Chapter 8
Choices and conjoint analysis: critical aspects and recent developments

Rossella Berni and Riccardo Rivello

8.1 Introduction

In the literature, a large number of researchers and practitioners are dealing with preference measurements which are considered as one of the most general methods in order to study and improve the consumer's behaviour intended as the consumer's decision about improving his/her utility in changing a service or a product. Nevertheless, a wide range of preference measurements' methods is defined according to the specific aim of the research, or of the application, and the basic theoretical elements involved therein.

In particular, the preference theory must be evaluated according to the nature and definition of preference, namely revealed or stated preferences and, in case of stated preferences, we may distinguish between Contingent Valuation (CV), Conjoint Analysis (CA) and Choice Modelling (CM), Hanley et al. (2001), Netzer et al. (2008). Nevertheless, by considering CA and CM, since the fundamental elements of distinctions are positively overlapped or interchanged, the classification is not so clearly definable; this can be observed when these methods are generally defined as multi-attribute methods.

However, the preference measurements about a product or a service are usually related to a new product/service and the main distinction between CA and CM is the monetary evaluation, namely the Willingness to Pay (WTP), which is the quantitative expression of the respondents about their willingness to accept a change in the product/service concerned or in a single attribute.

Rossella Berni
Department of Statistics "G. Parenti", University of Florence, Viale Morgagni 59, 50134 Florence, Italy,
e-mail: berni@ds.unifi.it

Riccardo Rivello
Department of Statistics "G. Parenti", University of Florence, Viale Morgagni 59, 50134 Florence, Italy,
e-mail: rivello@ds.unifi.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_8,

Furthermore, even though some steps and methods of these two techniques could be viewed as very similar, e.g. the experimental design, the basic elements of the experiments theory are defined and applied in both contexts taking into account, at the same time, that there exists many theoretical differentiations. Thus, the related statistical models were separately developed in the last decades, McFadden (1974), McFadden and Train (2000), Lenk et al. (1996), Greene and Hensher (2003); but, the recent developments in this field were mainly directed at improving common features, such as the heterogeneity of respondents and the complexity of the alternatives (profiles).

In this chapter, we focus on stated preferences (SP) and, namely, on CA and CM, carrying out a brief and critical review in order to clarify the distinctions, as well as to point out the common issues. In addition, we deal with the possibility of reaching the best profile in CA through the theory of statistical methods in the engineering field by considering the current situation and the user's preferences. Our proposal is discussed by showing an empirical example. In this context, we point out the presence in the literature of similar attempts, where the common issue is related to the statistical method applied, the Response Surface Methodology (RSM), Danaher (1997), Jiao et al. (2007), or to the general aim of creating a link between the needs of the manufacturer (design product/service stage) and the consumer/user's preferences, Michalek et al. (2005), Du et al. (2006).

This chapter is organized as follows: a literature review on CA and CM is presented in the second section, by pointing out the methods and recent developments related to CE and CA, respectively; in the third section, our proposal of applying RSM in a CA context is discussed in detail. Section four presents the data and results about the empirical example, while the concluding remarks are outlined in the final section.

8.2 Literature review

Many developments and improvements in consumer/user's preferences by considering the experimental design and the statistical modelling were achieved in the last two decades. Nevertheless, we mainly pay attention to the period 2000–2008, when methods and related applications gave an in-depth consideration to specific issues. Undoubtedly, a further and clear distinction must be made when we refer to preference measurements or, more in general, to the preference theory. Hence, we deal with Stated Preferences (SP), where we define as SP the preference of a respondent related to a hypothetical scenario shown as an alternative in a choice-set (CM) or presented as one of the suggested profiles (CA). However, in the literature, some recent developments are also reported in the Revealed Preference case, which is defined as the preference of the respondent about a real situation, such as in Scarpa et al. (2003).

Another more subtle differentiation is when we refer to CV, CA, CM. Contingent Valuation (CV) is defined as a method in which the respondent is asked to give his/her preference on a product by considering only its total price (mono-attribute method); on the other hand, when we refer to CM and CA, the respondent is asked to express his/her preference or choice about a product or a service by also evaluating the monetary impact of several attributes, and, therefore, the Willingness to Pay (WTP) may be estimated for each single aspect (multi-attribute valuation methods - MAVs).

In our context, we mainly consider the two CA and CM methods, by pointing out the further distinction within CM between Contingent Ranking (CR) and Choice Experiments (CE). The CE situation is related to a set of alternatives, called choice-set, which is selected from an experimental design; the respondent is asked to give his/her preference within each choice-set. The CR situation is applied when the respondent is asked not to give his/her preference, but he/she must rank or order the alternatives of the choice-set (obviously, in this case, each choice-set is comprised of more than two alternatives). The further distinction between the three types of response variables is a straightforward matter. In CA, rating (metric scale) and ranking (ordinal scale) are the preferred response variables, owing to the different framework of profiles. In the CM situation, choice (binary or not) and ranking are surely the conditioning response variables. Our expression "conditioning" means the corresponding statistical models involved in the analysis. Undoubtedly, CE is the preferred method in the literature and, consequently, the related theory has been largely developed in the last few years, by considering the experimental design with its optimality criteria and statistical models and first of all the class of Random Utility Models (RUM) and its variations, see Train (1998), McFadden and Train (2000), Boxall and Adamowicz (2002), Hynes et al. (2008), Wen and Koppelman (2001).

It is not irrelevant to point out that when a methodology is comprised of several theoretical steps, as CM and CA, these elements (mainly experimental designs and statistical models) are closely connected (Yu et al. (2009), Toubia and Hauser (2007)), and the properties of one design affect the corresponding model. When these properties do not exist in the design, this must be taken into account in the model. This is the case of an improvement in the design optimality specifically defined for a Mixed Multinomial Logit (MMNL), Sandor and Wedel (2002); on the other hand, when considering the respondents' heterogeneity, a specific design matrix for each respondent is planned (Sandor and Wedel (2005)), by including the heterogeneity evaluation directly in the design step instead of the model step.

