Fundamental characteristics and statistical analysis of ordinal variables: a review

Michele Lalla

Quality & Quantity
International Journal of Methodology

ISSN 0033-5177
Volume 51
Number 1

Qual Quant (2017) 51:435-458
DOI 10.1007/s11135-016-0314-5

Springer

Fundamental characteristics and statistical analysis of ordinal variables: A review

Michele Lalla

Abstract. The measurement of several concepts used in social sciences generates an ordinal variable, which is characterized by rawness of the output values and presents some much debated problems in data analysis. In fact, the need for effective analysis is easily satisfied with parametric models that deal with quantitative variables. However, the peculiarities of the ordinal scales, and the crude values produced by them, limit the use of parametric models, which has generated conflicting favourable and unfavourable views of the parametric approach. The main distinctive features of ordinal scales, some of which are critical points and nodal issues, are illustrated here along with the construction processes. Among the traditional procedures, the most common ordinal scales are described, including the Likert, semantic differential, feeling thermometers, and the Stapel scale. A relative new method, based on fuzzy sets, can be used to handle and generate ordinal variables. Therefore, the structure of a fuzzy inference system is exemplified in synthetic terms to show the treatment of ordinal variables to obtain one or more response variables. The nature of ordinal variables influences the interpretation and selection of many strategies used for their analysis. Four approaches are illustrated (nonparametric, parametric, latent variables, and fuzzy inference system), highlighting their potential and drawbacks. The modelling of an ordinal dependent variable (loglinear models, ordinary parametric models or logit and probit ordinal models, latent class models and hybrid models) is affected by the various approaches.

Keywords: measurement, fuzzy sets, feeling thermometer, semantic differential, Likert scale, Stapel scale

Department of Economics "Marco Biagi", CAPP (Centre for the Analysis of Public Policies), University of Modena and Reggio Emilia, Via Jacopo Berengario 51, 41121 Modena, Italy
e-mail: michele.lalla@unimore.it

1. Introduction

Attributes, also called concepts or characters, play an important role in the explanation of phenomena. However, especially in the social sciences, attributes are often magnitudes characterized at least by three elements: (*i*) a certain vagueness because they are derived from their context of use and standardized after a long processing path, as the concept of intelligence; (*ii*) a poor uniqueness of definitions because their construction is affected by the adopted reference value systems and by the underlying theories of interpretation; and (*iii*) reproducibility is not automatic because there are variations in space and time that can limit stability owing to the effect of the structural mutation that may be produced in the contingencies of the applications (Bernardi 1995). It follows that their measurement is difficult and often remains at an ordered qualitative level, i.e., an ordinal scale.

The processing of data from measurements with ordinal scales has always been problematic and it has been extensively discussed for at least a century. The solution has not yet been found because the problem is unsolvable in ontological terms in the context of the definitions adopted for the ordinal scale, compared to the interval scale, which is the subsequent hierarchical measurement level. Algebraic operations between different values of ordered data, obtained from several measurements and corresponding to levels of the ordinal scale, are prohibited by the definition of the scale, which admits the order relation between levels, but explicitly excludes equidistance between levels. In fact, if there were equidistance, it would be an interval scale (see below).

The key aspects of the attributes, measured at the ordinal level are outlined below, along with their development and approaches to their analysis. Section 2 describes the assumptions and the main characteristics of the measurement process. Other topics are also examined, including a synthetic description of the main ordinal scales – such as the Likert scale, the semantic differential, and the various forms of feeling thermometers – (§2.1), their construction method (§2.2), and the fuzzy approach aimed at obtaining an ordinal variable using a battery (§2.3). Section 3 briefly discusses the four main approaches (non-parametric, parametric, latent variables, and fuzzy inference systems) to the analysis of ordinal variables, highlighting their potential and drawbacks. Section 4 concisely illustrates some ideas on associations between ordinal variables and their impact on the interpretation of the models. A discussion of the formal and mathematical models has been avoided here, while the focus is centred on the identification of the methodological problems associated with approaches for the treatment of ordinal variables. Section 5 concludes with some comments and remarks.

2 Measurement characteristics of ordinal variables

The classical theory of measurement implies that the measured attribute may be expressed by a real number for multiplying the unit of measure. In formal terms, ω_i is the object evaluated with respect to some quantitative attribute *A*, also called the concept or character or variable, which represents a certain property. The attribute $A(\omega_i)$ is measured in conventional standard units *u*, so that it will result in: $A(\omega_i) = c_i\, u$ with $c_i \in \mathbb{R}$, where $\in$ stands for "belonging to" and $\mathbb{R}$ is the set of real numbers. Therefore, ordinal variables would not constitute a measurement coherent with such a definition because they do not provide a defined standard unit, *u*. The measurement process has been reviewed by Stevens (1951), who defined measurement as a procedure of assigning numbers to objects or events according to certain rules. The procedure is unique, homogeneous, consistent, valid, reliable, precise with respect to a unit of measurement, and meaningful. This approach states a representational theory of measurement because the numbers are representations of magnitudes or states of the attribute and the approach is based on the Platonic notion that there is a true reality (ιδέα) of the attribute. The measurement process is intended to evaluate the magnitude, intensity, strength or amount of an attribute in this true dimension, which is ideally a ratio or metric scale, and finally to translate it into a variable. The underlying assumptions are: (1) *unidimensionality* of the attribute constituting the object of measurement, implying that the various elements or statements used for measurement should refer to the same concept or property; (2) *continuity* of the magnitude concerning the attribute so that the measurement process, which often involves responses from individuals to the various statements used for its evaluation, generates a numeric value that expresses the intensity of the attribute in the measured objects/subjects; (3) *nonlinearity* or non-equidistance between the response categories. This latter assumption is the most controversial because the nature and definition of an ordinal variable specifically exclude linearity. However, linearity is implicitly assumed in data processing when the evaluation of the intensity of an attribute, as measured by an items' battery, is obtained through the sum of the numbers assumed as a representation of linguistic expressions adopted in naming the elements of the response set and this latter is used for each item of the battery. For example, in the Likert scale (Likert 1932), the assumption of linearity implies that the distance between "completely agree" and "agree" should be the same as the distance between "disagree" and "totally disagree". In addition, distances between the above and the "uncertain/ neutral" category should be equal.

The ordinal attribute is empirically determined by means of two operations. The first concerns the equality/inequality of two or more objects; the second refers to the sorting of objects. The first operation is characterized by the relation of coincidence (≡), which specifies or determines the nominal level of measurement. The second is characterized by the relation of precedence ($\prec$), which identifies or defines the ordinal level of measurement.

For all (∀) elements a, b, c belonging to ($\in$) the set of objects/subjects ($\Im$) sorted according to the measurement of the attribute under examination, in symbols $\forall\, a,b,c \in \Im$, the following properties hold in the relation of coincidence: (1) reflexivity, $a \equiv a$, that is, "*a* coincides with *a*"; (2) reciprocity or symmetricity, $a \equiv b \;\Rightarrow\; b \equiv a$, where the symbol $\Rightarrow$ indicates "implies/then", *i.e.*, "if *a* coincides with *b*, then *b* coincides with *a*"; (3) transitivity, $a \equiv b \wedge b \equiv c \;\Rightarrow\; a \equiv c$, where the symbol $\wedge$ is the logical connective conjunction "and", *i.e.*, "if *a* coincides with *b* and *b* coincides with *c*, then *a* coincides with *c*".

The ordering relation of precedence ($\prec$) has the following formal characteristics: (1) non-reflexivity, $a \nprec a$, *i.e.*, "*a* does NOT precede *a*, *i.e.*, itself", which is equivalent to "if $a \equiv b \;\Rightarrow\; a \nprec b \;\wedge\; b \nprec a$", and therefore the symbol $\nprec$ stands for "does not precede"; (2) non-reciprocity or anti-symmetricality, $a \prec b \;\Rightarrow\; b \nprec a$, *i.e.*, "if *a* precedes *b*, *b* cannot precede *a*"; (3) transitivity, $a \prec b \wedge b \prec c \;\Rightarrow\; a \prec c$, *i.e.*, "if *a* precedes *b* and *b* precedes *c*, then *a* precedes *c*"; (4) trichotomy, $a \not\equiv b \;\Rightarrow\; \vee\, a \prec b \;\vee b \prec a$, where the symbol $\not\equiv$ means "does not coincide", that is, "*a* does NOT coincide with *b*", and the symbol $\vee$ is the logical connective disjunction "or", *i.e.*, "if *a* does NOT coincide with *b*, then either *a* precedes *b* or *b* precedes *a*".

The attribute is thus classified as measured in different modalities expressing the class assignment, which are likely to be related to each other via the operators "less than" and "greater than" (for the relation of precedence, $\prec$) and it is possible to talk about an ordinal property (Coombs 1953; Hand 2004). Let *L* be the measurement of the attribute in the object ω_i, which generates the result $L(\omega_i)$. The set of all equivalence classes (*M*) resulting from the application of the two relations is called an *ordinal scale* because its representation is based on *M* categories or distinct modalities, which are logically connected by the order relation. Let ω_i and ω_j be two objects evaluated with respect to a specified quantitative attribute *A*. This latter is measurable at an ordinal scale level. If it satisfies all the assumptions of a nominal scale and if for all $A(\omega_i) > A(\omega_j)$, it follows that $L(\omega_i) > L(\omega_j)$. For the symbolism and relations with and between other scales, see Siegel and Castellan (1988), Khurshid and Sahai (1993), and Kampen and Swyngedouw (2000).

Each manifestation of the attribute is placed in one of the M possible categories (also called classes or levels), which satisfy the following properties and are similar to those at the nominal level: (1) *fundamentum divisionis*, which is unique and appropriate for obtaining a suitable criterion to determine categories (Marradi 2007) and that is related to the assumption of the underlying continuum; (2) *potential reducibility* of the observable states, when there are many, because there may be a need to decide how to convert them and into which categories of the attribute; (3) *classifying principles* for the creation of the categories; (4) *disjunction of different classes*, which is a consequence of the properties of the relationships of coincidence and precedence; (5) *exhaustiveness* or completeness of all identified categories; (6) *uniqueness* of belonging to classes because each element can belong to one and only one class; (7) *equivalence* among all elements belonging to a specified class; and (8) *insuppressibility of empty classes*, both intermediate and final, defined after any reduction, because a class can also be empty, unlike the attributes measured at the nominal level, in which the empty classes often do not make sense.

