Behavior Research Methods, Instruments, & Computers
2000, 32 (3), 379-388

ThurCatD: A tool for analyzing ratings on an ordinal category scale

MARTIN C. BOSCHMAN
IPO, Center for User–System Interaction, Eindhoven, The Netherlands

An algorithm for analyzing ordinal scaling results is described. Frequency data on ordinal categories are modeled for unidimensional psychological attributes according to Thurstone's judgment scaling model. The algorithm applies maximum likelihood estimation of model parameters. The Cramér–Rao bounds of the standard errors of the estimated parameters are calculated, and a stress measure and a goodness-of-fit measure are supplied.

This paper describes the underlying algorithm of the program ThurCatD (Thurstone categorical judgment, Condition D), which is meant to be used for the analysis of the results of rating experiments in which unidimensional attributes—such as image quality, brightness–contrast, sharpness, loudness, and so forth—are rated on an ordinal category scale. It assumes a specific Thurstone model (Thurstone, 1927): The attribute strength of each stimulus is measured on an internal psychological scale— that is, an interval scale—with Gaussian noise distribution. In ThurCatD, the internal noise spread parameter σ is assumed to be constant for all the rated conditions, and the stimulus values are assumed to be uncorrelated. This case of the Thurstone model is also known as Condition D according to the classification by Torgerson (1958), which explains the final letter in the acronym. In ThurCatD, a psychometric attribute scale is constructed with, for each condition, $N(S_i, \sigma^2)$ distributed strengths. On this continuum—that is, an interval scale— the categories are represented by intervals that, in general, may have different widths.

Over the years, many analytical procedures dealing with categorical judgment have been developed. Most of them apply linear regression techniques (e.g., Gulliksen, 1954; Tucker, 1952) for the estimation of scale values. Torgerson (1958) provided a review on the most important analytical procedures. Since the Thurstone model is a probabilistic model, we are able to apply a maximum likelihood approach to estimate model parameters. Although it demands some quite laborious processing power, the advantages are obvious. Even with sparse data matrices, we are able to construct the Thurstone scale and the corresponding attribute values for each stimulus. Moreover, we are able to calculate the asymptotic standard errors for each estimated parameter.

The author is indebted to Niek Versfeld for the fruitful discussions we had on this subject, to Jean-Bernard Martens for the useful suggestions he offered, and to Martijn Willemsen for reviewing my manuscript before submission. Correspondence concerning this article should be addressed to M. C. Boschman, IPO, Center for User–System Interaction, P. O. Box 513, 5600 MB Eindhoven, The Netherlands (e-mail: m.c.boschman@ tue.nl).

As input, ThurCatD needs frequency distributions per category for each stimulus that was presented in the experiment. These data are obtained from experiments in which subjects are repeatedly asked to scale the attribute under investigation. From the frequency distributions of ratings over the categories, ThurCatD calculates the stimulus scale values in σ-units and, also, the interval borders that define the intervals on the psychometrical scale.

THE THURSTONE MODEL

The basic assumption of the applied Thurstone model is that the strength of the stimulus attribute is measured on a psychometric interval scale Ψ (see Figure 1). Owing to internal noise, the strength x_i of stimulus i is stochastic with a Gaussian distribution

$$x_i \sim N(S_i, \sigma_i^2), \tag{1}$$

where S_i and σ_i^2 denote the position and variance of stimulus i on the Ψ-scale. Under Condition D of Torgerson's classification, which is equivalent to Case V of Thurstone's law of comparative judgment (Thurstone, 1927), the noise spread parameter is assumed to be constant for all the stimuli, or

$$\sigma_i = \sigma, \forall i = 1, \ldots, ns, \tag{2}$$

where *ns* denotes the number of stimulus conditions. Another assumption is that the strengths of different stimuli are uncorrelated, or

$$\rho_{ij} = 0, \tag{3}$$

where ρ_{ij} denotes the correlation coefficient between the perceived attributes of stimulus pair (i, j).

