SIMULTANEOUS TRANSFORMATION INTO INTERVAL SCALES FOR A SET OF CATEGORICAL VARIABLES

A. Zanella, G. Cantaluppi

1. INTRODUCTION

Nowadays the statistical study of a "structure" underlying a conceptual notion, assessed through the use of conventional rating scales, so far widely spread in the area of sociology and psychology, is acquiring a major interest for managerial sciences too, particularly, in connection with notions such as business excellence, quality system effectiveness, customer satisfaction, etc.; in the former context the recourse to a linear structural model with latent variables (LISREL) has become a current methodology.

In such a model the so-called manifest variables, which link a concept to empirical manifestations, play an essential role. Now the usual statistical analysis of a LISREL model assumes that the manifest variables, related to a certain concept, are measured on metric scales, that is, on an interval or on a ratio scale, while often the values of the above variables, typically obtained from the answers to a questionnaire by a group of respondents, are in fact expressed on conventional rating scales, which are at most ordinal scales only.

The former method is often used in practice for assigning numbers to the qualitative modalities of ordered categorical variables. The problem of associating meaningful numbers to objects or events on the basis of qualitative observations of attributes was extensively dealt with by Amato Herzel (1974 a, b), who called the process "quantification" and underlined the importance that it can at times have in statistical inference for a better description and understanding of a phenomenon under study.

The paper represents a contribution to the development of the National Research Project "*Statistical methods and models for Quality and Customer assessment: objective and subjective aspects in quality optimization of goods and services*" partially financed by Education, University and Research Ministry, Grant n° 2001138489, 2001.

Angelo Zanella is the author of the methodological part of the paper, especially of sections 1, 2, 3, 4.1; Gabriele Cantaluppi developed all the aspects of the simulation and numerical applications presented in the paper and of the corresponding tables and graphs and in particular contributed to section 4.2. All the work was discussed and agreed by both authors. A summary of the paper was also presented at the 54th Session of the International Statistical Institute, Berlin, August 2003.

Herzel's point of view can be summarized as follows. In an empirical investigation based on observations of ordered categorical variables, it may result that the reading of the aimed at objective would be more direct and easier if the variables were quantitative: for example if we are interested in assessing whether two aspects of the experimental units under study are connected, it would be easy to calculate the correlation coefficient to get a first idea of the state of things.

In general with respect to a given inference objective, which would require to obtain the numerical value of a particular statistical index in the case of quantitative variables, Herzel suggests to look for a transformation of the qualitative modalities into real values, which will be treated as belonging to an interval scale, so that, for the observed empirical frequencies, the index of interest assumes an extreme value (maximum or minimum). In the example of the correlation coefficient mentioned above, we should find out the transformation of each of the two categorical variables which allows one to obtain the maximum correlation coefficient between the numerical values corresponding to the original ordered categories, in particular under the conditions that their variances are 1 and the ordering is preserved. Herzel's proposal might be seen in the light of the axiomatic theory of measurement and the corresponding representation theorem which states how ordinal scales can be rationally constructed so that the properties of ordered categories can be faithfully represented by real numbers, e.g. see Krantz *et al.* (1971), Ch. 1.

If we can assume that the qualitative modalities of an attribute characterizing the elements of a finite set A under study is totally ordered, the representation theorem indicates how to construct a map φ: A $\rightarrow \Re$ from A to the real line $\Re$, which allows assigning numbers to the ordered qualitative modalities. However the theorem shows that the "faithful" assignment of numbers through the function $\varphi(\cdot)$ is indeterminate, since an ordinal scale is obtained even if the numbers $\varphi(\cdot)$ are transformed by means of a strictly increasing function $f(\cdot)$; then Herzel's proposal could be interpreted as aiming to make use of the said indeterminacy to obtain a possible transformation $f(\cdot)$ which most emphasizes the underlying trait which is of particular interest for the case under study.

The present paper concerns ordinal scales for a set of K qualitative ordered categorical variables (attributes). It is implicitly assumed that the I modalities of each categorical variable are totally ordered so that the representation theorem allows us to consider the corresponding conventional integer values $i = 1, 2, ..., I$ as a possible version of an ordinal scale. However we do not follow Herzel argument to obtain a data transformation into an interval scale but we complete the axiomatic framework, which justifies the original ordinal scales, by Thurstone's psychometric axiom which, for a psychological categorical variable, describing attitude, preference, etc., postulates the existence of a latent, that is not directly observable random variable. As we shall later explain in detail, this allows one to assign a value on an interval scale to each modality $i = 1, 2, ..., I$ of a categorical variable by transforming the cumulative probability (or relative frequency) of obtaining results not larger than i by means of the inverse of the distribution function of the corresponding latent variable. More precisely this contribution examines some implications and an extension of the method proposed by Jones (1986), see also Bock and Jones (1968), referring to the former basic structure, to simultaneously transform a set of observed categorical variables into interval scales, under the assumption that there exists a normal latent random variable corresponding to each of the categorical variables.

The paper in particular shows that the method does not allow to have a direct validation of a specific type of probability distribution assumed for the latent variables.

In this note we present the model which we believe to be at the basis of Jones' method – to which so far, if we are not wrong, little attention has been paid, however see Zanella *et al.* (2000) –, we discuss its implications and propose an extension to other families of latent random variables, besides the Normal one, when their probability distributions can be reduced to the location-scale type.

2. JONES' MODEL AND ITS EXTENSION

2.1. *Transforming the conventional scores into values on an interval scale*

Suppose we observe a K-dimensional random categorical variable $\mathbf{X} = (X_1, X_2, ..., X_K)'$, whose components assume the same number of ordered modalities I, denoted by the conventional integer values $x_{ki} = 1, 2, ..., I$, $k = 1, 2, ..., K$, for simplicity.

Let $P(X_k = i) = p_{ki}$, with $\sum_{i=1}^{I} p_{ki} = 1$, $\forall k$, be the corresponding marginal probabilities and put

$$F_k(i) = \sum_{\{j \le i\}} p_{kj} \tag{1}$$

to express the cumulative probability that we can observe a conventional value x_k for X_k not larger than i for X_k.

Furthermore assume that to each categorical variable X_k there corresponds an unobservable latent variable Z_k, which is represented on an interval or a ratio scale, with a continuous distribution function of the type:

$$P[(Z_k - \mu_k)/\sigma_k \le (\xi_k - \mu_k)/\sigma_k] \equiv \Psi_k[((\xi_k - \mu_k)/\sigma_k),\boldsymbol{\alpha}] = \Psi_k(\zeta_k,\boldsymbol{\alpha}), \tag{2}$$

where μ_k and $\sigma_k > 0$ are respectively a location and a scale parameter, which may depend on some other real positive parameters, $\alpha_1, ..., \alpha_K$, summarized by vector $\boldsymbol{\alpha}$: $\mu_k = \mu_k(\boldsymbol{\alpha})$, $\sigma_k = \sigma_k(\boldsymbol{\alpha})$, that is for any but fixed $\boldsymbol{\alpha}$ we assume a location scale family of distributions; when no ambiguity arises we shall preserve the simplified notations μ_k, σ_k. In (2):

$$\zeta_k = (\xi_k - \mu_k)/\sigma_k \tag{3}$$

$k = 1, 2, ..., K$, are real variables describing the possible values on the right side, which for a fixed $\boldsymbol{\alpha}$ are invariant for any scale-location change corresponding to a linear transformation of the type:

$$\tilde{\xi}_k = a\xi_k + b \tag{4}$$

with b and $a > 0$ real values, where a expresses the change of unit measurement of ξ_k and b a displacement of b new units of the origin of measurements. In fact note that the new measurement system implies that:

$$\tilde{\mu}_k = a\,\mu_k + b,\ \tilde{\sigma}_k = a\,\sigma_k, \tag{5}$$

when we assume that μ_k, σ_k are expressed in the same units as ξ_k and, in addition, the positive quantities σ_k do not depend on the origin of measurements. Thus according to (3) it follows from (4) that:

$$\left[\frac{(\tilde{\xi}_k - b)}{a} - \mu_k\right]/\sigma_k = [\tilde{\xi}_k - (a\mu_k + b)]/a\sigma_k = (\tilde{\xi}_k - \tilde{\mu}_k)/\tilde{\sigma}_k = (\xi_k - \mu_k)/\sigma_k,$$

in consequence of (5), which proves the invariance stated above. Correspondingly for a fixed $\boldsymbol{\alpha}$ we can identify $\Psi_k(\zeta_k,\boldsymbol{\alpha})$ in (2) with the *standard element* of the family defined by putting $\tilde{\mu}_k = 0$, $\tilde{\sigma}_k = 1$, i.e. $b = -\mu_k/\sigma_k$, $a = 1/\sigma_k$ in (5).

By inversion of (2) and with regard to (1) by putting the second member equal to $F_k(i)$ define:

$$\{[\xi_{ki}(\boldsymbol{\alpha}) - \mu_k(\boldsymbol{\alpha})]/\sigma_k(\boldsymbol{\alpha})\} = \Psi_k^{-1}[F_k(i), \boldsymbol{\alpha}] = \zeta_{ki}, \tag{6}$$

when needed with a suitable conventional or approximate definition of ζ_{ki} for $F_k(i) = 0$ or 1. Later on we shall examine the case when $\xi_{ki} = \xi_i$, $\forall k$, that is there exists a unique value ξ_i corresponding to each category i regardless of the random variable X_k. In this case it seems natural to substitute ξ_i, measured on a metric scale, for the conventional score i; with this assumption (6) becomes

$$\{[\xi_i(\boldsymbol{\alpha}) - \mu_k(\boldsymbol{\alpha})]/\sigma_k(\boldsymbol{\alpha})\} = \Psi^{-1}[F_k(i), \boldsymbol{\alpha}] = \zeta_{ki}, \tag{7}$$

$k = 1, 2,..., K$; $i = 1, 2,..., I$.

The following Fig. 1 illustrates for the normal case the interpolation of the cumulative probabilities $F_k(i)$ – which can be estimated through the observed cumulative proportions – obtained from the distribution function of the standard element of the assumed family (2) of continuous latent distributions and the corresponding new scores ξ_i representing the modalities i, $i = 1, 2, ..., I$, on an interval scale given by the inversion formula (7).

