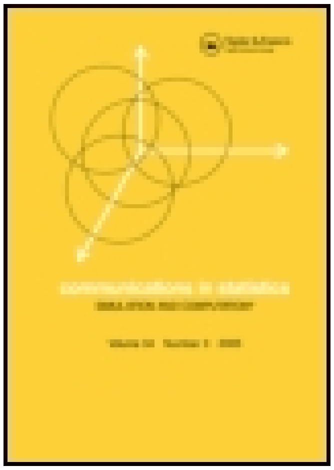

Communications in Statistics - Simulation and Computation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

Multivariate Approach For Comparative Evaluations Of Customer Satisfaction With Application To Transport Services

Stefano Bonnini[a]

[a] Department of Economics and Management, University of Ferrara Via Voltapaletto, 11-44121, Ferrara (Italy),

Accepted author version posted online: 05 Jun 2015.

MULTIVARIATE APPROACH FOR COMPARATIVE EVALUATIONS OF CUSTOMER SATISFACTION WITH APPLICATION TO TRANSPORT SERVICES

Stefano Bonnini
Department of Economics and Management
University of Ferrara
Via Voltapaletto, 11 - 44121 Ferrara (Italy)
stefano.bonnini@unife.it

Key Words: customer satisfaction; multivariate analysis; permutation test; ranking problem; pairwise comparisons.

ABSTRACT

In problems related to evaluations of products or services (e.g. in customer satisfaction analysis) the main difficulties concern the synthesis of the information, which is necessary for the presence of several evaluators and many response variables (aspects under evaluation). In this paper the problem of determining and comparing the satisfaction of different groups of customers, in the presence of multivariate response variables and using the results of pairwise comparisons is addressed. Within the framework of group ranking methods and multicriteria decision making theory, a new approach, based on nonparametric techniques, for evaluating group satisfaction in a multivariate framework is proposed and the concept of Multivariate Relative Satisfaction is defined. An application to the evaluation of public transport services, like the railways service and the urban bus service, by students of the University of Ferrara (Italy) is also discussed.

1. INTRODUCTION

In the evaluation of products or services in general and in customer satisfaction analysis in particular some mathematical and statistical issues arises. The main difficulties often consist in the synthesis of the information related to the satisfaction of a group of evaluators (multiplicity of subjects) and in the synthesis of the information in the presence of multivariate response variables, because more than one aspect or item is under evaluation (multiplicity of variables). Some studies are addressed to identify and measure the several service quality dimensions (Parasuraman et al. 1985, 1988). Bolton and Drew (1991) develop a multistage model of the determinants of perceived service quality and service value, and describe how expectations, perceptions and disconfirmation experiences of customers affect their satisfaction with a service.

Basically satisfaction is an abstract concept and varies among individuals and products or services. It depends on covariates which represent characteristics of individuals or objects. The main goal of customer satisfaction analysis is often studying the affect of individual factors on the overall satisfaction. The final purpose of the analysis usually consists in determining and comparing the satisfaction of different groups of customers (market segments), for adapting marketing strategies to market segments' characteristics through targeted communication strategies and product/service differentiation.

Some methods used in customer satisfaction analysis are based on ordinary least squares, conjoint analysis (see Arboretti et al., 2005), Shapley value regression, penalty and reward analysis, Kruskals relative importance, canonical correlation analysis, partial least squares and logistic regression, composite indicators (Marozzi, 2009a, 2012). Funa (2011) discusses and uses some of these tools for exploring the relation between explanatory variables and dependent variables and for identifying the greatest satisfiers and dissatisfiers influencing customer satisfaction. She applies some of the mentioned tools on a customer satisfaction survey regarding new car owners.

In the framework of the Total Quality Management and specifically of the so called early warning system, which provides signals for possible risks, Lombardo (2011) proposes exploratory tools for identifying at-risk customers and allowing for more timely interventions. She focuses on customer satisfaction in services to persons of public utility, like training services and health care

ACCEPTED MANUSCRIPT

services. The use of ordinal categorical variables for representing satisfaction, leads her to study the strength of variable associations by multiple correspondence analysis via polynomial transformations, obtaining clusters of individuals ordered with respect to satisfaction levels. Then she investigates on the dependence among variables through regression models by means of Boosting regression and partial least squares techniques.

Since the aspects of a service to be evaluated are less evident than those of a product, customer satisfaction analysis of services presents some specific difficulties. On the other hand, the need of assessing the quality of services, according to ISO 9000 standards (ISO, 2009), charters of services and quality management system, leads many companies to implement procedures for assessing the quality of services and in particular the customer satisfaction. This is true also in the public transport sector. In this area some scientific works have been tried to develop methodologies for studying the latent aspects of the customer satisfaction. Since the passenger satisfaction is related to cognitive emotional and psychological aspects (Oliver, 1993), its measurement is a complex problem because of its subjective nature. Hence these studies focus on what is perceived, rather than what is supplied.

Gallo et al. (2009) discuss the application of Rasch Analysis and Simple Components Analysis based on the RV coefficient, for studying passengers' satisfaction in the local public transport. Multidimensionality of customer satisfaction and the different nature of data are considered by Gallo and Ciavolino (2009), who analyze the spatial effects of the territorial dislocation of stations by means of a rating scale model and a spatial structural equation model. Satisfaction for urban public transportation is also studied by Bernini and Lubisco (2005, 2009), who propose and extended dynamic version of LISREL model to investigate possible changes over time of customer satisfaction, the main factors affecting satisfaction and the effects of customer covariates, with application to the Tram Service in Rimini (Italy). Local public transport is also studied by Rostirolla and Romano (2009), for evaluating the opportunity to move the mobility demand from private to public transport. For this purpose they perform a statistical analysis of satisfaction for the taxi

ACCEPTED MANUSCRIPT

service and an economic-financial analysis of costs and benefits.