In category scaling experiments, the subject's task is to rate the perceived attribute on a scale consisting of *nc* categories. In the experiment, these categories are labeled with integer numbers (1, 2, 3, 4, . . . , 8, 9, 10) or ordinal adjectives (*bad, poor, fair, good, excellent*). Categories are assumed to be represented by intervals on the Ψ-scale (Figure 2), of which the boundaries $(B_2 \ldots B_{nc})$ are unknown parameters in principle.

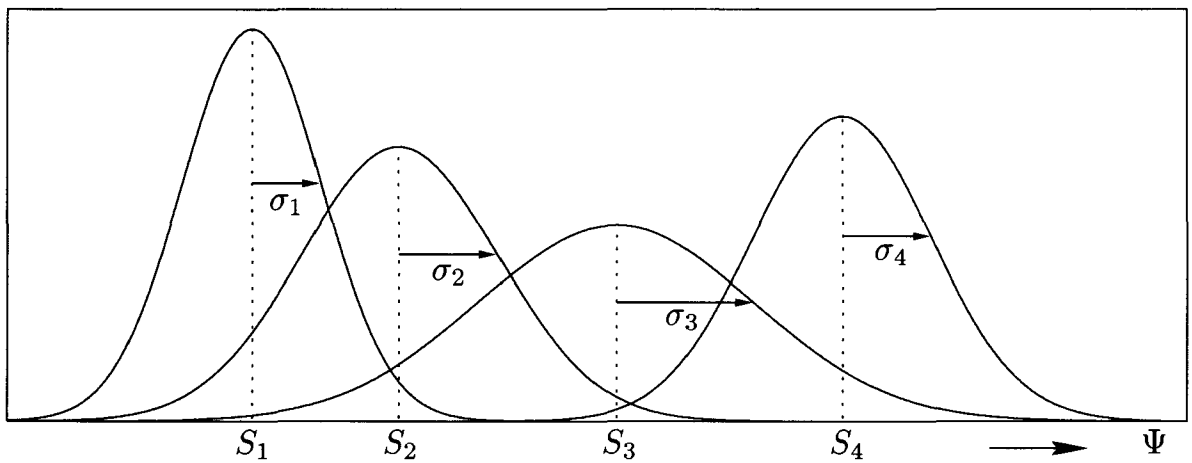

Figure 1. Gaussian probability distributions of the responses of four stimuli on the psychological continuum Ψ.

LIKELIHOOD FUNCTION

For the estimation of the parameters, a likelihood function needs to be derived, using the model assumptions and the outcomes of the scaling experiment: the frequency distribution of scale values f_{ik}. When B_k and B_{k+1} denote the lower and upper bounds of category k, the probability of a score in category k equals

$$p_{ik} = P(B_k < x_i < B_{k+1}) = \Phi\left(\frac{B_{k+1} - S_i}{\sigma}\right) - \Phi\left(\frac{B_k - S_i}{\sigma}\right),$$
$$\forall i = 1 \ldots ns, k = 2 \ldots nc - 1, \quad (4)$$

Φ denoting the standard normal distribution function. For the extreme categories, this is

$$p_{i1} = P(-\infty < x_i < B_1) = \Phi\left(\frac{B_2 - S_i}{\sigma}\right),$$
$$p_{inc} = P(B_{nc} < x_i < \infty) = 1 - \Phi\left(\frac{B_{nc} - S_i}{\sigma}\right). \quad (5)$$