The attributes measurable with ordinal scales are qualitative and ranked, and then they are still variates. Each modality can be represented by a whole number, expressing the level of the amount of an ordinal attribute. Therefore, the collected modality often indicates a position with respect to the other modalities, *i.e.*, the place occupied by an element in the sorted list and called the rank. The problem that arises in formal terms in the representational approach lies in the need to have a one-to-one relationship between the set of measures L and the set of attributes A, *i.e.*, L must be an isomorphism of A. For each scale measurement, there is an appropriate transformation. For the nominal level, the admissible transformations are those of all the invertible functions; for the ordinal level, the admissible transformations are those of all the invertible functions that are strictly monotone increasing.

The most incisive criticism of the representational approach focuses on the impossibility of knowing whether L is an isomorphism of A (Prytulac 1975). Given that the true value of an attribute cannot be observed, there is no way of verifying whether the measure obtained corresponds to the true value of the attribute. Moreover, when a qualitative attribute has unconnected modalities and is measured, then a nominal scale is obtained and the approach falls into a *contradictio in adjecto*, as the two terms “nominal” and “scale” contradict each other. Finally, another unresolved aspect of the representational approach regards the subjectivity of the measurement with respect to the agent or observer carrying out the measurement operations. In formal terms: “Does L belong to the observer or the observed?” Comparability of measures implies a need for a standard apparatus, which is obtained using either the same measuring instrument or calibrated instruments. In the social sciences, however, the respondent expresses only his/her degree of agreement according to a personal calibration, which makes the measures unfit for comparison, which, in turn, entails insurmountable problems in data analysis. For example, let a response scale be given to evaluate the saltiness of a food: 1=‘insipid’, 2=‘slightly salty’, 3=‘salty’, 4=‘fairly salty’ and 5=‘very salty’. The term ‘insipid’ may correspond to zero grams of salt for a subject and one or two or more grams for another subject. Therefore, the subjects are not categorizing the intensity of the attribute in the same manner and the values deriving from the measurement process may indicate different things, even when the numbers/labels are equal.

2.1 Subject-centred ordinal scales

The measurement of an attribute can consist of a single operation, often involving some form of a single assertion, typically administered to a subject, ω_i, and that generates a result $L(\omega_i)$, through a scale providing M modalities or an options response set, *i.e.*, a list of possible intensity/answers, often described in verbal terms and ordered in some way. More often, the attribute A cannot be measured satisfactorily by means of a single measuring operation because the magnitude of A is irreducible to a direct operational definition. The attribute is then disarticulated into the simplest sub-concepts, A_k $(k=1,\ldots,K)$, which are directly measurable. The gap between the attribute and the sub-concept is filled by a semantic relationship of indication, or a semantic representation of the attribute, between the concepts (which are translated operationally) and the more general attribute *A*. The *K* measurable concepts are termed indicators (Bernardi et al. 2004; Marradi 2007). The *K* assertions, statements or propositions (*items*), should then be formulated in a manner implying that they have a semantic connection with the attribute *A* and are monotonic, that is, formulated in a unidirectional form with respect to the object to be measured so that the increase in a subject's favourable attitude towards the object generates an increase in the score achieved for that proposition. If this is the case, the researcher must devise items with a favourable content and items with an unfavourable content towards the attribute *A*, each one having different intensities and so that the two sets have the same cardinality. The collection of statements *K* is termed a *battery*, but more frequently it is called a "scale", generating an ambiguity of meaning because the term "scale" refers to both the set of *M* choices prepared for the response to each item (hereinafter, often referred to as "response set") and the set of all *K* items designed to measure the attribute *A*.

The scaling technique was developed by the early 1920s to study psychophysical and psychological attitudes and behaviour (among others, see White 1926; Thurstone 1927a, 1927b, 1928). The ordinal scales most frequently examined in the literature and used in practical applications are described below.

The ***Likert scale*** (Likert 1932), an ideal scale conceived for measuring attitudes, is one of the most well known scales in use. In its usual format, it is constituted by a response scale with five-ordered categories (*M*=5). In its classical formulation, every item requires the respondent to express his level of agreement with the current statement, and the response set suggests five possible ordered alternatives: strongly agree, agree, neutral or uncertain, disagree, strongly disagree. For every item, the answers have the numeric labels 5, 4, 3, 2, 1 (or 4, 3, 2, 1, 0) with the specific function of ordering the alternative answers. The scores attributed to each subject for each item coincide with a number of labels for items in favour of the attribute and the numeric labels are inverted (1, 2, 3, 4, 5 or 0, 1, 2, 3, 4) for items against the attribute *A*. In a battery, the final score for each subject is generally given by the sum of all partial scores corresponding to each selected answer for the *K* items constituting the battery. In other words, "the individual score over the entire scale, [can also be] constituted by the [...] the sum of the numeric codes assigned to the chosen responses from one individual to the various items of the scale" (Cacciola and Marradi 1988, pp. 72-73) because the codes of the response set are almost always expressed with natural numbers.

The *semantic differential*, in its usual size or standards, consists of a response set of seven ordered categories (*M*=7), which may vary in number and are self-anchored to bipolar or opposite adjectives or statements. More specifically, in the horizontal arrangement of the response set, to the left of the seven adjacent boxes there appears a term, for example "low", and after the seventh box, to the right, the opposite term "high" appears, while the other boxes do not present indications. Therefore, the system is anchored to the two terms: the first at the beginning and the second at the end (Osgood 1952; Osgood et al. 1957). For every bipolar assertion, the respondent indicates the degree to which the descriptor represents his opinion of the concept under consideration. The semantic differential is aimed at measuring directly both the preferred direction between the two opposite terms (*e.g.* "useless" and "useful" or "unaffordable" and "affordable") and the extension or the entity of the direction emerging with the choice among categories expressing the intensity of preference. The number of items is generally high and the interpretation of the results is based primarily on three factors (*i.e.*, evaluation, potency, and activity), which require an analysis that is quite complex, involving burdensome and arduous data processing procedures. Therefore, the objectives of these scales can be achieved in the long term, limiting their applicability, or by subjecting them to simplified procedures, which reduce their potentiality (Yu et al. 2003).

The *Stapel scale*, in its standard form, offers a ten-point, non-verbal rating response set, ranging from −5 to +5 without the zero point (0) or the neutral/ central modality. As with the semantic differential, the Stapel scale is aimed at simultaneously measuring both the preferred direction and the intensity of preference and it avoids the problems of finding a usable antonym for each adjective, as it is aimed at measuring the extent to which the respondent believes that the proposed adjective describes the attribute under examination. In its presentation, it was explicitly stated that the intervals obtainable from the scale positions are not equal and the ratings are not additive, to avoid violation of the third assumption of ordinal scales (Crespi 1961). The use of a scale with ten modalities is more intuitive and common than a seven-point scale, but the absence of the zero-point creates a gap that breaks up the linearity of the labels. In fact, in practice non-additivity is often ignored and in a battery such as in the Likert scale, the final score for each subject is generally given by the sum of all partial scores corresponding to each selected answer in the K items constituting the battery. Compared to the semantic differential, the Stapel scale presents (measures) each adjective or phrase separately and modalities are identified by numbers. However, sometimes even the modalities of the semantic differential are numbered.

The *self-anchoring scale*, in its usual format, consists of a graphic non-verbal response set, consisting of a ten-point ladder scale or ten modalities (Kilpatrick and Cantril 1960; Cantril and Free 1962) associated with items which ask respondents to define their position (point) on the scale compared to the anchors. The best ranking modality is at the top, if the scale is in the vertical position (case 1), or at the extreme right, if the scale is in the horizontal position (case 2). The worst ranking modality is at the bottom in the first case, or at the extreme left in the second case. It was devised within the field of the transactional theory of human behaviour, according to which the reality of each individual is unique in some way and the results of their perceptions are conceived as a current extrapolation of the past related to the sensory stimulus. The self-anchoring scale can solve some problems and distortions typical of ordinal scales. In ordinary applications, it is often used with fixed anchors to obtain scores that are more homogeneous and coherent among subjects. However, where the anchor is already defined, it is implicitly assumed that there is an objective reality. In any case, self-anchoring and fixed anchors seem equivalent in psychometric terms, although there are obvious conceptual differences (see Hofmans et al. 2009, among others).

The *feeling thermometer scale*, in its usual format, is composed of a segment ranging from 0 to 100 degrees, which reports only some specific level values, resembling the centigrade scale of temperature. It was originally developed by Aage R. Clausen for the study of social groups and was used for the first time in the American National Election Study (American National Election Survey – ANES 1964). It was subsequently modified by Weisberg and Rusk (1970) and transformed into a card administered to the interviewed subject because they thought this tool might be useful in the evaluation of the ascendancy of a political candidate in the electorate. The card listed nine temperatures throughout the scale range and their corresponding verbal meanings using adjectives such as “hot or favourable” or “cold or unfavourable” to express the intensity of one’s attitude towards a candidate.

The *Juster scale*, in its usual format, is made up of eleven point-ordered items, as in the integer decimal scale ranging from zero to ten, with a verbal description for each scale value or modality. It was used to predict future purchase behaviour regarding durable goods (Juster 1960, 1966). Each question asks the respondent to assign a probability as to the possibility of adopting the behaviour described by the proposed statement.

In structural terms, the last four types of scales (Stapel, self-anchoring, feeling thermometer, and Juster) seem equivalent to the thermometer scale, used to measure temperature, which corresponds to the level of the interval scale. They have a long history, although they are ascribed to Crespi (1945a, 1945b), as quoted by Bernberg (1952), for example. It should be noted, however, that the thermometer scales used in the social sciences do not provide values corresponding to an interval scale because their construction cannot ensure invariance of the unit of measure or equidistance among the categories of the response set. Let ω_i be the object evaluated with respect to a quantitative attribute *A*. This attribute is said to be measurable at the level of an interval scale if all the properties and assumptions of an ordinal scale are valid, with the exception of the third assumption, which now states the equidistance, and meets both of the two following equalities, with $a > 0$:

$$(a) \quad L(\omega_i) = a\,A(\omega_i) + b, \qquad (b) \quad L(\omega_i) - L(\omega_j) = a[A(\omega_i) - A(\omega_j)]. \tag{1}$$

The *Guttman scale* does not have a specific response format, but it handles more than two statements generally having a binary response set and is a type of composite measure directed to represent some more general attributes (Babbie 2010). In other terms, it is a method that attempts to order subjects and statements/items simultaneously, where the latter are often also called stimuli. Its application is directed to discover and use the structure of the intensity of a set of empirical data indicators of an attribute (Guttman 1950, 1968). It is used to ascertain whether a set of terms or items forms a scale, that is, whether there is a hierarchy between them, as in the modalities of the "education level" and "qualification level" variables.