Except for airline industry, few empirical studies have been dedicated on satisfaction for transportation services (Sumaedi et al., 2012). Passengers' behavioral intentions in Taipei are studied by Jen and Hu (2003), Wen et al. (2005) and Lai and Chen (2011). Sumaedi et al. (2012) analyze the relationship between passengers' behavioral intentions and others latent factors affecting them in Jakarta city, by using structural equation modeling techniques.

This paper is dedicated to the problem of determining and comparing the satisfaction of different groups of customers, in the presence of multivariate response variables and using the results of pairwise comparisons. Section 2 concerns a literature overview and a discussion on some issues on methodological solutions for rating and ranking problems. In section 3 a new approach, based on nonparametric techniques, for evaluating group satisfaction in a multivariate framework is described and the concept of Multivariate Relative Satisfaction is introduced. Section 4 is devoted to the application of the proposed method to the evaluation of public transport services like the railways service and the urban bus service, using data from a statistical survey on living ad studying conditions of students of University of Ferrara (Italy). In section 5 the final conclusions of the study are summarized.

2. COMPARISONS OF GROUPS IN RATING AND RANKING PROBLEMS: LITERATURE OVERVIEW AND METHODOLOGICAL ISSUES

The problem of determining a global rating or a global ranking in the presence of multivariate response variables has been extensively dealt in the statistical, mathematical and economical literature under many methodological points of view and in several applications. The application frameworks vary from sports, to decision making problems, to machine learning, to performance analysis, etc. Specialized literature has been created and scientific studies has been dedicated to analyze and deepen specific aspects of this complex problem. The growing interest towards the problem of many researchers from different disciplines (engineering, mathematics, economics, statistics, ...) proves the importance and the multidisciplinary nature of this topic.

From the statistical point of view, the paper is focused on the general problem of comparing g multivariate populations for determining a global rating or a global ranking of these g populations. In particular methods based on pairwise comparisons are considered. In other words, by using a formalization typical of the ANOVA and MANOVA problems, we have n multivariate sample observations which can be grouped according to a factor representing g real or symbolic treatments. The number of observations is given by $n = n_1 + ... + n_g$ where n_j is the size of the sample drawn from population j, namely related to the j-th treatment, with $j = 1, \dots, g$.

Many scientific works have addressed the problem of defining a procedure for comparing and ranking a list of projets, services, teams, organizations, and other (according to the application framework), hereafter denoted by objects, in the presence of group-evaluations. Mainly this problem has been studied and framed under the generic label ”group ranking”. In ”group ranking” problems the main goal consists in consolidating and aggregating the individuals’ rankings or ratings to obtain a group-ranking or group-rating, useful for comparing the objects (Hochbaum and Levin, 2006). In these problems the factor denoting the g groups represents the objects (treatments) under evaluation and the n_j replications are the observations related to the evaluations of the j-th object. When the set of evaluators is the same for all the objects, we have balanced dependent samples of m observations and thus $n_1 = \dots = n_g = m$ and $n = gm$ (e.g. m customers express their satisfaction about g products or services). Otherwise, in the absence of matching techniques in the design of the survey, we have independent samples and each group of customers expresses his satisfaction for only one object.

In our specific problem we have just one object under evaluation and the treatment factor denotes groups of evaluators. We wish to compare g groups of evaluators and compute the relative satisfaction of each group with respect to all the others. Hence the design of the survey presents only the case of g independent samples. In this paper a general method is proposed. This method is suitable for multivariate observations, that is when the evaluations (satisfaction levels) are related

ACCEPTED MANUSCRIPT

to q aspects or items with $q > 1$, but applicable also to the simpler and special case of univariate observations, that is when $q = 1$, and easily adaptable both to numeric and categorical responses, for rating and ranking problems, in the presence of dependent or independent samples. In some works rating problems are denoted by ”intensity” ranking problems, hence hereafter, unless otherwise indicated, with ranking problems a general class of problems is denoted, including both rating and ranking problems.

Many studies on group ranking concern problems of voting and elections. An important result is given by the ”impossibility” theorem of Arrow (1963) who proves that specific natural fairness properties cannot be guaranteed by a voting scheme. An axiomatic approach based on the determination of the overall group ranking by minimizing the deviation from individual preference rankings is proposed by Kemeny and Snell (1962). Keener (1993) discusses the intensity ranking problem based on pairwise comparisons distinguishing between *direct method* and *nonlinear scheme*. In the *direct method* the overall score (rank) of the j-th object s_j is proportional to a linear combination of the results of the pairwise comparisons of the j-th object with the other objects, where the weights consist in the overall scores of the other objects. Such results of the pairwise comparisons a_{jr} are preference scores which quantify how much object j is preferred to object r with $r = 1, \ldots, g$. The preference scores should satisfy the following properties:

1. $a_{jr} \geq 0$

2. $a_{jr} + a_{rj} = 1$.

In other words the overall score of an object depends on the preference scores respect to the other objects weighted with the overall scores of the objects themselves. Formally

$$vs_j = \sum_r a_{jr} s_r, \tag{1}$$

where v is a constant of proportionality. With matrix notation:

ACCEPTED MANUSCRIPT

$$\nu\mathbf{s} = \mathbf{A}\mathbf{s}, \tag{2}$$

where $\mathbf{A} = [a_{jr}]$ is the preference matrix and $\mathbf{s}$ is the ranking (score) vector. The solution of this system of equations consists in determining a positive eigenvector $\mathbf{s}$ for the positive matrix $\mathbf{A}$. The Perron-Frobenius theorem (Keener, 1993) provides conditions to have a unique solution to this problem. A very common choice for the preference scores is

$$a_{jr} = \begin{cases} 1 & \text{if} \quad j \quad \text{is preferred to} \quad r \\ 0.5 & \text{in the case of tie (indifference)} \\ 0 & \text{if} \quad r \quad \text{is preferred to} \quad j. \end{cases}$$

Another possible choice is given by the product of the a_{jr} defined above and $1/n_j$.