The likelihood function describes the probability of finding results like those that were found in the experiment, given a particular parameter setting. The probability of finding a certain frequency distribution for condition i is described by a multinomial distribution

$$P_i = n_i! \prod_{k=1}^{nc} \frac{p_{ik}^{f_{ik}}}{f_{ik}!}, \quad (6)$$

where n_i denotes the sample size for condition i, or

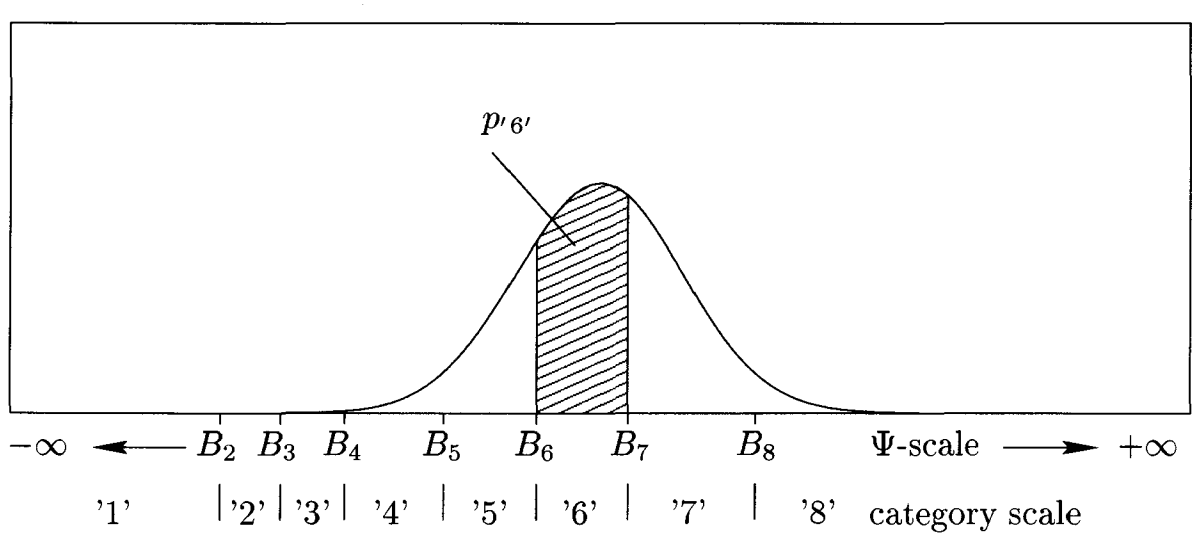

Figure 2. Relation between a category scale and the psychological (Ψ) scale in the case of eight numerical categories. The shaded area indicates the probability that the subject gives a rating in Category 6.

$$n_i = \sum_{k=1}^{nc} f_{ik}. \tag{7}$$

The probability of the entire set of results is then denoted by the likelihood function

$$L = \prod_{i=1}^{ns} P_i = \prod_{i=1}^{ns} \left\{ n_i! \prod_{k=1}^{nc} \frac{p_{ik}^{f_{ik}}}{f_{ik}!} \right\}, \tag{8}$$

which needs to be maximized for optimal parameter setting. For practical reasons, we prefer maximizing the log likelihood function

$$\ln(L) = \sum_{i=1}^{ns} \left\{ \ln(n_i!) + \sum_{k=1}^{nc} \left[f_{ik} \ln(p_{ik}) - \ln(f_{ik}!) \right] \right\}. \tag{9}$$

CONSTRAINTS AND ARBITRARY PARAMETERS IN THURCATD

In ThurCatD, the interval bounds are subject to the ordering constraint

$$B_2 < B_3 < \ldots < B_{nc}. \tag{10}$$

This constraint is checked after each iteration step. If the bound values do not comply with Equation 10, they will be sorted. The user will be notified if and when this occurred in the history output (see the Appendix). This will, however, not often happen, since the start configuration is already compliant with Equation 10 (see the next section). The Ψ-scale unit must be set arbitrarily. In ThurCatD, the noise parameter σ is used as scale unit and is therefore arbitrarily set to 1. Equation 1 becomes

$$x_i \sim N(S_i, 1). \tag{11}$$

Since it is an interval scale, we also need to define an arbitrary origin for the Ψ-scale. A logical choice is to set the mean stimulus position to zero, or

$$\sum_{i=1}^{ns} S_i = 0. \tag{12}$$

After each iteration step, the latest set of S-values will be translated in order to comply with Equation 12.