As mentioned above, the term scale often means both the set of responses to a single assertion and the set of outcomes obtainable from all possible combinations of answers to all items used to measure an attribute. In this second sense, there are many scales aimed at measuring many concepts used in the social sciences. By way of illustration, the following scales are cited: the social distance scale by Bogardus, designed to measure the degree to which a person would want to be associated with a certain category of people, such as an ethnic minority, and also used to measure racism; the Srole scale, devised to measure the concept of anomie (Babbie 2010); the hope scale for adults (Snyder et al. 1991); and the scale of organizational awareness (Weick and Sutcliffe 2007).

2.2 The construction of ordinal variables

The scales more frequently used and described above, are used as summative instruments. Let *i* be the index for subjects: $i = 1, \ldots, n$. Let *j* be the index for the attributes under examination: $j = 1, \ldots, J$. Let *k* be the index for the statements prepared for the *j*-th attribute: A_j : $k = 1, \ldots, K_j$. The score obtained by the *i*-th individual, in the *j*-th attribute and in the *k*-th statement is $x_{ijk} \in \{1, \ldots, M\} \subset \mathbb{N}$, where $\mathbb{N}$ is the set of natural numbers. In theory, *M* may vary from one statement to another, M_k, but in practice the response format always has *M* modalities for all statements. In general, the sum (x_{ij}) or the average ($\bar{x}_{ij}$) of K_j numbers measures the intensity of the *j*-th concept in the *i*-th subject:

$$(a) \quad x_{ij} = \sum_{k=1}^{K_j} x_{ijk}, \qquad (b) \quad \bar{x}_{ij} = \left(1/K_j\right) \sum_{k=1}^{K_j} x_{ijk}. \tag{2}$$

The sum (x_{ij}) is often scaled to one (or ten) with the following transformation $x_{ij} = (x_{ij} - x_{\min;\, j}) / (x_{\max;\, j} - x_{\min;\, j})$, for the *i*-th subject and the *j*-th attribute, and where $x_{\max;\, j}$ and $x_{\min;\, j}$ are the maximum and minimum of x_{ij}, respectively (Aiello and Attanasio 2004; Hoaglin et al. 1983; Ricolfi 1984). The rankings by sum are obtained through x_{ij} as in equation (2). However, although they are used in the applications in very pragmatic terms, the operations of sum and average are not admissible because the result depends on how the variables are coded at the origin. If the response choices are binary or dichotomous, then the problems are partially reduced simultaneously with the reduction of the location precision of the attitude of the respondent. Purists prohibit the use of the average even when limited to its descriptive function of a distribution to summarize, with an index of central tendency, if its concentration or centre of mass is more to the left or more to the right of the median. This information can be achieved in part by associating the value of cumulative or retro-cumulative distribution with the median. In general, the use of the average as an indicator function of the centre of gravity of the distribution on the domain of an ordinal scale, with its conventional numeric system and when it is well known at the origin, could perhaps be admitted in a purely informative/descriptive function, but purity is never concessive. For example, the median is the correct tool for the synthesis of teaching evaluations in the current system used in Italy, which has items in a Likert format. However, in cases of crowded courses or combinations of courses, it can happen that the median provides no difference between the items under examination, as can be seen from Lalla et al. (2014, p. 28), where the median is always equal to the modality "more yes than no", evaluated numerically as seven, while the average varies from statement to statement. In fact, the mean informs us as to how the centre of gravity moves in the field of the adopted numeric labels, which are the same for everyone. Without indices of synthesis, the person dealing with the results of evaluations finds him/herself forced to specifically analyse the distribution of each question.

Pragmatics may have its reasons and logic, but it must be recognized that formally, in theory, there is no reason or logic in the practice of treating the numbers assigned to the modalities of an ordinal scale as real numbers, given the assumption of non-equidistance between the categories of the ordinal variables. The practice formalised in equation (2) remains problematic even in the case of an accepted conventional assessment of the intensity values for the modalities of the scale. Within the scope of an axiomatic approach, however, the latter strategy comes close, in point of fact, to the level of the interval scale. The modalities constituting the scale format can be evaluated by experts and by subjects randomly selected from the target population, according to a predetermined unit of measure. In the latter case, the median or the average may be assumed as the numeric value of the corresponding modalities. Furthermore, it is often possible to conceive the selection of a modality as the result of a discriminatory process governed by an underlying normal random variable. This might justify the use of the sum and average operators, exploiting the properties of normal random variables (see below). However, if the modalities are subjected only to an order relation, the use of the sum and the average remains problematic because they are inconsistent with the assumption of non-equidistance between the response categories. In fact, when the equidistance is valid, the obtained variable transits with good approximation in the level of the interval scale. Some interesting considerations on these focal points can be found in Niederée (1994) and Hand (2004), among many others.

The previous arguments reveal an apparent logical incongruity, as an attribute disarticulated into a set of sub-attributes, each one operationally translated in a binary item format, could be evaluated (measured) through a sum of the item values without strong theoretical and statistical discrepancies, whereas when the answer scale format has a finer graduation of the level of a binary scale format the sum is forbidden, in which case it is possible to dichotomise the same scale formats and proceed with the sum and obtain an evaluation that is rawer than that obtained through the sum of the original ordinal scale formats.

2.3 The fuzzy approach

The approach used by the Fuzzy Inference System (FIS) is based on the theory of fuzzy sets (Zadeh 1965; Dubois and Prade 2000) and it solves the problem arising from use of the sum and the average with a coherent and consistent operating procedure, when an overall indicator of the attribute A needs to be obtained using a battery or measured by transforming it into an index. If the attribute A is measured by a single statement, and only that attribute is to be analysed, then the two approaches (classical and fuzzy) generally provide the same results. If the attribute A is measured by more than one statement and has a certain vagueness or there are arduous measurement difficulties, as is the usual case in the social sciences and often reported in the literature (Bernardi 1995; Marradi 2007), then the fuzzy approach would be almost the natural choice.

The attribute *A* is often measured by more than one statement. In fact, there is an interest in obtaining an overall assessment of its intensity or strength. In this case, the FIS may constitute an interesting alternative compared to the traditional process, which violates the assumptions of the ordinal scale. The FIS could be developed according to a process that involves six steps. The first four are required steps, while the other two, (*v*) and (*vi*), are optional according to the specific requirements of the current application. The steps are described briefly and in abstract terms below. An example with further details can be found in Lalla et al. (2004). Moreover, hereinafter the attribute, character, statement, assertion, and item are prevailingly referred to as the variable.

The *identification of the problem* (*i*) is the process that identifies the relationship between the input variables and output variables and may start from the top down or from the bottom up. In the first case, it follows definite steps, which are similar to the process of a social survey, that is, it starts from the output variables, which may be one output variable, and all the corresponding macro-indicators, which semantically represent the given output variable(s), are identified for each one. Each macro-indicator is in turn disjointed into indicators of the same reduced semantic area and the recursive procedure continues until the indicators compounded only by the input variables are achieved. The final product has a modular tree-patterned structure, consisting of several levels or stages, where each level in the various fuzzy modules constituting the level of the tree is disarticulated downwards, so that the fuzzy modules are interlinked in the vertical direction. In the bottom-up procedure, the initial starting level (zero) is constituted by the input variables, which are already available, e.g. the assertions of the battery. These variables are aggregated to each other according to rational, random or convenient rules to create the first level or stage of aggregations, which consists of fuzzy modules of new construction. The latter are aggregated between each other and/or with the input variables that should be still aggregated, always according to rational, random or convenient rules to construct the fuzzy modules of the second level. The recursive part of the procedure continues in the same way until the fuzzy modules of the output level are obtained. The final product looks like a pyramidal tree having more than one level with a single or multiple outputs at the last level (see Fig. 1). The variables entering the system at a higher level than others, affect the output more heavily than those entering at lower levels. For example, X_{i9} affects the output FM4 of the fourth stage more than the variables $X_{i1}, \ldots, X_{i8}$.

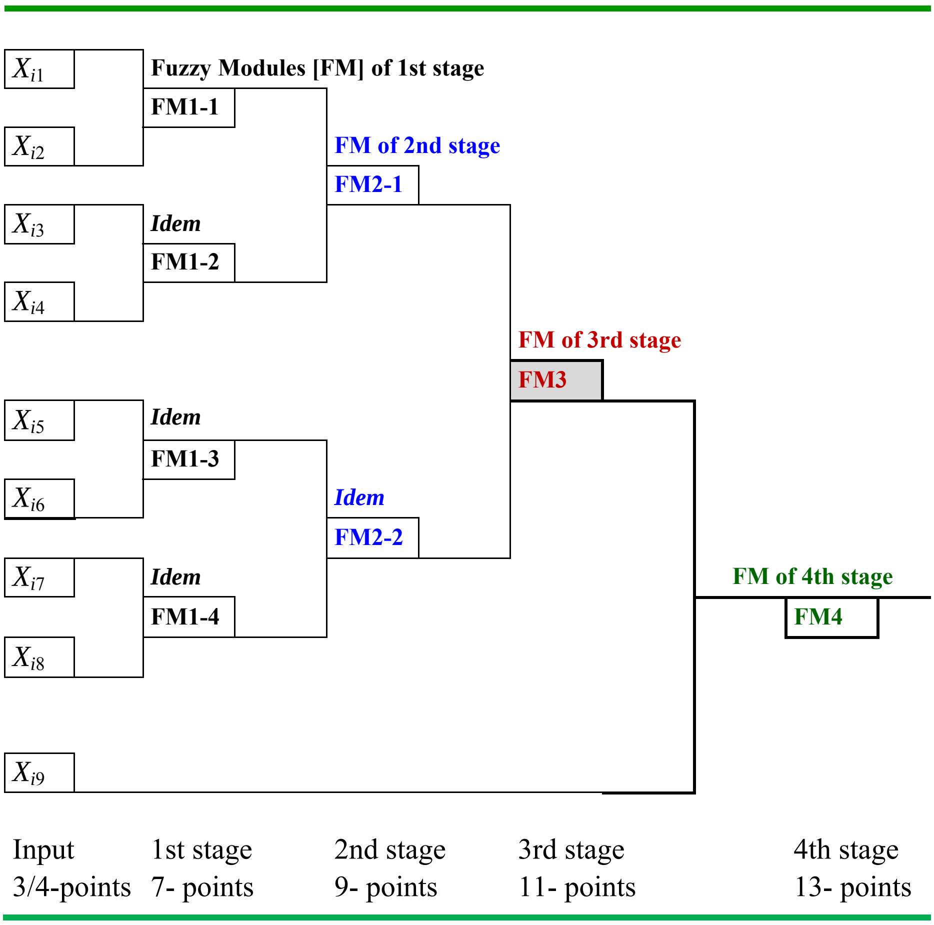

Fig. 1 Structure of a fuzzy inference system with nine input variables

The *fuzzification of the input variables* (*ii*) concerns the definition of the shape and number of the membership functions for each input variable. The membership function defines the degree to which a numeric value or lexical definition of a modality belongs to some specific category of response or destination scale. The fuzzification distributes the values of the input variables described as lexical or numeric labels on a segment broader than that expressed by the crisp value arising from the traditional scaling. In fact, the fuzzy procedure involves the two modalities adjacent to the one selected by respondent in order to model the vagueness of the response. The fuzzification of input variables should be understood in this sense. The scale formats of the fuzzy modules FM1.*i* (in Fig. 1), for example, will have seven modalities in the case of five-point Likert scale input variables because for the fuzzy modules, the central modality can be used without problems. The number of modalities for the scale formats of fuzzy modules increases usually by two units at each stage, but it can also increase by one unit or remain unchanged. The final outcome of the FIS is one (or more) fuzzy module(s), generally corresponding to one (or more) ordinal variable(s) with a determined number of modalities. For the example appearing in Fig. 1, the output is given by FM4, which is an ordinal variable with thirteen (13) modalities.