For concreteness we shall refer to the real case presented by Jones (1986), who examines the conventional preference scores, ranging from 1 to 9, which we might as well interpret as a degree of customer satisfaction, obtained from a sam-

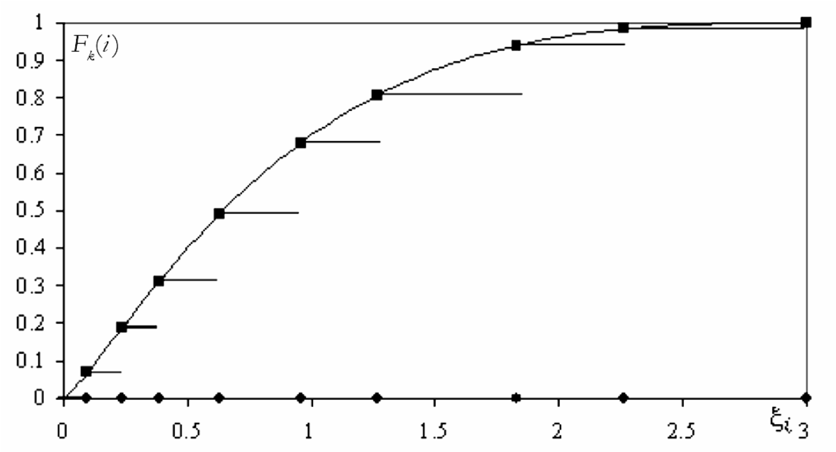


Figure 1 – Transforming the conventional scores i into values ξ_i on an interval scale by inverting the cumulative distribution function of the unitary element of the latent family of distributions, ref. (8) below for $\alpha_k = 1.18$.

ple of 255 army enlisted men with respect to 12 food items, or categories, $k = 1, 2, ..., 12$, $i = 1, 2, ..., 9$, and treated by the author assuming 12 underlying normal distributions. Later on we shall simulate some data corresponding to Jones', see § 4. Theorem 1 shows that if Jones' model (7) is valid, the unique values ξ_i can be defined as $\xi_i = \sum_{k=1}^{K} \zeta_{ki} / K = \bar{\zeta}_i$, $i = 1, 2, ..., I$, which are the arithmetic means respectively obtained for each modality i by taking the average of the theoretical values ζ_{ki}, $k = 1, 2, ..., K$, over the K categorical variables. If we should establish a diet by choosing some of the 12 food items, with scores, say, $i_1, i_2, ..., i_s \in \{1, 2, ..., I\}$ and we could assume that customer satisfaction scores are additive, *the scale unification leading to an interval scale* would allow us to represent the overall customer satisfaction of a subject as the sum $\xi_{i1} + \xi_{i2} + ... + \xi_{is}$ of the transformed scores assigned to the s food items of the diet. Note that this advantage of scale unification would likewise ensue in general when we are concerned with the evaluation of a subset of K attributes of a good or a service.

We can finally note that, if Jones' model (7) holds, the unified scale would allow us to study the possible dependence between any pair of categorical variables X_i, $X_{i'}$, $i \neq i'$ by having resort to the correlation coefficient ρ obtained from the corresponding scores transformed to an interval scale by using the model.

Table 1 illustrates graphically the transformation of conventional scores i, $i = 1, 2, ..., 9$, with reference to Jones' example, into values $\xi_i = \bar{\zeta}_i$ on an interval scale which hold for any categorical variables X_k, $k = 1, 2, ..., K$, if Jones' model is true.

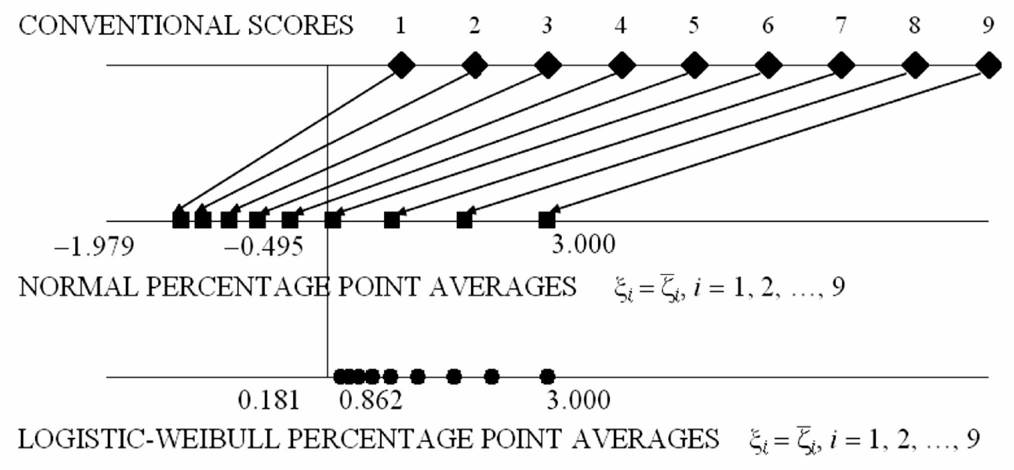
2.2. *Generalization of Jones' model*

Jones (1986) treats the case when $\Psi_k(\zeta)$ is the distribution function of the Normal standard deviate and α is not present.

TABLE 1

Percentage point averages related to class i, i = 1,2,..., 9, for the latent variables referring to the Normal and Logistic-Weibull distribution functions.

	i	1	2	3	4	5	6	7	8	9
Normal	$\bar{\zeta}_i = \xi_i$	−1.979	−1.683	−1.335	−0.936	−0.495	0.089	0.886	1.876	3.000
Logistic-Weibull	$\bar{\zeta}_i = \xi_i$	0.181	0.300	0.428	0.615	0.862	1.225	1.725	2.236	3.000



To illustrate the generalization here proposed we consider a Logistic Weibull family of distributions, see Zanella (1998), (1999), with the following definition:

$$\Psi_k = 1 - \exp\left\{ -\left[\frac{2}{\alpha_k} ln\left(\frac{3 + (\xi_{ki} - \mu_k)/\sigma_k}{3 - (\xi_{ki} - \mu_k)/\sigma_k} \right) \right]^{\alpha_k} \right\} = F_k(i), \tag{8}$$

$\alpha_k > 0$, for given μ_k, $\sigma_k > 0$, with conventional value 0 if $\zeta = (\xi_{ki} - \mu_k)/\sigma_k$ is such that $\zeta \le 0$, conventional value 1 if $3 \le \zeta$, which are based on the limits for $\zeta \to 0^+$, $\zeta \to 3^-$. We shall examine the case $\mu_k = 0$, $k = 1, 2, ..., K$, which according to (7) leads to:

$$\frac{\xi_{ki}}{\sigma_k} = \zeta_{ki} = 3 \cdot \frac{1 - \exp\{-\alpha_k/2[-\ln(1 - F_k(i))]^{1/\alpha_k}\}}{1 + \exp\{-\alpha_k/2[-\ln(1 - F_k(i))]^{1/\alpha_k}\}}, \tag{9}$$

with conventional values of $\zeta = \xi_{ki}/\sigma_k \notin (0, 3)$: 0 for $\zeta < 0$, 3 for $\zeta \ge 3$.

Fig. 2 and Fig. 3 respectively give plots of relationship (8) for $\mu_k = 0$, $\sigma_k = 1$, $\alpha_k = 1.18, 2.72$ and the corresponding probability densities.

For other values of the location parameter μ_k the graphs are the same but with the origin placed at μ_k.

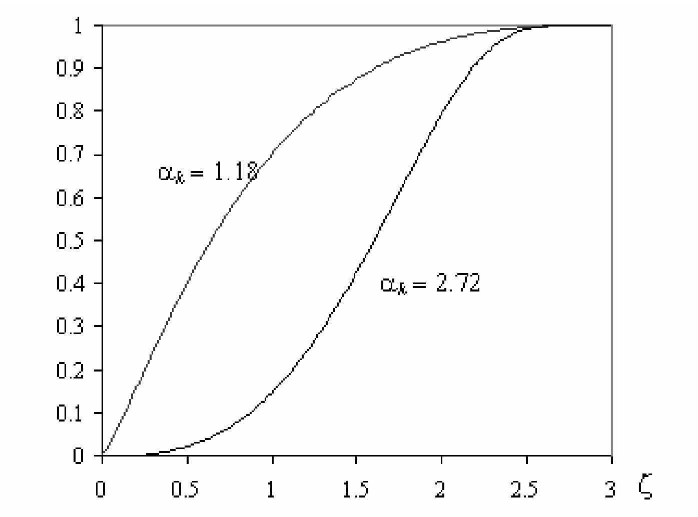


Figure 2 – Logistic-Weibull probability distribution function, $\mu_k = 0$, $\sigma_k = 1$.

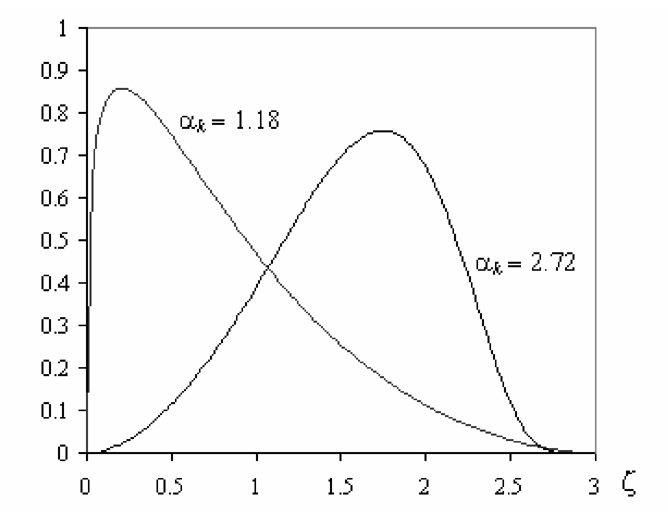


Figure 3 – Logistic-Weibull probability density function, $\mu_k = 0$, $\sigma_k = 1$.

3. NECESSARY AND SUFFICIENT CONDITIONS SO THAT JONES' MODEL HOLDS

This section presents the main results of the paper; a preliminary version of Theorem 1 is given in Zanella and Cerri (2000).

