_1, f_2) = \int_{-\infty}^{\infty} f_1(x, \theta_1) \ln \frac{f_1(x, \theta_1)}{f_2(x, \theta_2)} dx = E_1 \left( \ln \frac{f_1(x, \theta_1)}{f_2(x, \theta_2)} \right) \quad (4)$$

is the mean information, with respect to  $f_1$ , for discrimination in favor of the first hypothesis against the second one. The other term in (3),  $I(f_2, f_1)$  is similarly defined. Of course, the case of a discrete random variable can be easily introduced by extending (4) accordingly.

Note that the KL divergence is not a metric. As a matter of fact, it satisfies the following relationships:  $J(f_1, f_2) \geq 0$ , the equality holds if and only if  $f_1$  and  $f_2$  coincides;  $J(f_1, f_2) = J(f_2, f_1)$ ; but it doesn't satisfy the triangular inequality.

However, due to its statistical properties, it represents a very interesting tool for establishing the comparison of CUB models as a problem of hypotheses testing.

For this aim, we illustrate a general result derived from [17]. Consider two discrete populations each characterized by a probability distribution function having the same functional form  $p(x, \theta_i)$  with unspecified parameters  $\theta_i$ ,  $i = 1, 2$ . Also assume that, for all points in the random variable support,  $p(x, \theta_i) > 0$ . Suppose that two samples of  $N_1$  and  $N_2$  observations have been randomly drawn from the specified  $i$ -th population and we wish to decide if they were in fact generated from the same population. In order to test the hypothesis  $H_0 : \theta_1 = \theta_2$  against  $H_1 : \theta_{1j} \neq \theta_{2j}$  (for at least one of the distribution parameters), the KL divergence statistic is defined by:

$$\hat{J} = \frac{N_1 N_2}{N_1 + N_2} \left[ \sum_x (p(x, \theta_1) - p(x, \theta_2)) \ln \frac{p(x, \theta_1)}{p(x, \theta_2)} \right]_{\theta_1 = \hat{\theta}_1, \theta_2 = \hat{\theta}_2}, \quad (5)$$

where the vector parameters  $\theta_1$  and  $\theta_2$  have been replaced by the maximum likelihood estimators. It can be shown that  $\hat{J}$  is asymptotically distributed as a  $\chi_g^2$  random variable when the null hypothesis is true, being  $g$  the dimension of the vector parameter [16]. In the case under investigation, the objects of comparison are CUB distributions, each characterized by  $g = 2$  parameters, then the  $100\alpha\%$  critical region for hypotheses testing is simply given by:  $\hat{J} > \chi_{(2,\alpha)}^2$ .

## 5.4 Clustering

Although CUB models only describe univariate distributions of judgements, they may help to give further insights into data originated by complex surveys.

In literature several approaches have been proposed for clustering ordinal data. The problems related to the choice of an adequate measure of dissimilarity between ordinal data and the necessary techniques for producing clustering have been investigated (see, for instance, [23] for a review).

Moreover, model-based approaches which use estimated membership probabilities to classify cases into the appropriate cluster have been introduced and widely studied (see [5, 6, 14] for a general discussion). In this respect, a well established technique is the mixture-model clustering, where each latent class represents a hidden cluster ([19, 24] and references therein).

The approach that we discuss in the present chapter moves from a different point of view since the elements which are object of comparison are the distributions of ratings. In other words, the focus is not on the judgements that each individual expresses, but on the shape of the overall rating distribution that their judgements originate for a given item. Moreover, it is worth noting that we are not postulating that the population is clustered in two groups behaving according to one of the two unobserved components of the mixture distribution (1). The latter is only a probability distribution which results being flexible enough to represent observed ratings, but any other distribution which ensures a good fit for the data may be used.

Multivariate approaches, which account for the dependence among judgments expressed by subjects, exploit data information more efficiently than CUB models which, at this stage of the work, simply represent univariate distributions. However, CUB models have proved to be effective in numerous real applications arising in various fields such as social analysis [15], medicine [11], marketing [22], linguistics [1] and others (see [10] for a discussion) and, for this reason, they deserve further attention.

