Available online at www.sciencedirect.com
ScienceDirect
Food Quality and Preference 19 (2008) 407–411
Food Quality and Preference
www.elsevier.com/locate/foodqual

Thurstonian scales obtained by transformation of beta distributions

L. Carbonell, J.M. Sendra, S. Bayarri, L. Izquierdo *, A. Tárrega

Instituto de Agroquímica y Tecnología de Alimentos (IATA-CSIC), P.O. Box 73, Burjassot, Valencia, Spain

Received 21 February 2007; received in revised form 10 December 2007; accepted 17 December 2007
Available online 5 February 2008

Abstract

According to Thurstone Case V, the quantiles of the normal standard distribution corresponding to the proportion of answers from a set of paired comparisons between several products are used to compute scales. This paper analyses two alternative approaches based on Bayesian inference by which the normal quantiles are obtained not only from the exact proportion of answers actually observed for each paired comparison but from all potentially observable proportions, continuously distributed in the range from 0 to 1 according to a beta distribution. Using the first approach a normal distribution is assumed for the obtained quantiles and so the properties of normal distributions were applied to estimate scale scores and confidence intervals. The second approach is based on a simulation process that avoids the assumption of normality. Both approaches give similar results that, for a low number of respondents, differ from those obtained by applying Thurstone Case V in its traditional way.

Keywords: Thurstonian scales; Bayesian inference; Beta distributions

1. Introduction

Thurstonian scaling is a common procedure in social and behavioural sciences and also in sensory analysis of foods (Ashby & Ennis, 2002; Duineveld, Arents, & King, 2002; Ennis & O'Mahony, 1995; Jeon, O'Mahony & Kim, 2004; Koo, Kim & O'Mahony, 2002; Lawless & Schlegel, 1984; McEwan & Colwill, 1996; Suzuki & Satake, 1969). A usual method in food analysis is based on case V of Thurstone's law of comparative judgement (Thurstone, 1927) and consists in performing, by a group of assessors, all possible paired comparisons between several products. The observed proportion of answers from each paired comparison is considered as the point estimate of the parameter π of a binomial distribution, and the quantile z of the normal standard distribution corresponding to this point estimate is determined. Next, a linear scale is developed in which the score of each product is obtained by averaging the quantiles corresponding to the paired comparisons including this product. Montag (2006) proposed Monte Carlo simulations to obtain confidence intervals of the scores whereas Lipovetsky and Conklin (2004) and Luker, Beaver, Leinster, Owens, Degner, and Sloan (1995) used t-tests to analyse the significance of the differences between scores.

By using Thurstone's case V procedure only the normal quantiles corresponding to the observed proportions are computed what implies obtaining the same scale scores whatever the number of respondents may be (if the observed proportion of answers to the paired comparisons remains constant). Nevertheless, although an observed proportion is the maximum likelihood (ML) estimate of the parameter (π) of the corresponding binomial distribution, the same proportion could have been observed for other values of the parameter.

The purpose of this paper is to develop Thurstonian scale scores by computing the normal quantiles not only from the observed proportion of answers but from all possible values of π, according to their distribution of probabilities. In these conditions a continuous distribution of z values instead of a single one is obtained for each comparison. Confidence limits of scale scores are obtained either by assuming normality of this distribution or by a

* Corresponding author. Fax: +34 963 63 63 01.
E-mail address: luisiz@iata.csic.es (L. Izquierdo).

0950-3293/$ - see front matter
doi:10.1016/j.foodqual.2007.12.002

simulation process that avoids the assumption of normality. The procedure is presented and discussed with the help of an artificial example. The classical procedure of obtaining Thurstonian scales is briefly described as an introduction to the mathematical justification of the proposed procedure, which is then applied to the example.

2. Computing case V Thurstonian scales: A brief description of the method

Let us suppose that 10 assessors performed all possible paired comparisons of 3 products indicating which one of the two products forming each pair presented higher acceptability and that the results were: 7 assessors preferred A to B, 9 assessors preferred A to C, and 5 preferred B to C.