3.1. *Necessary conditions to ensure Jones' model*

Theorem 1. (Necessary conditions for Jones' model)

With reference to the family of distribution functions (2) suppose that: a) for fixed values α_k, the values ζ_{ki} are percentage points – assumed to be distinct and thus by hypothesis increasing for $i = 1, 2, ..., I$ – of the standard element of the distribution functions relative to each latent variable Z_k, $k = 1, 2, ..., K$, which are not functions of a single random variable Z, and that b) there exist some values of the location and scale parameters σ_k^*, μ_k^*, say, and some unifying transformed scores ξ_i, $i = 1, 2, ..., I$, so that Jones' model (7) holds, which means that the following systems of linear equations are satisfied:

$$\begin{cases} \xi_1/\sigma_k^* = \mu_k^*/\sigma_k^* + \zeta_{k1} \\ \xi_2/\sigma_k^* = \mu_k^*/\sigma_k^* + \zeta_{k2} \\ \cdots \\ \xi_I/\sigma_k^* = \mu_k^*/\sigma_k^* + \zeta_{kI} \end{cases}, \quad (10)$$

$k = 1, 2, ..., K$.

i. It follows that it must be:

$$\xi_i = (\overline{\zeta}_i + \overline{\mu}) / \overline{\sigma}, \quad i = 1, 2, ..., I, \quad (11)$$

where $\overline{\zeta}_i = \sum_{k=1}^{K} \zeta_{ki}/K$, $i = 1, 2, ..., I$,

i.e. the unifying scores must coincide with a linear function of the average $\overline{\zeta}_i$ of the percentage points ζ_{ki} over the K latent variables, $\overline{\sigma}, \overline{\mu}$ being the arbitrary functions of σ_k^*, μ_k^* defined below in (16).

ii. It also ensues that the averages $\overline{\zeta}_i$ must also satisfy the K linear equations:

$$\overline{\zeta}_i / \sigma_k^* = (\mu_k^*/\sigma_k^*) + \zeta_{ki} \quad (12)$$

$i = 1, 2, ..., I$ and $k = 1, 2, ..., K$, that is, from a geometric point of view, the points $(\zeta_{ki}, \overline{\zeta}_i)$, $i = 1, 2, ..., I$, must align on K straight lines if (and only if):

$$\sum_{k=1}^{K} \frac{1}{K\sigma_k^*} = 1, \qquad \sum_{k=1}^{K} \frac{\mu_k^*}{K\sigma_k^*} = 0, \quad (13)$$

with $\sigma_k^* > 0$.

iii. Suppose that the latent variables Z_k, $k = 1, 2, ..., K$, have distribution functions of the location-scale type defined by (2) and that, for some fixed values α_k, the necessary relationships (11), (12), (13) required by the unifying model are satisfied. Now define a location-scale transformation $\tilde{\mu}_k = \tilde{\sigma}_k \mu_k^* + \tilde{\mu}$, $\tilde{\sigma}_k = \tilde{\sigma}\sigma_k^*$, with $\tilde{\sigma} > 0$ and $\tilde{\mu}$ some real values. It is shown that the transformed random variables $\tilde{Z}_k = \tilde{\sigma} Z_k + \tilde{\mu}$, $k = 1, 2, ..., K$, can still satisfy Jones' model and define some other unifying scores as:

$$\tilde{\xi}_i = \tilde{\sigma}\overline{\zeta}_i + \tilde{\mu}, \tag{14}$$

$i = 1, 2, ..., I$, which thus are defined except for a location-scale transformation.

Proof. i. Consider the i^{th} equation of each of the K systems (10) obtaining:

$$\left(\frac{\xi_i}{\sigma_k^*} - \frac{\mu_k^*}{\sigma_k^*}\right) = \zeta_{ki},\ i = 1, 2, ..., I. \tag{15}$$

If we take the average of both sides on varying k we get:

$$\left(\sum_{k=1}^{K} \frac{1}{K\sigma_k^*}\right)\xi_i - \left(\sum_{k=1}^{K} \frac{\mu_k^*}{K\sigma_k^*}\right) = \overline{\zeta}_i$$

and defining

$$\overline{\sigma} = \sum_{k=1}^{K} \frac{1}{K\sigma_k^*}, \qquad \overline{\mu} = \sum_{k=1}^{K} \frac{\mu_k^*}{K\sigma_k^*}, \tag{16}$$

$$\overline{\sigma}\xi_i - \overline{\mu} = \overline{\zeta}_i,$$

from which it follows:

$$\xi_i = (\overline{\zeta}_i + \overline{\mu})/\overline{\sigma}.$$

We remark that, by the way, if we now insert this expression into (15) and take once more the average of both sides on varying k we obtain:

$$\left(\sum_{k=1}^{K} \frac{1}{K\overline{\sigma}\sigma_k^*}\right)\overline{\zeta}_i - \left(\sum_{k=1}^{K} \frac{\mu_k^*}{K\sigma_k^*} - \frac{\overline{\mu}}{\overline{\sigma}} \sum_{k=1}^{K} \frac{1}{K\sigma_k^*}\right) = \overline{\zeta}_i, \quad i = 1, 2, ..., I,$$

which is identically satisfied since, in force of (16), the first expression in round brackets has value 1, the second expression has value 0.

ii. Note that relationships (12) are obtained by those defining systems (10) when we take into account the necessary condition (11), that for $\overline{\sigma} = 1$, $\overline{\mu} = 0$, i.e. for (13), gives $\xi_i = \overline{\zeta}_i$. Let us consider the i-th equation of any of the K systems (10); with vector-matrix notation we have the I systems:

$$\begin{bmatrix} \left(\frac{1}{K\sigma_1^*}-1\right) & \frac{1}{K\sigma_1^*} & \cdots & \frac{1}{K\sigma_1^*} \\ \frac{1}{K\sigma_2^*} & \left(\frac{1}{K\sigma_2^*}-1\right) & \cdots & \frac{1}{K\sigma_2^*} \\ \vdots & \vdots & & \vdots \\ \frac{1}{K\sigma_K^*} & \frac{1}{K\sigma_K^*} & \cdots & \left(\frac{1}{K\sigma_K^*}-1\right) \end{bmatrix} \cdot \begin{bmatrix} \zeta_{1i} \\ \zeta_{2i} \\ \vdots \\ \zeta_{Ki} \end{bmatrix} = \begin{bmatrix} \frac{\mu_1^*}{\sigma_1^*} \\ \frac{\mu_2^*}{\sigma_2^*} \\ \vdots \\ \frac{\mu_K^*}{\sigma_K^*} \end{bmatrix} \qquad (17)$$

that is

$$\mathbf{M}_K \cdot \boldsymbol{\zeta}_i = \boldsymbol{\mu},$$

$i = 1, 2, ..., I$, where $\mathbf{M}_K$ is the $K \times K$ square matrix on the first side, $\boldsymbol{\zeta}_i$, $\boldsymbol{\mu}$ are $K \times 1$ column vectors so that their transpose are $\boldsymbol{\zeta}'_i = [\zeta_{1i}, \zeta_{2i}, ..., \zeta_{Ki}]$, $\boldsymbol{\mu}' = [\mu_1^*/\sigma_1^*, \mu_2^*/\sigma_2^*, ..., \mu_K^*/\sigma_K^*]$, and note that, in particular, $\boldsymbol{\mu}$ is a vector whose components do not depend on the index i. If $\mathbf{M}_K$ should be non singular and, thus, admitting an inverse $\mathbf{M}_K^{-1}$, we should have:

$$\boldsymbol{\zeta}_i = \mathbf{M}_K^{-1} \boldsymbol{\mu}$$

for $i = 1, 2, ..., I$, that is the percentage points ζ_{ki} would be the same *whichever* is $i = 1, 2, ..., I$, which implies $\zeta_{ki} = \zeta_k$, $k = 1, 2, ..., K$, and $\overline{\zeta}_i = \overline{\zeta}$, $\forall i$. Since ζ_{ki} are arbitrary for varying i, this would lead, according to (1), to

$$P[(Z_k - \mu_k^*)/\sigma_k^* = Z_k^* \leq (\overline{\zeta} - \mu_k^*)/\sigma_k^* = \zeta_k] = \sum\nolimits_{\{j \leq i\}} p_{kj}, \qquad \forall i, \qquad (18)$$

for any ζ_k and $k = 1, 2, ..., K$, which would mean that the random variables $(Z_k - \mu_k^*)/\sigma_k^* = Z_k^*$, $k = 1, 2, ..., K$, could be reduced to a single random variable Z, $\forall k$. This is owing to the fact that being relationships (18) simultaneously satisfied for $k = 1, 2, ..., K$ implies that the probabilities of all Z_k^* are ruled by the same distribution function and thus all of them have the same probability of belonging to a given arbitrary interval. It ensues that there they can differ only on a set of probability zero.

Now the former case is excluded by hypothesis and the equations of a system (17) must be compatible but with the rank of matrix $\mathbf{M}_K$ less than K. For obtaining this condition in explicit form notice that matrix $\mathbf{M}_K$ has the following structure:

$$\mathbf{M}_K = \begin{bmatrix} \frac{1}{K\sigma_1^*} & 0 & \cdots & 0 \\ 0 & \frac{1}{K\sigma_2^*} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \frac{1}{K\sigma_K^*} \end{bmatrix} \cdot \left\{ \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} - \begin{bmatrix} K\sigma_1^* & 0 & \cdots & 0 \\ 0 & K\sigma_2^* & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & K\sigma_K^* \end{bmatrix} \right\} =$$

$$= \mathrm{Diag}(1/K\sigma_1^*, 1/K\sigma_2^*, ..., 1/K\sigma_K^*) \cdot [\mathbf{uu}' - \mathrm{Diag}(K\sigma_1^*, K\sigma_2^*, ..., K\sigma_K^*)] \qquad (19)$$

where **u** is a $K \times 1$ vector with elements all equal to 1.