Coming to the clustering problem, the strategy that we propose relies on the following steps:

- firstly, the mixture distribution (1) is fitted to the observed rating distributions concerning the items object of evaluation;
- secondly, each estimated CUB model is compared with the others by means of the KL divergence and a symmetric square matrix of KL divergences between all models is evaluated;

- finally, a hierarchical clustering technique (complete or simple linkage method) is applied to the matrix of divergences.

Regarding the last step, we suggest that the  $100\alpha\%$  critical value, derived for hypotheses testing, be used for sectioning the dendrogram and for the subsequent identification of groups.

As is well known, both clustering methods impose a fixed hierarchical rule in order to decide when a new group has to be created: simple linkage method allows for elongated clusters whereas complete linkage method tends to recognize compact clusters. This fact reduces the flexibility of the approach with respect to other approaches, such as the BEA algorithm which has been studied by [9, 18].

Notwithstanding this limit, hierarchical clustering still represents an effective device to provide a very simple graphical display of data.

Moreover, having used the  $100\alpha\%$  critical value derived for hypotheses testing as threshold makes the interpretation of resulting clusters more meaningful. As a matter of fact, with reference to complete linkage method, the suggested criterion for dendrogram's sectioning ensures that all ratings distributions belonging to the same cluster have been generated by the same population. Instead, as regards simple linkage method, the criterion ensures that, given a certain group, there exists a single link path along clustered elements which joins ratings distributions which are similar according to the KL divergence test.

## 5.5 The Analysis of Students' Opinions

In this section, by means of the analysis of an empirical data set, we will illustrate how the proposed technique can be used in practice for clustering rating distributions. In particular, the study refers to a real data set from the yearly survey on students' opinions about teaching quality at the University of Naples Federico II.<sup>2</sup>

### 5.5.1 The Data Set

According to the CNVSU guidelines [8], the questionnaire aims at assessing the students' opinions about various elements which characterize teaching activity: (1) quality of lecture halls and teaching equipments; (2) several features of the specific course the interviewees are attending; (3) instructor's abilities: clear explanations, ability to inspire and motivate students' interest in the course content, instructor's availability for consultation outside of class, time-table respect, instructor's concern for students' learning problems and adequacy of textbooks and other material.

---

<sup>2</sup> The problems related to the evaluation of university teaching and services has been widely discussed (see, for instance, [3, 4, 13]). Further references can be found at the following websites: <http://dssm.unipa.it/divago/> for the *DIVAGO* project, <http://www2.stat.unibo.it/prin2006/> for the PRIN2006 project on "Metodi e modelli statistici per la valutazione dei processi formativi", <http://valmon.ds.unifi.it> for the *VALMON* project.

**Table 5.1** Students' profiles and participation

Year	(%)	HS qualification	(%)	Attendance	(%)	N.courses	(%)
1	35.9	Classical studies	19.8	<20%	0.7	1	5.3
2	29.6	Scientific education	53.3	20–50%	2.9	2	10.8
3	21.7	Technical education	17.1	50–80%	20.1	3	22.0
4	4.0	Professional education	9.8	>80%	76.3	4	30.2
5	1.8					5	21.3
6	0.5					>5	10.5
>6	6.5						

The data set consists of 34,507 valid records. It was gathered at the end of 2005–2006 academic year from the 13 Faculties belonging to the University Federico II (Medicine, Veterinary medicine, Pharmacy, Agricultural Science, Biotechnology, Engineering, Architecture, Mathematics and Natural Science, Classics and Modern Studies, Law, Economics, Political Sciences, Sociology). The students' ratings are expressed using a 7 point Likert scale where 7 relates to the highest positive judgement.<sup>3</sup>

Table 5.1 illustrates the students profile and participation. More than half of the interviewees have a scientific education at senior high school level; only 20% have specialized in classical studies and the remaining 27% have a technical or professional education.

The attendance of courses is generally high; only 3.6% of the students show low attendance of courses. Since the questionnaires are submitted in the last weeks of the term, this result confirms that most students have a solid knowledge of the course they are requested to evaluate and repute it important for their curricula. In this respect, it is worth noting that students usually are not requested to attend courses in order to be admitted to final exams. Then, it is more likely that those who attend lectures regularly are satisfied about teaching. Moreover, about 84% of students attend more than two courses; in other words, interviewees prefer to attend most courses offered in each term within their curricula. This is in part justified by the fact that most students (87%) are enrolled in the first 3 years of their university degree course.