The *nonlinear scheme* consists in the following generalization of the direct method:

$$s_j = \sum_r h(e_{jr}), \tag{3}$$

where $e_{jr} = a_{jr}s_r$ are partial scores for the pairwise comparisons and $h(x)$ is a suitable nonnegative continuous increasing function. When $h(x) = cx$, for a given constant c, we have the direct method, otherwise s_j is a nonlinear function of $s_1, \ldots, s_g$. This theory can be easily adapted to our problem, where instead of comparing g different objects we compare g different groups of evaluators.

The solution connected with the Perron-Frobenius theorem is important for the Analytic Hierarchy Process (AHP) proposed by Saaty (1977, 1980). The AHP has a key role in *Multicriteria Decision Making* theory (MCDM), based on the idea that the choices of the evaluators may include a degree of uncertainty and thus based on fuzzy preference models (see Brans and Vincke, 1985, and Fuller and Carlsson, 1996). The goal of the MCDM is to determine a ranking of a set of actions (decisions) $D = \{d_1, \ldots, d_g\}$ according to k criteria. This general procedure is introduced as a ranking problem or choice problem in the decision making framework and widely studied and applied by economists and experts in business economics and management, but it can be easily extended

ACCEPTED MANUSCRIPT

and adapted to many applications related to other disciplines. Considering our problem of customer satisfaction, the actions correspond to the groups of customers, the criteria correspond to the response variables (satisfactions related to partial aspects) and the goal consists in ranking these groups according to their overall satisfaction. Given the specific application framework, for our problem the concept of preference among two actions or decisions can be more suitably replaced by the concept of dominance between two groups or populations.

Let us consider a set of actions $D = \{d_1, \ldots, d_g\}$ and q criteria $f_1, \ldots, f_q$ differentiating these actions, with $f_k : D \rightarrow \Re$, $k = 1, \ldots, q$. According to classical theory, d_j dominates (is preferred to) d_r if and only if $f_k(d_j) \geq f_k(d_r)$ $\forall k \in \{1, \ldots, q\}$ and $\exists k \in \{1, \ldots, K\}$ such that $f_k(d_j) > f_k(d_r)$. In this way we introduce a partial order (transitive relation) on D called *dominance order*, based on the ”unanimity of the points of view”, in the sense that d_j dominates d_r if and only if, for each criterium, we have dominance of the former on the latter or indifference, and for at least one criterium we have strict dominance. The ”unanimity” is not very common in real problems, hence three types of solutions where proposed for overcoming this limit: (1) aggregation methods using utility functions; (2) interactive methods; (3) outranking methods. The type (1) methods are included in the *Multi Attribute Utility Theory* (MAUT) which focuses on problems of choices among multiattribute alternatives in the presence of risk and uncertainty and considers methods for assessing individual values and subjective probabilities (see Dyer et al., 1992). The models commonly used to represent the preferences of the decision makers are based on linear additive or multiplicative nonlinear utility functions (Keeney and Raiffa, 1976). The interactive methods at point (2) require a progressive articulation of preferences, in the sense that they require information about preferences from the decision maker, throughout the solution process (see Geoffrion et al., 1972; Köksalan and Sagala, 1995). The type (3) methods are based on a majority principle, instead of the unanimity principle, and their application is divided into two phases: in the first phase an outranking relation is determined; in the second phase an exploitation of this relation is made for decision making (see Brans and Vicke, 1985).

The most important outranking solutions are the ELECTRE methods proposed by Roy (1973, 1977). An evolution of this solution is the PROMEETHE method, easier to be understood for economists and simpler to be applied for users (Brans and Vicke, 1985). In the classical notion of criterion, d_j is preferred to d_r according to the k-th criterion if and only if $f_k(d_j) - f_k(d_r) > 0$ and we have indifference between d_j and d_r if and only if $f_k(d_j) - f_k(d_r) = 0$. The outranking methods introduce the notions of quasi-criterion (with a larger area of indifference) and pseudo-criterion (with an area of hesitation between indifference and preference). They define the *preference function* for the k-th criterion when comparing two decisions as a function which can take values between 0 and 1,

$$p_k : D \times D \to [0, 1].$$

For small values the relation tends to indifference; for large values the relation tends to dominance, and value 1 correspond to strict dominance. Formally:

$$p_k(d_j, d_r) = u[f_k(d_j) - f_k(d_r)] = {}_k a_{jr} \in [0, 1], \quad (4)$$

where ${}_k a_{jr}$ is the preference value according to the k-th criterion.

Table 1 shows the main criterion types defined in the PROMETHEE method. The threshold parameters l, m, l_1, l_2 and the parameter σ have positive values that should be defined by the decision makers. The quasi-criterion consider an interval instead of a point (the zero value) as area of indifference. The criterion with linear preference, the level criterion, the criterion with linear preference and indifference area and the gaussian criterion admit that the preference value ${}_k a_{jr}$ may vary as increasing function of the difference $f_k(d_j) - f_k(d_r)$ and (except the gaussian criterion) they distinguish between simple dominance (when $0 < {}_k a_{jr} < 1$) and strict dominance (when ${}_k a_{jr} = 1$). The values aggregation respect to the criteria is obtained by computing the mean.

$$p(d_j, d_r) = q^{-1} \sum_k p_k(d_j, d_r) = q^{-1} \sum_k ({}_k a_{jr}) = a_{jr}. \quad (5)$$

The second phase of the PROMETHEE method consists in an exploitation of the outranking relation establishing a partial or total preorder on D on the base of the scores $s_j^+ = \sum_r a_{jr}$ and $s_j^- = \sum_r a_{rj}$. A graph representation, where the nodes are $d_1, \ldots, d_g$ and each arc (d_j, d_r) has the value a_{jr}, is used in this second phase.

