Qual Quant (2015) 49:1859–1872
DOI 10.1007/s11135-014-0063-2

Methods for quantifying ordinal variables: a comparative study

Sara Casacci · Adriano Pareto

Published online: 25 July 2014

Abstract The solution to the problem of 'quantification' or *scoring*, i.e., assigning real numbers to the qualitative modalities (categories) of an ordinal variable, is of primary relevance in data analysis. The literature offers a wide variety of quantification methods, all with their pros and cons. In this work, we present a comparison between an univariate and a multivariate approach. The univariate approach allows to estimate the category values of an ordinal variable from the observed frequencies on the basis of a distributional assumption. The multivariate approach simultaneously transforms a set of observed qualitative variables into interval scales through a process called *optimal scaling*. As an example of application, we consider the Bank of Italy data coming from the "Survey on Household Income and Wealth" in order to 'quantify' a self-rating item of happiness. A simulation study to compare the performance of the two approaches is also presented.

Keywords Data analysis · Quantification · Scoring · Optimal scaling

1 Introduction

Most of the social and psychological phenomena are measured by ordered categorical or simply ordinal variables. Typically, an ordinal variable denotes the *rank order* of a particular attribute, but does not necessarily represent its actual magnitude on a substantively meaningful scale. So, the main problem facing a researcher analyzing an ordinal variable is one of incomplete information (Powers and Xie 2000).

Examples of ordinal qualitative data include the Likert scale for measuring attitudes and opinions in questionnaires (Likert 1932). In this case, respondents are asked to indicate their level of agreement with a given question or statement by choosing from a limited, predeter-

S. Casacci · A. Pareto (✉)
Italian National Institute of Statistics, Viale Oceano Pacifico 171, 00144 Rome, Italy
e-mail: pareto@istat.it

S. Casacci
e-mail: casacci@istat.it

mined set of responses (e.g., “strongly disagree”, “disagree”, “neither”, “agree”, “strongly agree”). Such response categories have a rank order, but the intervals between values cannot be presumed equal. In particular, it is ‘illegitimate’ to infer that the intensity of feeling between “strongly disagree” and “disagree” is equivalent to the intensity of feeling between “disagree” and “neither” or between other consecutive categories on the Likert scale (Jamieson 2004).

How to analyse ordinal variables has always been the subject of considerable controversy among researchers. The question is ultimately about the usage of parametric statistics (Stevens 1946, 1959; Labovitz 1972; Doering and Hubbard 1979; Knapp 1990; Kampen and Swyngedouw 2000).

The most common methods of statistical analysis, such as correlation and regression analysis, structural equation analysis, exploratory and confirmatory factor analysis, assume that the variables of interest possess continuous interval level measurement. In other words, it is assumed that the intervals between adjacent variable values are nonarbitrary. However, in case of ordinal variables, the ‘distances’ between adjacent categories are entirely arbitrary.

Because of the evident advantages and flexibility of interval methods of statistical analysis, sociologists and psychologists prefer to use them even when strictly interval measurement has not been achieved. As a consequence, they often face the practical problem of quantifying ordinal variables by assigning numerical scores to the categories (Labovitz 1970; Allen 1976). The conversion of an ordinal variable into an interval variable through a set of real numbers is based on the assumption that the ordinal variable is an observable surrogate for an unobserved or latent interval variable (Pearson 1909; Thurstone 1928). In such a case, a *scoring system* that minimizes the distortion of the latent variable should be used. Nevertheless, researchers frequently avoid the question of quantification merely by using pre-existing code values without considering that categories values are principally assigned for coding convenience or to facilitate categorical and cross-tabular analysis, not to represent appropriately the underlying variable.

Thus, current practice typically uses the so-called *integer scoring*, an equal-interval strategy of assigning rank-order numbers as category values, without recognizing that rank-order assignment implies equal inter-categories intervals (e.g., we can assign the following scores: “strongly disagree”=1, “disagree”=2, “neither”=3, “agree”=4, and “strongly agree”=5, with higher numbers representing greater levels of agreement). This practice steams from the mistaken conviction that using ordinal data as interval will not greatly distort the underlying relationships among variables. But using rank category values to represent continuous variables can have potentially critical statistical consequences and can lead to an unacceptably large degree of distortion (O’Brien 1981a).

As an alternative to this simple procedure, different quantification strategies can be used for estimating category values of ordinal variables. The literature offers a wide number of methods, such as transformation by expert ratings or estimation from item text, estimation from criterion variables, estimation from the observed frequencies based on a distributional assumption, and *optimal scaling* algorithms (Hensler and Stipak 1979; Băltătescu 2002). These techniques represent flexible and useful tools, adaptable to social researches’ needs, and their sensible application can avoid undesired measurement errors and resulting statistical bias (Herzel 1974).

The aim of this work is to compare an univariate and a multivariate approach to the problem of quantification. The univariate approach allows to estimate the category values of an ordinal variable from the observed frequencies on the basis of a distributional assumption. The multivariate approach simultaneously quantifies a set of categorical variables through an *optimal scaling* algorithm that maximizes the proportion of variance accounted for by the principal components in the quantified variables.