### 5.5.2 The Results

The rest of this study refers to the assessment of 6 instructor's abilities. Specifically, CUB models have been fitted to the rating distributions of each items observed for the 13 Faculties. The parameter estimates were significant in all the examined cases.

Figure 5.1 displays the estimated parameters in the  $(\pi, \xi)$  parameter space. Since,  $\hat{\pi} \in (0.57, 0.99)$  and  $\hat{\xi} \in (0.18, 0.37)$  only a sub-region of the unit square is shown.

<sup>3</sup> The study presented in the chapter refers to Faculties and no investigation at a lower level of aggregation (such as curricula) has been carried out. As a matter of fact, the use of the database was restricted by the University of Naples Federico II which also requested that no identifier establishing the identity of the Faculties could be included in any publication.

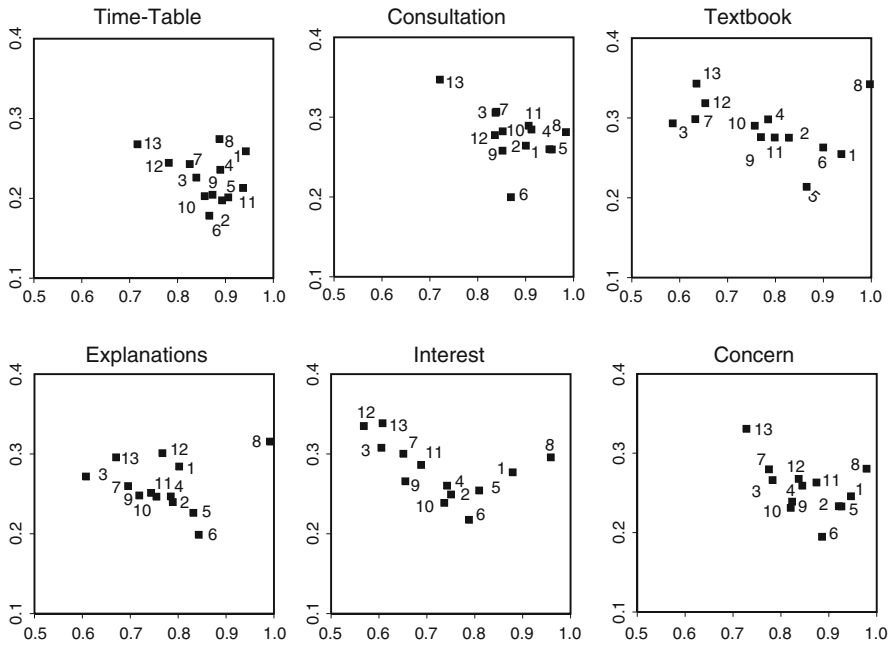


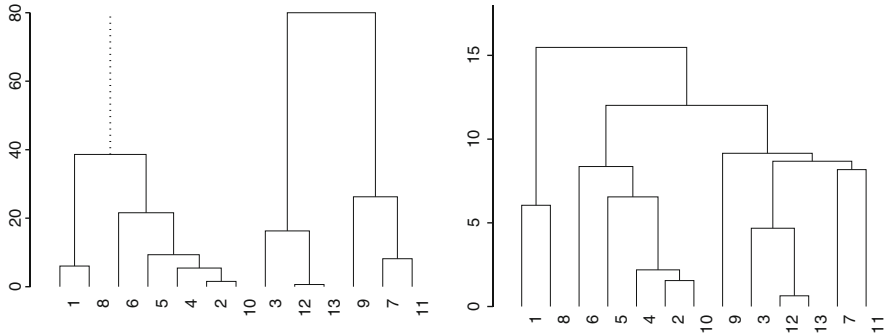
Fig. 5.1 CUB models of the rating distributions of instructor's abilities

In general, students express a rather positive evaluation of the teacher's abilities as confirmed by the high values of  $(1 - \hat{\xi})$ . But, at least at this stage of the analysis, this consideration does not allow a clear discrimination among Faculties since the estimated values,  $\hat{\xi}$ , appear very close. Interestingly, the *uncertainty*, measured by  $(1 - \hat{\pi})$ , is generally low but it seems to vary more widely among Faculties, and this behaviour is common to all the items. This implies that the students of some Faculties (for instance (8)), have a stronger attitude towards the assessment with respect to others and they express a more convinced opinion.