However, as was said hereinabove, a different evolution has characterized the experimental CA and CM designs, even though some features are in common, such as specific methods, algorithms and models, in order to select alternatives, by considering the planning step (De Bruyn et al. (2008) or Toubia et al. (2007)) or the analysis of collected data, such as in Netzer and Srinivasan (2007), where a dynamic evaluation of the questionnaire through an Adaptive Self Explication method is performed in a Multi-Attribute context.

In Table 8.1, differentiations are summarized between CA and CM, namely CE and CR, by considering these first issues and the time developments.

As is shown in the summary (Table 8.1), where some specific features such as status-quo are not yet included, the preference theory is more articulated when

Table 8.1 MAV methods- A summary of recent developments

Steps	Conjoint Analysis-CA	Choice Experiments-CE	Contingent Ranking-CR
Preference	profile	alternative-choice	alternative-order
Dep.var.	rate; rank	choice	rank
Exp. design	factorial; frac. factorial	D-optimal; Local Bay. optimal; optimality ad-hoc	D-optimal
Stat. models	linear model; Hyerarchical Bayesian	Random Utility Model (RUM): Nested Logit (NL), Generalized NL	RUM: Rank Ordered Logit-Asc, Kernel logit
	finite-mixture model	RUM: Mixed-MNL; Latent Class Model (LCM)	RUM: Rank Ordered Logit-LC

considering all the steps within the three methods. Furthermore, having previously outlined the differentiations related to the type of preference and to the dependent variable, we may now observe that the experimental design step could be varied within these methods. Undoubtedly, CA is an easier task at this point: the theory and applications in the literature present above all developments and studies about the complexity and selection of profiles in the model step, (De Bruyn et al. (2008), Netzer and Srinivasan (2007)), i.e. some problems of complexity, such as preference uncertainty and conflicts solved through the evaluation of judgement time and response error in a rating task (Fischer et al. (2000)). The design of experiments is involved in order to create an orthogonal design (sometimes optimal) where all the created profiles, according to the set of attributes considered, are eventually reduced by applying a fractional factorial design of high Resolution.

The complexity of statistical models developed in the recent years, like finite-mixture models and hierarchical Bayes models, such as in Gilbride and Allenby (2004) and Lenk et al. (1996), in order to take account of the respondents' heterogeneity or the complexity of alternatives, or in Bradlow et al. (2004) for imputing missing levels of profiles, has not yet received in the literature an adequate response when considering the properties of the experimental design. Instead, a different situation is presented in the CM sector, namely in the CE method, where optimality criteria, above all D-optimality, ad-hoc algorithms and specified information matrices for the experimental design involved were entirely defined in 1990s (Zwerina et al. (1996)). Recent developments are related to the construction of optimal or near optimal designs with two-level attributes for binary choices in the presence of the first order interactions, Street and Burgess (2004), or when optimal designs are defined with mixed-level attributes, Burgess and Street (2005). Furthermore, a new optimum criterium is suggested, the M-optimality (Toubia and Hauser (2007)), where attempts in order to focus the planning by considering the manager's need were introduced; in fact, M-optimality means optimality manager. Note that a common feature of recent years is to create a link among designs and models together with the need of a guiding thread between manufacturers and consumers. In addition, it is not so irrelevant to quote the paper of Sandor and Wedel (2002) which reflects the strict connection between experimental designs and statistical models, because they suggest an experimental design with ad-hoc properties for a Mixed Multinomial Logit. This model, belonging to the class of Random Utility models, is certainly the most widely applied and developed model in recent years for the CE situation. Its success is easily explained when considering the theoretical results of McFadden and Train (2000), Train (1998) and the possibility, by adding additional random parameters, to study respondents' heterogeneity and the correlation structures due to repeated choices. The last developments of this model include its relationship with the latent class model, in order to create a finite number of respondent groups (Greene and Hensher (2003), Hynes et al. (2008), Boxall and Adamowicz (2002), Scarpa and Thiene (2005)). Furthermore, a distinct class (anyhow, close enough) is that of Generalized Nested Logit (GNL) models, Wen and Koppelman (2001). This class of models, which generalizes the Nested Logit (NL) model, impose an a-priori tree structure with nests and nodes. The relationship between the NL and the Multinomial Logit model (MNL) is very strong because an NL model can be viewed as the product of a series of MNL models, each MNL for each node. The main issue of the GNL model could be its flexibility due to the nesting structure; undoubtedly, this can also be viewed as a limit because an a-priori tree-structure must be imposed.

Finally, before entering into details related to Conjoint Analysis and Choice Experiments (Sects. 8.2.1 and 8.2.2), we briefly outline some features about Contingent Ranking. In this situation, the ordinal response variable conditions the respondent's interview (repeated and ordered choices) and, therefore, the statistical models to be apply. The repeated and ordered choices, called also panel, create a correlation between choices which can not be adequately treated through the Rank Ordered Logit (ROL) also when including the Alternative Specific Constant (ROL-ASC). An improvement may be obtained through the Kernel Logit (KL) model, which allows to take care of heteroschedasticity and correlations; in general, in this case, an Alternative Specific Constant (ASC) is introduced in order to discriminate, during the model estimation, for the status-quo (Herriges and Phaneuf (2002)). A recent study (Van Dijk et al. (2007)) introduces the concept of latent segments (Latent Class) jointly with a Rank Ordered Logit model (ROL-LC), in order to treat the heterogeneity of respondents due to their difficulty at ranking.