The matrix $\mathbf{M}_K$ is non-singular and it admits an inverse if the same holds for the matrix within brackets, say $\mathbf{M}^*$, on the right side of the second expression (19). As it is known by matrix algebra (see for instance Bodewig (1959) p. 39), the inverse of $\mathbf{M}^*$ is

$$\mathbf{M}^{*-1} = -\begin{bmatrix} \frac{1}{K\sigma_1^*} & 0 & \cdots & 0 \\ 0 & \frac{1}{K\sigma_2^*} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \frac{1}{K\sigma_K^*} \end{bmatrix} - c \begin{bmatrix} \frac{1}{K\sigma_1^*} \\ \frac{1}{K\sigma_2^*} \\ \cdots \\ \frac{1}{K\sigma_K^*} \end{bmatrix} [1/K\sigma_1^*, 1/K\sigma_2^*, ..., 1/K\sigma_K^*] \qquad (20)$$

with

$$c = 1 / \left(1 - \mathbf{u}' \begin{bmatrix} \frac{1}{K\sigma_1^*} & 0 & \cdots & 0 \\ 0 & \frac{1}{K\sigma_2^*} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \frac{1}{K\sigma_K^*} \end{bmatrix} \mathbf{u} \right) = 1 / \left(1 - \sum_{k=1}^{K} \frac{1}{K\sigma_k^*} \right)$$

which shows that $\mathbf{M}^*$ is singular – and thus its rank is less than K – if and only if:

$$\sum_{k=1}^{K} \frac{1}{K\sigma_k^*} = 1. \qquad (21)$$

If (21) is satisfied and thus the rank of **M*** and of **M** is less than K, in order to have a solution for the systems (17) there must exist values $\zeta_{1i}, \zeta_{2i}, ..., \zeta_{Ki}$, $i = 1, 2, ..., I$, so that for a given i they can satisfy all equations (17). If we consider a set of such values and sum up all K equations (17) we find:

$$\left(\sum_{k=1}^{K}\frac{1}{K\sigma_k}-1\right)\zeta_{1i}+\left(\sum_{k=1}^{K}\frac{1}{K\sigma_k}-1\right)\zeta_{2i}+\ldots+\left(\sum_{k=1}^{K}\frac{1}{K\sigma_k}-1\right)\zeta_{Ki}=\sum_{k=1}^{K}\frac{\mu_k^*}{\sigma_k^*}, \qquad (22)$$

which in force of (21) implies

$$0=\sum_{k=1}^{K}\frac{\mu_k^*}{\sigma_k^*} \qquad (23)$$

from which dividing by K we finally see that condition (13) is as well necessary as sufficient.

In conclusion the property (12) of comma *ii.* has been proved.

iii. It is assumed that according to (12):

$$P[(Z_k-\mu_k^*)/\sigma_k^* \le (\overline{\zeta}_i-\mu_k^*)/\sigma_k^*=\zeta_{ki}]=\Psi_k[F_k(i),\alpha], \qquad (24)$$

$i = 1, 2, ..., I$, $k = 1, 2, ..., K$. Remember that in the statement of point *iii.* under consideration some new variables $\tilde{Z}_k=\tilde{\sigma}Z_k+\tilde{\mu}$ are defined with $\tilde{\mu},\tilde{\sigma}>0$ any real values. By replacing Z_k by the corresponding expressions $Z_k=(\tilde{Z}_k-\tilde{\mu})/\tilde{\sigma}$ we obtain from (24):

$$P\left\{\left[\frac{(\tilde{Z}_k-\tilde{\mu})}{\tilde{\sigma}}-\mu_k^*\right]/\sigma_k^* \le \left[\frac{(\tilde{\zeta}_i-\tilde{\mu})}{\tilde{\sigma}}-\mu_k^*\right]/\sigma_k^*=\frac{(\overline{\zeta}_i-\mu_k^*)}{\sigma_k^*}=\zeta_{ki}\right\}=\Psi_k[F_k(i),\alpha],$$

$i = 1, 2, ..., I$, $k = 1, 2, ..., K$, where $\tilde{\zeta}_i$ are the percentage points of $\tilde{Z}_k$ corresponding to $\zeta_{ki}=(\overline{\zeta}_i-\mu_k^*)/\sigma_k^*$ and it can be easily shown that they have values $\tilde{\zeta}_i=\tilde{\sigma}(\sigma_k^*\zeta_{ki}+\mu_k^*)+\tilde{\mu}$, which justifies all equalities given above in brackets (remember (4), (5) for the location-scale parameters transformation). Thus the equality expression on the left side ensures the existence of a unique percentage point $\tilde{\zeta}_i$ whichever is k and it follows that the unifying percentage point $\tilde{\zeta}_i$ must be:

$$\tilde{\zeta}_i=\tilde{\sigma}\overline{\zeta}_i+\tilde{\mu},$$

which establishes (14).

Thus the theorem has been completely proved. □

3.2. *Sufficient conditions to ensure Jones' model*

With reference to families of distributions defined by relationships (2), which mean that they can be reduced to be of the location scale-type for any chosen values of α_k, $k = 1, 2, ..., K$, it will be shown as one can construct percentage points ζ_{ki}, $i = 1, 2, ..., I$, $k = 1, 2, ..., K$, which are increasing with i and satisfy Jones' model requirements.

Theorem 2. (Sufficient conditions for Jones' model)

For fixed values α_k, $k = 1, 2, ..., K$, consider K families of continuous distribution functions of the type defined by (2) and choose some real values σ_k^*, μ_k^*, in a way that $\sigma_k^* > 0$ are finite: $\sum_{k=1}^{K} (1/K\sigma_k^*) = 1$ with $K\sigma_K^* > 1$, $\sum_{k=1}^{K} (\mu_k^* /K\sigma_k^*) = 0$, and some values $\zeta_{Ki}(\alpha_K)$, $i = 1, 2, ..., I$, that are assumed to be distinct and strictly increasing, which fix the percentage points for the standard element of the K-th family of distributions.

It is shown that one can construct other percentage points ζ_{ki}, $k = 1, 2, ..., K - 1$, $i = 1, 2, ..., I$, increasing with i which allow a unified representation of all percentage points as is expressed by systems (10), with $\xi_i = \overline{\zeta}_i$, and consequently equations (12) are also satisfied.

Proof. For any $i = 1, 2, ..., I$, consider system (17) and note that the K-th equation is always satisfied after we were able to find out some values ζ_{ki}, $k = 1, 2, ..., K - 1$ which satisfy the first $(K - 1)$ equations. This follows from (22) since by hypothesis assumptions (21) and (23) hold true.

Now consider the following systems which are obtained from (17) when we consider the first $(K - 1)$ equations only and remember that ζ_{Ki}, $i = 1, 2, ..., I$, are fixed:

$$\begin{bmatrix} \left(\frac{1}{K\sigma_1^*}-1\right) & \frac{1}{K\sigma_1^*} & \cdots & \frac{1}{K\sigma_1^*} \\ \frac{1}{K\sigma_2^*} & \left(\frac{1}{K\sigma_2^*}-1\right) & \cdots & \frac{1}{K\sigma_2^*} \\ \vdots & \vdots & & \vdots \\ \frac{1}{K\sigma_{K-1}^*} & \frac{1}{K\sigma_{K-1}^*} & \cdots & \left(\frac{1}{K\sigma_{K-1}^*}-1\right) \end{bmatrix} \cdot \begin{bmatrix} \zeta_{1i} \\ \zeta_{2i} \\ \vdots \\ \zeta_{K-1i} \end{bmatrix} = \begin{bmatrix} \frac{\mu_1^*}{\sigma_1^*} - \frac{\zeta_{Ki}}{K\sigma_1^*} \\ \frac{\mu_2^*}{\sigma_2^*} - \frac{\zeta_{Ki}}{K\sigma_2^*} \\ \vdots \\ \frac{\mu_{K-1}^*}{\sigma_{K-1}^*} - \frac{\zeta_{Ki}}{K\sigma_{K-1}^*} \end{bmatrix}. \quad (25)$$

The $(K - 1)\times(K - 1)$ matrix $\mathbf{M}_{K-1}$, say, of the former systems has exactly the same form as (17), where, however, the terms in $K\sigma_k^*$ are lacking and $\mathbf{u}$ is a $1\times(K - 1)$ vector of unitary values. Since by hypothesis we have $0 < \sum_{k=1}^{K-1} \frac{1}{K\sigma_k^*} = 1 - \frac{1}{K\sigma_K^*} < 1$ it follows from (19) and (20), with $\tilde{c}$ =

$1/\left(1-\sum_{k=1}^{K-1}\frac{1}{K\sigma_k^*}\right)$, that the inverse $\mathbf{M}_{K-1}^{-1}$ of $\mathbf{M}_{K-1}$ is well-defined and multiplying both members of (25) by it we can obtain the unique solutions:

$$\begin{bmatrix}\zeta_{1i}\\ \zeta_{2i}\\ \vdots\\ \zeta_{K\text{-}1i}\end{bmatrix}=\left\{-\begin{bmatrix}\frac{1}{K\sigma_1^*} & 0 & \cdots & 0\\ 0 & \frac{1}{K\sigma_2^*} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \frac{1}{K\sigma_{K\text{-}1}^*}\end{bmatrix}-\tilde{c}\begin{bmatrix}\frac{1}{K\sigma_1^*}\\ \frac{1}{K\sigma_2^*}\\ \vdots\\ \frac{1}{K\sigma_{K\text{-}1}^*}\end{bmatrix}[1/K\sigma_1^*\ 1/K\sigma_2^*\cdots 1/K\sigma_{K\text{-}1}^*]\right\}\cdot$$

$$\cdot\operatorname{Diag}(K\sigma_1^*, K\sigma_2^*, \ldots, K\sigma_{K\text{-}1}^*)\begin{bmatrix}\frac{\mu_1^*}{\sigma_1^*}-\frac{\zeta_{Ki}}{K\sigma_1^*}\\ \frac{\mu_2^*}{\sigma_2^*}-\frac{\zeta_{Ki}}{K\sigma_2^*}\\ \vdots\\ \frac{\mu_{K\text{-}1}^*}{\sigma_{K\text{-}1}^*}-\frac{\zeta_{Ki}}{K\sigma_{K\text{-}1}^*}\end{bmatrix}=$$