The representation of CUB models in the parameter space can not be used for clustering purposes, since, as discussed in [9], the Euclidean distance among points, established by visual inspection of the graph, does not reflect the true dissimilarity between the shape of the underlying distribution. Even a small change in position of a point in reference to the horizontal and vertical axis implies very different consequences in terms of the shape of the related rating distribution. For this reason we use the KL divergence to compare CUB models and proceed according to the strategy previously described.

First, we examine in some detail the instructor's ability to raise student interest.

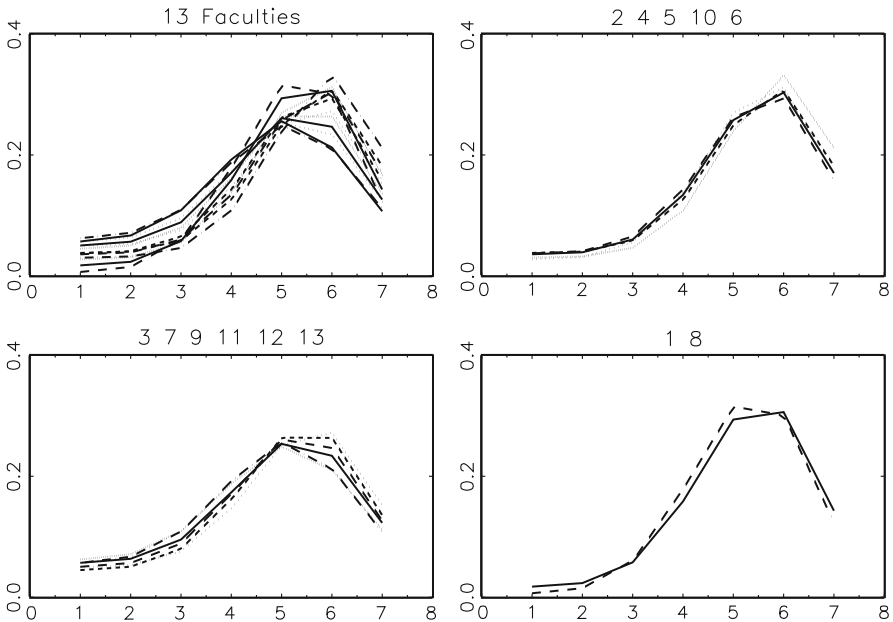
The dendrogram derived by complete linkage method (Fig. 5.2) helps the identification of clusters of Faculties which students assign similar ratings to the considered item. The threshold used for ending the aggregation process is the percentile  $\chi^2_{(2,0.01)}$ . The dotted line in the graph corresponds to a very extreme divergence.



**Fig. 5.2** Ability to raise student interest: complete linkage (*lhs*), single linkage method (*rhs*)

From the inspection of the dendrogram, the following elements are immediately connected: (1, 8) (5, 4, 2, 10) (12, 13) (7, 11).

If elongated clusters are allowed (see the single linkage dendrogram in Fig. 5.2), other elements merge so that only three clusters are classified:  $G_1 = (3, 12, 13, 7, 11, 9)$  which refer to the Faculties strongly related to professional skills and vocational education,  $G_2 = (2, 4, 5, 10, 6)$  which includes the Faculties related to humanities and arts, and  $G_3 = (1, 8)$  which includes two Faculties with a marked specialization. Figure 5.3 shows the estimated rating distributions belonging to each group.



**Fig. 5.3** Clustered estimated CUB distributions



It is evident that KL-divergence is able to discriminate between distributions which apparently are characterized by similar overall pattern.

The representation of the CUB models in the parameter space (Fig. 5.1) shows that the resulting clusters (from  $G_1$  to  $G_3$ ) are ordered according to the  $\pi$  estimates, that is in terms of decreasing uncertainty. Furthermore, due to the specific shape of the rating distributions, this fact reflects the ordering of the groups in terms of increasing probability of positive judgements (which are 0.62, 0.73 and 0.74, respectively).