3. MULTIVARIATE RELATIVE SATISFACTION OF CUSTOMER GROUPS

Let us consider the general problem of determining a vector of g values which represent the overall satisfactions of g different groups of evaluators. For denoting this satisfaction the acronym *MRS* (Multivariate Relative Satisfaction) is used. The g overall values depend by the pairwise comparisons between the g groups.

The main difficulty consists in the multivariate nature of the problem. Several aspects are under evaluation hence the satisfaction is multidimensional. The solution to this problem include a suitable method for reducing the dimension of the response variable from q (number of aspects under evaluation) to 1, that is to synthesize the information provided by the q-variate response variable.

A second difficulty is related to the multiplicity of evaluators involved. In the j-th group we have n_j evaluators. The goal consists in determining a value for each group to represent the overall satisfaction of the group, by synthesizing the information respect to the evaluators. Our method is based on the pairwise comparisons between groups, hence this synthesis is made in the phase of pairwise comparisons.

Let us denote with X_{ijk} the satisfaction expressed by the i-th evaluator in the j-th group for the k-th aspect, with $i = 1, \ldots, n_j$, $j = 1, \ldots, g$ and $k = 1, \ldots, q$. Let us assume the observed values $X_{1jk}, \ldots, X_{n_j jk}$ are independent realizations of the random variable Z_{jk}. According to the evaluation scale, Z_{jk} can be numerical or ordinal categorical. The q variables representing the satisfactions for the q aspects may be not of the same type: some of them can be numerical and some others can be categorical.

ACCEPTED MANUSCRIPT

The proposed solution is divided into two phases: in the first phase $g(g-1)/2$ multivariate tests on stochastic dominance are performed as pairwise comparisons between the g groups; in the second phase the $g \times g$ matrix of p-values of the pairwise comparisons is considered and, by combining the values for each row of the matrix, the vector of values representing the satisfactions of the g groups is obtained.

In the first phase, for the couple of groups (j, r), a nonparametric inferential procedure is applied to test the null hypothesis

$$H_0 : [\mathbf{Z}_j =^d \mathbf{Z}_r] \equiv [\bigcap_{k=1}^{q} Z_{jk} =^d Z_{rk}] \tag{6}$$

against the alternative hypothesis

$$H_1 : [\mathbf{Z}_j >^d \mathbf{Z}_r] \equiv [\bigcup_{k=1}^{q} Z_{jk} >^d Z_{rk}], \tag{7}$$

where $\mathbf{Z}_j = (Z_{j1}, \ldots, Z_{jq})'$ is the q-variate variable which represents the evaluations of the j-th group, with $j = 1, \ldots, g$ and the symbols $=^d$ and $>^d$ denote ”equality in distribution” and ”stochastic dominance” respectively (see Arboretti and Bonnini, 2009). The use of the intersection symbol in (6) means that the null hypothesis is true if, for each of the k component variables, the equality in distribution is true. The use of the union symbol in (7) means that the alternative hypothesis is true if, for at least one of the q component variables, the stochastic dominance is true and for the other component variables the equality in distribution is true.

This testing problem can be solved with the application of a multivariate permutation test for stochastic dominance (see Pesarin, 2001; Pesarin and Salmaso, 2010). An univariate permutation test for stochastic dominance is applied for each variable and a suitable combination of the p-values provides a univariate statistic for the multivariate testing problem. For the combination, different functions can be used, thus for example the test statistic for the pairwise comparison for multivariate stochastic dominance between j-th and r-th group may be

1. $T_{jr} = -2 \sum_k \ln(\lambda_{jrk})$ (Fisher’s combination)

2. $T_{jr} = \sum_k \phi^{-1}(1 - \lambda_{jrk})$ (Liptak's combination)

3. $T_{jr} = \max_k(1 - \lambda_{jrk})$ (Tippett's combination)

where λ_{jrk} is the p-value of the univariate test, adjusted in order to take values strictly included in the interval $(0, 1)$ (for more details see Pesarin, 2001) and ϕ is the gaussian cumulative distribution function. This test is distribution-free and may be applied to numeric, categorical or mixed multivariate variables. No assumption is needed about the multivariate distribution. Neither parameter values (e.g. correlations) nor type of dependence (e.g. linear or nonlinear) have to be known or explicitly assumed.

Each combined p-value λ_{jr} can be used to determine suitable preference scores. A possible choice is to consider λ_{jr}^{-1} as dominance score between group j and group r. Of course, for computing the matrix of the p-values, only $g(g-1)/2$ tests for pairwise comparisons are needed because $\lambda_{jr} = 1 - \lambda_{rj}\ \forall j, r = 1, \ldots, g$.

In the second phase of the procedure the following combination of the values in the j-th row (based on the combining function chosen for the multivariate test on stochastic dominance) is applied for computing an overall (satisfaction) score for the j-th group:

1. $s_j = -\sum_{r=1}^{g} \ln(\lambda_{jr})$. (Fisher's combination)

2. $s_j = \sum_{r \neq j}\{\phi^{-1}(1 - \lambda_{jr}) - \phi^{-1}[\min_t(1 - \lambda_{jt})]\} + \epsilon$ (Liptak's combination)

3. $s_j = \max_r(1 - \lambda_{jr})$ (Tippett's combination)

where the change in the Liptak's combination, respect to the formula of the multivariate test statistic for pairwise comparisons, is necessary to have positive scores and ϵ is a small positive quantity (e.g. 10^{-5}).