$$=\begin{bmatrix}-\frac{\mu_1^*}{\sigma_1^*}+\frac{\zeta_{Ki}}{K\sigma_1^*}\\ -\frac{\mu_2^*}{\sigma_2^*}+\frac{\zeta_{Ki}}{K\sigma_2^*}\\ \vdots\\ -\frac{\mu_{K\text{-}1}^*}{\sigma_{K\text{-}1}^*}+\frac{\zeta_{Ki}}{K\sigma_{K\text{-}1}^*}\end{bmatrix}-\tilde{c}\begin{bmatrix}\frac{1}{K\sigma_1^*} & \frac{1}{K\sigma_1^*} & \cdots & \frac{1}{K\sigma_1^*}\\ \frac{1}{K\sigma_2^*} & \frac{1}{K\sigma_2^*} & \cdots & \frac{1}{K\sigma_2^*}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{1}{K\sigma_{K\text{-}1}^*} & \frac{1}{K\sigma_{K\text{-}1}^*} & \cdots & \frac{1}{K\sigma_{K\text{-}1}^*}\end{bmatrix}\begin{bmatrix}\frac{\mu_1^*}{\sigma_1^*}-\frac{\zeta_{Ki}}{K\sigma_1^*}\\ \frac{\mu_2^*}{\sigma_2^*}-\frac{\zeta_{Ki}}{K\sigma_2^*}\\ \vdots\\ \frac{\mu_{K\text{-}1}^*}{\sigma_{K\text{-}1}^*}-\frac{\zeta_{Ki}}{K\sigma_{K\text{-}1}^*}\end{bmatrix},$$

that is, since $\tilde{c}=K\sigma_K^*$ and $\sum_{k=1}^{K-1}\mu_k^*/\sigma_k^*=-\mu_K^*/\sigma_K^*$ respectively and in force of the conditions (21), (23):

$$\begin{cases}\zeta_{1i} = \dfrac{\mu_K^*-\mu_1^*}{\sigma_1^*}+\dfrac{\sigma_K^*}{\sigma_1^*}\zeta_{Ki}\\ \zeta_{2i} = \dfrac{\mu_K^*-\mu_2^*}{\sigma_2^*}+\dfrac{\sigma_K^*}{\sigma_2^*}\zeta_{Ki}\\ \vdots\\ \zeta_{K-1i} = \dfrac{\mu_K^*-\mu_{K-1}^*}{\sigma_{K-1}^*}+\dfrac{\sigma_K^*}{\sigma_{K-1}^*}\zeta_{Ki}\end{cases} \tag{26}$$

which show how the percentage points ζ_{ki} are increasing functions of ζ_{Ki}, that is they increase with ζ_{Ki} for $i = 1, 2, ..., I$, and any $k = 1, 2, ..., K-1$, because this is assumed to be true for ζ_{Ki}, $i = 1, 2, ..., I$.

Now remember that the values (26) are solutions of the first $(K-1)$ equations of systems (17), but, since we assumed that conditions (21), (23) are satisfied, expression (22) must have a zero value, which implies that also the K-th equations are satisfied, that is for any i the values $\zeta_{1i}, \zeta_{2i}, ..., \zeta_{K-1i}$, (26), together with ζ_{Ki}, satisfy the K-th equation of system (17). If we introduce the averages $\overline{\zeta}_i$ of the percentage points ζ_{ki} which are solutions of systems (17) the latter are equivalent to relationships (12), that is to

$$\overline{\zeta}_i / \sigma_k^* = \mu_k^* / \sigma_k^* + \zeta_{ki}$$

$i = 1, 2, ..., I$, $k = 1, 2, ..., K$, which ensure that relationships (10) hold with $\overline{\zeta}_i$ as unifying percentage points.

Thus the proof of the theorem is completed. □

4. EXPLAINING THE MEANING OF JONES' MODEL AND CORRESPONDING SIMULATION RESULTS

4.1. *A summary of Jones' model main aspects*

As we stated in §1, according to the methods proposed and used by Thurstone in the case of the normal distribution, here it is assumed in general that, to each stimulus corresponding to an attribute which gives rise to an observed ordinal categorical variable, there corresponds a latent variable Z_k, $k = 1, 2, ..., K$, of the continuous type with a distribution function described by (2). We explained Jones' model on the basis of the following hypotheses by having regard to the Logistic-Weibull family (9) for the general case:

a) the K latent random variables Z_k can not be reduced to one single variable Z (see Theorem 1, *i.*) so that in general there is a latent random vector $\mathbf{Z} = (Z_1, Z_2, ..., Z_K)'$;

b) the corresponding marginal distributions relative to the random variables Z_k, are subjected to the two parametric constraints (13), which are repeated below for clarity, that is:

$$\sum_{k=1}^{K} \frac{1}{K\sigma_k^*(\alpha)} = 1, \qquad \sum_{k=1}^{K} \frac{\mu_k^*(\alpha)}{K\sigma_k^*(\alpha)} = 0, \tag{13}$$

which impose two *structural relationships* on the location and scale parameters, $\mu_k^*(\alpha)$ and $\sigma_k^*(\alpha)$ respectively for a specific value α;

c) if we refer to the marginal latent variables Z_k and, for each of the latter to the standard element of the distribution functions attached to it, we should have to note that when Jones' model is valid the percentage points ξ_{ki}, which establish the latent thresholds corresponding to the conventional scores i of the observed categorical variable X_k, $i = 1, 2, ..., I$, $k = 1, 2, ..., K$, are defined in such a way that:

- the percentage points, with respect to the pertinent standard element, of one of the latent random variables Z_k can assume whatever value – for simplicity assume that is the case for Z_K – so that ζ_{Ki}, which we shall call *leading percentage points*, can take on arbitrary values within the domain $\Re_1$ of the corresponding standard element $\Psi_k(\zeta_k,\alpha)$, which, recall, is based on a fixed value α_K; thus $\Re_1$ is split up into $I + 1$ parts by the points ζ_{K1}, ζ_{K2}, ..., ζ_{KI}, which correspond to the conventional scores $i = 1, 2, ..., I$;

- by (26) we have that the other percentage points ζ_{ki} are linear functions – we shall call them *constrained percentage points* – of the leading values ζ_{Ki} and for each k and fixed α_k they split up the domain $\Re_1$ of the k-th standard element $\Psi_k(\zeta_k,\alpha)$ into $I + 1$ parts by the points ζ_{k1}, ζ_{k2}, ..., ζ_{kI}, which correspond to the conventional scores of the k^{th} categorical variable;

- altogether the percentage points ζ_{ki}, $i = 1, 2, ..., I$, $k = 1, 2, ..., K$, satisfy the K alignment conditions (10), see next Fig. 4.

d) When we consider the conventional scores i, $i = 1, 2, ..., I$, assigned to the K categorical variables by a sample of n independent subjects, we assume that we are concerned with a simple random sample of n elements (stochastically independent) from the K-dimensional random vector $\mathbf{Z}$ with marginal distribution functions of the type (2).

The latent mechanism which corresponds to the realization of said sample according to Jones' model is the following: the random assignment of score i to the categorical variable X_k on behalf of a respondent implicitly corresponds to the random selection of a value, say $\zeta_{ki\cdot s}$, $s \in \{1, 2, ..., n\}$, according to the probabilistic law established by the relevant latent standard distribution $\Psi_k(\zeta_k,\alpha_k)$ and to its implicit comparison either with the constrained percentage points ζ_{ki}, $i = 1, 2, ..., I$, if $k < K$ or with the leading percentage points ζ_{Ki} if $k = K$, in order to allocate $\zeta_{ki\cdot s}$ to the i-th element of the partition in $I + 1$ elements that we saw is induced by the model.

4.2. *A Montecarlo simulation of Jones' model*

A Monte Carlo simulation procedure, run on a Personal Computer Acer Veriton 7600G, was set up, which is appropriate to obtain results comparable with the ones of the real case presented by Jones (1986) and to check the interpretation of Jones' model outlined above.

1) A 12-dimensional Normal distribution was considered with all marginal univariate components Z_k following a normal standard distribution, i.e. of type $N(0,1)$; the correlation matrix $\mathbf{R}$, 12 × 12, was assumed to be of the form:

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} \tag{27}$$

with ρ, which represents a constant correlation coefficient between pair of components, so that – (1/11) < ρ < 1 in order that **R** is non-singular and thus positive definite.

In the example given here we put ρ = 0.9. To obtain sequences of random vectors $\mathbf{z}_j$, 12 × 1, ruled by the 12-dimensional normal distribution considered above (j = 1, 2, ..., 500 in the example) we used the following procedure. Assuming that $\mathbf{Z}^\circ = (Z_1^\circ, Z_2^\circ, ..., Z_K^\circ)'$ is a random vector of independent normal random variables with standard distribution we considered a transformation **A**, that is a matrix **A**, 12 × 12, for which:

$$\mathrm{E}(\mathbf{A}'\mathbf{Z}^\circ\mathbf{Z}^{\circ\prime}\mathbf{A}) = \mathbf{A}'\mathbf{A} = \mathbf{R}$$

where E(·) is the expected value and the assumed stochastic independence ensures that $\mathrm{E}(\mathbf{Z}^\circ\mathbf{Z}^{\circ\prime}) = \mathbf{I}_{12}$, with $\mathbf{I}_{12}$ the identity matrix of dimensions 12, while $\mathbf{A}'\mathbf{A}$ represents the Cholesky decomposition of **R**, (27), with $\mathbf{A}'$ an upper triangular matrix.

Once the elements of $\mathbf{A}'$ are determined, the random vector $\mathbf{A}'\mathbf{Z}^\circ$ will follow the required probabilistic law. Correspondingly by the software GAUSS for Windows NT/95, Version 3.2.38, Copyright 1984-1999, Aptech Systems, Inc. Maple Valley, WA, we produced 12 × 500 pseudorandom values from a normal standard distribution and in sequence from each vector relative to sub-group of 12, $\mathbf{z}_j^\circ$ say, we could obtain vectors $\mathbf{z}_j = \mathbf{A}'\ \mathbf{z}_j^\circ$, j = 1, 2, ..., 500, representing a pseudorandom sequence of 500 elements, drawn from the desired multinormal distribution.