The shape of the membership functions may be determined using different methods (Smithson 1987; Smithson 1988), which are omitted here for the sake of brevity, but some of them could also be applied in the construction of classical ordinal scales to make them more manageable with the operators of sum and average (Lalla et al. 2004).

In formal terms, consider a single attribute to eliminate the index j of x_{ijk}. Let x_{ik} (for $k = 1,\ldots,K$) be the k-th input variables over K provided by the i-th respondent, each one with range U_k. Let y be the output variable with range V. Let M_k be the number of the categories of x_{ik}. In general, M_k may vary from one variable to another, but for a standard Likert scale and a battery of K statements having the same response set, it will be $M_k = 5$ (or $M_k = 4$ if it does not contain the central modality) for all $k = 1,\ldots,K$. It follows that an effective fuzzification of the input variables requires a number of membership functions greater than one and less than M_k. Additionally, each category of x_{ik} is described by a fuzzy number, $A^k_{m_k}$, $\forall\, m_k \in \{1, \ldots, M_k\}$, and denotes the set $A^k = \left\{A^k_1,\ldots,A^k_{M_k}\right\}$ of fuzzy inputs of x_{ik}, while the fuzzy output set is defined by $B = \left\{B_1,\ldots,B_{M_y}\right\}$, where M_y denotes the number of membership functions, categories or modalities of y. Each set has its own membership function:

$$(a) \quad \mu_{A^k_{m_k}}(x) : U_k \;\rightarrow\; [0,1], \qquad (b) \quad \mu_{B_{m_y}}(x) : V \;\rightarrow\; [0,1]. \tag{3}$$

The *construction of rule-blocks* (*iii*) is carried out by the relations established between the input ordinal variables and the output ordinal variable. Relationships can be identified considering situations involving multi-criteria decision-making processes and are expressed through rules, R_s, such as the following:

$$R_s: \quad IF\left[x_{i1} \text{ is } A^1_{m_1} \otimes \cdots \otimes x_{iK} \text{ is } A^K_{m_K}\right] \quad \text{THEN} \quad \left(y \text{ is } B_{m_y}\right), \tag{4}$$

for all combinations of $m_k \in \{1, \ldots, M_k\}$ and $m_y \in \{1, \ldots, M_y\}$. The expression on the left of "THEN" is the protasis, antecedent or premise, while the expression on the right is the apodosis, or consequent or conclusion. The symbol $\otimes$ (otimes) denotes an operator of aggregation, one of several t-norms (if the aggregation is carried out with an AND operator) or t-conorms (if the aggregation takes place with an OR operator). For example, the aggregation operator AND yields a numeric value $\alpha_{s,m_y} \in [0,1]$ that represents the execution of an antecedent in the rule. The number α_{s,m_y} should operate with the membership function of the consequent B_{m_y} in order to calculate the output of each rule. You can still apply the AND operator, but in a somewhat different manner: the $\otimes$ operator works on a number and the membership function fuzzy set B_{m_y}, while the R_s rule is applied on two numbers (Von Altrock 1997). An example of the rule-block is presented in Table 1 for the fuzzy module FM1-1 of Fig. 1 with numeric/symbolic values, instead of the actual labels, and with three membership functions in input and five membership functions in output, for the sake of brevity. The rules can be generated automatically with an algorithm or formulated by an expert.

Table 1 Example of the rule-block for the fuzzy module FM1-1 of Fig. 1

	Protasis			Apodosis
Statement 1	$\otimes$	Statement 2		
– IF X_{i1} is mf1	AND	X_{i2} is mf1,	THEN	FM1-1 is mf1
– IF X_{i1} is mf1	AND	X_{i2} is mf2,	THEN	FM1-1 is mf2
– IF X_{i1} is mf1	AND	X_{i2} is mf3,	THEN	FM1-1 is mf3
– IF X_{i1} is mf2	AND	X_{i2} is mf1,	THEN	FM1-1 is mf2
– IF X_{i1} is mf2	AND	X_{i2} is mf2,	THEN	FM1-1 is mf3
– IF X_{i1} is mf2	AND	X_{i2} is mf3,	THEN	FM1-1 is mf4
– IF X_{i1} is mf3	AND	X_{i2} is mf1,	THEN	FM1-1 is mf3
– IF X_{i1} is mf3	AND	X_{i2} is mf2,	THEN	FM1-1 is mf4
– IF X_{i1} is mf3	AND	X_{i2} is mf3,	THEN	FM1-1 is mf5

The *aggregation of block rules* (*iv*) incorporates the unification process of the outputs of all the rules in a single output, *Y*. For each rule, R_s, involved in the numeric input values, $\mu(\alpha_{s,m_y} \otimes B_{m_y})$, a different output is obtained. These membership functions of fuzzy sets should be aggregated by an OR operator using a t-conorm: those more frequently used are the maximum, the probabilistic, and the Lukasiewicz t-conorm known as the limited sum. Considering Fig. 1 again, the response fuzzy module (FM1-1) is ready, but it is still in a fuzzy form. One proceeds in the same manner for the other input variables to generate all the fuzzy modules of the first stage. The procedure is applied following a recursive form, in the subsequent stages. At each level of the tree diagram, the fuzzy modules generated in the previous levels or variables, which are not yet aggregated, are aggregated in the same manner. The process continues up to completion of all the operations provided by the tree diagram and reaching the top level containing the final fuzzy module (or fuzzy modules if there is a multiple output) corresponding to the variable (or variables) of interest, which constitutes the response of the FIS. The output may be of the ordinal level. Therefore, it remains within the sphere of the same measurement level of the input variables, without violating the assumptions of ordinal scales. In some applications, especially when the input variables are continuous, there is a need to summarize the results with a numeric/crisp value to achieve an easy and more direct understanding of the result of measurement. Then, in the latter case, the following step is necessary.

The *defuzzification of output* (*v*) is the process that maps the fuzzy set obtained as output $\mu_B(y)$ in a real numeric value, *y*, so that for the *i*-th respondent and the *q*-th output fuzzy module, the crisp output will be y_{iq}, where $q = 1, \ldots, Q$, if the output variables are *Q*. The defuzzification of output is the inverse operation of the fuzzification of input. Therefore, this operation concentrates the vagueness or fuzzy response of the system, represented by the polygon resulting from the activations of the membership functions in the last stage of output, into a number, which expresses the central tendency of the entire polygon. There is no universal technique to perform defuzzification, *i.e.*, to summarise the output polygon with a number, because each algorithm presents interesting properties for particular classes of applications (Van Leekwijck and Kerre 1999). The selection of a suitable method requires an understanding of the process that underlies the mechanism generating the output fuzzy module and the meaning of the different possible responses. Two criteria are used to choose the most suitable method: (1)

the "best compromise" and (2) the "most plausible outcome" (Von Altrock 1997). The first criterion resolves situations, typically for which the average makes sense: the most frequently used techniques are the centre-of-maxima (CoM) and centre-of-area/gravity (CoA/G). As regards the CoM, for the q-th output fuzzy module, let $M_{\mathrm{F};iq}$ be the number of activated output membership functions because the i-th respondent may activate more than one membership function. Let y_{iqm} be the abscissa of the maximum of the m-th activated output membership function. If the latter corresponds to a maximising interval, then y_{iqm} is assumed to be equal to the median value of this interval. The numeric real crisp value, $y_{\mathrm{CoM};\,iq}$, is given by the weighted average of the maximum of the membership functions, each weighted with the corresponding level of activation, $\mu_{\mathrm{out};\,m}$,

$$y_{\mathrm{CoM};\,iq} = \sum_{m=1}^{M_{\mathrm{F};iq}} \mu_{\mathrm{out};\,m}\, y_{iqm} \Big/ \sum_{m=1}^{M_{\mathrm{F};iq}} \mu_{\mathrm{out};\,m}\,. \tag{5}$$

The second criterion deals with the situations, generally for which the average of the maxima does not produce the most plausible result, as the average value may often not be observable, so that the set-up techniques determine the system's output only for those membership functions with the highest resulting degree of support, always corresponding to an observable value. If the maximum is not unique because it corresponds to a maximising interval, then y_{iqm} takes the value of the median of the maximising interval. There are many techniques, but in the case described above, the response of the FIS becomes a sort of maximum of maxima (MoM), $y_{\mathrm{MoM};\,iq}$, although the acronym is also used for other techniques:

$$y_{\mathrm{MoM};\,iq} = \max_{1 \le m \le M_{\mathrm{F};iq}} (y_{iqm})\,. \tag{5}$$

The method thus selects the terms most suited to the problem at hand, instead of mediating between the different results of inference. Therefore, it is frequently used for the recognition of structures and classifications, as in the case of outputs with an ordinal level of measurement. The modalities of such outputs can also be described by linguistic expressions because the intrinsic nature of the result with respect to the phenomenon under examination implies that the more plausible solution is more suitable than the average. Finally, note that should the final result be an ordinal variable with modalities described by linguistic expressions, then the actual corresponding data values are always given by the membership functions and therefore their comprehension plays a key role.

The calibration of the model (*vi*) or sensitivity analysis of the model is carried out to adapt the FIS to real situations that it should represent. The operator handles the FIS as an ordinary model, where he/she may change the input variables, the membership functions of the input variables, the fuzzy rules, the hedge operations, the aggregations, and so on. The tuning of the performances of the FIS can be achieved following a procedure of four steps: (1) definition of the objective function for the output fuzzy variables; (2) changes in the various elements constituting the system, such as the parameters of the input data and/or the membership functions and/or rules, and/or aggregation operators; (3) validation of the results by comparing the objective functions and output functions; (4) repetition of steps (2) and (3) until the differences between the target and output functions are not below the chosen error criterion.