8.2.1 Choice Experiment: theory and advances

As shown hereinabove (Table 8.1), the CE theory considers the experiments and statistical models as main theoretical elements; nevertheless, further issues should be evaluated in order to completely discuss this methodology, such as the estimation methods and simulation algorithms to solve the model's expression, Bhat (2001). In this brief section, we mainly focus on the model step, by evaluating the solutions suggested in the literature in order to solve the effective problems when this method is applied.

The role of the experimental design is not irrelevant when we consider its properties; broadly speaking, the search of a D-optimal design implies the maximization of the determinant of information matrix and, therefore, this directly influences the variances of parameter estimates and, obviously, the volume of the ellipsoid, confidence region for the parameters, which is strictly connected to the precision of the design. This implies a larger efficiency in the estimates. In the specific literature about CE, the consideration of a D-optimal design, from the fractional factorials to more complex designs (Zwerina et al. (1996), Yu et al. (2009)) built through specific algorithms of trial-point selections, has been replaced in recent years by using Bayesian optimal designs (Kessels et al. (2004)). However, the experimental planning through D-optimal designs can not be considered as a limited tool because it guarantees optimal properties jointly with a notable manageability in comparison with the implementation of Bayesian designs.

By considering the experimental planning for a choice or conjoint experiment, some features are general common rules for a valid experimental planning, independently of the application field. Therefore, the attributes must be accurately defined in their number and in the number of levels. Surely, an experimental design formed by attributes with the same number of levels is more easy to treat; at the same time, a great attention must be paid to the distance among levels. Undoubtedly, the inclusion of a large number of attributes with distant levels increases the complexity of the design and the decision of the user/consumer becomes more difficult and implies a response error; Swait and Adamowicz (2001) face this problem from the point of view of the choice capability and its difficulty through a heteroschedastic Multinomial Logit Model.

In Scott (2002) a problem of dominant preferences is focused in the health care system, by considering the consumer's decision task and its complexity when evaluating the defined levels and the presence of a lexicographic preference- i.e. when the consumer always prefers the same alternative, independently of the other alternative settings. In addition, a relationship between a general alternative and the status-quo or current situation, is created according to these general criteria. If alternatives are very distant, a problem of a dominant alternative could be found; on the contrary, when alternatives have close level values, the presence of the status-quo alternative could be much more appealing and the respondent tends to prefer the current situation without changing. Nevertheless, the inclusion of status-quo alternative cannot be disregarded in Choice Modelling (CM) for the interpretation and estimation of economical concepts, first of all the Willingness To Pay (WTP) for a relative change in each single attribute. Furthermore, the complexity of choice and the planning issues outlined previously must be evaluated together with the number of choice-sets given to a single respondent and with the kind of response variable adopted, ranking or binary-choice variable. Undoubtedly, in the CR field, the complexity is increased by the ranking task; on the other hand, for example, in CE environmental situation, a choice-set is usually comprised of two alternatives and the status-quo alternative.
In the literature, several studies attempted to improve these issues by starting from the planning phase or by considering improvements in the model step; in DeShazo and Fermo (2002) the sources of variability are studied in order to identify the impact of complexity on the consistency of choice, by introducing measures of complexity and studying the effect of complexity as in the variance-components field.

In fact, the authors analyze the problem by defining an heteroschedastic logit model according to the five complexity measures defined; thus, the dependent variable is the variability due to the characteristics of the choice-set. In a previous study, on the same subject, (Dellaert et al., 1999), the consistency is evaluated by considering the specific attribute of cost. Recent developments about the complexity of choice and related problems, such as discontinuity, where discontinuity could be defined as a "break point" in the likelihood function due to extreme situations of the consumer/user's behaviour, are studied in Gilbride and Allenby (2004), where discontinuity points are evaluated in the estimation step by introducing the concept of consideration-set and screening-rules for consumers. Thus, threshold values are defined by discriminating according to specific rules of consumer's utility. In Campbell et al. (2008) the impact of discontinuous preferences, from the point of view of respondents, is evaluated on the WTP estimates; the respondents with discontinuous preferences are identified during the decision process through a multinomial error component logit model which includes the constant term, namely the ASC, in order to consider the status-quo situation. These authors deal with the correlation between the utility of changing alternatives and the status-quo aspect together with the heterogeneity due to the different type of the respondent's preference. A very interesting remark is the consideration of different scale parameters according to the number of respondents' discontinuities; this allows to treat differently the sets of respondents owing to their preferences.

Furthermore, recent developments about the WTP estimates are in Garrod et al. (2002), Strazzera et al. (2003), Scarpa et al. (2007), Sonnier et al. (2007); in Scarpa et al. (2007) this theoretical problem is faced by defining a parallel Willingness To Pay (WTP) space where parameter estimates are evaluated by considering a more specific economic definition of the WTP, in order to improve its interpretation; in Sonnier et al. (2007) a Bayesian approach for the WTP estimates is introduced. Willingness To Pay estimates and zero values, according to the typology of response motivations, are studied in Strazzera et al. (2003).

In order to deal with the above features, the general class of Random Utility Models (RUM) is defined. In general, every alternative is indicated by j $(j = 1,...,J)$, while i denotes the consumer/user $(i = 1,...,I)$; thus, the following expression is characterized by a stochastic utility index U_{ij}, which may be expressed, for each unit i, as:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{8.1}$$

where V_{ij} is the deterministic part of utility, while ε_{ij} is the random component, independent and Gumbel distributed. The class of RUM, which aims to achieve the utility maximization, enlarges the characteristics of Logit and Nested Logit (NL) models where the Independence of Irrelevant Alternatives (IIA) is hypothesized. The relaxation of this assumption is undoubtedly a very substantial improvement because the IIA means that the choosing probability in one choice-set is independent of the presence of other attribute values or any other alternative; on the other hand, we may say that IIA derives from the hypothesis of independence and homoschedasticity of the error terms. In addition, this can also be interpreted by considering the cross-elasticity term. In fact, IIA implies an equal proportional substitution between alternatives.