2) In order to comply in the reading of the simulation results with the need of specifying the latent structural constraints (13) – in the example **α** is absent – we considered the values μ_k^*, σ_k^*, listed in the Table 2, in comparison with their estimates obtained by fitting the regression models (12) – which express the alignment of the percentage points discussed above – by the constrained least-squares method to the empirical percentage points resulting from simulation (see below). Note that these estimates allow us to recover the latent parametric structure through the standard elements of the marginal distributions, which appear to be invariant but only with respect to linear transformations of the location and scale parameters, see § 2. Table 2 indicates the values of the leading percentage points also required for a complete specification of the model.

3) Table 3 presents the theoretical percentage points ζ_{ki}, k = 1, 2, ..., 12, i = 1, 2, ..., 9, as they are defined by Jones' model according to the equation systems (12), which are satisfied in force of the structural relationships (13) ensured by the chosen values μ_k^*, σ_k^*. The latter led to the values given in Table 3, calculated through relationships (26) by using the leading percentage points ζ_{12i} also given in Table 2. These theoretical percentage points are shown in Table 3 in comparison with those obtained by means of the 500 simulated vector data. This was done by considering for each component Z_k the cumulative proportions, say $\hat{F}_k(i)$, of simulated cases leading to values z_{ki} not larger than ζ_{ki} and by taking as corresponding estimate $\hat{\zeta}_{ki}$ the value of $\Psi^{-1}[\hat{F}_k(i)]$ where $\Psi^{-1}(\cdot)$ denotes the inverse of the distribution function of the standard normal distribution.

TABLE 2

Latent parametric structure assumed in the Monte Carlo simulation: values assumed for μ_k^*, σ_k^* *and corresponding estimates* $\hat{\mu}_k^*$, $\hat{\sigma}_k^*$, *obtained by the constrained least-squares method – Leading percentage points* ζ_{Ki}

k	1	2	3	4	5	6	7	8	9	10	11	12
μ_k^*	−0.80	0.50	0.60	0.40	−0.60	−0.80	−0.90	−0.30	1.00	1.00	−0.30	−0.37
$\hat{\mu}_k^*$	−0.62	0.47	0.54	0.33	−0.52	−0.61	−0.80	−0.26	0.95	0.94	−0.25	−0.35
σ_k^*	1.83	1.00	0.87	0.93	0.86	1.88	0.99	0.83	1.00	1.01	0.83	0.89
$\hat{\sigma}_k^*$	1.66	1.03	0.89	0.94	0.88	1.68	0.95	0.82	1.06	1.07	0.83	0.87
μ_k^* / σ_k^*	−0.44	0.50	0.69	0.43	−0.70	−0.43	−0.91	−0.36	1.00	0.99	−0.36	−0.42

Leading percentage points, relative to the standard normal distribution, ζ_{12i}	1	2	3	4	5	6	7	8	9
	−2.333	−1.667	−1.000	−0.333	0.333	1.000	1.667	2.333	3.000

TABLE 3

Theoretical percentage points ζ_k, $k = 1, \ldots, 12$ *corresponding to the marginal normal latent distributions (in parentheses) and their estimates* $\hat{\zeta}_{ki}$ *obtained from a random vector sample of 500 units drawn by Monte Carlo simulation as specified above. The food items are those indicated in Jones (1986)*

Food Item	i	1	2	3	4	5	6	7	8	9
Sweetbreads	ζ_1	−0.978	−0.656	−0.342	−0.030	0.337	0.619	1.003	1.282	1.728
		(−0.900)	(−0.576)	(−0.252)	(0.072)	(0.396)	(0.720)	(1.044)	(1.368)	(1.692)
Cauliflower	ζ_2	−2.409	−2.097	−1.799	−1.195	−0.631	−0.075	0.490	1.216	1.728
		(−2.947)	(−2.354)	(−1.761)	(−1.168)	(−0.575)	(0.018)	(0.611)	(1.204)	(1.797)
Fresh pineapple	ζ_3	−2.878	−2.512	−1.977	−1.555	−0.856	−0.197	0.490	1.329	1.977
		(−3.502)	(−2.820)	(−2.139)	(−1.457)	(−0.776)	(−0.094)	(0.587)	(1.269)	(1.951)
Parsnips	ζ_4	−2.512	−2.144	−1.774	−1.136	−0.565	−0.015	0.719	1.433	2.097
		(−3.061)	(−2.423)	(−1.786)	(−1.148)	(−0.511)	(0.127)	(0.765)	(1.402)	(2.040)
Baked beans	ζ_5	−1.977	−1.447	−0.856	−0.166	0.607	1.293	1.977	2.512	2.652
		(−2.147)	(−1.458)	(−0.768)	(−0.079)	(0.611)	(1.300)	(1.990)	(2.679)	(3.369)
Wieners	ζ_6	−0.946	−0.668	−0.342	−0.010	0.269	0.625	0.954	1.341	1.665
		(−0.876)	(−0.560)	(−0.245)	(0.070)	(0.386)	(0.701)	(1.017)	(1.332)	(1.647)
Chocolate cake	ζ_7	−1.522	−1.063	−0.440	0.181	0.800	1.432	2.144	2.512	2.878
		(−1.562)	(−0.963)	(−0.364)	(0.235)	(0.833)	(1.432)	(2.031)	(2.630)	(3.229)
Salmon loaf	ζ_8	−2.326	−1.881	−1.237	−0.496	0.207	0.986	1.728	2.197	2.878
		(−2.586)	(−1.872)	(−1.158)	(−0.443)	(0.271)	(0.986)	(1.700)	(2.414)	(3.129)
Blueberry pie	ζ_9	−2.878	−2.409	−2.054	−1.665	−1.117	−0.553	−0.010	0.674	1.305
		(−3.447)	(−2.854)	(−2.261)	(−1.668)	(−1.075)	(−0.482)	(0.111)	(0.704)	(1.297)
Turnips	ζ_{10}	−2.878	−2.326	−2.054	−1.572	−1.071	−0.542	0.045	0.631	1.282
		(−3.413)	(−2.825)	(−2.238)	(−1.651)	(−1.064)	(−0.477)	(0.110)	(0.697)	(1.284)
Liver	ζ_{11}	−2.326	−1.881	−1.216	−0.462	0.181	0.970	1.607	2.409	2.652
		(−2.586)	(−1.872)	(−1.158)	(−0.443)	(0.271)	(0.986)	(1.700)	(2.414)	(3.129)
Spaghetti	ζ_{12}	−1.977	−1.626	−1.071	−0.412	0.202	0.946	1.685	2.257	2.878
		(−2.333)	(−1.667)	(−1.000)	(−0.333)	(0.333)	(1.000)	(1.667)	(2.333)	(3.000)
	$\bar{\zeta}$	−2.134	−1.726	−1.264	−0.710	−0.136	0.457	1.069	1.649	2.143

The results of Table 3 appear to be very satisfactory since they confirm thoroughly, from a descriptive point of view, Jones' model under study. In particular this is seen, with regard to the basic "alignment conditions" (12), which we transcribe below in the equivalent form:

$$\bar{\zeta}_i = \mu_k^* + \sigma_k^* \zeta_{ki} \qquad (29)$$

i = 1, 2, ..., I, k = 1, 2, ..., K, and which require that the unifying percentage points, represented by the arithmetic means $\bar{\zeta}_i = \sum_{k=1}^{K} \zeta_{ki} / K$, must belong to K, in the simulation example K = 12, different straight lines on varying ζ_{ki}, i = 1, 2, ..., I, I = 9 in the example. These conditions result in being true not only for the theoretical values ζ_{ki}, which have to be the case by construction, but they hold with a good approximation even when we consider the corresponding estimates $\hat{\bar{\zeta}}_{ki}$, $\hat{\zeta}_{ki}$ obtained by simulation.

This is shown by Table 4, which presents the intercepts $\hat{\mu}_k^*$ and the slopes $\hat{\sigma}_k^*$ obtained from the fitting by the ordinary least-squares criterion to the mean values $\hat{\bar{\zeta}}_{ki}$ of the 12 linear models in ζ_{ki} summarized in (29). The corresponding coefficients of determination R^2 are very close to 1 and thus ensure a good alignment.

TABLE 4

Ordinary least-squares estimates of the slopes and intercepts of the 12 straight lines (29) to which the percentage points arithmetic means $\hat{\bar{\zeta}}_i = \sum_{k=1}^{K} \hat{\zeta}_i / K$, *i = 1, 2, ..., 9, are constrained by Jones' model on varying* $\hat{\zeta}_{ki}$ *and corresponding coefficient of determination* R^2

Food Item	*intercept* ($\hat{\mu}_k^*$)	*slope* ($\hat{\sigma}_k^*$)	*det. coeff.* R^2
Sweetbreads	–0.61642	1.65358	0.9967
Cauliflower	0.46952	1.02168	0.9965
Fresh pineapple	0.53342	0.88222	0.9953
Parsnips	0.33152	0.93217	0.9961
Baked beans	–0.51794	0.87296	0.9889
Wieners	–0.61058	1.67859	0.9986
Chocolate cake	–0.80085	0.94719	0.9932
Salmon loaf	–0.25872	0.81618	0.9978
Blueberry pie	0.94721	1.05385	0.9946
Turnips	0.93718	1.07068	0.9955
Liver	–0.24998	0.82684	0.9960
Spaghetti	–0.35090	0.87043	0.9994

Figure 4 gives a graphical picture of some of the straight lines summarized in Table 4 with the points of coordinates ($\hat{\bar{\zeta}}_i$, $\hat{\zeta}_{ki}$), i = 1, 2, ..., 9, which were obtained as a result of the considered Monte Carlo simulation.