The FIS easily solves the problem of the subjectivity of the measure through fuzzification: with his/her answer, the subject activates not only an individual modality, but partially the two adjacent modalities as well. The analogy between fuzzy sets and probability theory may suggest possible use of a similar procedure with the probability functions in place of the membership functions, but the random process does not seem easily manageable in its current state. A method of construction of the membership functions is based on empirical distributions of numeric evaluations of the modalities of a response set, constructed with specific surveys of the target population. In such cases, in the traditional approach, the surveyed values of the modalities of the response set may be assumed to be equal to the median or the average of the evaluations obtained with the survey. Using the same previous symbols: $x_{ijk} \in \left\{m_1, \ldots, m_{M_k}\right\} \subset \mathbb{R}$, the model becomes $x_{ijk} = m_{ijk} + \varepsilon_{ijk}$ with somewhat simplified indices, where m_{ijk} is the value of the level chosen by the i-th respondent in the k-th assertion of the j-th battery, and ε_{ijk} is the corresponding error term. For example, in conditions of homoskedasticity and $\varepsilon_{ijk} \sim N\left(0, \sigma^2\right)$, the sum becomes $x_{ij} = m_{ij} + \varepsilon_{ij}$. Note that, in general, the values of the modalities of the response set, $\left\{m_1, \ldots, m_{M_k}\right\}$, depend on the lexical label used to describe the modalities and on the question.

The FIS does not solve the problem of the ordinal level of measurement of the variable, even when the output can provide a real number. This is because the output real number always depends on the support of the input variables, which are ordinal, and numeric values attributed to their lexical labels are also labels themselves and not numbers. Therefore, the final number does not express the intensity in the continuum if the numbers at the origin or input do not express it. In fact, this specific issue has been investigated (Domingo-Ferrer and Torra 2002). The FIS solves the problem concerning the use of the sum and average operators, but it generates several other problems, no less relevant than those involved in the conventional use of the sum and average of the numeric labels to obtain the final evaluation of the attribute under examination. In fact, the two approaches yield very comparable results within the same numeric assignments to the response set modalities (Lalla et al. 2004).

Attempts developed in the sphere of feeling thermometers to obtain measures belonging to an interval scale have not provided encouraging outcomes because they involve both epistemological issues and a strong empirical tendency to reduce the number of modalities (see among others, Hofacker 1984; Marradi 1998). Therefore, the attempt to achieve an interval level has in fact been brought back to the ordinal level.

To take advantage of the various models available for data analysis requiring variables with a level of measurement higher than the ordinal level, the alternative is to work on the procedures for assigning numbers to scale format modalities, which remains problematic. For instance, to pass the level of ordinal measurement, offering the respondent a self-anchored segment on which to indicate the intensity of the attribute corresponding to his/her feelings has been suggested. The respondent marks his/her evaluation on the segment and the distance of the mark from the extreme left of the segment can be measured by optical technologies to obtain a real number. Although the outcome of the measure may seem to be at the interval level, the respondent's mark is almost surely subject to error as it is a rough estimate. This essential imprecision might be much higher than the error corresponding to a feeling thermometer and therefore, the proposed technique solution would fall within the cases of uncertainty and difficulties that are well known in the literature. Moreover, the same distance may have different meanings for different subjects.

3 Analysis of ordinal variables

3.1 The nonparametric approach

Nonparametric statistics developed because it was recognized that the internal structure of data makes the models usually applied to variables measured with an interval or ratio scale unusable. Their unusability depends on the assumption that the variables of each level of a scale have a set of admissible transformations, which in theory preserve the meaning of the propositions in unchanged terms with respect to their truth or falsity. For ordinal variables such transformations lead to ambiguous and/ or improper results (Stevens 1951; Wilson 1971; Siegel and Castellan 1988), *i.e.*, the results of the parametric models depend on the numeric coding system adopted for the modalities of the ordinal variables. In short, formal models are very effective and elegant, but they impose strict constraints on the measurement of the variables involved and not satisfied by the ordinal level. However, the nonparametric approach also has some conceptual and operational limits. To eliminate the uncertain invariance of propositions deriving from scaling, the value labels of modalities in the scale formats are fixed in more or less consistent modes and forms and this is what often happens in real applications. This practical strategy nullifies the representational measurement approach in favour of an operational approach in which the procedures for the assignment of the numeric codes are fixed in a conventional manner, perhaps even when this convention is reasonable, rational, and appropriate. Therefore, the scores become numbers with all the properties of numbers.

The univariate analysis of ordinal variables can use some statistical techniques that make this investigation possible without many problems because there is a large set of tests that satisfy all possible questions that may emerge from the data. The chi-square, sign, and Kolmogorov tests are part of this set, although the Kolmogorov test may present some difficulties (Siegel and Castellan 1988; Landenna and Marasini 1990; Conover 1999). All nonparametric tests are based on the absence of assumptions about the distribution shape of the ordinal variable (distribution-free tests), making these tests less efficient, but suitably applicable to data as they require a reduced number of restrictions.

The bivariate analysis immediately raises insurmountable theoretical and practical problems. Two cases can be distinguished: (a) an ordinal variable with respect to a dichotomous one, which divides the set of observations into two independent samples, and (b) two ordinal variables, which first pose the question of what association exists between them and how to determine it. If they are associated, then the knowledge of one (independent) variable reduces the uncertainty about the other (dependent) variable or it allows a more accurate prediction of the dependent variable, narrowing the range of expected values. In case (a), the solution of rankings is the strategy more frequently used in practice, as expressed by the Wilcoxon-Mann-Whitney test: in a data set ranked without regard to the sample to which they belong, it considers the sum of the ranks for the observations in one sample compared with that expected for the entire data set. The Kolmogorov-Smirnov test is less problematic than others and is based on the differences between the two empirical distribution (cumulative) functions observed in the two samples. In case (b), there are different measures of association such as Spearman's rho, Kendall's tau, the polychoric correlation, and other representative parameters such as the Somer's *d* and Goodman and Kruskal's *gamma* (Siegel and Castellan 1988; Landenna and Marasini 1990; Jöreskog 1990). If the number of modalities of the two variables is limited, a contingency table is easily obtained and it can be analysed using various techniques, including loglinear models (Agresti 1990) or the analysis of variance with the Kruskal-Wallis test, the median test or the Nemenyi test, especially if the independent variable belongs to a level of a nominal scale.

Multivariate analysis generally involves a model that is discussed in the subsequent section (§4). It should be noted here that the model is limited by various elements such as the number and nature of the explanatory variables or the number of cases. For example, the two-way analysis of variance can be still performed with the Friedman test, among others. The application of the loglinear models is limited by the high number of modalities of variables in the model and by the number of cases. The development of permutation tests has opened up several interesting prospects for the application of multivariate analysis (Pesarin 2001). However, it should be noted that as in many other methods of data analysis, not all inferential problems in real situations can be treated with the permutation approach because this approach also requires the satisfaction of some conditions. If these conditions are not met, it can lead to erroneous results (Pesarin and Salmaso 2010).

3.2 The parametric approach

For both monovariate and multivariate analysis, the parametric approach starts from the identification of a statistical data model generator, before data are observed, *i.e.* the first step is constituted by the formulation of the statistical model, which requires the specification of a probability function, involving one or more parameters to be estimated from the observed data of the variables. The formulation of the model starts from a simple random sampling, $X_1, \ldots, X_n$ representing the observational vector, and states that $X_i \sim f\left(x_i | \boldsymbol{\theta}\right)$, where $i = 1, \ldots, n$ and $\boldsymbol{\theta}$ is the vector of the parameters. After carrying out the measurements, the observed vector, $x_1, \ldots, x_n$, is obtained. The variables involved in the model should belong to at least an interval scale, but there have been many attempts to use parametric analysis for ordinal variables justified by general and specific arguments. In general, the nonchalant use of quantitative techniques assumes the apodictic truth of the numbers, because "numbers do not know where they come from" (Lord 1953), arguing that measurement scales are irrelevant in statistical analysis (in the vast literature on the controversy, see Savage 1957; Boneau 1961; Gaito 1980; Gaito and Yokubynas 1986). In specific cases, for some techniques such as in factor analysis (Atkinson 1988), the robustness of the results with different scales and different distributions has been demonstrated.

The limits and inappropriateness of the parametric analysis performed on ordinal variables are evident to all. However, the upholders of parametric approach argue that some of the nonparametric techniques proposed for ordinal data processing, generally applied to the numbers indicating the positions on the ordered list of observed data (subjects), have formulae that are simple variants of the parametric forms. And this argument may be embarrassing for advocates of the nonparametric approach. For example, the formula for Spearman's rank correlation coefficient is similar to that for the Bravais-Pearson correlation coefficient for quantitative variables (Binder 1984). Another example is the formula for the Wilcoxon-Mann-Whitney test, which is used to ascertain the equality of the amount of an attribute in two groups.The Wilcoxon-Mann-Whitney formula corresponds to that of Student's t test, used to assess the equality of two means of a quantitative variable in two groups. Such correspondence comes from the fact that by placing the subjects in a ranking and working with numbers indicating their positions, the equidistance of the categories is implicitly assumed in the response set, violating the third assumption (see above) characterizing the ordinal scales. Specifically, the third assumption is violated when the final ordinal variable, representing an attribute measured by a battery of statements, is obtained by the sums of the numbers, which are the labels of the modalities in the scale format selected or provided by respondents.

In any case, the question concerning the distances between the modalities of the response set cannot be avoided, as the ordinal variables do not have an objective standard of reference for the absence of the unit of measure. In fact, the difference between the means of an ordinal variable observed in the two groups could be relevant or irrelevant. For example, if an observed difference is always equal to two, then is it always more relevant than an observed difference equal to one? Without a unit of measurement, which determines the distance between the modalities and gives them meaning, it is impossible to answer this question with certainty. Therefore, the use of parametric techniques for the analysis of ordinal variables must be rejected a priori. Notwithstanding this, the apodictic conclusion is disputed in principle because for contesters of the parametric technique: (a) it is excessively simplifying and purist; (b) the level of measure cannot in itself dictate the process of data analysis, but the analyst should start from the data to discover their existing internal structures; and (c) often the assumptions do not precede the data (Tukey 1977; Hoaglin et al. 1983; Velleman and Wilkinson 1994). Therefore, it is not possible to admit reasons to limit the statistical procedures only to those variables that involve "arithmetic operations consistent with the scale properties of the observed quantities" (Savage 1957, p. 333), *i.e.*, coherent with the interval or ratio scales.