Furthermore, these models cannot take account of a different behaviour of the consumer; i.e. each respondent, with different baseline characteristics, is treated in a similar way (the same estimate values of attributes) according only to their judgement, exclusively.

In the literature, a first contribution to improving these issues is in Train (1998), where a Random Parameter Logit (RPL) model is introduced. At present, this model is more precisely called Mixed Multinomial Logit (MMNL). In fact, this RUM model allows to evaluate the respondents' heterogeneity or, better, the consumer/user's variability is estimated by considering the attributes as random variables and not fixed variables, i.e. as random variables across respondents; in addition, just because more choice-sets are supplied to the respondents, the repeated choices (during time) imply a correlation which is confounded with the consumer/user's variability (unobserved utility).

In this case, an appropriate example is in Train (1998) where, in a fishing case, the unobserved utility of the consumer is identified in the difference for each fisherman, when he must choose the fishing site. Further, according to repeated choices, this unobserved utility is confounded with the correlation due to several sites and trips; so, a correlation over trips and over sites for each fisherman's decision must be taken into account.

A general formulation for a single decision, according to McFadden and Train (2000), for a MMNL is:

$$
\begin{aligned}
Pr_C(i \mid \mathbf{x}; \boldsymbol{\lambda}) &= \int_{\Re^I} L_C(i; \mathbf{x}, \alpha) G(d\alpha; \boldsymbol{\lambda}) \qquad (8.2)\\
L_C(i; \mathbf{x}, \alpha) &= \frac{\exp(x_i \alpha)}{\sum_{j \in C} \exp(x_j \alpha)}
\end{aligned}
$$

where $C = (1, ..., j, ..J)$ is the general choice-set; $\mathbf{x}$ is the vector of attributes $(\mathbf{x} = x_1, ..., x_j, ...x_J)$, and x_i is the observed value of the decision i; α is the vector $(Ix1)$ of random parameters which expresses the respondents' heterogeneity. The term $L_C(\cdot)$ is the general expression for a Multinomial Logit (MNL), where the $G(\cdot)$ is the mixing component. It is very important to note that the random parameter α varies in the mixing term, where the differential over the integration is performed over α, because is $G(d\alpha; \boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is the vector of parameters related to the mixing distribution.

By considering expression (8.2), we may further assume that an individual i belongs to the s group or segment, $(s = 1, ..., S)$, i.e. we assume a finite number of groups identified through the consumer baseline characteristics; from a theoretical point of view it is like assuming that the mixing term $G(\cdot)$ is defined on a finite support. Therefore, the probability for the unit i of belonging to the set s is included in $[0, 1]$ and $\sum_s Pr_{is} = 1$.

The deterministic term of the utility function may be expressed through a function of attributes and the characteristics of the s group; thus, the utility function defined by (8.1) may be now formulated as:

$$U_{ij|s} = V_{ij|s} + \varepsilon_{ij|s} \tag{8.3}$$
$$V_{ij|s} = \alpha_s x_{ij} \tag{8.4}$$

Note that formula (8.3) expresses the deterministic term conditional to the belonging to the group s and the specific choice is weighted through the utility characteristics of the set s.

Therefore, for each segment s, the probability to choose the alternative $j^\star$ for the unit i belonging to s is:

$$Pr_{i|s}(j^\star) = \frac{\exp(\mu_s \alpha_s \mathbf{x}_{j^\star})}{\sum_{j\in C} \exp(\mu_s \alpha_s \mathbf{x}_j)} \tag{8.5}$$

where α_s is the specific utility parameter for the segment s and μ_s is the specific scale factor, usually re-scaled to one, and here generically assumed. Note that if $\mathbf{x}_{j^\star}$ is an alternative-specific value, as defined in Sect. 8.2.1, then α includes the alternative specific variable, in this case evaluated as a random effect.

Furthermore, we have the following relation:

$$Pr_i(j^\star) = \sum_s Pr_{is} Pr_{i|s}(j^\star) \tag{8.6}$$

Formula (8.6) expresses the global likelihood for a generic individual i who prefers $j^\star$ as the sum of products of two terms: the probability of the unit i of belonging to the group s is multiplied by the probability of the unit i belonging to s to choose the alternative $j^\star$.

Formula (8.6) may be explicitly written as:

$$Pr_i(j^\star) = \sum_s \frac{\exp(\beta \gamma_s \mathbf{z}_i)}{\sum_s \exp(\beta \gamma_s \mathbf{z}_i)} \frac{\exp(\boldsymbol{\mu}_s \alpha_s \mathbf{x}_{j^\star})}{\sum_{j\in C} \exp(\boldsymbol{\mu}_s \alpha_s \mathbf{x}_j)} \tag{8.7}$$

where $\mathbf{z}_i$ is the vector of baseline individual characteristics, γ_s is the vector of parameters for the group s, while β is the scale factor, usually re-scaled to one as in (8.2) when considering L_C.

The last formula (8.7) could be interpreted as the model expression for a Latent Class Model (LCM) and it is also called as finite-mixture model (Boxall and Adamowicz (2002)), in comparison with the MMNL, formula (8.2), where the mixing term is assumed as distributed according to a continuous distribution (normal or log-normal, for example); by referring to formulas (8.2) and (8.7), in this case, the probability to choose the $j^\star$ alternative of the choice-set C is multiplied by the probability to choose given that i belongs to the group s.