4) As mentioned before in order to recover the latent parametric structure of Jones' model we had resort to the constrained least-squares criterion to obtain the estimates $\hat{\mu}_k^*$, $\hat{\sigma}_k^*$ given in Table 2. Regarding the empirical percentage points $\hat{\zeta}_{ki}$, obtained by simulation in the present study, and their arithmetic means $\hat{\bar{\zeta}}_i = \sum_{k=1}^{K} \hat{\zeta}_{ki} / K$ we considered the regression models (29) looking for values $\hat{\mu}_k^*$, $\hat{\sigma}_k^*$ which minimize:

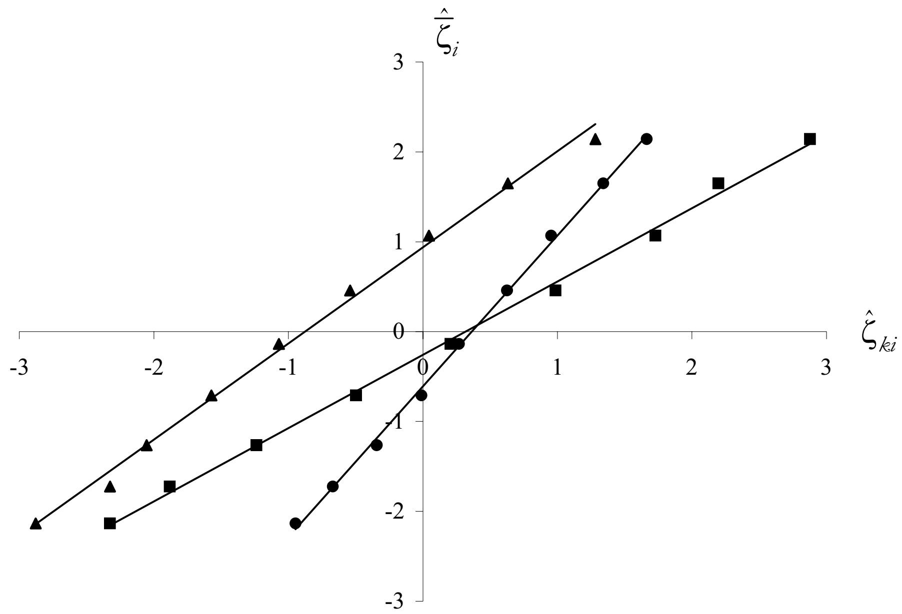


Figure 4 – Example of the alignment of the unifying percentage points $\hat{\bar{\zeta}}_i$ considered as a function of the percentage points $\hat{\zeta}_{ki}$ relative to the various categories, $i = 1, 2, ..., 9$, $k = 6, 8, 10$.

$$Q^2 = \frac{1}{K}\sum_{k=1}^{K}\left[\sum_{i=1}^{I}(\hat{\bar{\zeta}}_i - \mu_k^* - \sigma_k^*\hat{\zeta}_{ki})^2\right] \tag{30}$$

under the constraints (13):

$$\varphi_1(\cdot) = \left(\sum_{k=1}^{K}\frac{1}{K\sigma_k^*} - 1\right) = 0, \quad \varphi_2(\cdot) = \sum_{k=1}^{K}\frac{\mu_k^*}{K\sigma_k^*} = 0 \tag{31}$$

where, as the simulation concerns the normal distribution, the further parameter α is absent. In order to apply the method of Lagrange multipliers we referred to the Lagrange function

$$L = Q^2 + \lambda_1\varphi_1(\sigma_1^*, ..., \sigma_K^*) + \lambda_2\varphi_2(\sigma_1^*, ..., \sigma_K^*, \mu_1^*, ..., \mu_K^*),$$

with λ_1, λ_2 the real variables called Lagrange multipliers and solved the non-linear system:

$$\begin{cases} \dfrac{\partial L}{\partial \sigma_k^*} = 0 \\ \dfrac{\partial L}{\partial \mu_k^*} = 0 \ , \ k = 1, 2, ..., K \\ \varphi_1 = 0 \\ \varphi_2 = 0 \end{cases} \tag{32}$$

with the help of the software GAUSS.

Details relative to system (32) and to its solution will be given in a following paper.

We can conclude this section by remarking that the percentage points $\hat{\zeta}_{ki}$ could have been estimated directly, without having recourse to the inverse of the normal standard distribution, as $z_{ki}(n_{ki})$, where $n_{ki} = [n\pi_{ki}] + 1$, with n the sample size $n = 500$ in our case, $[\cdot]$ indicates the integral part, π_{ki} is the theoretic known probability level of the percentage point ζ_{ki}, $k = 1, 2, ..., 12$, $i = 1, 2, ..., 9$, in the simulation, $z_{ki}(n_{ki})$ are the ordered observations regarding the k-th latent variable.

We replaced the estimates $z_{ki}(n_{ki})$ for $\hat{\zeta}_{ki}$ to check the alignments resulting from Table 4 and we came to quite similar results.

5. THE EXTENSION OF JONES' METHOD: AN EXAMPLE

The preceding Theorem 2 shows that, without further restrictions, the percentage points "alignment" required by conditions (10) and (12) might not be sufficient to validate the assumed latent distributions.

To underline this point we considered the distribution model (9) to examine the table of 9 × 12 empiric cumulative proportions $\hat{F}_k(i)$, given by Jones (1986) and reported in Table 5, pertaining, as we already said, to conventional preference scores, ranging from 1 to 9, obtained from a sample of 255 army enlisted men with respect to 12 food items, $k = 1, 2, ..., K = 12$, $i = 1, 2, ..., I = 9$, and treated by the author assuming 12 underlying normal distributions.

TABLE 5

Empiric cumulative proportions $\hat{F}_k(i)$, given by Jones (1986)

Food Item *i*	1	2	3	4	5	6	7	8	9
Sweetbreads	0.099	0.226	0.333	0.453	0.646	0.798	0.930	0.984	1.000
Cauliflower	0.059	0.083	0.122	0.224	0.331	0.575	0.858	0.976	1.000
Fresh pineapple	0.016	0.020	0.028	0.055	0.119	0.277	0.573	0.877	1.000
Parsnips	0.068	0.188	0.312	0.492	0.680	0.808	0.940	0.984	1.000
Baked beans	0.016	0.035	0.059	0.126	0.232	0.469	0.752	0.972	1.000
Wieners	0.000	0.008	0.036	0.071	0.182	0.482	0.842	0.992	1.000
Chocolate cake	0.000	0.004	0.012	0.039	0.118	0.248	0.539	0.874	1.000
Salmon loaf	0.032	0.067	0.111	0.202	0.344	0.636	0.889	0.988	1.000
Blueberry pie	0.008	0.008	0.020	0.075	0.171	0.333	0.611	0.889	1.000
Turnips	0.075	0.194	0.324	0.514	0.652	0.810	0.937	0.996	1.000
Liver	0.083	0.146	0.186	0.245	0.336	0.502	0.783	0.964	1.000
Spaghetti	0.004	0.008	0.039	0.075	0.169	0.409	0.772	0.953	1.000

In the case of the distribution model (9) (of the Logistic-Weibull type), since it assumes that $\mu_k = 0$, $k = 1, 2, ..., 12$, relationships (12) become

$$\overline{\zeta}_i = \sigma_k^* \zeta_{ki}, \tag{33}$$

where we recall that $\overline{\zeta}_i = \sum_{k=1}^{K} \zeta_{ki}/K$, $i = 1, 2, ..., I$.

Having regard to the left side of expression (9) referring to the inverse of the standard element of the Logistic-Weibull family of distributions, defined by putting $\mu_k = 0$, $\sigma_k = 1$ in (8), we obtained the empiric percentage points $\tilde{\zeta}_{ki}$, say, by replacing $F_k(i)$ with the empiric cumulative frequency ratio $\hat{F}_k(i)$, given in Table 5, after we had chosen some positive starting values $\boldsymbol{\alpha}_0 = (\alpha_{01}, \alpha_{02}, ..., \alpha_{012})'$. We set up a numerical procedure so that for any given $\boldsymbol{\alpha}$ vector and corresponding $\tilde{\zeta}_{ki}$ values, (9) leads to find out the values $\hat{\sigma}_k^*$ which minimize the expression:

$$Q^2 = \sum_{k=1}^{K}\sum_{i=1}^{I} \hat{\sigma}_k^{*2}[\bar{\tilde{\zeta}}_i(\boldsymbol{\alpha}) / \hat{\sigma}_k^* - \tilde{\zeta}_{ki}(\boldsymbol{\alpha})]^2 / IK,$$

under the constraint $\sum_{k=1}^{K} 1/(K\sigma_k^*) = 1$, (stepwise constrained least-squares criterion). In fact starting from an initial value $\boldsymbol{\alpha}_0$ we improved minimization on varying $\boldsymbol{\alpha}$ by using the numerical gradient minimization algorithm available in the software GAUSS.

The minimum value $Q^2 = 0.0052$ was obtained (in the normal case); the corresponding percentage points $\tilde{\zeta}_{ki}$ are given in Table 6 and the estimates $\hat{\sigma}_k^*$ with the optimum α_k are reported in Table 7, while Fig. 5 shows that the alignment conditions are also very well satisfied.