These objections, however, do not resolve the fundamental issue: if the data do not have a real correspondence, what one discovers is just unreal and therefore the dispute will remain in a vicious circle with no way out of it. Perhaps following a pragmatic approach, in many cases it is possible to admit the application of parametric techniques, with the awareness that the results could require further investigations. In fact, the obtained output may give some idea of the existing potential structures of relationships in the data (Amemiya 1981). Indeed, the empirical experience of data analysis shows invariance of the statements and therefore the existence of their truth/ interpretability, when the relationships among attributes are very strong. If the relationships among attributes are weak, then the results can be influenced by the variation of the values of the variables, the number of observations, and the selection of the techniques used to ascertain the existence of the multiple and multivariate relationships. Different possibilities for analysis and potential development of new strategies arise from this approach and they can lead to adaptations of existing methods as well as new methods.

3.3 The latent variables approach

Parametric statistics can be applied if it is assumed that the observed ordinal variable is the result of a crude and approximate measurement process, which evaluates a continuous underlying variable. The ordinal variable is then a kind of categorization of a latent continuous variable (Pearson 1909; Coombs 1950; Jöreskog 1990; Kampen and Swyngedouw 2000). In the approach with latent variables, the assumptions are as follows (the indices *i* and *j* concerning the subjects and the various concepts, respectively, are omitted for the sake of simplicity): (a) for each manifest random ordinal variable, X_k, there is a continuous latent random variable, $\tilde{X}_k$, with a normal mean μ_k and variance σ_k^2, i.e. $\tilde{X}_k \sim N\left(\mu_k, \sigma_k^2\right)$; (b) for each X_k, there are M_k categories so that the following relationship holds: $x_k = m \iff \tau_{m-1}^{(k)} < \tilde{x}_k < \tau_m^{(k)}$ where $m = 1, \ldots, M_k$ and the $\tau_m^{(k)} \in \mathbb{R}$, such that $\tau_0^{(k)} = -\infty < \tau_1^{(k)} < \tau_2^{(k)} < \cdots < \tau_{M_k-1}^{(k)} < \tau_{M_k}^{(k)} = +\infty$, are unknown parameters, called threshold values; (c) for each $\tilde{X}_k$, only the ordinal values are known and therefore μ_k and σ_k^2 are not identified, so that they are assumed to be equal to zero and one, respectively; (d) there exists a function h of $\tilde{x}_k$, $h(\cdot): \mathbb{R} \to \mathbb{R}$, such that $x_k = h(\tilde{x}_k) = m$ if $\tau_{m-1}^{(k)} < \tilde{x}_k < \tau_m^{(k)}$. Usually, it is assumed that $h(\cdot)$ is a many-to-one function to generate the categorization of $\tilde{x}_k$, implying that the domain of $h(\cdot)$ does not coincide with the domain of $\tilde{x}_k$ and in turn that $h^{-1}(\cdot)$ does not exist. In fact, if $h^{-1}(\cdot)$ exists, it is possible to calculate $\tilde{x}_k = h^{-1}(x_k)$ and to have a continuous variable starting from an ordinal variable.

The latent variable approach is not beyond reproach: (1) the existence of the latent random variable $\tilde{X}_k$ cannot always be proven, given that in many situations the ordinal variable is the only possible measurement that can be carried out and therefore, the assumption is unfalsifiable (Kampen and Swyngedouw 2000) according to the principle of Popper; (2) given the existence of the latent random variable $\tilde{X}_k$, there are difficulties in verifying the assumption concerning its distribution, $F\left(\tilde{x}_k\right)$; (3) the assumption of the normality of $\tilde{X}_k$ is not always verifiable and reasonable; and (4) little or nothing can be said about the robustness of the inference results in the presence of a violation of the assumptions and, therefore, this aspect is similar and related to the previous point (3). In some cases, the latent variable approach is immediate. For example, the survey data on some continuous variables such as income or saving, are often collected through classes for multiple and complex reasons (Moore et al. 2000).

It is claimed that the variables obtained from the level of agreement is an example of the inconceivability of continuum mathematics for $\tilde{X}_k$ (Kampen and Swyngedouw 2000), arguing that this is partly due to the presence of a multidimensional nature, which is difficult to eliminate from concepts used in the social sciences. A respondent may have some reasons for agreeing and other reasons for not agreeing, some of which could be additive and others may not be additive. However, this type of argument applies to many other types of measurements in the social sciences and also applies to the traditional interval scales, as obtained by Thurstone's procedures (1927a, 1927b, 1928) because the actions of the judges carried out to build up the scales are no more immune from the latter issues than those of the respondents. The variability of the threshold parameters from one subject to another can be a problem, but it is also present in the usual accepted procedures for the construction of interval scales. Some formal considerations can be found in Westermann (1983). In other cases, the approach does not have immediate justification or the ordinal variable is so rough that does not even make sense to think about the underlying continuum. The latter case, *i.e.*, the roughness of ordinal variable, was dealt with in an experiment carried out by Price and others (Agresti 1990, p. 320), proposed by Kampen and Swyngedouw (2000) as an example of the inconceivability of a continuous scale. Each pregnant mouse in the experimental group was exposed to one of five concentration levels (which belongs to a ratio scale) of some toxic substance, while each pregnant mouse in the control group was exposed to a concentration level of zero. After two days, each fetus of the pregnant mice was examined for defects and classified as 1="dead", 2="malformed", and 3="normal". A variable of this type seems to belong more to the nominal level than to the ordinal level, and the example becomes misleading. In fact, the order of the modalities is unsustainable or very weak and the continuous scale for the attribute "state or integrity of fetus" is inconceivable because the seriousness of the malformation was not measured at all, but only classified as "yes" or "no", the modality "no" being subsequently distinguished between two extreme cases, with or without (a lethal) malformation, actually implying a nominal level of measurement.

3.4 The fuzzy approach

Only recently has the fuzzy approach become a subject of interest in social applications, although several studies have been carried out since the beginning of the founding of the discipline and specifically in the measurement process (Nowakowska 1977). Applications and studies of fuzzy techniques in the social sciences are increasing (see among others, Smithson 1987; Das 2002, 2006): The ongoing development of various methods are aimed at addressing many unresolved issues, but also issues that have been solved so as to address them again from a different point of view.

Inference is the part more in the making, however, because it presents complex and arduous issues. Advancements have been made in classification techniques to build typologies, in fuzzy linear models (Tanaka 1987; Yang and Lin 2002), and in inference (Dubois et al. 2008; Viertl 2011), although the latter remains a field requiring further extension and the development of user-friendly routines. In fact, it is challenging because the fuzzification process partially overlaps the concept of probability function entailing various sensitive issues, without easy and simple solutions. Moreover, the fuzzy approach is not immediately understood and tends to appear as a black box with an input and an output, in which it is hard to inspect the inside. Data analysis involving testing tasks is problematic, as in the end one needs to know if two distributions or two means are similar or definitely different. This field seem to be very promising for future inquires and growth of our knowledge, especially for processing data coming from ordinal variables. At present, the crisp values obtained by means of defuzzification steps, are often handled through usual statistical procedures.

4 Modelling ordinal variables

A model expresses the formal structure of associations and interdependence among variables derived from theoretical beliefs about the relationships in reality, *i.e.*, as a substantive or phenomenological function, or derived from empirical operations to summarise the data set statistically in a convenient form, *i.e.*, as a descriptive or representational function. The former is useful for understanding the underlying interactive process, while the latter is useful for prediction and decision-making (Hand 2004). In both cases, the structure of association among the variables is the starting point. For ordinal variables, their measurement uses the specific property of having an order, but there are many modes of using the order relationships and each mode generates a correlation index; those most used are listed above. Therefore, the general interpretation is weakened by a lack of generality concerning the different kinds of measurements, depending on the type of definition used for the association between two variables, which determines an understanding of the role of the variables under examination, which, in turn, is not always clear as in general two different kinds of ordinal associations describe two underlying families of models (Gilula et al. 1988). For metric variables, the measurement of the associations is expressed by the Bravais-Pearson coefficient correlation and their structure is commonly represented by the correlation or variance-covariance matrix in the case of linear relationships.

In data modelling, the ordinal variables should be distinguished according to the function that they serve, as they may enter the model as explanatory/independent or response/dependent variables (Greene 2003; Hand 2004).

4.1 Modelling independent ordinal variables

As an explanatory variable, the ordinal variable enters the model, which is often linear, as a factor or regressor, with at least four possibilities.

The first possibility is direct entry of an independent ordinal variable into the model, which is not recommended in general, owing to the potential non-linearity of its scale of measurement. Tthe results cannot be interpreted in terms of quantity or impact of the independent variable on the dependent variable, but more just as a trend line. In other words, when the relationship of an ordinal independent variable with the dependent variable is the point of interest, only the sign of the relationship might be considered, admitted that at least the sign of the correlation has an indicative value or function and that the analogy with the use of polychoric correlations in the structural equation is acceptable (Jöreskog 1994).

The second possibility is the transformation of the ordinal variable into a set of binary variables, one for each modality of the scale format, with the exception of one arbitrarily selected modality defining the reference unit. The binary variables assume a value equal to one in the presence of the corresponding modality and equal to zero otherwise, therefore more often qualified by the adjective ‘dummy’ instead of binary. The method is almost natural for some variables such as the level of education, but for others it is at the very least inelegant. When the range of values is ample, the number of binary variables increases, becoming unmanageable and hence, the first or the fourth possibility becomes a necessity.

The third possibility has been proposed in a Bayesian approach (see among others, Aitchison and Silvey 1957; Farebrother 1977; Alvarez et al. 2011). The effects of the different modalities are evaluated through a Bayesian shrinkage estimator restricted to the general additive patterns.

The fourth possibility was devised by methodologists who sustain the use of the conversion of ordinal variables into quantitative variables either with the projection of the scores on a hypothetical underlying continuum or with the polychoric correlations (among others, see Coombs 1950, 1953). In fact, they generally belong to the latent variables approach. However, the conversion remains theoretically questionable leaving the issue of the non-equidistance of modalities unresolved, but in the latent models they are handled satisfactorily in the equations.

4.2 Modelling dependent ordinal variables

The invariance properties of a model concern the stability of the conclusions with respect to various categorizations of ordinal variables and some methodologists sustain that it is an essential requirement. The decision to use a scale format with three, five or seven modalities can then become a crucial issue, as can the scoring of modalities (Agresti 1990, p. 294). There are models that are sensitive to the scores assigned to the modalities of ordinal variables, such as linear regression, and models that are insensitive to them, such as the proportional likelihood ratios and latent variable approaches. In cases of sensitivity to the score values, it is convenient to perform a specific analysis to verify the extent and the stability of the results, especially when there is a possibility of assigning immediate and reasonable scores. In fact, often assigning scores to the ranks has no rational basis in the practice of the construction of ordinal variables because perhaps the same procedures are carried out for reasons relating to efficiency and established practice, so that their assignment still remains arbitrary (Wilson 1971).