It is not surprising that students attending Faculties in  $G_1$  find the judgement of this item is more problematic with respect to the other two groups. As a matter of facts, students of those Faculties are generally more demanding. In addition, on the one hand the related disciplines are very technical and less fascinating, on the other a number of instructors are not specifically trained for classroom teaching practice since they are involved in professional activities.

In Figure 5.4, the dendrograms of the remaining items are illustrated. In order to facilitate the reading, due to the presence of extreme divergences, the joining level of extreme points has been rescaled. The horizontal line helps to detect the groups identified by cutting the unscaled dendrogram according to the criterion described in the previous section.

Some behaviours are worth of comments. Firstly, the clustering of the rating distributions concerning instructors' ability to explain concepts clearly reveals the presence of several isolated Faculties and few small groups. This may be due to the fact that "clarity" is a very personal trait of instructors, and in addition it may depend to some extent by the nature of disciplines. It is, therefore, unlikely that large clusters can be recognized at the considered level of aggregation.

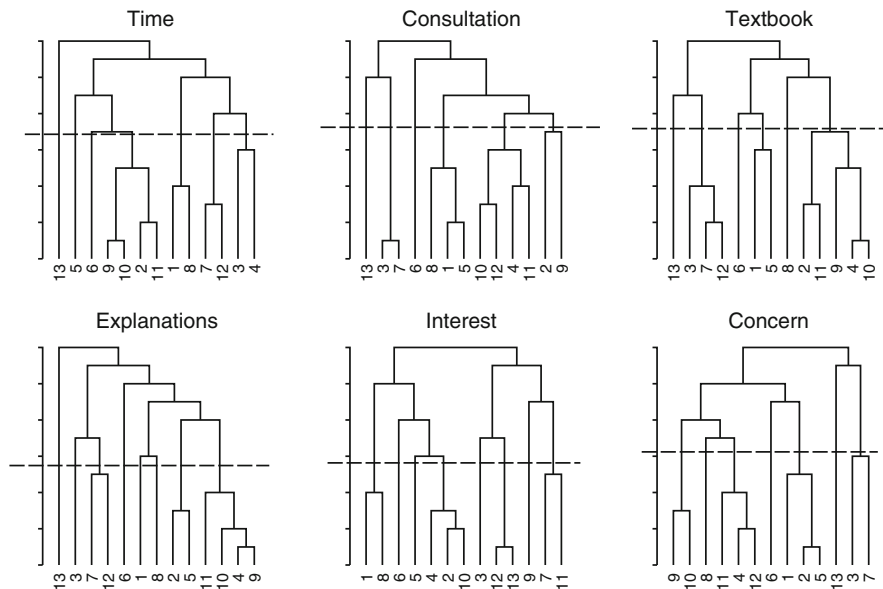


Fig. 5.4 Dendrogram by complete linkage method

From the comparison, other distinctive characteristics emerge. Faculties (6) and (13) are generally isolated. The former achieves the highest probability of positive judgements with respect to the others whereas the latter attain the lowest one. Moreover, (3) is generally close to (7) for most items confirming the partial overlapping of the two Faculties with respect to the scientific areas which the instructors are related to. Furthermore, (4), (9), (10) and (11) show many similarities with respect to various items. The Faculties (8) and (1) seem to share a common behaviour of the students rating distributions on two organizational aspects, “time table respect” and “availability outside of class”. Finally, (2) and (10) receive similar evaluation on “clarity”, “time table respect” and “adequacy of textbooks”, but they belong to different groups with respect to the other two items which instead concern the interaction with students.

## 5.6 Final Remarks

In this chapter an approach to ordinal data clustering based on KL divergence has been presented. The proposed technique helps the identification of similarities in the behaviour of groups of judges when they are asked to express their ratings on a set of items. Specifically, the technique is able to discriminate the scores distributions even when the latter are apparently characterized by similar overall patterns. Moreover, combining the hypotheses testing based on KL divergence with the clustering technique adds further strength to the identification of groups.