TABLE 6

Optimized percentage points for the Logistic-Weibull distribution

Food Item	*i*	1	2	3	4	5	6	7	8	9
Sweetbreads	ζ_1	0.138	0.292	0.425	0.587	0.904	1.255	1.772	2.250	3.000
Cauliflower	ζ_2	0.280	0.340	0.427	0.616	0.788	1.166	1.716	2.200	3.000
Fresh pineapple	ζ_3	0.358	0.391	0.447	0.590	0.801	1.137	1.583	2.060	3.000
Parsnips	ζ_4	0.095	0.236	0.385	0.630	0.957	1.271	1.830	2.261	3.000
Baked beans	ζ_5	0.229	0.335	0.427	0.616	0.834	1.216	1.659	2.247	3.000
Wieners	ζ_6	0.000	0.218	0.417	0.563	0.852	1.354	1.918	2.449	3.000
Chocolate cake	ζ_7	0.000	0.267	0.399	0.619	0.924	1.215	1.654	2.134	3.000
Salmon loaf	ζ_8	0.205	0.313	0.416	0.591	0.824	1.282	1.815	2.323	3.000
Blueberry pie	ζ_9	0.258	0.258	0.374	0.645	0.904	1.206	1.613	2.068	3.000
Turnips	ζ_{10}	0.125	0.279	0.446	0.716	0.953	1.319	1.823	2.523	3.000
Liver	ζ_{11}	0.320	0.450	0.521	0.621	0.765	1.020	1.513	2.104	3.000
Spaghetti	ζ_{12}	0.167	0.225	0.446	0.587	0.838	1.260	1.801	2.212	3.000
	$\bar{\zeta}$	0.181	0.300	0.428	0.615	0.862	1.225	1.725	2.236	3.000

TABLE 7

Parameter estimates for the Logistic-Weibull distribution

	k	1	2	3	4	5	6	7	8	9	10	11	12
Logistic-Weibull	$\hat{\sigma}_k^*$	0.994	1.012	1.037	0.984	1.007	0.956	1.018	0.983	1.029	0.949	1.042	0.998
	$\hat{\alpha}_k$	1.206	1.790	2.503	1.184	2.132	2.328	2.725	1.816	2.460	1.261	1.742	2.351

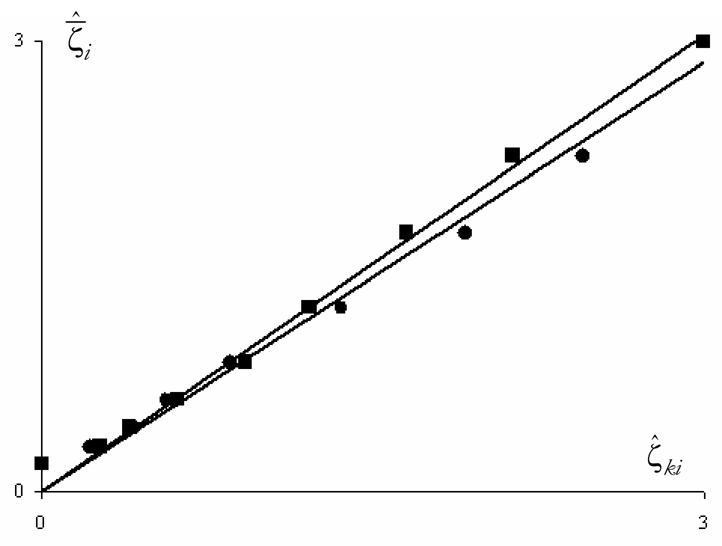


Figure 5 – Logistic-Weibull distribution function: example of the alignment of the unifying percentage points $\hat{\hat{\zeta}}_i$ considered as a function of the percentage points $\hat{\zeta}_{ki}$ relative to the various categories, i = 1, 2, ..., 9, k = 6, 7, and obtained from Table 5 using the numerical method mentioned above which led to the values of Table 6.

6. CONCLUSIONS AND FURTHER RESEARCH DEVELOPMENTS

1) The paper has shown which are the probabilistic and statistical implications of the model proposed by Jones (1986) in the framework of Thurstone psychometric approach to scaling. The latter assumes that stimuli produced by the discriminant process aimed at the evaluation of a given object or a situation are associated in the "judge" to a latent, that is not directly observable, random variable, which follows over occasions, either with a single judge or over judges, a normal distribution. As we saw Jones' model assumes that regarding an empirical manifestation characterized by K aspects or attributes, a judge (subject) is asked to assign to each of them a "preference score" on a conventional integer rating scale with points $i \in \{1, 2, ..., I\}$. Thus a K-dimensional categorical variable $\mathbf{X} = (X_1, X_2, ..., X_K)'$ can be attached to the K attributes, with components X_k, k = 1, 2, ..., K, which are ordinal categorical variables. To these in Jones' model there corresponds a latent random vector $\mathbf{Z} = (Z_1, Z_2, ..., Z_K)'$ whose components orderly match with those of $\mathbf{X}$ and are assumed to follow the Normal distribution. In § 4.1 the meaning of Jones' model when we consider a sample of respondents is illustrated.

As we widely discussed in sections 3 and 4 the essential trait of Jones' model is the existence of *unifying scores* ξ_i which together with appropriate values μ_k^*, σ_k^* of the location and scale parameters satisfy the relationships:

$$\frac{\xi_i - \mu_k^*}{\sigma_k^*} = \zeta_{ki} = \Psi^{-1}[F_k(i)]$$

$i = 1, 2, ..., I$, $k = 1, 2, ..., K$, where, according to (1), $F_k(i)$ is the cumulative probability that we may observe a conventional value x_k of X_k not larger than i, $\Psi^{-1}(\cdot)$ denotes the inverse of the distribution function of the standard normal distribution, ζ_{ki} are the percentage points or quantiles corresponding to $F_k(i)$. In Theorem 1 it is shown that the unifying scores can be chosen as $\xi_i = \overline{\zeta}_i = \sum_{k=1}^{K} \zeta_{ki}/K$, which thus by hypothesis are defined on an interval scale. Since whatever is the attribute, the values $\overline{\zeta}_i$ correspond to the conventional rating scores $i = 1, 2, ..., I$, they give a unitary representation of the latter on a same interval scale and can be assumed as a metric version of the conventional scores i which is valid for all attributes. Once it has been ascertained that such a scale can be constructed, which we actually did, both the observations already available and the possible new observations can be recorded as values on this interval scale with the advantage that now all algebraic operations required by the usual statistical methodology become fully justified. We considered appropriate to mention this point again in the final remarks to underline the practical relevance of the subject dealt with in the paper.

Theorem 2 shows that statistical models with Jones' requirements exist and the way they can be constructed.

2) We recall that in this paper the distributional context of Jones' model has been widened by showing that it can be extended beyond the Normal distribution to other types of latent probability distributions, provided that they can be led back to location-scale family. The case of the Logistic-Weibull family of distributions, which seems to be particularly appropriate to describe the preference process underlying "customer satisfaction" assessments, is given as an example in §6. It is shown as the choice of the levels of some other parameters, which are present in the distribution model besides the location-scale parameters, can help to ensure that Jones' model is valid.

3) As it is specifically discussed in the comments of section 4, Jones' model presupposes a latent parametric structure among the location and scale parameters μ_k^*, σ_k^*, see (13), which ensures the "alignment" of the unifying scores $\overline{\zeta}_i$ considered in function of the quantile ζ_{ik}, $i = 1, 2, ..., I$, $k = 1, 2, ..., K$, which allows an evident geometric image of the model, see Fig. 4,5. Correspondingly in the paper the constrained least squares criterion is proposed to estimate the parameter μ_k^*, σ_k^* and it was used in the included simulation study.

More details and a discussion on this point will be given in a following paper.

4) The further development of the research also envisages the setting up of statistical tests which may be appropriate to check whether Jones' model is justified to interpret a given set of data on the basis of an assumed family of latent probability distributions and, in view of Theorem 2, which of two or more competing latent distribution models has to be considered the best.

In this regard the asymptotic theory of quantile estimates already available, see David (1981), Ch. 9, appears to represent a promising starting point.

Istituto di Statistica ANGELO ZANELLA
Università Cattolica del S. Cuore, Milano GABRIELE CANTALUPPI

REFERENCES

E. BODEWIG (1959) *Matrix calculus*, North-Holland Publishing Company, Amsterdam.

R.D. BOCK, L.V. JONES (1968) *The measurement and prediction of judgement and choice*, Holden Day, San Francisco.

H.A. DAVID (1981) *Order Statistics*, 2nd Ed., John Wiley, New York.

A. HERZEL (1974, a) *Un criterio di quantificazione: Aspetti statistici*, "Metron", XXXII, pp. 3-54.

A. HERZEL (1974, b) *Un criterio di quantificazione: Aspetti matematici*, "Metron", XXXII, pp. 55-121.

L.V. JONES (1986) *Psychological scaling*, in: S. KOTZ & N.L. JOHNSON (eds), *Encyclopedia of Statistical Sciences*, John Wiley, New York, Vol. 7, pp. 340-343.

D.H. KRANTZ, R.D. LUCE, P. SUPPES, A. TVERSKY (1971) *Foundations of Measurement*, Vol. 1, Academic Press, San Diego.

L.L. THURSTONE (1959) *The measurement of values*, University of Chicago Press, Chicago.

A. ZANELLA (1998) *A statistical model for the analysis of customer satisfaction: some theoretical and simulation results*, "Total Quality Management", 9, pp. 599-609.

A. ZANELLA (1999) *A Stochastic Model for the analysis of customer satisfaction: Some theoretical aspects*, "Statistica", LIX, pp. 3-40.

A. ZANELLA, M. CERRI (2000) *La misura di Customer Satisfaction: qualche riflessione sulla scelta delle scale di punteggio*, in: Atti della giornata di studio *Valutazione della qualità e Customer Satisfaction: il ruolo della Statistica*, Vita e Pensiero, Milano, pp. 217-231.

RIASSUNTO

Trasformazione simultanea su scale ad intervalli di un insieme di variabili categoriche

Nel contributo sono esaminate alcune implicazioni e un'estensione del metodo proposto, su base euristica, da Jones (1986) con riferimento alla trasformazione simultanea di un insieme di variabili qualitative ordinate in scale per intervalli, sotto l'ipotesi che esista, in corrispondenza a ciascuna delle variabili qualitative, una variabile latente normale. Nell'articolo – che si collega al problema delle trasformazioni di scale ordinali, ampiamente trattato da Amato Herzel – viene presentato e discusso il modello statistico-probabilistico alla base del metodo di Jones e si propone l'estensione dello stesso ad altre famiglie di variabili latenti, con distribuzione di probabilità dipendente solo dai parametri di posizione e di scala. Si presenta un esempio di applicazione alla famiglia di distribuzioni di probabilità del tipo Logistico-Weibull.

SUMMARY

Simultaneous transformation into interval scales for a set of categorical variables

The paper – related to the problem of ordinal scale transformations, extensively dealt with by Amato Herzel – examines some implications and an extension of the method heuristically proposed by Jones (1986) to simultaneously transform a set of observed categorical ordinal variables into interval scales, under the assumption that there exists a normal latent random variable corresponding to each of the categorical variables. The ar-

ticle, on the one hand, presents and discusses the statistical-probabilistic model at the basis of Jones' method and on the other hand proposes its extension to other families of latent variables, besides the Normal distribution, when their probability distributions can be reduced to a location-scale type. An example of application to the Logistic-Weibull family of distributions is also illustrated.