Multivariate analyses can be performed with various approaches and different models, taking into account their specific conditions of applicability.

As already mentioned, within the non-parametric framework, there are several possibilities: permutation tests (Pesarin 2001; Pesarin and Salmaso 2010); loglinear models proposed by Goodman (1979), describing the associations between ordinal variables; models that are similar to the regression model, such as the proportional odds and proportional hazard models (McCullagh 1980) and logistic models (Agresti 1990). Except for permutation tests, which also assume certain conditions for their application, the other models are based on certain assumptions regarding the ordinal variable, involving specific types of associations (which are omitted here for the sake of brevity): for a synthesis see Kampen and Swyngedouw (2000).

Within the parametric framework, the ordinal dependent variables generally cannot be processed as continuous variables, even if their direct estimation could be performed to achieve important suggestions regarding heuristics and guidance, representing the horizon on which to address the prospect of subsequent inspections and assessments. In a linear model, for example, the linearity check may not serve an effective purpose, but it may serve to ascertain a potential association (Mantel 1963; McKelvey and Zavoina 1975; Amemiya 1981; Agresti 1990). A linear model, which uses a dependent equispaced variable, *describes* the state of the association between the regredend and the regressors (Amemiya 1981; Agresti 1990), with no claim of quantitatively evaluating the exact impact of the regressors on the regredend. If that assessment is necessary, then it is also necessary to perform an exact quantitative measure of the regredend.

Scores can also be introduced as parameters in the models, rather than assigning them as default and conventional numbers to the various modalities, but the number of parameters to be estimated increases dramatically, reducing parsimony and efficiency. There are ordered logit models for multiple choices (Greene 2003) to process these types of data, but they generate results that are a bit laborious to describe. Some of these models are semiparametric and belong to a borderline area between the different approaches already illustrated above with the exception of the fuzzy approach.

Within the latent variable approach, the attributes are operationalised as random variables underlying the scores observed on the manifest variables and generally the data are analysed by means of a structural equation model (Jöreskog 1973; Jöreskog and Sörbom 1979; Bollen 1989) known by the acronym LISREL, which stands for linear structural relations. These models identify groups of covarying variables and consider the common part of their covariation as a latent factor or variable, one for each group, using the polychoric correlations when the manifest variables are ordinal ones. A different approach consists of the latent class model for ordinal variables (Goodman 1974; Agresti 1990; Hagenaars and McCutcheon 2002),which assumes that the relationship between any two manifest variables is represented by the latent variables (axiom of local independence) and starts from the conditional distribution probability of data overriding the difficulties arising from the correlations used in the LISREL approach. The parameters of the latent class model are usually estimated through the EM (Expectation-

Maximisation) algorithm or through the maximum likelihood method. A critical aspect of the latent class model concerns the assumed equidistance between the ranks of the ordinal variables. Moreover, the two approaches (LISREL and latent class model) applied to same data set may lead to different conclusions.

5 Conclusions

Data analysis is an art that is performed in each case by looking for strategies, techniques, and models plus agents: their evolution feeds on the specific problems of each application and therefore, data processing cannot be carried out in an automatic mode. In the case of ordinal data, the choice of analysis plans and operating rules derived directly from theories of measurement is useful for interpreting the data and navigating the panorama of techniques, but it may sacrifice the peculiarity of each analysis, harnessing every possible exploratory path. In short, the restrictions imposed by a theory of measurement can lead to a flat analysis and perhaps to bad results as well.

Given the observations reported above, it is therefore possible to formulate a conclusion that is somewhat heretical and flexible, allowing greater freedom in the application of parametric models to some ordinal scales, even with the awareness that the results are approximate and not always robust, in accordance with the various authors previously cited, especially in the case of concepts for which it is possible to think of them with an underlying continuum and/or with a measurement through a battery having an identical response set for all items. Even Stevens (1951, p. 26) wrote: "for this 'illegal' statisticizing there can be invoked a kind of pragmatic sanction: in numerous instances it leads to fruitful results". A prudent liberalization can therefore offer useful tools for discovering structures and relationships, which might otherwise remain unexplored.

The ordinal variables can be grouped approximatively into three groups. The first group concerns the ordinal variables, which are originally continuous, such as income, and are surveyed directly through a series of class intervals defined in the response set. In this case, the thresholds of the classes are known by respondents and the ordinal data could be represented by the class marks given by the exact middles (midpoints) of the classes and the latter could be introduced directly in the models as independent and dependent variables. The second group regards the ordinal variables involving a wide range and might include metric or latent variables with unknown thresholds or discrete variables often summarising the answers obtained in a battery. Handling the ordinal data then becomes arduous and the suggested flexibility could help in the analysis, refining and adapting the model to the specific problems and fields of application for both dependent and independent variables. The third group includes ordinal variables having few modalities and it could once again include metric or latent variables with coarse categorisation (such as low, middle, and high) or discrete variables. The reduced number of categories implies the use of suitable models considering all the remarks and limits reported above and/or highlighted in the literature for both dependent and independent variables.

When possible, the goal of any scientific investigation must be directed towards improving the measurement process to substantiate the apodictic value of numbers. Therefore, a progressive emptying of the class of ordinal variables can be foreseen, with a transition to the interval scale class intervening at various levels, including the system of conventional values. However, Galileo Galilei's words of warning should always be kept in mind: "We should measure that which is measurable and make measurable that which is not".

Acknowledgements Part of this paper, specifically Section 2, has been previously published (Lalla 2015) in the volume edited by Sefano Campostrini, Giulio Ghellini, and Arjuna Tuzzi (2015), a collection of papers written by pupils and colleagues in honour of Professor Lorenzo Bernardi, who was a fine,

versatile, and brilliant academic and social statistician. Overall, he was a master, a mentor and a friend for many of us.

References

Agresti, A.: Categorical Data Analysis. John Wiley & Sons, New York (1990)

Aiello, F., Attanasio, M.: How to transform a batch of simple indicators to make up a unique one?. In: Atti della XLII Riunione Scientifica: Sessioni Plenarie e Specializzate, pp. 327-338. SIS, 9-11 giugno. University of Bari, Bari, IT (2004)

Aitchison, J., Silvey, S.D.: The Generalization of Probit Analysis to the Case of Multiple Responses. Biometrika **44** (1/2), 131-140 (1957)

Alvarez, R.M., Bailey, D., Katz, J.N.: An Empirical Bayes Approach to Estimating Ordinal Treatment Effects. Polit. Anal. **19** (1 Winter), 20-31 (2011)

Amemiya, T.: Qualitative Response Models: A survey. J. Econ. Lit. **19**(4), 1483-1538 (1981)

Atkinson, L.: The measurement statistics-controversy: Factor analysis and subinterval data. Bull. Psychon. Soc. **26**(4), 361-364 (1988)

ANES, American National Election Studies: Pre- post- election study. Survey Research Center (S473) (1964) http://www.electionstudies.org/studypages/1964prepost/int1964.txt

Babbie, E.: The Practice of Social Research, twelfth edition. Cengage Learning, Wadsworth, Belmont, CA (2010)

Bernardi, L.: Misurazione e valutazione: le difficoltà di una coppia alle prime esperienze in comune. In: Bertin, G. (ed.) Valutazione e sapere sociologico. Metodi e tecniche di gestione dei processi decisionali, pp. 69-82. Franco Angeli, Milano, IT (1995)

Bernardi, L., Capursi, V., Librizzi, L.: Measurement Awareness: The Use of Indicators between Expectations and Opportunities. In: Atti della XLII Riunione Scientifica: Sessioni Plenarie e Specializzate, pp. 315-326. SIS, 9-11 giugno. University of Bari, Bari, IT (2004)

Bernberg, R.E.: Socio-psychological factors in individual morale: I. The prediction of specific indicators. J. Soc. Psychol. **36**(1), 73-82 (1952)

Binder, A.: Restrictions on statistics imposed by method of measurement: some reality, much methodology. J. Crim. Just. **12**(5), 467-481 (1984)

Bollen, K.A.: Structural Equations with Latent Variables. John Wiley & Sons, New York (1989)

Boneau, C.A.: A note on measurement scales and statistical tests. Am. Psychol. **16**(5), 260-261 (1961)

Cacciola, S., Marradi, A.: Contributo al dibattito sulle scale Likert basato sull'analisi di interviste registrate. In: Marradi, A. (ed.) Costruire il dato. Sulle tecniche di raccolta delle informazioni nelle scienze sociali, pp. 63-102. Franco Angeli, Milano, IT (1988)

Campostrini, S., Ghellini, G., Tuzzi, A.: Con senso di misura. Riflessi statistici da alcuni allievi di Lorenzo Bernardi. CLEUP, Padova, IT (2015)

Cantril, H., Free, L.A.: Hopes and Fears for Self and Country: The Self-Anchoring Striving Scale in Cross-Cultural Research. Am. Behav. Sci. **6**(2, Supplement: Oct.), 1-30 (1962)

Conover, W.J.: Practical Nonparametric Statistics, third edition. John Wiley & Sons, New York (1999)

Coombs, C.H.: Psychological scaling without a unit of measurement. Psychol. Rev. **57**(3), 145-158 (1950)

Coombs, C.H.: Theory and Method of Social Measurement. In: Festinger, L., Katz, D. (eds.) Research Methods in the Behavioral Sciences, pp. 471-535. Dryden, New York (1953)

Crespi, I.: Use of a Scaling Technique in Surveys. J. Marketing **25**(July), 69-72 (1961)

Crespi, L.P.: Public opinion toward conscientious objectors: II. Measurement of national approval-disapproval. J. Psychol. **19**(2), 209-250 (1945a)

Crespi, L.P.: Public opinion toward conscientious objectors: III. Intensity of social rejection in stereotype and attitude. J. Psychol. **19**(2), 251-276 (1945b)

Das, S.: Quantifying fuzziness due to the scale of measurement in response systems. Fuzzy Sets Syst. **132**(3), 317-333 (2002)

Das, S.: On measuring imprecision in human response due to respondent and attribute and its utility in questionnaire design. Int. J. Uncertain. Fuzz. **14**(2), 155-173 (2006)

Domingo-Ferrer, J., Torra, V.: Extending Microaggregation Procedures using Defuzzification Methods for Categorical Variables. Proc. 1.st International IEEE symposium on intelligent systems, Verna, Sept. 44-49 (2002)

Dubois, D., Prade, H. (eds.): Fundamentals of Fuzzy Sets. Kluwer Academic Publ., Boston, MA (2000)