**Acknowledgements** This research is part of PRIN 2006–2008 project on: “Metodi e modelli statistici per la valutazione dei processi formativi” and has benefited from support of MiPAAF ex CFEPSR (Portici). The author thanks the University of Naples Federico II, and especially the Nucleo di Valutazione di Ateneo and UPSV for kindly providing the data set which has been analyzed in this chapter.

## References

1. Balirano G, Corduas M (2008) Detecting semiotically expressed humor in diasporic TV productions. *Humor: Int J Humor Res* 3:227–251
2. Bernardi L, Capursi V, Librizzi L (2004) Measurement awareness: the use of indicators between expectations and opportunities. In: *Atti della XLII Riunione Scientifica SIS*. Cleup, Padova, pp 315–326
3. Biggeri L (2000) Valutazione: idee, esperienze, problemi. Una sfida per gli statistici. In: *Atti della XL Riunione Scientifica SIS*. CS2p, Firenze, pp 31–48
4. Biggeri L, Bini M (2001) Evaluation at University and State level in Italy: need for a system of evaluation and indicators. *Tertiary Educ Manage* 7:149–162
5. Bock HH (1996) Probabilistic models in cluster analysis. *Comput Stat Data Anal* 23:5–28
6. Bock HH (1998) Probabilistic aspects in classification. In: Hayashi C, Yajima K, Bock HH, Oshumi N, Tanaka Y, Baba Y (eds) *Data science, classification and related methods*. Springer, New-York, NY, pp 3–21
7. Capursi V, Porcu M (2001) La didattica universitaria valutata dagli studenti: un indicatore basato su misure di distanza fra distribuzioni di giudizi. In: *Atti della Riunione Intermedia SIS on “Processi e Metodi Statistici di Valutazione”*, Roma

8. CNVSU (2002) Proposta di un insieme minimo di domande per la valutazione della didattica da parte degli studenti frequentanti. Doc 9/02, <http://www.cnsvu.it>
9. Corduas M (2008) A testing procedure for clustering ordinal data by CUB models. Proceedings of SFC-CLADAG 2008 Meeting. ESI, Napoli, pp 245–248
10. Corduas M, Iannario M, Piccolo D (2009) A class of statistical models for evaluating services and performances. In: Bini M, Monari P, Piccolo D, Salmaso S (eds) Statistical methods for the evaluation of educational services and quality of products. Contribution to Statistics. Physica-Verlag, Heidelberg, pp 99–117
11. D'Elia A (2007) A statistical modelling approach for the analysis of TMD chronic pain data. *Stat Methods Med Res* 16:1–15
12. D'Elia A, Piccolo D (2005) A mixture model for preference data analysis. *Comput Stat Data Anal* 49:917–934
13. Fabbris L (ed) (2006) Effectiveness of university education in Italy: employability. Competences, human capital. Springer, Heidelberg
14. Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. *JASA* 97:611–631
15. Iannario M (2007) A statistical approach for modelling Urban Audit Perception Surveys. *Quaderni di Statistica* 9:149–172
16. Kullback S (1959) Information theory and statistics. Dover Publ., New York, NY
17. Kupperman M (1957) Further applications of information theory to multivariate analysis and statistical inference. George Washington University, Washington, DC
18. McCormick WT, Schweitzer PJ, White TW (1972) Problem decomposition and data reorganization by a clustering technique. *Oper Res* 20:993–1009
19. McLachlan GJ, Basford KE (1988) Mixture models: inference and application to clustering. Marcel Dekker, New York, NY
20. Piccolo D (2003) On the moments of a mixture of Uniform and shifted Binomial random variables. *Quaderni di Statistica* 5:85–104
21. Piccolo D (2006) Observed information matrix for MUB models. *Quaderni di Statistica* 8:33–78
22. Piccolo D, D'Elia A (2008) A new approach for modelling consumers' preferences. *Food Qual Prefer* 19:247–259
23. Podani J (2005) Multivariate exploratory analysis of ordinal data in ecology: pitfalls, problems and solutions. *J Veg Sci* 16:497–510
24. Vermunt JK, Magidson J (2002) Latent class cluster analysis. In: Hagenaars JA, McCutcheon AL (eds) Advances in latent class analysis. Cambridge University Press, Cambridge