The process of obtaining Thurstonian scales has three steps:

1. From the observed frequencies mentioned above a square matrix of proportions, shown in Table 1, is computed. A proportion of 0.5 is assigned to the comparison of each product with itself (main diagonal of the matrix).
2. A matrix of z values is obtained by determining the normal quantiles corresponding to each proportion. Table 1 shows these quantiles.
3. The matrix of these z values is averaged by columns. The obtained average values constitute the scores (shown in the table) of the products on the Thurstonian scale.

The scale values do not depend on the observed frequencies of answers to the paired comparisons but only on the proportions. The same results would have been obtained with 100 respondents instead of 10 in case the proportions were identical.

This procedure implies several assumptions, some of them common to Thurstone cases I to IV. Basically, it is assumed that stimuli produce discriminal processes with values on an unidimensional psychological continuum. These values vary according to normal distributions when presentations are repeated. The means and the standard deviations of these distributions determine the scale values of the stimuli and their discriminal dispersion. In case V these distributions are considered to be uncorrelated and to have equal variances.

Table 1
Process of obtaining scales according to Thurstone case V

	Products	A	B	C
Matrix of proportions corresponding to the frequencies mentioned in the text	A	0.5	0.7	0.9
	B	0.3	0.5	0.5
	C	0.1	0.5	0.5
Matrix of quantiles (z) from the proportions above	A	0	0.524	1.281
	B	−0.524	0	0
	C	−1.281	0	0
Scale scores		−0.602	0.175	0.427

3. Mathematical justification of the proposed method

3.1. Binomial and beta distributions

The observed results from a paired comparison, y out of n assessors preferring one product and $n–y$ preferring the other product, can be interpreted as observations from two symmetric binomial distributions with unknown parameters, π and $1-\pi$. The maximum likelihood (ML) estimate of π is the observed proportion y/n, but the true parameter can really have any other value in the continuous interval from 0 to 1. According to Bayesian statistics theory (see Gelman, Carlin, Stern, & Rubin, 2000), given an observed result from a binomial experiment the density function of its parameter π follows a beta distribution with parameters $y+1$ and $n-y+1$ ($\beta_{y+1,n-y+1}$) if a uniform prior distribution, $\beta_{1,1}$, is assumed. This assumption implies the lack of previous information ($y=n=0$).

3.2. Transforming values of the beta distributions into z values of the standard normal distribution

Step 2 of obtaining Thurstonian scales consists in the transformation:

$$z = \theta^{-1}(\text{prop}) \tag{1}$$

where θ is the standard normal distribution function and prop the observed proportion of answers to the paired comparison.

The basic idea of this paper is that the transformation into z values of the standard normal distribution should be applied not only to the observed proportion but to all possible π values, with probabilities distributed according to a beta distribution. The density function of the Z distribution obtained by transformation of a beta distribution can be expressed (see Gelman et al., 2000) as

$$\begin{aligned} f_z(z) &= |J| f_\pi(\theta(z)) \\ &= \left(\frac{1}{\sqrt{2\Pi}} e^{\frac{-z^2}{2}}\right)\left(\frac{(n+1)!}{y!(n-y)!}(\theta(z))^y(1-\theta(z))^{n-y}\right) \end{aligned} \tag{2}$$

where $\theta(z)$ (=π) is the inverse of the transformation shown in expression 1, $|J|$ is the determinant of the Jacobian:

$$J = \frac{d\pi}{dz} = \frac{d(\theta(z))}{dz} = \frac{1}{\sqrt{2\Pi}} e^{\frac{-z^2}{2}}$$

and f_π is the density function of a beta distribution with parameters $y+1$ and $n-y+1$.

Table 2 shows the means, variances and modes of the Z distributions obtained by transformation according to expression 2 of the beta distributions related with the proportions in Table 1.