The paper is organized as follows. Section 2 outlines the proposed methods for category value estimation: the first is based on marginal distributions, the second requires the use of the Nonlinear Principal Component Analysis (NLPCA). In Sect. 3 we present an application of these methods to a self-rating item of happiness from the Bank of Italy "Survey on Household Income and Wealth", while a set of simulations has been carried out in Sect. 4 in order to highlight advantages and limits of the two approaches and to compare their performances. Finally, conclusions are drawn in Sect. 5.

2 Methods for category value estimation

2.1 Estimation based on a distributional assumption

The approach based on a distributional assumption is used when the researcher supposes that the ordinal variable is a crude measure of an underlying continuous variable with a particular distribution, for example uniform or normal (for the first works on this topic, see Thurstone 1925, 1927).

The procedure allows to estimate the category values of an ordinal variable on the basis of the observed frequencies. This is possible by assuming that the observed categories correspond to separate segments under the density function of a latent variable. In other words, we consider the sample proportion in each category as estimate of area under the probability density function of the underlying variable.

The expected value of the latent variable X^* conditioned on category j having been observed, $\mathrm{E}(X^*|j)$, is then the mean value for the segment of the density function corresponding to category j. Hence, the researcher needs only to compute those means and assign them as category values (Torgerson 1958).

If we assume that the underlying variable is uniformly distributed, then we simply assign to each category the average of the two endpoints of the cumulative probability function for the corresponding segment. That is, the value for the j-th rank category is $(\mathrm{P}_{j-1} + \mathrm{P}_j)/2$, where P_j is the cumulative frequency for the category j.

Nevertheless, an assumption of a normal distribution may be more often applicable than a uniform one. In fact, by the Central Limit Theorem, variables that can be thought as the sum of a number of roughly independent factors will tend to be approximately normal.

To assign values to the categories, under the assumption of normality, we take the following steps (Guilford 1954; Hensler and Stipak 1979):

1. Calculate the sample proportion or relative frequency in each category.
2. Accumulate relative frequencies across categories.
3. For each category, estimate the upper boundary of the corresponding segment of the density function by transforming the cumulative frequency to a z-score based on the standardized normal distribution. The lower boundary of each segment coincides with the upper boundary of the previous one. The lower boundary of the first category and the upper boundary of the last category extend to $-\infty$ and $+\infty$, respectively.
4. Compute the ordinates for the boundaries. The ordinates for the lower boundary of the first category and the upper boundary of the last category are zeros.
5. Quantify each category by subtracting the ordinate for the upper boundary from the ordinate for the lower boundary, and dividing the result by the proportion of cases in the category.

For example, if the first category contains 5 percent of the cases and the next contains 45 percent, the boundaries are $-\infty$, -1.64, and 0. The corresponding ordinates are 0.0000, 0.1031, and 0.3989. So the estimated value for the first category is $(0.0000-0.1031)/0.05=-2.0627$, and for the second category is $(0.1031-0.3989)/0.45=-0.6573$. These values minimize the distortion when using category values to represent scores from an underlying normal distribution (O'Brien 1981b).

Hensler and Stipak (1979) suggest that the researchers should be aware of the logical relationship among observed frequencies, inter-category distances and the shape of the underlying variable, even if they do not use formal methods for quantifying category values. For instance, if the analyst supposes that the underlying variable is uniform and the category frequencies are about equal, he can reasonably assign rank-orders numbers; but if he assumes that the latent variable is normally distributed, then equal category frequencies imply unequal inter-category intervals. Researchers conscious of these relationships can at least assign category values that are logically consistent with their assumption and the data.

2.2 Estimation based on Nonlinear Principal Component Analysis

The multivariate approach simultaneously transforms a set of observed qualitative variables into interval scales through a process called *optimal scaling* (Young 1981).

A typical example of this approach is the Nonlinear Principal Component Analysis (also referred to Categorical Principal Component Analysis or CATPCA) that optimizes the properties of the correlation matrix of the transformed variables (for a mathematical formulation, see Gifi 1990; Michailidis and Leeuw 1998). Specifically, the method maximizes the first p eigenvalues of the correlation matrix of the transformed variables, where p indicates the number of components that are chosen in the analysis. In other words, the aim of *optimal scaling* is to assign numeric values to categorical variables in order to maximize the variance accounted for by the principal components in the quantified variables (Linting et al. 2007a).

NLPCA is the nonlinear equivalent of standard Principal Component Analysis (PCA) and it can deal with all types of variables—nominal, ordinal and numeric—simultaneously. The principal components in linear PCA are weighted sums (linear combinations) of the original variables, whereas in NLPCA they are weighted sums of the quantified variables. In contrast to the linear PCA solution, the NLPCA solution is not derived from the correlation matrix, but iteratively computed from the data itself, using the *optimal scaling* process to quantify the variables according to their analysis level. Transformation of a variable treated as numeric results in its *standardization* (i.e., z-scores, as in PCA), and therefore, the order of the original categories as well as the spacing between the categories will be maintained.[1] Quantification of the categories of an ordinal variable maintains the same order as the categories (i.e., a category quantification is always less than or equal to the quantification for the category that has a higher rank number in the original data) and finds an optimal spacing between the categories. For nominal variables, both an optimal order and an optimal spacing will be obtained.

The steps for performing NLPCA are summarized below (Gifi 1990):

1. Quantify the categories of each variable in such a way that the correlations among variables can increase.
2. Update the case scores.
3. Compute a PCA of the matrix of the case scores.