Dubois, D., Prade, H., Gil, M.A., Grzegorzewsky, P., Hryniewicz, O. (eds.): Soft Methods for Handling Variability and Imprecision. Springer-Verlag, Heidelberg, DE (2008)

Farebrother, R.W.: A Class of Shrinkage Estimators. J. R. Stat. Soc. B **40**(1), 47-49 (1977)

Gaito, J.: Measurement Scales and Statistics: Resurgence of an Old Misconception. Psychol. Bull. **87**(3), 564-567 (1980)

Gaito, J., Yokubynas, R.: An empirical basis for the statement that measurement scale properties are irrelevant in statistical analysis. Bull. Psychon. Soc. **24**(6), 449-450 (1986)

Gilula, Z., Krieger, A.M., Ritov, Y.: Ordinal association in contingency tables: some interpretive aspects. J. Am. Stat. Assoc. **83**(402), 540–545 (1988)

Goodman, L.A.: Simple models for the analysis of association in cross-classifications having ordered categories. J. Am. Stat. Assoc. **74**(367), 537-552 (1979)

Greene, W.H.: Econometric Analysis, fifth edition. Prentice Hall, Upper Saddle River, NJ (2003)

Guttman, L.A.: The Basis for Scalogram Analysis. In: Stouffer, S. (ed.), Measurement and Prediction, pp. 60-90. Princeton University Press, New York (1950)

Guttman, L.A.: A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration of Points. Psychometrika **33**(4), 469-506 (1968)

Hand, D.J.: Measurement theory and practice. The world through quantification. Arnold, London (2004)

Hoaglin, D.C., Mosteller, F., Tukey, J.W.: Understanding Robust and Exploratory Data Analysis. John Wiley & Sons, New York (1983)

Hofacker, C.F.: Categorical Judgment Scaling with Ordinal Assumptions. Multivar. Behav. Res. **19**(1), 91-106 (1984)

Hofmans, J., Theuns, P., Van Acker, F.: Combining quality and quantity. A psychometric evaluation of the self-anchoring scale. Qual. Quant. **43**(5), 703-716 (2009)

Jöreskog K.G.: A General Method for Estimating a Linear Structural Equation System. In: Goldberger, A.S., Duncan, O.D. (eds.) Structural Equation Models in the Social Sciences, pp. 85-112. Seminar Press, New York (1973)

Jöreskog, K.G.: New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. Qual. Quant. **24**(4), 387-404 (1990)

Jöreskog, K.G.: On the estimation of polychoric correlations and their asymptotic covariance matrix. Psychometrika **59**(3), 381-389 (1994)

Jöreskog, K.G., Sörbom, D.: Advances in Factor Analysis and Structural Equation Models. Abt books, Cambridge, MA (1979)

Juster, F.T.: Prediction and consumer buying intentions. Am. Econ. Rev. **50**(2), 604-617 (1960)

Juster, F.T.: Consumer buying intentions and purchase probability: An experiment in survey design. J. Am. Stat. Assoc. **61**(315), 658-696 (1966)

Kampen, J., Swyngedouw, M.: The Ordinal Controversy Revisited. Qual. Quant. **34**(1), 87-102 (2000)

Khurshid, A., Sahai, H.: Scales of measurement: An introduction and selected bibliography. Qual. Quant. **27**(3), 303-323 (1993)

Kilpatrick, F.P., Cantril, H.: Self-Anchoring Scaling: A Measure of Individuals' Unique Reality Worlds. J. Individual Psychol. **16**(2 Nov), 158-173 (1960)

Lalla, M.: Le scale ordinali e i relativi problemi operativi. In: Campostrini, S., Ghellini, G., Tuzzi, A. (eds.) Con senso di misura. Riflessi statistici da alcuni allievi di Lorenzo Bernardi, pp. 35-52. CLEUP, Padova, IT (2015)

Lalla, M., Facchinetti, G., Mastroleo, G.: Ordinal Scales and Fuzzy Set Systems to Measure Agreement: An Application to the Evaluation of Teaching Activity. Qual. Quant. **38**(5), 577-601 (2004)

Lalla, M., Ferrari, D., Pirotti, T.: Fuzzy Inference Systems to Analyze Ordinal Variables – The Case of Evaluating Teaching Activity. In: Proceedings of the International Conference on Fuzzy Computation Theory and Applications, pp. 25-36. SciTePress – Science and Technology Publications Digital Library, Setubal, PT (2004)

Landenna, G., Marasini, D.: Metodi statistici non parametrici. il Mulino, Bologna, IT (1990)

Likert, R.: A Technique for the Measurement of Attitudes. Archives of Psychology, monograph no. 140, pp. 1-55 (1932)

Lord, F. M.: On the Statistical Treatment of Football Members. Am. Psychol. **8**(12), 750-751 (1953)

Marradi, A.: Metodologia delle scienze sociali. il Mulino, Bologna, IT (2007)

Marradi, A.: Termometri con vincolo di ordinabilità: il «gioco della torre» consente di aggirare la tendenza alla desiderabilità sociale?. Sociologia e ricerca sociale, **57**, 49-59 (1998)

Mantel, N.: Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. J. Am. Stat. Assoc. **58**(303), 690-700 (1963)

McCullagh, P.: Regression models for ordinal data (with discussion). J. R. Stat. Soc. B **42**(2), 109-142 (1980)
McKelvey, R.D., Zavoina, W.: A Statistical Model for the Analysis of Ordinal Level Dependent Variables. J. Math. Sociol. **4**(1), 103-120 (1975)
Moore, J.C., Stinson, L.L., Welniak, E.J. Jr.: Income measurement error in surveys: a review. J. Off. Stat. **16**(4), 331-361 (2000)
Niederée, R.: There Is More to Measurement than Just Measurement: Measurement Theory, Symmetry, and Substantive Theorizing. J. Math. Psychol. **38**(4), 527-594 (1994)
Nowakowska, M.: Methodological Problem of Measurement of Fuzzy Concepts in the Social Sciences. Behav. Sci. **22**(2), 107-115 (1977)
Osgood, C.E.: The Nature of Measurement and Meaning. Psychol. Bull. **49**(3), 197-237 (1952)
Osgood, C.E., Suci, G.J., Tannenbaum, R.H.: The Measurement of Meaning. University of Illinois Press, Urbana, IL (1957)
Pearson, K.: On a new method of determining correlation between a measured character A, and a character B, of which only the percentage of cases wherein B exceeds (or falls short of) a given intensity is recorded for each grades of A. Biometrika **7**(1/2), 96-105 (1909)
Pesarin, F.: Multivariate Permutation Tests: With Application in Biostatistics. John Wiley & Sons, Chichester, UK (2001)
Pesarin, F., Salmaso, L.: Permutation Tests for Complex Data: Theory, Applications and Software. John Wiley & Sons, Chichester, UK (2010)
Prytulac, L.S.: A critique of S. S. Stevens' theory of measurement scale classification. Percept. Motor Skills **41**(1), 3-28 (1975)
Ricolfi, L.: Operazioni di ricerca e scale. Rassegna italiana di sociologia **XXVI**(2), 189-227 (1985)
Savage, I.R. Nonparametric statistics. J. Am. Stat. Assoc. **52**(279), 331-344 (1957)
Smithson, M.J.: Fuzzy Set Analysis for Behavioral and social sciences. Springer-Verlag, Heidelberg, DE (1987)
Smithson, M.J.: Fuzzy set theory and the social sciences: the scope for applications. Fuzzy Sets Syst. **26**(1), 1-21 (1988)
Siegel, S., Castellan, N.J. Jr.: Nonparametric Statistics for the Behavioral Science. McGraw-Hill, New York (1988)
Snyder, C.R., Harris, C., Anderson, J.R., Holleran, S.A., Irving, L.M., Sigmon, S.T., et al.: The will and the ways: Development and validation of an individual-differences measure of hope. J. Pers. Soc. Psychol. **60**(4), 570-585 (1991)
Stevens, S.S.: On the Theory of Scale Measurement. Science **103**(2684), 677-680 (1946)
Stevens, S.S.: Mathematics, Measurement, and Psychophysics. In: Stevens, S.S. (ed.) Handbook of Experimental Psychology, pp. 1-49. Wiley & Sons, New York (1951)
Tanaka, H.: Fuzzy data analysis by possibilistic linear models. Fuzzy Sets Syst. **24**(3), 363-375 (1987)
Thurstone, L.L.: A Law of Comparative Judgment. Psychol. Rev. **34**(4), 273-286 (1927a)
Thurstone, L.L.: The Method of Paired Comparison for Social Values. J. Abnorm. Soc. Psychol. **21**(4), 384-397 (1927b)
Thurstone, L.L.: Attitudes Can Be Measured. Am. J. Sociol. **33**(4), 529-554 (1928)
Tukey, J.W.: Exploratory Data Analysis. Addison-Wesley, Reading, MA (1977)
Van Leekwijck, W., Kerre, E.E.: Defuzzification: criteria and classification. Fuzzy Sets Syst. **108**(1), 159-178 (1999)
Velleman, P.F., Wilkinson, L.: Ordinal, Interval, and Ratio Typologies Are Misleading. Am. Statistician **47**(1), 65-72 (1993)
Viertl, R.: Statistical Methods for Fuzzy Data, John Wiley and Sons, New Delhi, India (2011)
Von Altrock, C.: Fuzzy Logic and Neurofuzzy Applications in Business and Finance. Prentice Hall PTR, Upper Saddle River, NJ (1997)
Weisberg, H.F., Rusk, J.G.: Dimensions of Candidate Evaluation. Am. Polit. Sci. Rev. **64**(4), 1167-1185 (1970)
Weick, K.E., Sutcliffe, K.M.: Managing the Unexpected. Resilient Performance in an Age of Uncertainty. John Wiley & Sons, New York (2007)
Westermann, R.: Interval-scale measurement of attitudes: Some theoretical conditions and empirical testing methods. Br. J. Math. Stat. Psychol. **36**(2), 228-239 (1983)
White, M.: Psychological technique and social problems. Southwestern Polit. Social Sci. Quarterly **7**, 58-73 (1926)
Wilson, T.P.: Critique of ordinal variables. Soc. Forces **49**(3), 432-444 (1971)

Yang, M., Lin, T.: Fuzzy least-squares linear regression analysis for fuzzy input-output data, Fuzzy Sets Syst. **126**(3), 389-399 (2002)
Yu, J.H., Albaum, G., Swenson, M.: Is a central tendency error inherent in the use of semantic differential scales in different cultures?. Int. J. Market Res. **45**(2), 213-228 (2003)
Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338-353 (1965)