The choice of the combination function must take into account the following properties: due to the logarithmic transformation, in case of Fisher's combination, the overall score s_j tends to be large in case of large dominance of the j-th group even only on one or few groups (reward effect),

ACCEPTED MANUSCRIPT

while to be dominated by one or more groups does not particularly penalize in the overall score; using Liptak's combination the reward effect is attenuated and there is a symmetrical penalty effect for each comparison where the group is dominated; Tippet's combination considers only the "best" result of the pairwise comparisons, ignoring all the others. Hence the choice of the combination function must be based on the importance given to the reward effect and to the penalty effect.

Finally, for interpretation convenience, the Multivariate Relative Satisfaction (MRS) can be computed by comparing each overall score with the maximum observed value. The relative satisfaction of a group takes values in $(0, 1]$ and 1 corresponds to the most satisfied group. If the relative satisfaction is equal to 0.8, the satisfaction of the corresponding group is 80% of that of the most satisfied group. Formally

$$\mathbf{MRS} = \varphi(\mathbf{s}), \ \varphi : (0, \infty)^g \to (0, 1]^g, \tag{8}$$

and specifically

$$\varphi(\mathbf{s}) = c\mathbf{s}, \ \text{with } c = max\{s_1, \ldots, s_g\}^{-1}. \tag{9}$$

Let us consider the following dominance values

$$e_{jr} = 1 - \lambda_{jr}.$$

They satisfy the two typical properties of the preference (or dominance) scores. As a matter of fact $e_{jr} \geq 0$ and $e_{jr} + e_{rj} = 1$. Like in the PROMETHEE method, the results of the pairwise comparisons are not simply indifference ($e_{jr} = 0$) or strict dominance ($e_{jr} = 1$), but a measure of dominance e_{jr} between 0 and 1. Note that the proposed method is consistent with the nonlinear scheme ed by Keener (1993), because the combination satisfies (3) when:

1. $h(x) = -ln(1 - x)$ (Fisher's combination)

2. $h(x) = \{\phi^{-1}(x) - \phi^{-1}[\min(x)] + \epsilon/(g - 1)\}I_{(0,1)}(x)$ (Liptak's combination)

3. $h(x) = \max(x)$ (Tippett's combination),

where *min* and *max* correspond to the observed minimum and maximum value respectively, I_A is the indicator function of the set A and $(0, 1)$ is the open interval of real numbers between 0 and 1.

4. CUSTOMER SATISFACTION ABOUT TRANSPORT SERVICES: THE CASE OF THE STUDENTS OF THE UNIVERSITY OF FERRARA

In the spring 2012 a statistical survey about living and studying conditions of university students was performed in Ferrara (Italy). A stratified random sample of 747 students was drawn, with *gender* and *type of enrollment* as stratification variables. The sample size corresponds to 4.2% of the whole population size. Through the *CATI* method (Computer-Assisted Telephone Interviewing), the students in the sample were asked to answer some questions related to their habits, spare time activities, used services and working and economic conditions. A section of the questionnaire was dedicated to means and services used by the students for moving from house to University. In particular, among the other considered means of transport, the questionnaire was focused on the public rail service and on the urban bus service. In the following subsections the described approach to determine the *MRS* of groups of students defined according to *gender*, *type of enrollment* and *housing condition* is applied.

4.1. CUSTOMER SATISFACTION FOR THE PUBLIC RAIL SERVICE

For medium-long distances the main mean of transport used by the students for moving from their house to University, for attending lessons or doing examinations, is the train. In this analysis, postgraduate students, who have to attend lessons and do examinations much less often than undergraduate students, are excluded. Specifically the focus is on the commuters, undergraduate students who live outside Ferrara, who do not want or cannot buy or rent a house in Ferrara, and daily or almost daily go to university.

In the sample of the survey on the students of Ferrara, among the undergraduate students the commuters are 204. By considering the factors *gender* and *type of enrollment* (freshmen or other students), it is possible to define four groups of students. The global percentage of commuters who usually take the train for moving from their house to University is 41.7%. The major users of the public rail service correspond to the group of male-freshmen (48.0% of users); the minor users are the group of female-freshmen (33.3%)(see Table 2). The 85 commuters who use the service were asked to give an opinion about six different aspects of the service, by rating them according to a numeric scale from 1 =*worst evaluation* (maximum dissatisfaction) to 10 =*best evaluation* (maximum satisfaction). The considered aspects are: price, punctuality, times, comfort, cleanliness and geographical distribution. In this rating system, the minimum value corresponding to a satisfaction level, that is a positive evaluation, is 6. Hence, according to the mean values reported in Table 2, almost all the groups are unsatisfied for almost all the aspects. The worst mean evaluation 4.1 is related to punctuality (given by the group male-others) and to cleanliness (given by the group female-others). The only positive evaluations (slightly exceeding 6) are related to comfort and given by male- and female-freshmen.

The application of the multivariate directional tests for pairwise comparisons to the satisfaction data for the rail service, gives the results described in Table 3. We use the Fisher's rule for combination because we wish the overall score s_j tends to be large in case of large dominance of the j-th group even only on one or few groups (reward effect), while we are not interested in including a penalty effect in computation of the overall score. In the 4×4 table of the p-values, the minimum (0.113) corresponds to the test which compare the satisfaction of male-freshmen and that of male-others for testing if the former is greater than the latter. Hence this is the hypothesis most supported by the empirical evidence. The most similar groups, in terms of satisfaction, are female-others and male-others, because the p-values related to the two directional tests of these groups (0.373 and 0.627) are the nearest, respect to all the possible couples of groups.