[1] Note that if all variables are at a numeric analysis level, the analysis is analogous to standard PCA, since no optimal quantification is required, and the variables are merely standardized.

4. Repeat steps 1–3 until the sum of the p largest eigenvalues derived in step 3 is maximized.

NLPCA has been developed for handling categorical variables and modelling nonlinear relationships. If nonlinear relationships between variables exist, and nominal or ordinal analysis levels are specified, NLPCA leads to a higher variance accounted for than linear PCA, because it allows for nonlinear transformations.

An empirical method to assess if a quantification of ordinal variables with equal inter-category intervals is acceptable consists in comparing the variance accounted for by the first p components of a standard PCA with rank-order numbers as category values, and the one of a NLPCA with *optimal scaling*. If the proportion of variance accounted for in the NLPCA is greater, then a quantification with unequal inter-category intervals may be preferable.

The stability of the NLPCA results can be assessed by resampling techniques (Markus 1994). For example, Linting et al. (2007b) used the nonparametric balanced bootstrap to construct the 90 % confidence intervals for the eigenvalues, the component loadings, the quantified variables, and the case scores.

3 An application to real data

In order to compare the two quantification methods, we estimated the category values for a self-rating item of happiness (“Happy”) from the 2008 Bank of Italy “Survey on Household Income and Wealth”.[2] A national random sample of about 3,900 heads of household was used (Banca d’Italia 2010). The question taken into consideration is the following: “Considering all aspects of your life, how happy would you say you are?”. The item is answered on a ten-point scale that ranges from 1 (“Very unhappy”) to 10 (“Very happy”).

Figure 1 shows the bar diagram of the item. The distribution of the original variable is slightly negatively skewed with a mean of 6.89 and a standard deviation of 1.72. The frequency distribution points out that low-order categories present the lowest percentage frequencies, while the highest percentage of observations (about half) fall into categories 7 and 8 considered together. For a description of the happiness distribution in a population in terms of the parameters of the distribution of a continuous random variable, see Kalmijn (2013).

The results of the estimation from the observed frequencies under the assumption of normality are reported in Table 1. The table provides, for each category of the original variable, percentage and cumulative percentage frequencies, z-value corresponding to the area under the normal curve, ordinate of the normal curve for the value $z - \mathrm{f}(z)$—and estimated category value obtained under the assumption of normality.

Overall, there is less distinction among the quantifications of the low-order categories (2–5). This may be due to a ‘scholastic’ use of the ten-point scale.[3] So, perceptually, the ‘distance’ 2–3 ($-1.94+2.22=0.28$) is considerably smaller than the ‘distances’ 8–9 ($1.29-0.65=0.64$) and 9–10 ($1.97-1.29=0.68$).

For the quantification by NLPCA, we selected a set of variables which are normally considered on literature as the causes and correlates of happiness (Cantril 1965; Argyle 1999; Easterlin 2001; Veenhoven 2012). From these, we included in the analysis seven ordinal and quantitative variables which showed a significant empirical association with the happiness level: “Age”, “Education” (educational qualification), “Health” (state of health), “Npersons” (number of persons living in the household), “Holiday” (holiday making), “Gencond”

2 www.bancaditalia.it/statistiche/indcamp/bilfait.

3 In Italy, high-school grades are assigned on a scale from 1 (“impossible to assess”) to 10 (“excellent”).

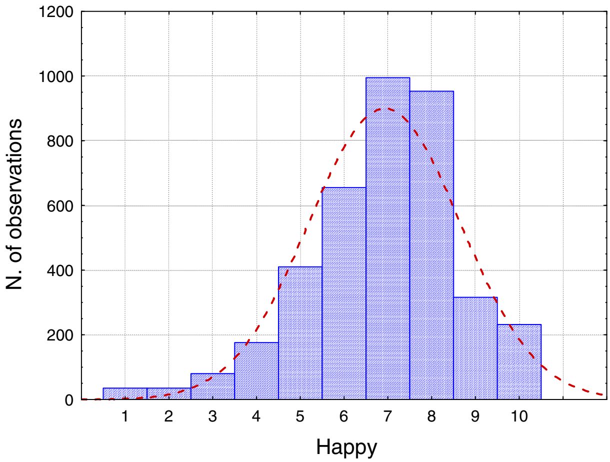

Fig. 1 Bar diagram of a self-rating item of happiness (“Happy”)

Table 1 Quantification by distributional assumption of normality

Category	frequency (%)	Cumulative frequency (%)	z	f(z)	Estimated category value
1	0.9	0.9	−2.37	0.024	−2.67
2	0.9	1.8	−2.10	0.044	−2.22
3	2.1	3.9	−1.76	0.085	−1.94
4	4.5	8.4	−1.38	0.154	−1.54
5	10.5	18.9	−0.88	0.271	−1.11
6	16.9	35.8	−0.37	0.373	−0.61
7	25.6	61.4	0.29	0.383	−0.04
8	24.5	85.9	1.08	0.223	0.65
9	8.1	94.0	1.56	0.118	1.29
10	6.0	100.0	–	–	1.97

(perceived income adequacy) and “Income” (household’s income). Each variable was given an ordinal optimal scaling level, except for the number of persons in the household and the age, for which was specified a numerical level. The chosen solution is two-dimensional, since only the first two components account for a meaningful amount of variance (in the three-dimensional solution, the eigenvalue of the third dimension is smaller than one). So, the quantification is based on the maximum number of significant components, according to the “eigenvalue greater than one criterion” (Fabrigar et al. 1999). NLPCA leads to a higher variance accounted for (56.7 %) than linear PCA (53.4 %), as it considers nonlinear relationships too (e.g., the relationship between “Income” and “Happy”).