It is not so irrelevant to remark that the scale factors $\boldsymbol{\mu}_s$ should be posed equal to one in order to avoid imposing parameter values. This formula (8.7) expresses a flexible range of situations: if $S = 1$ there are not differences between baseline characteristics of the consumers; otherwise, if $S = I$ a group is defined for each unit, assuming the extreme situation of a total differentiation between individuals.

8.2.2 Conjoint Analysis: theory and advances

Conjoint Analysis (CA), (Netzer et al., 2008), can be defined, in our opinion, as the historical MAV method, where the term conjoint means the measurement of relative attribute values jointly. The first studies were made in 1970s where the basic fundamental theory of CA was posed (Johnson, 1974; Green and Rao, 1971; Green and Srinivasan, 1978). In Johnson (1974) trade-offs among alternatives were evaluated by considering a pair of attributes at-a-time and the respondent (consumer) must rank his/her preference as to these two attributes. The empirical example reported, (Johnson, 1974), is about the car-market, and the author assumes the independence of attributes and his analysis does not include the interaction first order terms. A not irrelevant point is a first introduction of the individual's characteristics, through a suggested weighting.

In Green and Srinivasan (1978), a further improvement was introduced, by considering a full profile (and not paired) evaluation where the global consumer/user's utility is then decomposed in order to estimate each single attribute's utility. Rating and ranking are the response variables preferred and the suggested statistical analysis usually applies the linear regression model. In the paired comparisons case, logit and probit are applied. Then, further studies have developed these issues, by pointing out model definition and estimation, (Green, et al., 1981; Green, 1984), where a hybrid utility estimation model for CA is suggested. Here, a self-explicated model, based on a procedure of measuring preference functions, is used with a conjoint model, with the inclusion of interaction (I order) terms. This model, which also takes account of differences (through clusters) of respondents, by evaluating their similarities in the self-explicated model, may be considered a basic model of CA. An enlargement of this model has different parameters for each attribute within each cluster. However, the correlation due to the evaluation of the same attributes in the two model steps is not completely assessed. A further note relates to the burden of respondent in this context; the respondent, in CA and more precisely for participating to a hybrid CA model, is asked to perform a heavy task, because he/she must participate two times to a judgement procedure.

Undoubtedly, the first attempts of respondents' segmentation are found in CA method. Currim (1981) and Moore (1980) applied strategies in order to satisfy the need of an intermediate level of consumer's aggregation through cluster analysis, individual a-priori information and preferences.

In fact, the multivariate statistical analysis played a relevant role in CA during 1980s, and the first half of 1990s (Punj and Stewart, 1983; Hagerty, 1985), above all the cluster analysis.

After dealing with this historical picture of fundamental CA elements, and, in general, with the evolution of MAV methods, we point out the different configurations of CA and the following developments in order to gain the respondent's coherency and reliability.

Starting from Green and Srinivasan (1990), CA had many differentiations according to a decompositional or compositional or mixed approach. All of these methods are related to an easy-to-treat multi-attribute situation; the self-explicated technique, just cited, requires the respondent to have a two-step evaluation; the Adaptive Conjoint Analysis (ACA) also applies the mixed approach and a paired comparison is performed through a computer-assisted interview. It is important to note that the attempts are directed towards a mitigation of the respondent's task, especially when the number of attributes and/or levels is high (Netzer and Srinivasan, 2007). In this respect, some studies are in common with Choice Modelling (CM) methods, such as De Bruyn et al. (2008).

A dynamic evaluation of CA was implemented in Bradlow et al. (2004); a consumer's learning phase is suggested through partial conjoint profiles in order to avoid the missing levels problem, which may exist when the experimental planning is conducted by a fractional factorial design, which is a reduced design of the corresponding full factorial design. Surely, optimal designs and specific algorithms are further solutions in order to overcome this problem.

In Bradlow et al. (2004), as in Lenk et al. (1996), the respondent's heterogeneity is taken into account through the application of a hierarchical Bayes model. Furthermore, in Lenk et al. (1996), the reduction of the number of profiles supplied to each respondent is studied in order to improve the estimation accuracy.

Therefore, three issues are variously combined in order to solve the CA problems: the reduction of the consumer's task; the complexity of data collection (the experimental planning step); the respondents' heterogeneity. Even though some features are strictly connected with CM methods, as was said previously, the respondent's stimulus, also introduced in Green and Srinivasan (1990), was also studied in recent years. Conjoint Analysis often appears in mixed techniques, such as in Barone et al. (2007) and in Schütte and Eklund (2005), where a Kansei Engineering (KE) is applied; KE is a multidisciplinary approach where the consumer is stimulated through real perceptions of existing products in order to give a weight to the technical and performance characteristics.

The link between consumer's preferences and the engineering field is largely used in recent years. The Response Surface Methodology (RSM) is also applied in the Customer Satisfaction (CS) field, see, for example Danaher (1997), but the strict relation between CA and the quality measures is in Kazemzadeh et al. (2008), Du et al. (2006) and Jiao and Tseng (2004). In Jiao and Tseng (2004), an Adaptive CA is applied jointly with the RSM. Two issues must be outlined: the concept of mass customization, i.e. the product design performed to attract the consumer's attention (through cost and customization value), and the consideration of a cost variable. A remark is the evaluation of RSM in a discrete context. An approach similar to Jiao and Tseng (2004) is in Du et al. (2006), where a more in-depth analysis is performed by considering the unit costs. Indexes about quality performance, costs and satisfaction are defined and applied in Kazemzadeh et al. (2008).