The mode of the Z distribution is:

$$\left(\frac{1}{\sqrt{2\Pi}} e^{-\frac{z^2}{2}}\right)\left(\frac{y}{\theta(z)} - \frac{n-y}{1-\theta(z)}\right)$$

Table 2
Means, variances and modes of the transformed Z distributions obtained from beta distributions

Z distribution transformed from	Mean	Variance	Mode
$\beta_{8,4}$	0.458	0.149	0.440
$\beta_{10,2}$	1.06	0.205	1.01
$\beta_{6,6}$	0	0.137	0

what can be proved by equating the first derivative of expression 2 (or its logarithm) to 0.

The Z function is symmetric when obtained from a symmetric beta distribution. In this case $y = n - y$ and mode and mean of the Z distribution equal 0.

4. Thurstonian scales obtained from transformed distributions

4.1. First approach. Scale scores computed from mean z values

Table 3 shows, for our example, the z matrix formed by means of the transformed distributions and the corresponding expected scale values of products A, B and C. These values are different from those shown in Table 1 from the observed proportions of answers to the paired comparisons. The differences affect not only the absolute scale values, but also to the relative distances between points. The ratio

$$\frac{|A - B|}{|B - C|}$$

is 3.08 when computed from Table 1 and 3.30 when computed from Table 3.

Table 3
Scales obtained by using as quantiles the means of the Z transformed distributions shown in Table 2

	Products	A	B	C
z values	A	0	0.458	1.06
	B	−0.458	0	0
	C	−1.06	0	0
Scale estimates		−0.506	0.153	0.353
Variances of z	A	0	0.149	0.205
	B	0.149	0	0.137
	C	0.205	0.137	0
Variances of the scale estimates		0.0393	0.0318	0.038
Confidence intervals (95%)		A ± 0.389	B ± 0.350	C ± 0.382

To estimate the confidence intervals of the scale values in Table 3, the properties of the normal law can be assumed with minimal loss of precision since the transformed Z distributions are almost normal even when obtained from extremely asymmetric beta distributions as Fig. 1 shows. According to the properties of normal distributions, the step of averaging matrix z by columns to obtain the scale estimates is equivalent to a sum of independent normal distributions divided by a constant (the number of products), resulting in a new a normal distribution. In Table 3 the first zero of the z matrix (and all zeros in its main diagonal) is a conventional constant corresponding to the hypothetical comparison of a product with itself, for which a proportion of $p = 0.5$ is considered. The scale value −0.506 is the mean of the three z values corresponding to product A (0, −0.458 and −1.06). The variances of the Z distributions mentioned in Table 2 are also shown in Table 3 in the form of a symmetric matrix. The main diagonal of this matrix is constituted by zeros since constants have no variance. The sum