Table 2 shows, for each category of the original variable, the absolute frequency, the estimated category value[4] and the vector coordinates[5] for the two dimensions.

[4] Note that the one-dimensional solution accounts for 39.5 % of the variance, but the results of the quantification for the target variable do not change significantly.

[5] The vector coordinates are the coordinates of the categories when they are required to be on a line, representing the variable in the principal component space.

Table 2 Quantification by NLPCA

Category	Absolute frequency	Estimated category value	Vector Coordinates	
			Dimension 1	Dimension 2
1	35	−2.52	−1.60	−0.15
2	35	−2.52	−1.60	−0.15
3	80	−2.26	−1.43	−0.13
4	176	−1.80	−1.14	−0.11
5	410	−1.34	−0.85	−0.08
6	655	−0.66	−0.42	−0.04
7	995	0.21	0.14	0.01
8	953	0.90	0.57	0.05
9	316	1.08	0.68	0.06
10	232	1.08	0.68	0.06

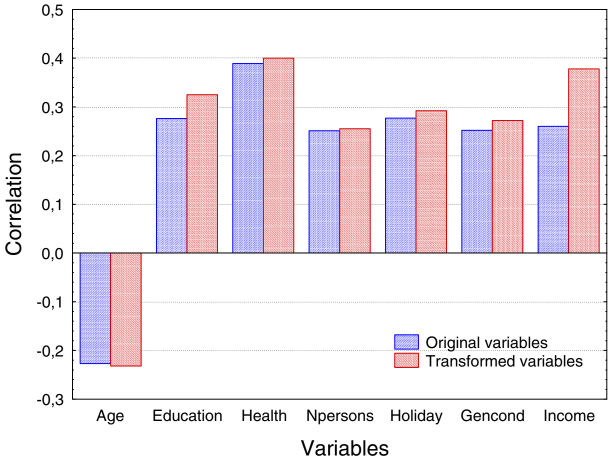

Fig. 2 Correlations between “Happy” and the other variables

The quantified values for the two highest and the two lowest categories are equal. In particular, the estimated value for categories 1 and 2 is −2.52, while the one for 9 and 10 is 1.08. This result shows that scores of 1 and 2 (9 and 10) do not differentiate between individuals and suggests that such categories could be aggregated by using a scale with a less number of response categories.

Correlations between “Happy” and the other variables, for the original and transformed values, are shown in Fig. 2.

Positive correlations occur among happiness and all the other variables, but the age. Note that the correlation between “Income” and “Happy” for the transformed values is substantially greater than the correlation for the original values, as a logarithmic relationship between the variables has been modeled.

Figure 3 allows to compare the results of the two approaches. The estimated category values are given on the y-axis versus the original category values on the x-axis. The lines connecting category quantifications indicate the variable’s transformations.

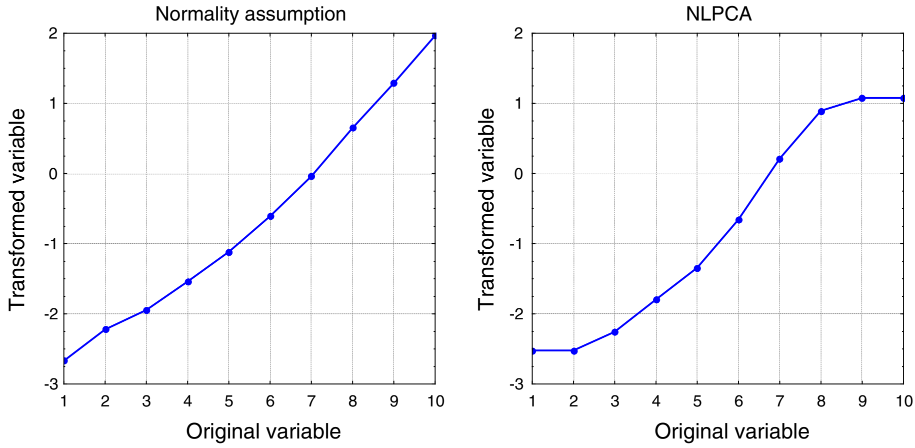

Fig. 3 Estimated category values versus original values

In both cases the original spacing between the categories is not maintained in the quantifications. Nevertheless, results of quantification by NLPCA are quite different from results of quantification under the assumption of normality.

The transformation obtained with the univariate approach presents a little distinction between the low-order categories and increases the distance between scores assigned to 8, 9 and 10. The transformation obtained with the multivariate approach, instead, shows a sigmoidal trend. It can be clearly seen that, in accordance with the variable's ordinal analysis level, both the quantifications are (non-strictly) increasing with the original category values (the transformation is monotonically non-decreasing). As noted above, some consecutive categories (1–2 and 9–10) receive the same quantification with NLPCA, as indicated by horizontal segments in the line. A possible reason for this is that persons scoring in the tied categories do not structurally differ in their patterns on the other variables, and therefore the categories cannot be distinguished from each other.