8.3 Our proposal: conjoint analysis and response surface methodology

The aim of the present study is the proposal of a modified Conjoint Analysis (CA) in order to establish an optimal solution for the product/service from the point of view of the user/consumer. The subsequent procedure is performed through the Response Surface Methodology (RSM) theory, by considering the quantitative judgement of each respondent for each profile with respect to the assessed score about the status-quo, and taking into account the individual information. The final result is achieved by carrying out an optimization procedure on the estimated models, and defining an objective function in order to reach the optimal solution for the revised (or new) service/product. Furthermore, it is relevant to point out the modifying 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. Therefore, two remarks must be made: the former is the consideration of the status-quo as the current situation for a revised product, otherwise the status-quo may be interpreted as the center of design or, alternatively, the full profile which identifies the medium situation; the latter is that the search of the best profile for the respondent is performed on a surface delimited by the range of attributes and centered on the status-quo.

8.3.1 The outlined theory

In this and in the following sections we briefly explain the RSM theory and the general optimization measures applied, according to a robust design approach (for details see Khuri and Cornell, 1987). Note that we focus our attention on the statistical models and optimization in the RSM; the fundamental elements of the experimental design (Box et al., 1978) are, however, indirectly introduced through the experimental planning related to CA.

In this case, the concept of a robust design approach is used for optimizing the service/product as more insensitive as possible with respect to the respondents heterogeneity or in order to adjust the service/product by considering those respondents characteristics which are relevant for the product/service studied. 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: $\mathbf{z} = [\mathbf{z}_1,..,\mathbf{z}_s,..,\mathbf{z}_S]$. 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. Note that, in general, if J are the profiles and I the respondents, the observations are IxJ. The general RSM model can be written as:

$$\mathbf{Y}_{ij}(\mathbf{x},\mathbf{z}) = \beta_0 + \mathbf{x}'\beta + \mathbf{x}'\mathbf{B}\mathbf{x} + \mathbf{z}'\delta + \mathbf{z}'\boldsymbol{\Delta}\mathbf{z} + \mathbf{x}'\boldsymbol{\Lambda}\mathbf{z} + \mathbf{e}_{ij} \qquad i = 1,..,I; j = 1,..,J \quad (8.8)$$

where $\mathbf{x}$ and $\mathbf{z}$ are the vectors of judgements attributes 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.

In general, a noise variable may be defined as a categorical or quantitative variable which is also controllable and measurable. In the technological context, a noise effect which has these characteristics is introduced in the experimental design to reduce the pure experimental error and to set the variables controlling the process variability in order to find the experimental run which is the most insensitive to noise, through the first order interaction effect. In this situation, the set $\mathbf{z}$ is comprised of measurable categorical or quantitative variables which measure the baseline respondents characteristics. Therefore, the best profile is reached through the estimation of (8.8) conditional to the heterogeneity of respondents, taking into account judgements and individual data through the interaction terms. Furthermore, the response variable is comprised of the individual scores for each hypothetical profile and this information is used to gain an optimal solution on the surface around the status-quo (the attribute judgements $\mathbf{x}$) and conditionally to $\mathbf{z}$. In addition, it is not irrelevant to observe that the individual characteristics are an external source of variability with respect to an ideal design of service or product. In order to perform this procedure, it is necessary to effect a combined interview, with three steps: (1) gathering information about baseline variables; (2) quantitative judgements about each attribute in the status-quo when the service/product is revised, or, when the service/product is new, the judgements about each attribute in the medium profile: $\mathbf{x}_0 = (0,..,0)$; (3) the quantitative judgement on each full profile for each respondent. Note that the set $\mathbf{x}$ is the same by considering either the attributes involved in the experimental design (profiles) and the attributes in the status-quo.

Therefore, the prospective evaluation of the new or revised product/service is obtained by computing the optimal hypothetical solution through the status-quo. Note that, as explained hereinafter (Sect. 8.3.2), the estimated surface is subsequently optimized in order to gain the best preference on the basis of the attributes and judgments involved. Nevertheless, if the service/product studied is new, the status-quo is the centered scenario (always hypothetical); if the service/product studied is under revision, then the status-quo represents the real and current scenario in comparison with the other hypothetical scenarios supplied in the CA step, as usually happens in a Choice Experiments context (Sect. 8.2.1).

A further issue about the baseline variables must be outlined. In general, there are some aspects we wish to examine and which may have an influence on the expressed judgement of the respondent. We refer to those aspects related to social and demographic characteristics such as gender, age, educational level, income, job status. As described in Sect. 8.2, the heterogeneity of respondents plays a central role in MAV methods and this is confirmed by recent developments in the literature. In fact, there are sensible reasons to believe that such features affect the final results. In this proposal, we suggest and compare two analyses which differently include the individual information. In the first analysis, baseline variables are included in the model (8.8) as explained before; if a baseline variable is categorical, as the gender, this must be considered both in the estimation of the model and in the optimization step, carrying out an optimal surface for each level of the categorical variable.

This proposal is compared with the consideration of building a-priori strata according to the baseline level variables. In this case, the response surface model (8.8) does not include the set of variables **z** which are used to build the strata. The comparison is not trivial, just because in the first case we may estimate the interaction effects which may add useful information to obtain the full optimal solution; in the second case, where the problem of a categorical baseline variable is not relevant, the stratification allows to carry out the optimization process within every a-priori stratum.

8.3.2 The searching of the best profile through optimization

As was said in the previous section, our aim is the optimization of the the model (8.8) according to the status-quo situation. The expressed rate for each conjoint profile is considered as the response or dependent variable (formulated on a continuous scale); for example, a vote expressed according to the metric scale [0,100] may lead to a valid evaluation of the response as a continuous variable. Therefore, in general, the optimal target score may be defined as the maximum value of the metric scale; in the above example, this is equal to 100.