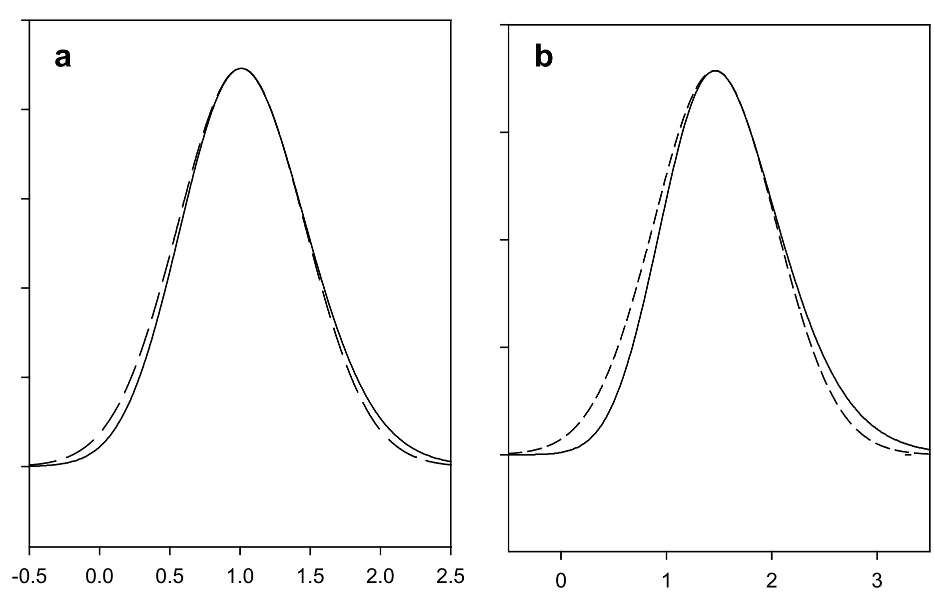

Fig. 1. Density functions of Z distributions (solid), transformed by expression 2 from beta distributions, compared with normal distributions (dotted) with the same modes and variances (a) from $\beta_{10,2}$, mode and variance of Z distribution, 1.01 and 0.205 and (b) from $\beta_{11,1}$, mode and variance of Z distribution 1.46 and 0.333.

by columns of this matrix, divided by the square of the number of products (9), gives the variances of the scale estimates. Symmetric confidence intervals can be computed from these variances by:

$$\text{Interval} = \text{e} \pm zs_{\alpha/2}\sqrt{\text{var}} \quad (3)$$

where e is the scale estimate, var its variance, and $zs_{\alpha/2}$ the quantile of the standard normal distribution corresponding to the desired level of type I error. Table 3 shows the confidence intervals for a confidence level of $1-\alpha = 0.95$.

On the other hand, p values for the observed differences between two scale estimates, e_1 and e_2, assuming a null hypothesis of no differences between populations, can be approximated by the probability in the normal standard distribution of:

$$-\frac{|e_1 - e_2|}{\sqrt{\text{var}_1 + \text{var}_2}} \quad (4)$$

For the differences A–B, A–C, and B–C this expression equals −2.47, −3.09 and −0.76, respectively, with corresponding p values of 0.007, 0.001, and 0.225.

As mentioned above, Luker et al. (1995) analysed differences between scale scores by t-tests. They applied a SAS computer program developed by Sloan, Doig, and Yeung (1994) by which the scale scores are obtained by applying Thurstone case V approach and the confidence intervals are estimated from the standard errors of the individual normal deviates. Also, t-statistic values of the differences between scores are computed from the observed proportions of answers. The program and a complete example are detailed on the Manitoba University website (Sloan, Doig, & Yeung, 2006). In this example 14 assessors expressed their preferences of a total of 9 items presented in pairs. The whole example is not shown here but only results related with items 1 and 5. Item 1 was preferred by 9, 3, 7, 3, 7, 8, 6, and 12 assessors when compared with the remaining items whereas item 5 was preferred by 11, 13, 10, 13, 9, 10, 11, and 12 assessors. The program from Sloan et al. (1994) applied to these data gave scale scores estimates of -0.017 ± 0.372 and 0.781 ± 0.316 (95% confidence level) for items 1 and 5, respectively. The t-statistic of the difference between scores is −3.086. By our approach (expression 3) the scale scores obtained were -0.019 ± 0.226 and 0.690 ± 0.206. By applying expression 4, a value of −4.54 was obtained. If a t-test instead of a normal test (expression 3) were applied in our approach considering 13 degrees of freedom (14 assessors), the estimates would be -0.019 ± 0.249 and 0.690 ± 0.228 still different from those obtained with the program from Sloan et al. (2006). The differences between both approaches in scale scores, confidence intervals and statistic values to compare scores are reasonable considering that the methods of estimation also differ.