Combining the p-values with the Fisher's function according to the rows of the matrix, we obtain the global score which measures the relative satisfaction of each group. The greatest relative satisfaction is related to male-freshmen (combined score equal to 5.284), whose normalized value is then equal to 1. The second greatest combined score is 3.699 and corresponds to the group female-freshmen. The normalized satisfaction or *MRS* of this group is then 0.700, namely the satisfaction of female-freshmen is equal to 70% of the satisfaction of male-freshmen. The *MRS* of female-others is equal to 0.253 and finally, the less satisfied group is male-others with a *MRS* equal to 0.152.

4.2. CUSTOMER SATISFACTION FOR THE URBAN BUS SERVICE

For short distances, that is for moving within the city, the main public mean of transport of students is the bus. Also in this analysis, for the same reason of the previous one, we do not consider postgraduate students. In this case the most important factor affecting students' behavior is the *housing condition*, hence it is useful to distinguish between residents, offsite students (who usually reside in another city but live in Ferrara during the study period) and commuters. Thus by jointly considering *gender* and *housing condition* we can distinguish six groups of students and analyze their relative satisfaction.

The total number of students with these characteristics in the sample is 567. The global percentage of undergraduate students who usually take the urban bus for moving from their house to University is 8.6%. The major users of the urban bus service correspond to the group of female-commuters (13.8%); the minor users are the group of male-commuters (4.2%)(see Table 4). The 49 students who use the service were asked to give an opinion about the same six aspects considered in the previous subsection regarding the public rail service. Also the rating method, according to a numeric scale from 1 =*worst evaluation* (maximum dissatisfaction) to 10 =*best evaluation* (maximum satisfaction), is the same. According to the mean values reported in Table 4, the mean satisfactions seem to be slightly greater than for the public rail service. The worst mean evaluation 5.3 is related to price (given by the group male-offsite and male-commuters), cleanliness and geographical distribution (given by the group male-commuters). The best evaluation (8.7) is related to comfort and given by male-residents.

The application of the multivariate directional tests for pairwise comparisons to the satisfaction data for the urban bus service, gives the results described in Table 5. In the 6×6 table of the p-values, the minimum (0.025) corresponds to the test which compare the satisfaction of male-residents and that of female-offsite for testing if the former is greater than the latter. At significance α level 0.05, this is the only significant test together with the one regarding the comparison between female-commuters and female-offsite. The most similar satisfactions concern the couple of groups female-resident and male-offsite. The most satisfied group is that of male-residents (combined score equal to 9.836), whose normalized value is then equal to 1. The second most satisfied group is that of female-commuters and presents a very similar combined score (9.436), that is the MRS is equal to 0.959 or, in other words, their satisfaction is 95.9% of the maximum observed satisfaction. The minimum satisfaction ($MRS = 0.106$) is that of female-offsite.

5. CONCLUSIONS

For the problem of computing the overall satisfaction of different groups of customers considering several aspects of a product or service, in other words Multivariate Relative Satisfaction (MRS), a procedure based on multivariate directional permutation tests, a suitable p-value combination and a final normalization of the scores is proposed. The new method satisfies some properties typical of well known solutions for group ranking problems and multicriteria decision making problems. It also presents some advantages typical of other methods of the specific literature, like the time saving algorithms of some group ranking solutions and the ease of application and interpretation of the PROMETHEE methods.

Some additional advantages characterize the proposed solution. Firstly it is a method based on inferential techniques, hence the application results can be extended to populations as they are not purely based on descriptive statistics.

Then the inference is not based on likelihood methods or assumptions about the underlying multivariate distribution law. The dependence structure among the components of the multivariate response variable needs not to be explicitly assumed because it is implicitly taken into account

ACCEPTED MANUSCRIPT

by the adopted permutation strategy. It is not even required the use of asymptotic properties for assuming specific null distributions of the test statistics, hence the method is suitable in presence of small sample sizes.

Furthermore the procedure is very flexible and applicable to numeric, categorical and mixed multivariate variables, useful for ordering objects under evaluations or groups of evaluators, and appropriate both with dependent or independent samples. The proposed combination method does not assume linear relations between the overall scores and the pairwise dominance scores or the original variables. It follows that the existence of one solution is not dependent on the application of the Perron-Frobenius theorem or other mathematical propositions.

Finally the results of the described procedure do not depend on specified parameters, threshold or specific rules for establishing the dominance relations, as in the case of other group ranking or multicriteria methods.

A possible direction of future research may be the consideration of other tests for comparing the groups like those for the location/scale problem (see Marozzi 2009b, 2013).

BIBLIOGRAPHY

Arboretti Giancristofaro, R., Bonnini, S. (2009). Some new results on univariate and multivariate permutation tests for ordinal categorical variables under restricted alternatives. *Statistical Methods and Applications - Journal of the Italian Statistical Society*, **18**, 2, 221-236.

Arboretti Giancristofaro, R., Marozzi, M., Salmaso, L. (2005). Nonparametric Pooling and Testing of Preference Ratings for Full-Profile Conjoint Analysis experiments. *Journal of Modern Applied Statistical Methods*, **4**, 545–552.

Arrow, K. J. (1963). *Social Choice and Individual Values*. New York: Wiley.

Bernini, C., Lubisco, A. (2005). A new extension of dynamic simplex model for the public transport customer satisfaction. *Statistica*, **LXV**, 4, 367–386.