Two optimization measures are defined for the optimization process, with only one dependent variable; both measures allows to consider the optimization within a specific delimited surface defined by the range of attribute scores. The first measure is formulated by considering the quadratic deviation of the estimated surface model $\hat{Y}$ from the maximum score τ. Therefore, the formula to be minimized is the following:

$$F_1 = (\hat{Y}(\mathbf{x},\mathbf{z}) - \tau)^2 \tag{8.9}$$

The second optimization measure is defined by considering the approaching of $E(\hat{Y})$ to the maximum score; thus, we carry out the minimization of the model variance of $\hat{Y}$ jointly with the approaching to the ideal maximum score.

The formula is:

$$minF_2 = V(\hat{Y})(\mathbf{x},\mathbf{z}) = E(\hat{Y} - E(\hat{Y}))^2 \tag{8.10}$$

according to the following decomposition of the Mean Square Error (MSE):

$$MSE = E(\hat{Y} - E(\hat{Y}))^2 + (\tau - E(\hat{Y}))^2 \quad (8.11)$$

$$B(\tau) = (\tau - E(\hat{Y}))^2 \quad (8.12)$$

where (8.12) explains the adjustment of the expected score value to the target score. A further issue is about the computation of $E(\hat{Y})$ which is calculated by considering the expected value of the estimated surface model.

The optimization procedure, carried out through the Statistical Analysis System (SAS) and the procedure NLP, is preferably computed using non-coded data, just because we are not in a technological context (for further details, see Berni and Gonnelli, 2006). Note that, as regards optimization, the final result expresses the optimal score for each attribute involved in the model according to the respondents' preferences. Furthermore, the final optimal score for an attribute may be explained as the importance/utility of that variable in order to reach the best profile when considering the judgement of the respondent about the current situation (status-quo). A further consideration may concern the inclusion of a categorical baseline information in the optimization process by including the proportion of units belonging to each level of the baseline variable, or belonging to level combinations for several baseline variables (strata), as in Robinson et al. (2006) where this case is studied in a technological field; nevertheless, the optimization must be always performed taking account of different surfaces according to these different strata.

8.4 Case study

The main aim is the evaluation of an interdisciplinary degree course of the University of Florence. As regards the data collection, a "questionnaire" is planned and submitted to a sample of students of the II-nd and III-rd year. The questionnaire is articulated on three parts according to the three different sets of information: (i) baseline variables; (ii) judgements about status-quo; (iii) the specific planned experimental design for the basic CA.

Every judgement is expressed on the metric scale $[0,100]$. The first set of variables is related to the social and demographical data for each student: gender, age, exam average, enrolment status, job status. In the second part, the current situation is analyzed according to the specific five attributes: contents of the basic subjects (cb); practice/laboratory (pl); intermediate exam (ie); exam modalities (me); professional subjects for the future job (prof), see Table 8.2.

The third part contains the conjoint study planning through a fractional factorial design 2_V^{5-1}. Note that the profiles are 16 and the students are 46; therefore, the total number of observations is 736. Furthermore, in the following application, we consider as noise categorical variable the job status of the student (job), identified also by case (i) working, and case (ii) non-working.

Table 8.2 Attributes and levels

Attributes	-1	1
cb	basic subjects with lower theoretical deepening	basic subjects with higher theoretical deepening
pl	practice and laboratory as compulsory part of typical courses	practice and laboratory only as two distinct courses
ie	one intermediate exam	no intermediate exam
me	oral test with practice	written and oral test
prof	a general degree course in order to continue studies	a more specifical degree course, in order to seek a job

8.4.1 Optimization results

The general response surface model (8.8) is applied by considering judgements of the full profiles and judgements on the attributes in the current situation. Parameter estimates with standard error and p-values are displayed in Table 8.3. Note that all the variables are significant, except "prof", which has a non significant p-value. However, this main effect must be inserted given that it is relevant when considering the interaction effects of "prof" with the other variables, and, above all, with the "ie" variable. In addition, a highly significant p-value results for the interaction effect of "prof" with the noise variable "job". The same observation can be made considering "ie" and "cb". The optimization procedure is performed applying the two measures (8.9) and (8.10) defined above. The optimization results are described also by considering diagnostic results such as: the objective function value (of), the infinity norm of the gradient ($\|x\|_\infty$), the determinant of the Hessian matrix ($|\mathbf{H}|$). We have also checked the max-step, i.e. a specified limit for the step length of the line search algorithm, during the first r iterations. Two surfaces are optimized, according to the two levels of the job variable: working and non-working. The results are shown in Tables 8.4 and 8.5, related to the results about the measures (8.9) and (8.10), respectively. We must point out that, in this case, even though convergency is always reached and diagnostic results are quite satisfactory, the starting diagnostic results are not perfect. The reason of this problem may be leaded to the kind of data, so different with respect to technological data, where the experimental trials are usually conducted with high accuracy.

In this respect, we must remark that a non controllable variability due to the respondent is implicitly inserted in our data. In fact, in this context, the optimization measure (8.10) is more precise with respect to measure (8.9) just because the computation of $E(\hat{Y})$ takes care of non orthogonal data and of moments values. This is also confirmed when selecting the best fitted models; in this context, by including or not a model term may be very relevant for the following optimization procedure. By considering the optimization measure (8.10), the best solution considers "cb" and "ie" (Table 8.5) as relevant attributes for the non-working students. The scores are very high for case (ii): 84.98 and 99.90, respectively. The attribute "ie" is included in the final solution also for the working students. The scores for the opti-