4.1.1. Effect of the number of assessors

As the number of respondents (n) increases the scale estimates tend to those obtained when only the observed proportions were considered (Table 1) and the width of the confidence intervals decreases. For instance, if $n = 20$ (and for a constant proportion y/n of answers) the scale estimate of A in our example with 3 products is -0.547 ± 0.292. In the limit, for an infinite number of assessors, the same scale estimates would be obtained by both methods. The reason is that the mean of the beta distribution, $\frac{y-1}{n-2}$, tends to the value of the mode $\frac{y}{n}$ (observed proportion) as y and n increase, thus obtaining the same matrix of z values either from means or from modes.

4.1.2. Results when all assessors give the same response to a paired comparison

It is not unusual, mainly when the number of assessors is low, that all of them give the same response to one or more of the performed paired comparisons. In our example with 3 products let us suppose that in the comparisons A–B and B–C the same results commented above are observed but that in the comparison A–C all assessors prefer product A. The corresponding proportions are 1 for A–C and 0 for C–A for which no normal quantiles exist. A solution is to consider minimum and maximum limits for the proportion. For instance, the commercial program we used for comparative purposes (PC-MDS, 1990) fixed lower and upper limits of the proportion at 0.025 and 0.975 with quantiles −1.96 and 1.96, whereas in the program developed by Sloan et al. (1994) the limits are fixed at 0.02 and 0.98. Table 4 shows in its upper section the z matrix and the scales obtained by fixing the limits at 0.025 and 0.975. But, specially when the number of assessors is low, these limits look unrealistic. For instance, with 10 assessors an observed proportion of 1 is totally compatible with a real proportion in the whole population of potential assessors of 0.9 or even lower, too far from the conventional value of 0.975. By considering transformations of beta distributions it is not necessary to fix limits for the proportions. The distribution related with the comparison A–C, $\beta_{11,1}$, is transformed into a Z distribution whose mean is 1.586. The whole z matrix is shown in the lower part of Table 4 as well as the scale estimates, clearly different from those obtained with fixed limits.

Table 4
Estimation of scales from observed proportions and from Z distributions when in the pair comparisons A–C all assessors (10) give the same response

	Products	A	B	C
z Matrix from observed proportions	A	0	0.524	1.96[a]
	B	−0.524	0	0
	C	−1.96[a]	0	0
Scale estimates		−0.828	0.175	0.653
z Matrix from means of transformed distributions	A	0	0.458	1.586
	B	−0.458	0	0
	C	−1.586	0	0
Scale estimates		−0.681	0.153	0.529

[a] The quantiles corresponding to proportions of 0.975 and 0.025 are conventionally used since the observed proportions of 1 and 0 have no quantiles.

Table 5
Scales obtained as average of 1000 simulations of values from $\beta_{8,4}$, $\beta_{10,2}$, and $\beta_{6,6}$

	Products	A	B	C
Average matrix of simulated π values	A	–	0.665	0.835
	B	0.335	–	0.499
	C	0.165	0.501	–
Average matrix of z values	A	0	0.463	1.060
	B	−0.463	0	−0.009
	C	−1.060	0.009	0
Scale estimates		−0.508	0.157	0.351
Lower confidence limit (2.5%)		A – 0.420	B – 0.367	C – 0.356
Upper confidence limit (97.5%)		A + 0.364	B + 0.337	C + 0.398

4.2. Second approach. Scale scores obtained by simulation of beta distributions

In the previous section the confidence intervals of the scale estimates have been obtained assuming a normal distribution of Z. In fact, this assumption of normality is not necessary. An alternative consists in obtaining the scale values, their confidence intervals and the significance of the differences between scores, by direct simulation of values from beta distributions. For the example with 10 assessors a total of 1000 simulated draws were performed, obtaining in each draw a π value for each of the three distributions $\beta_{8,4}$, $\beta_{10,2}$ and $\beta_{6,6}$. A square matrix of proportions was constructed, the upper half matrix being formed by the simulated π values and the lower half by the corresponding $1-\pi$ values. The obtained quantiles constituted a square matrix of z values that was averaged by columns. In this way, an estimate of the scale scores for the three products was obtained in each simulation step.