ACCEPTED MANUSCRIPT

Bernini, C., Lubisco, A. (2009). Modelling dynamic customer satisfaction in urban public transportation. *Global and Local Economic Review*, **13**, 2, 87–107.

Bolton, R.N., Drew, J.H. (1991). A Multistage Model of Customers' Assessments of Service Quality Value. *Journal of Consumer Research*, **17**, 375–384.

Brans, J.P., Vincke, P. (1985). A preference ranking organization method. *Management Science*, **31**, 6, 647–656.

Dyer, J.S., Fishburn, P.C., Steuer, R.E., Wallenius, J., Zionts, S. (1992). Multiple Criteria Decision Making, Multiattribute Utility Theory: the Next Ten Years. *Management Science*, **38**, 5, 645–654.

Fuller, R., Carlsson, C. (1996). Fuzzy multiple criteria decision making: Recent developments. *Fuzzy Sets Systems*, **78**, 139–153.

Funa, L. (2011). *Customer Satisfaction Analysis*. Master Thesis in Statistics, Data Analysis and Knowledge Discovery. Department of Computer and Information Science, Linköping University, Linköping, Sweden.

Gallo, M., Ciavolino, E. (2009). Multivariate statistical aproaches for customer satisfaction into the transportation sector. *Global and Local Economic Review*, **13**, 2, 55–70.

Gallo, M., D'Ambra, A., Camminatiello, I. (2009). The evaluation of passenger satisfaction in the local public transport: a strategy for data analysis. *Global and Local Economic Review*, **13**, 2, 71–86.

Geoffrion, A.M., Dyer, J.S., Feinberg, A. (1972). An interactive approach for Multi-Criterion Optimization, with application to the Operation of an Academic Department. *Management Science*, **19**, 4, 357–368.

Hochbaum, D.S., Levin, A. (2006). Methodologies and Algorithms for Group-Rankings Decision. *Management Science*, **52**, 9, 1394–1408.

ISO - International Organization for Standardization (2009). *Selection and use of the ISO 9000 family of standards*. Geneve: ISO Central Secretariat.

Jen, W., Hu, K.C. (2003). Application of perceived value model to identify factors affecting passengers' repurchase intentions on city bus: a case of the Taipei metropolitan area. *Transportation*, **30**, 307–321.

Keener, J.P. (1993). The Perron-Frobenius Theorem and the Ranking of Football Teams. *Society for Industrial and Applied Mathematics*, **35**, 1, 80–93.

Keeney, R., Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Trade-offs*. New York: Wiley.

Kemeny, J.G., Snell, J.L. (1962). Preference ranking: axiomatic approach *Mathematical Models in the Social Sciences*. Boston: Ginn.

Köksalan, M.M., Sagala, P.N.S. (1995). Interactive Approaches for Discrete Alternative Multiple Criteria Decision Making with Monotone Utility Functions. *Management Science*, **41**, 7, 1158–1171.

noindent Lai, W.T., Chen, C.F. (2011). Behavioral intentions of public transit passengers - the role of service quality, perceived value, satisfaction and involvement. *Transport Policy*, **18**, 2, 318–325.

Lombardo, R. (2011). Data Mining and Explorative Multivariate Data Analysis for Customer Satisfaction Study, Koyuncigil, A.S. and Ozgulbas, N. (Eds), *Surveillance Technologies and Early Warning Systems: Data Mining Applications for Risk detection*. Hershey: Information Science Reference.

Marozzi, M. (2009a). A Composite Indicator Dimension Reduction Procedure with Application to University Student Satisfaction. *Statistica Neerlandica*, **63**, 258–268.

Marozzi, M. (2009b). Some Notes on the Location-Scale Cucconi Test. *Journal of Nonparametric Statistics*, **21**, 629–647.

Marozzi, M. (2012). Tertiary Student Satisfaction with Socialization: a Statistical Assessment. *Quality and Quantity*, **46**, 1271–1278.

Marozzi, M. (2013). Nonparametric Simultaneous Tests for Location and Scale Testing: a Comparison of Several Methods. *Communication in Statistics: Simulation and Computation*, **42**, 1298–1317.

Oliver, R.L. (1993). A conceptual model of service quality and service satisfaction: compatible goals, different concepts, in *Advances in Service Marketing and Management: Research and Practice, 2*. JAI Press.

Parasuraman, A., Zeithaml, V.A., Berry, L.L. (1985). A Conceptual Model of Service Quality and Its Implications for Future Research. *Journal of Marketing*, **49**, 4, 41–50.

Parasuraman, A., Zeithaml, V.A., Berry, L.L. (1988). SERVQUAL: A Multiple Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, **64**, 1, 12–37.

Pesarin, F. (2001). *Multivariate Permutation Tests with Applications in Biostatistics*. Chichester: Wiley.

Pesarin, F., Salmaso, L. (2010). *Permutation Tests for Complex Data: Theory, Applications and Software*. Chichester: Wiley.

Rostirolla, P., Romano, O. (2009). Taxi service and sustainable mobility: a people satisfaction survey. *Global and Local Economic Review*, **13**,2, 109–124.

Roy, B. (1973). How Outranking Relations Help Multicriteria Decision Making, Cochran and Zeleny (Eds), *Multiple Criteria Decision Making*. Columbia: University of South Carolina.

Roy, B. (1977). Conceptual Framework for a Normative Theory of Decision Aid, Starr and Zeleny

(Eds), *Multiple Criteria Decision Making*. North Holland, Amsterdam.

Saaty, T. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, **15**, 234–281.

Saaty, T. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.