4 A simulation study

In this Section we empirically compare the performance of the univariate and multivariate approaches to the problem of quantification.

Treating the sample as the population, we repeatedly draw same-size random samples from the original sample and estimate the sampling distributions of the category values from the many samples. Three simulations are performed increasing the sample size n: simulation 1 for $n = 250$, simulation 2 for $n = 500$, simulation 3 for $n = 1000$. Each simulation consists in drawing, for every approach, 100 random samples which provide estimations of the category values for the self-rating item of happiness.

Table 3 shows the summary statistics (minimum, maximum, mean, standard deviation and skewness) for the distributions of the estimated category values.

As we can observe, in both approaches the variability of the sampling distributions decreases from the lowest categories (greatest variability) to the highest categories (smallest variability). This is mainly due to the low frequency of the categories 1–4, compared to the others. The result is most evident for NLPCA, where the small frequency of the lowest

Table 3 Estimated category values by normality assumption and by NLPCA

Category	Normality assumption					NLPCA				
	Min	Max	Mean	Std.dev.	Skews.	Min	Max	Mean	Std.dev.	Skews.
$n = 250$										
1	−2.96	−2.35	−2.71	0.17	−0.18	−4.93	−1.58	−2.93	0.77	−0.33
2	−2.52	−1.86	−2.21	0.16	−0.23	−4.87	−1.58	−2.57	0.64	−0.94
3	−2.35	−1.59	−1.94	0.15	−0.51	−3.85	−1.01	−2.08	0.51	−1.17
4	−1.89	−1.30	−1.58	0.11	−0.45	−2.93	−0.95	−1.69	0.36	−0.81
5	−1.32	−0.92	−1.13	0.08	−0.17	−1.90	−0.57	−1.30	0.27	0.13
6	−0.84	−0.46	−0.62	0.07	−0.11	−1.27	0.05	−0.65	0.25	0.02
7	−0.20	0.10	−0.05	0.06	0.02	−0.42	0.66	0.22	0.25	−0.37
8	0.49	0.81	0.63	0.07	0.39	0.36	1.06	0.77	0.16	−0.55
9	1.07	1.53	1.27	0.08	0.09	0.63	1.69	1.08	0.22	0.59
10	1.72	2.20	1.96	0.08	0.02	0.72	2.04	1.14	0.25	0.74
$n = 500$										
1	−3.17	−2.46	−2.73	0.15	−1.07	−4.91	−1.82	−2.92	0.70	−0.82
2	−2.60	−2.02	−2.24	0.13	−0.54	−3.51	−1.64	−2.48	0.40	−0.38
3	−2.18	−1.75	−1.93	0.09	−0.37	−3.02	−1.21	−2.15	0.34	0.05
4	−1.75	−1.41	−1.57	0.07	−0.16	−2.40	−1.02	−1.74	0.27	−0.14
5	−1.26	−0.98	−1.12	0.05	−0.08	−1.75	−0.85	−1.31	0.20	0.03
6	−0.73	−0.48	−0.61	0.05	0.31	−1.05	−0.17	−0.64	0.18	−0.01
7	−0.14	0.12	−0.05	0.05	0.45	−0.23	0.63	0.19	0.17	−0.25
8	0.53	0.78	0.64	0.05	0.48	0.52	1.11	0.84	0.11	−0.33
9	1.15	1.42	1.28	0.05	0.08	0.80	1.47	1.09	0.16	0.44
10	1.83	2.12	1.98	0.06	0.02	0.80	1.51	1.12	0.16	0.32
$n = 1000$										
1	−3.05	−2.50	−2.72	0.10	−0.53	−4.42	−1.94	−2.76	0.49	−0.76
2	−2.53	−1.99	−2.23	0.09	−0.44	−3.51	−1.81	−2.46	0.32	−0.70
3	−2.10	−1.74	−1.93	0.07	−0.17	−2.91	−1.60	−2.16	0.26	−0.20
4	−1.68	−1.47	−1.56	0.05	−0.21	−2.21	−1.32	−1.74	0.21	−0.19
5	−1.19	−1.03	−1.11	0.04	0.14	−1.59	−0.98	−1.33	0.14	0.08
6	−0.68	−0.53	−0.61	0.03	0.51	−0.97	−0.36	−0.66	0.12	−0.07
7	−0.10	0.05	−0.04	0.03	0.43	−0.07	0.42	0.18	0.09	−0.10
8	0.58	0.72	0.64	0.03	0.09	0.68	1.09	0.88	0.08	−0.18
9	1.19	1.40	1.29	0.04	−0.26	0.88	1.42	1.09	0.12	0.50
10	1.89	2.09	1.98	0.04	−0.02	0.88	1.42	1.10	0.12	0.51

categories does not allow to adequately model the relationships between the variables, giving rise to wide ranges of estimated values.