Table 5 shows the matrix of the average π values of the beta distributions, the average z values of the transformed distributions, and the average scale estimates. All these values can be considered the same as those shown in Table 3, taking into account that they have been obtained in a simulation process. The confidence limits at 95% level are also shown in Table 5. These confidence limits were determined by ordering all (1000) scale estimations for each product and identifying those occupying the 25th and 976th positions. The intervals are not symmetric since normality has not been assumed but the total width of the intervals almost coincides with the symmetric intervals in Table 3. The proportion of draws in which the score of one product is higher than the score of another product is an estimate of the significance of the differences between scale scores. These proportions were 0.022, 0.004 and 0.258 for the comparisons A–B, A–C, and B–C, respectively, slightly different from the p values obtained above by the normal approach.

5. Conclusion

Two Bayesian approaches based on the properties of beta distributions have been applied to obtain scale scores of products tested by paired comparisons. The scale scores obtained by these approaches depend on the number of assessors and, for a low number of them, differ from the scores obtained by Thurstone case V scaling method. When testing the differences between scale scores our results also differ from those obtained from the observed proportions of answers to the paired comparisons (Sloan et al., 2006).

References

Ashby, F. G., & Ennis, D. M. (2002). A Thurstone-Coombs model of concurrent ratings with sensory and liking dimensions. *Journal of Sensory Studies, 17*, 43–59.

Duineveld, C. A. A., Arents, P., & King, B. M. (2002). Log-linear modelling of paired comparison data from consumer tests. *Food Quality and Preference, 11*, 63–70.

Ennis, D. M., & O'Mahony, M. (1995). Probabilistic models for sequential taste effects in triadic choice. *Journal of Experimental Psychology-Human Perception and Performance, 21*, 1088–1097.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2000). *Bayesian Data Analysis*. Boca Raton: Chapman and Hall/CRC.

Jeon, S. Y., O'Mahony, M., & Kim, K. O. (2004). A comparison of category and line scales under various experimental protocols. *Journal of Sensory Studies, 19*, 49–66.

Koo, T. Y., Kim, K. O., & O'Mahony, M. (2002). Effects of forgetting on performance on various scaling protocols: Magnitude estimation and labelled magnitude scale (green scale). *Journal of Sensory Studies, 17*, 177–192.

Lawless, H., & Schlegel, M. P. (1984). Direct and indirect scaling of sensory differences in simple taste and odour mixtures. *Journal of Food Science, 49*, 44–46.

Lipovetsky, S., & Conklin, W. (2004). Thurstone scaling via binary response regression. *Statistical Methodology, 1*, 93–104.

Luker, K., Beaver, K., Leinster, S., Owens, R., Degner, L., & Sloan, J. (1995). The information needs of women newly diagnosed with breast cancer. *Journal of Advanced Nursing, 22*, 134–141.

McEwan, J. A., & Colwill, J. S. (1996). The sensory assessment of the thirst-quenching characteristics of drinks. *Food Quality and Preference, 7*, 101–111.

Montag, E. D. (2006). Empirical formula for creating error bars for the method of paired comparison. *Journal of electronic imaging 15*, Art. No. 010502.

PC-MDS. Multidimensional statistics package (1990). Institute of Business MGT. Brigham Young University, Provo, UT, USA.

Sloan, J. A., Doig, W., & Yeung, A. (1994). A manual to carry out thurstone scaling and related analytic procedures. Manitoba Nursing Research Institute Technical Report #11, University of Manitoba, Canada.

Sloan, J. A., Doig, W., & Yeung, A. (2006). The Thurstone manual. Manitoba Centre for Health Policy website. Manitoba University. 15-September-2006. (http://www.umanitoba.ca/centres/mchp/concept/thurstone/index.shtml).

Suzuki, R., & Satake, K. (1969). Determination of the optimum temperature for drinking beer, drinks and water. *Journal of the Society of Brewing Japan, 64*, 551–554.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286.