Sumaedi, S., Bakti, G.M.Y, Yarmen, M. (2012). The empirical study of public transport passengers' behavioral intentions: the roles of service quality, perceived sacrifice, perceived value, and satisfaction (case study: paratransit passengers in Jakarta, Indonesia). *International Journal for Traffic and Transport Engineering*, **2**, 1, 83–97.

Wen, C.H., Lawrence, W.L., Cheng, H.L. (2005). Structural equation modeling to determine passenger loyalty toward intercity bus service. *Journal of the transportation research board*, **1927**, 249–255.

ACCEPTED MANUSCRIPT

Table 1: Types of criteria defined in the PROMETHEE method, described in Brans and Vicke (1985)

criterion type	${}_{k}a_{jr}$	condition	relation
usual criterion	0	$f_k(d_j) - f_k(d_r) < 0$	d_j is dominated by d_r
	0	$f_k(d_j) - f_k(d_r) = 0$	indifference
	1	$f_k(d_j) - f_k(d_r) > 0$	d_j dominates d_r
quasi-criterion	0	$f_k(d_j) - f_k(d_r) < -l$	d_j is dominated by d_r
	0	$-l \leq f_k(d_j) - f_k(d_r) \leq l$	indifference
	1	$f_k(d_j) - f_k(d_r) > l$	d_j dominates d_r
linear preference	0	$f_k(d_j) - f_k(d_r) < 0$	d_j is dominated by d_r
	0	$f_k(d_j) - f_k(d_r) = 0$	indifference
	$\frac{f_k(d_j)-f_k(d_r)}{m}$	$0 < f_k(d_j) - f_k(d_r) < m$	d_j dominates d_r
	1	$f_k(d_j) - f_k(d_r) \geq m$	d_j strictly dominates d_r
level criterion	0	$f_k(d_j) - f_k(d_r) < -l_1$	d_j is dominated by d_r
	0	$-l_1 \leq f_k(d_j) - f_k(d_r) \leq l_1$	indifference
	0.5	$l_1 < f_k(d_j) - f_k(d_r) \leq l_2$	d_j dominates d_r
	1	$f_k(d_j) - f_k(d_r) > l_2$	d_j strictly dominates d_r
linear preference	0	$f_k(d_j) - f_k(d_r) < -l_1$	d_j is dominated by d_r
and indifference area	0	$-l_1 \leq f_k(d_j) - f_k(d_r) \leq l_1$	indifference
	$\frac{[f_k(d_j)-f_k(d_r)]-l_1}{l_2-l_1}$	$l_1 < f_k(d_j) - f_k(d_r) < l_2$	d_j dominates d_r
	1	$f_k(d_j) - f_k(d_r) \geq l_2$	d_j strictly dominates d_r
gaussian criterion	0	$f_k(d_j) - f_k(d_r) < 0$	d_j is dominated by d_r
	0	$f_k(d_j) - f_k(d_r) = 0$	indifference
	$1 - e^{\frac{-[f_k(d_j)-f_k(d_r)]^2}{2\sigma^2}}$	$f_k(d_j) - f_k(d_r) > 0$	d_j dominates d_r

Table 2: Users of the public rail service and mean satisfaction for the six considered aspects

Group	Sample size	Users *n*	Users %	Mean satisfaction of users: price	punctuality	times	comfort	cleanli-ness	geograph. distrib.
female-freshmen	24	8	33.3	5.5	4.5	5.1	6.3	5.5	5.6
male-freshmen	25	12	48.0	5.4	5.3	5.8	6.1	4.8	5.9
female-others	85	36	42.4	5.0	4.7	5.0	5.0	4.1	5.2
male-others	70	29	41.4	4.5	4.1	5.1	5.7	4.2	5.0
Total	204	85	41.7						

ACCEPTED MANUSCRIPT

Table 3: Matrix of multivariate pairwise comparison *p*-values and satisfaction scores for the public rail service

	female-freshmen	male-freshmen	female-others	male-others	combined score	normalized score (MRS)
female-freshmen	1	.641	.197	.196	3.699	0.700
male-freshmen	.359	1	.125	.113	5.284	1.000
female-others	.803	.875	1	.373	1.339	0.253
male-others	.804	.887	.627	1	0.805	0.152

Table 4: Users of the urban bus service and mean satisfaction for the six considered aspects

Group	Sample size	Users		Mean satisfaction of users					
		n	%	price	punctuality	times	comfort	cleanliness	geograph. distrib.
female-residents	54	7	13.0	5.4	6.9	5.6	7.0	6.6	6.6
female-offsite	134	11	8.21	5.4	5.5	5.5	6.0	6.0	5.5
female-commuters	109	15	13.8	6.7	7.2	7.2	6.6	6.5	6.7
male-residents	54	3	5.6	5.7	7.3	8.0	8.7	7.0	7.7
male-offsite	121	9	7.4	5.3	6.7	6.2	6.2	6.9	6.2
male-commuters	95	4	4.2	5.3	5.8	6.3	7.8	5.3	5.3
Total	567	49	8.6						

ACCEPTED MANUSCRIPT

Table 5: Matrix of multivariate pairwise comparison *p*-values and satisfaction scores for the urban bus service

	female-resid.	female-offsite	female-commut.	male-resid.	male-offsite	male-commut.	combined score	normalized score (MRS)
female-residents	1	.183	.736	.811	.443	.378	4.001	0.407
female-offsite	.817	1	.970	.975	.791	.574	1.047	0.106
female-commuters	.264	**.030**	1	.617	.115	.142	9.436	0.959
male-residents	.189	**.025**	.383	1	.128	.231	9.836	1.000
male-offsite	.557	.209	.885	.872	1	.342	2.410	0.245
male-commuters	.622	.426	.858	.769	.658	1	2.162	0.220

ACCEPTED MANUSCRIPT