Overall, distributions of the estimates obtained under the assumption of normality are more concentrated than distributions of the estimates obtained by NLPCA. For example, for $n = 250$, category 1 has a standard deviation of 0.17 under the assumption of normality, against a standard deviation of 0.77 obtained using NLPCA; similarly category 10 shows a standard deviation of 0.08 (normality assumption) against 0.25 (NLPCA). In both

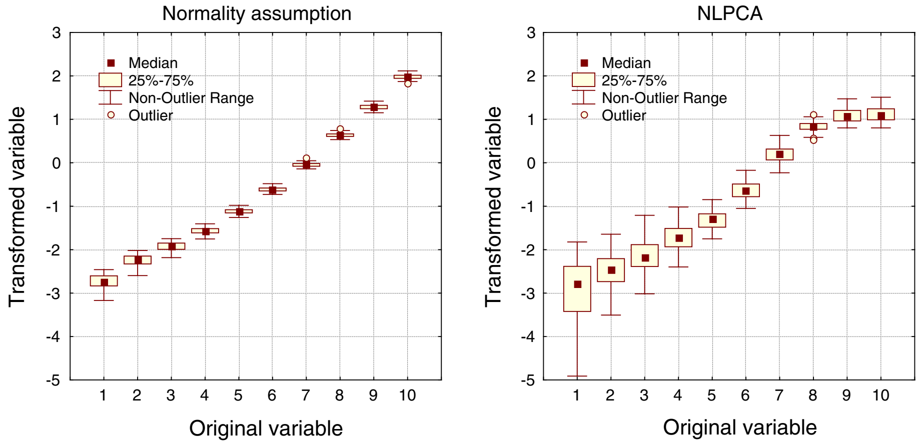

Fig. 4 Box-plot of the estimated category values ($n = 500$)

approaches, variability decreases as sample size increases. In fact, standard deviation of category 1 decreases from 0.17 ($n = 250$) to 0.10 ($n = 1000$) in the univariate approach, and from 0.77 to 0.49 in the multivariate approach, whereas standard deviation for category 10 drops from 0.08 to 0.04 in the first case, and from 0.25 to 0.12 in the second.

An interesting aspect concerns the asymmetry of the distributions which is negative for the lowest categories (1–3) in both methods. In the univariate case, this asymmetry tends to fade in the other categories (4–10), resulting in rather symmetrical distributions, while in the multivariate case, it is almost 'compensated' by a positive asymmetry in the highest categories (9–10). For a graphical representation see Fig. 4.

In order to assess the reliability of sample quantifications in different simulations, we calculated the Mahalanobis distances between the vectors of estimated values and the vector of the values of the 'population'. The choice of this kind of distance is due to the correlation between the ten categories of the variable.

Table 4 shows the main characteristics and the frequency distributions of the Mahalanobis distances. Note that the results of quantification under the assumption of normality are, on average, closer to the values of the population.

Quantifications by NLPCA are more variable: although some of them are very close to the population values, there are cases with large discrepancies. For an example, see Fig. 5 ($n = 500$), in which 6.1 % of samples provides a quantification with a distance greater than five.

Finally, we calculated correlation matrices of the estimated category values, for the two methods. Table 5 shows the results for $n = 500$. The most interesting aspect concerns the relationship between contiguous categories, which is particularly high under the assumption of normality. At a greater distance between categories corresponds a lower correlation. For example, category 1 has a correlation of 0.90 with category 2, a correlation of 0.63 with category 3, a correlation of 0.40 with category 4, etc. On the contrary, quantification by NLPCA allows to obtain adjacent category values not correlated. For example, category 1 has a correlation of 0.48 with category 2, category 2 has a correlation of 0.22 with category 3, category 3 shows a correlation of 0.10 with category 4, etc.

In this case, all the correlations are smaller in absolute value than 0.5, except for the correlation between category 9 and category 10. The strong positive correlation between the

Table 4 Mahalanobis distances from each sample to the population

Statistics and classes	$n = 250$		$n = 500$		$n = 1000$	
	Normality assumption	NLPCA	Normality assumption	NLPCA	Normality assumption	NLPCA
Absolute values						
Min	1.62	1.65	1.00	1.25	1.44	1.61
Max	5.37	6.72	4.73	5.99	5.53	9.30
Mean	3.10	3.30	3.10	3.21	3.11	3.13
Std. dev.	0.80	1.01	0.78	0.99	0.76	1.03
Skews.	0.84	0.93	0.08	0.62	0.25	2.30
Frequencies (%)						
0–1	0.0	0.0	1.0	0.0	0.0	0.0
1–2	5.0	6.3	7.1	10.1	9.0	7.0
2–3	46.3	40.0	36.4	33.3	35.0	40.0
3–4	37.5	30.0	40.4	40.4	45.0	43.0
4–5	7.5	16.3	15.2	10.1	10.0	6.0
5–6	3.8	5.0	0.0	6.1	1.0	3.0
over 6	0.0	2.5	0.0	0.0	0.0	1.0
Cumulative frequencies (%)						
0–1	0.0	0.0	1.0	0.0	0.0	0.0
1–2	5.0	6.3	8.1	10.1	9.0	7.0
2–3	51.3	46.3	44.4	43.4	44.0	47.0
3–4	88.8	76.3	84.8	83.8	89.0	90.0
4–5	96.3	92.5	100.0	93.9	99.0	96.0
5–6	100.0	97.5	100.0	100.0	100.0	99.0
over 6	100.0	100.0	100.0	100.0	100.0	100.0