Table 8.3 Model estimates; job status as noise variable

Parameter	Estimate	Stand. Error	t-value	p-value
Intercept	−281.20	47.863	−5.87	0.0001
cb	2.33	0.751	3.10	0.0020
pl	−1.54	0.705	−2.18	0.0293
me	3.05	0.843	3.62	0.0003
ie	4.04	0.667	6.05	0.0001
prof	0.36	1.069	0.34	0.7375
job	−44.53	24.862	−1.79	0.0737
cb*pl	−0.01	0.004	−2.23	0.0262
cb*ie	0.02	0.005	4.43	0.0001
cb*me	−0.05	0.011	−4.17	0.0001
cb*prof	0.01	0.008	1.30	0.1958
cb*job	−2.69	0.359	−7.49	0.0001
pl*ie	−0.03	0.004	−5.76	0.0001
pl*me	0.067	0.012	5.41	0.0001
pl*prof	−0.01	0.006	−2.99	0.0029
me^2	0.02	0.007	3.76	0.0002
me*ie	−0.08	0.011	−7.02	0.0001
me*prof	−0.02	0.011	−1.91	0.0565
me*job	0.24	0.154	1.54	0.1232
ie^2	0.01	0.004	1.31	0.1909
ie*prof	0.02	0.005	3.76	0.0002
ie*job	−1.04	0.218	−4.79	0.0001
prof*job	3.52	0.430	8.18	0.0001

Table 8.4 Optimization through measure (8.9). Case (i) working; case (ii) non-working

Results	measure (8.9); case (i)	measure (8.9); case (ii)
Best score:cb	cb = 26.76	cb = 7.01
Best score:pl	pl = 0.00	pl = 2.00
Best score:me	me = 0.00	me = 4.00
Best score:ie	ie = 34.50	ie = 46.09
Best score:prof	prof = 0.00	prof = 59.47
o.f.	5.0e-27	2.0e-28
$\|\|x\|\|_\infty$	8.6e-13	1.3e-13
$\|H\|$	$<$ 10e-8	$<$ 10e-8

Table 8.5 Optimization through measure (8.10). Case (i) working; case (ii) non-working

Results	measure(8.10); case (i)	measure (8.10); case (ii)
Best score:cb	cb = 42.04	cb = 84.98
Best score:pl	pl = 0.00	pl = 0.00
Best score:me	me = 0.00	me = 0.06
Best score:ie	ie = 56.85	ie = 99.90
Best score:prof	prof = 0.00	prof = 22.10
o.f.	3.2e-27	3.2e-25
$\|\|x\|\|_\infty$	7.5e-13	1.0e-11
$\|H\|$	$<$10e-8	$<$ 10e-8

mization measure (8.9) show very low values for all variables involved, except "ie" and "prof" in case (ii), Table 8.4. This may be viewed as a higher interest of non-working students versus professional learning. As regards case (i), (Tables 8.4 and 8.5), "cb" and "ie" are the only relevant attributes; however, in table 8.4, scores are low for both variables, 26.76 and 34.50 respectively; while, by considering measure (8.10), "ie" and "cb" achieve higher scores (42.04 and 56.85, respectively). These solutions allow us to hypothesize a larger consideration of the professional elements by the non-working student in comparison with the one who works. The optimal solution obtained through measure (8.10) highlights the importance of "cb", "ie", "prof", by confirming the results obtained applying the (8.9) and the previous considerations about the relevance of computing $E(\hat{Y})$.

Furthermore, we compare these results with those obtained by using the baseline variables, in particular the job status of the student, for setting a-priori strata.

Two response surface models are estimated within each level of the job variable (estimates are not shown); "prof", "cb" and "pl" are significant attributes for the working students. The estimated surface model related to the non-working students allows us to confirm a large interest towards "cb" and "prof". Furthermore, "prof" is a common relevant attribute within each stratum; "pl" is relevant when considering the working students, while "cb" is more relevant for the students without a job, which express a great interest towards the basic courses in conjunction with more professional tools.

By considering the optimization results, (Tables 8.6 and 8.7), the diagnostic measures are always good, even though the results obtained through measure (8.10) have a high objective function value; however, the values of $|\mathbf{H}|$ are very good. Optimization measures (8.9) and (8.10) highlight "pl" and "prof" as the attributes with the highest scores for the working students. Within non-working students, "prof" and "cb" result as relevant attributes, confirming the propensity of the non-working student towards studying.

Table 8.6 Optimization through measure (8.9); a-priori strata; case (i) working; case (ii) non-working

Results	measure (8.9); case (i)	measure (8.9); case (ii)
Best score:cb	cb = 0.13	cb = 65.77
Best score:pl	pl = 73.57	pl = 0.00
Best score:me	me = 0.00	me = 0.00
Best score:ie	ie = 0.00	ie = 0.00
Best score:prof	prof = 84.21	prof = 50.99
o.f.	8.1e-28	1.5e4
$\|\|x\|\|_\infty$	1.3e-13	2.1e-14
$\|H\|$	<10e-8	8.4e-1

Table 8.7 Optimization through measure (8.10); a-priori strata; case (i) working; case (ii) non-working

Results	measure (8.10); case (i)	measure (8.10); case (ii)
Best score:cb	cb = 0.00	cb = 65.77
Best score:pl	pl = 56.49	pl = 0.00
Best score:me	me = 0.00	me = 0.00
Best score:ie	ie = 0.00	ie = 0.00
Best score:prof	prof = 100.00	prof = 51.00
o.f.	2.7e4	3.4e2
$\|\|x\|\|_\infty$	2.1e-9	3.6e-15
$\|H\|$	$<$10e-8	1.9e-2

8.5 Concluding remarks

By concluding, the main feature of this empirical example is the application of RSM jointly with CA in order to establish the best profile according to the judgements, expressed in metric scale, on the full profiles and on the status-quo. With this approach it is possible to take into account both a new service/product and a revised one. In addition, baseline variables of respondents, evaluated as noise variables, are introduced in the optimization procedure, by also considering their categorical nature. Note that in this case (Sect. 8.4.1) an only one surface is estimated and two optimal solutions are evaluated in the optimization step. The empirical results confirm the relevance of our proposal, also when comparing these results with the optimization within a-priori strata and the working situation.