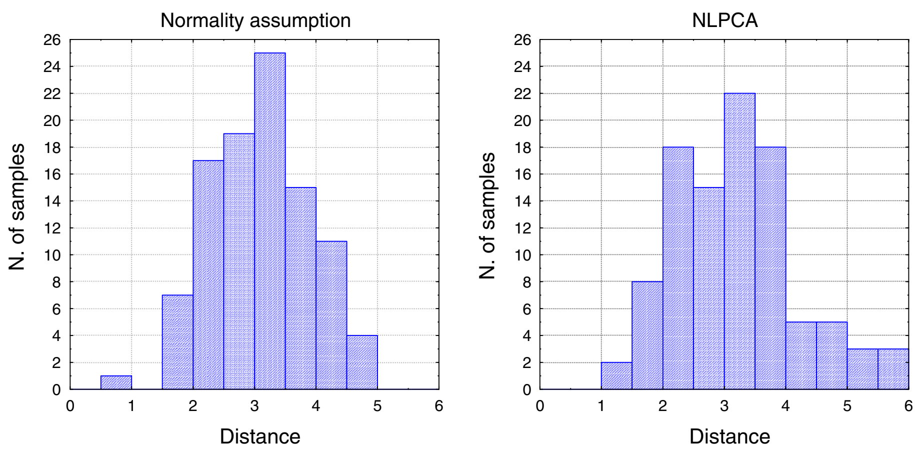

Fig. 5 Histogram of the Mahalanobis distances ($n = 500$)

Table 5 Correlation matrix of the estimated category values ($n = 500$)

Category	1	2	3	4	5	6	7	8	9	10
Normality assumption										
1	1.00	0.90	0.63	0.40	0.12	−0.03	−0.08	−0.12	−0.13	−0.16
2	0.90	1.00	0.83	0.51	0.18	−0.02	−0.07	−0.09	−0.08	−0.09
3	0.63	0.83	1.00	0.80	0.39	0.12	0.01	0.01	0.06	0.08
4	0.40	0.51	0.80	1.00	0.76	0.39	0.22	0.17	0.15	0.14
5	0.12	0.18	0.39	0.76	1.00	0.78	0.49	0.43	0.33	0.21
6	−0.03	−0.02	0.12	0.39	0.78	1.00	0.81	0.56	0.32	0.14
7	−0.08	−0.07	0.01	0.22	0.49	0.81	1.00	0.79	0.32	0.13
8	−0.12	−0.09	0.01	0.17	0.43	0.56	0.79	1.00	0.72	0.41
9	−0.13	−0.08	0.06	0.15	0.33	0.32	0.32	0.72	1.00	0.83
10	−0.16	−0.09	0.08	0.14	0.21	0.14	0.13	0.41	0.83	1.00
NLPCA										
1	1.00	0.48	0.05	−0.25	−0.23	−0.06	−0.06	0.36	−0.03	−0.02
2	0.48	1.00	0.22	−0.08	−0.24	−0.26	0.12	0.11	−0.05	0.00
3	0.05	0.22	1.00	0.10	−0.26	−0.23	−0.06	0.13	0.09	0.07
4	−0.25	−0.08	0.10	1.00	−0.08	−0.17	−0.16	−0.15	0.29	0.39
5	−0.23	−0.24	−0.26	−0.08	1.00	−0.19	−0.08	−0.03	0.23	0.24
6	−0.06	−0.26	−0.23	−0.17	−0.19	1.00	−0.18	0.00	0.05	0.00
7	−0.06	0.12	−0.06	−0.16	−0.08	−0.18	1.00	−0.35	−0.44	−0.43
8	0.36	0.11	0.13	−0.15	−0.03	0.00	−0.35	1.00	−0.37	−0.32
9	−0.03	−0.05	0.09	0.29	0.23	0.05	−0.44	−0.37	1.00	0.85
10	−0.02	0.00	0.07	0.39	0.24	0.00	−0.43	−0.32	0.85	1.00

two highest categories (0.85) is due to fact that they often receive the same quantification (idem for the two lowest categories).

Researchers should take into consideration this property, especially if they intend to include estimated category values in further analyses.

5 Conclusions

The 'quantification' of notions, such as statements of attitude or belief, assessed through the use of conventional rating scales is a central issue in data analysis, particularly in the area of sociology and psychology. Researcher cannot avoid this question simply by using positive integers since they are typically used for coding convenience and do not represent properly the underlying metric.

In this paper we compared two different approaches for estimating category values. The univariate approach requires no data beyond the marginal frequencies of the variable of interest, while the multivariate approach uses some auxiliary information from other variables observed in the same survey.

When the underlying variable can be assumed to be normally distributed, an univariate approach with estimation from the observed frequencies is the most straightforward method. Otherwise, an *optimal scaling* method, based on a set of variables associated with the latent variable of interest, may be more convenient.

Computer-intensive resampling techniques, such as the bootstrap, are a useful tool for comparing different quantification methods and finding the 'best' solution in terms of stability and robustness. In particular, the effect of small marginal frequencies should be investigated in order to determine its impact on the category quantifications (Linting et al. 2007b). If some categories lead to unstable results (sampling distributions with high variability), they could be merged with the adjacent ones, and the simulation study should be repeated to check whether the results are acceptable after recoding.

Because the estimated category values may significantly differ from a sample to another, we think that the use of auxiliary information in the quantification process is very important to obtain reliable results.

Acknowledgments The paper is the result of combined work of the authors: Sara Casacci has written Sects. 2.1, 3 and 5; Adriano Pareto has written Sects. 1, 2.2 and 4.

References

Allen, M.P.: Conventional and optimal interval scores for ordinal variables. Sociol. Method. Res. **4**, 475–494 (1976)
Argyle, M.: Causes and Correlates of Happiness. In: Kahneman, D., Diener, E., Schwarz, N. (Eds) Well-Being. Foundations of Hedonic Psychology, pp. 353–373. Russell Sage Foundation, New York (1999)
Băltătescu, S.: Problems of Transforming Scales of Life Satisfaction. Euromodule Workshop Wissensfchaft Zentrum, Berlin (2002)
Banca d'Italia: Household Income and Wealth in 2008. Supplements to the Statistical Bulletin - Sample Surveys, Year XX, No. **8** (2010)
Cantril, H.: The Pattern of Human Concerns. Rutgers University Press, New York (1965)
Doering, T.R., Hubbard, R.: Measurement and statistics: the ordinal-interval controversy and geography. Area **11**, 237–243 (1979)
Easterlin, R.A.: Income and happiness: towards a unified theory. Econ. J. **111**, 465–484 (2001)
Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., Strahan, E.J.: Evaluating the use of exploratory factor analysis in psychological research. Psychol. Methods **4**, 272–299 (1999)
Gifi, A.: Nonlinear Multivariate Analysis. Wiley, New York (1990)
Guilford, J.P.: Psychometric Methods. McGraw-Hill, New York (1954)
Hensler, C., Stipak, B.: Estimating interval scale values for survey item response categories. Am. J. Polit. Sci. **23**, 627–648 (1979)
Herzel, A.: Un criterio di quantificazione. Aspetti statistici. Metron **32**, 3–54 (1974)
Jamieson, S.: Likert scales: how to (ab)use them. Med. Educ. **38**, 1212–1218 (2004)
Kalmijn, W.: From discrete 1 to 10 towards continuous 0 to 10: the continuum approach to estimating the distribution of happiness in a nation. Soc. Indic. Res. **110**, 549–557 (2013)
Kampen, J., Swyngedouw, M.: The ordinal controversy revisited. Quality & Quantity **34**, 87–102 (2000)
Knapp, T.: Treating ordinal scales as interval scales: an attempt to resolve the controversy. Nurs. Res. **39**, 121–123 (1990)
Labovitz, S.: The assignment of numbers to rank order categories. Am. Sociol. Rev. **35**, 515–524 (1970)
Labovitz, S.: Statistical usage in sociology: sacred cows and ritual. Sociol. Method Res. **1**, 13–37 (1972)
Likert, R.: A technique for the measurement of attitudes. Arch. Psychol. **140**, 1–55 (1932)
Linting, M., Meulman, J.J., Groenen, P.J.F., Van der Kooij, A.J.: Nonlinear principal components analysis: introduction and application. Psychol. Methods **12**, 336–358 (2007a)
Linting, M., Meulman, J.J., Groenen, P.J.F., Van der Kooij, A.J.: Stability of nonlinear principal components analysis: an empirical study using the balanced bootstrap. Psychol. Methods **12**, 359–379 (2007b)
Markus, M.T.: Bootstrap Confidence Regions in Nonlinear Multivariate Analysis. DSWO Press, Leiden (1994)
Michailidis, G., De Leeuw, J.: The gifi system of descriptive multivariate analysis. Stat. Sci. **13**, 307–336 (1998)
O'Brien, R.M.: Using rank category variables to represent continuous variables: defects of common practice. Soc. Forces **59**, 1149–1162 (1981a)
O'Brien, R.M.: Reducing grouping distortion in rank category variables. Am. J. Polit. Sci. **25**, 605–616 (1981b)
Pearson, K.: On a new method of determining the correlation between a measured character A and a character B. Biometrika **7**, 96–105 (1909)

Powers, D.A., Xie, Y.: Statistical Methods for Categorical Data Analysis. Academic Press, San Diego (2000)
Stevens, S.S.: On the theory of scales of measurement. Science **103**, 677–680 (1946)
Stevens, S.S.: Measurement, psychophysics and utility. In: Churchman, C.W., Ratoosh, P. (eds.) Measurement: Definitions and Theories, pp. 18–36. Wiley, New York (1959)
Thurstone, L.L.: A method of scaling psychological and educational tests. J. Educ. Psychol. **16**, 443–451 (1925)
Thurstone, L.L.: The method of paired comparisons for social values. J. Abnorm. Soc. Psychol. **21**, 384–400 (1927)
Thurstone, L.L.: Attitudes can be measured. Am. J. Sociol. **33**, 529–554 (1928)
Torgerson, W.S.: Theory and Methods of Scaling. Wiley, New York (1958)
Veenhoven, R.: Correlates of happiness, World Database of Happiness. Erasmus University Rotterdam. http://worlddatabaseofhappiness.eur.nl/hap_cor/cor_fp.htm (2012)
Young, F.W.: Quantitative analysis of qualitative data. Psychometrika **46**, 357–388 (1981)