It is shown that we need to estimate two types of parameters: interval bounds $B_2, \ldots, B_{nc}$) and stimulus scale values ($S_1, \ldots, S_{ns}$). Hence, the total number of parameters is equal to

$$np = ns + nc - 1. \tag{13}$$

ESTIMATION OF PARAMETERS

Iteration With the Method of Gradients

The parameter vector $\vec{\theta}$ is defined by the set of bound parameters $B_k (k = 2 \ldots n_c)$ and the stimulus position parameters $S_i (i = 1 \ldots n_s)$:

$$\vec{\theta} = [S_1\ S_2 \cdot S_{ns}\ B_2 \cdot B_{nc}]^T. \tag{14}$$

ThurCatD applies the method of gradients to estimate the parameters. The basic idea is to maximize the log likelihood function by iteratively changing the parameter vector $\vec{\theta}$ in the direction of the gradient of $\ln L$:

$$\vec{\theta}_{t+1} = \vec{\theta}_t + \alpha_t \frac{\nabla_{\vec{\theta}}(\ln L_t)}{\left\| \nabla_{\vec{\theta}}(\ln L_t) \right\|}, \tag{15}$$

where

$$\nabla_{\vec{\theta}}(\ln L) = \sum_{i=1}^{ns} \sum_{k=1}^{nc} \frac{f_{ik}}{p_{ik}} \nabla_{\vec{\theta}}(p_{ik}) \tag{16}$$

denotes the gradient of the log likelihood function with respect to $\vec{\theta}$. The denominator of the right term of Equation 15 denotes the norm of the gradient vector, and α_t denotes the step size during iteration t. This step size depends on the progress of the iteration process and is determined according to the method proposed by Kruskal (1964), who applied the gradient method for the minimization of stress in a nonmetric multidimensional scaling algorithm. The step size is denoted by

$$\alpha_t = \alpha_{t-1} \cdot (\text{angle factor}) \cdot (\text{relaxation factor}) \cdot (\text{good luck factor}), \tag{17}$$

with

$$\text{angle factor} = 4.0^{(\cos \phi)^{3.0}}, \tag{18}$$

where ϕ denotes the angle between the present and the previous gradient vector. Accordingly

$$\cos \phi = \frac{\nabla_{\vec{\theta}}(\ln L_t) \cdot \nabla_{\vec{\theta}}(\ln L_{t-1})}{\left\| \nabla_{\vec{\theta}}(\ln L_t) \right\| \cdot \left\| \nabla_{\vec{\theta}}(\ln L_{t-1}) \right\|}. \tag{19}$$

The other factors,

$$\text{relaxation factor} = \frac{1.3}{1 + (\text{5-step ratio})^{5.0}}, \tag{20}$$

with

$$\text{5-step ratio} = \min\left(1.0, \frac{\ln L_t}{\ln L_{t-5}}\right), \tag{21}$$

and

$$\text{good luck factor} = \min\left(1.0, \frac{\ln L_t}{\ln L_{t-1}}\right), \tag{22}$$

depend on the history of the log likelihood value.

Trivial Stimulus Conditions and Empty Categories

Before ThurCatD starts the iteration process, it checks the frequency matrix for trivial stimulus conditions and empty categories. A trivial condition is a condition that always scores in one of the two extreme categories. Since the corresponding intervals on the psychological continuum have one infinite boundary, the method of gradients will fail to

converge. Apparently, the information of such a stimulus condition is insufficient for finding a stable maximum likelihood solution. It will be removed from the set before the iteration process starts. Similarly, if a category is found empty—that is, no scores are found there for any condition—it will be removed before the iteration. Whenever these operations occur, they will be logged in the history file.

Start Configuration

The iteration process requires some start configuration, which in principle, could be any set of parameter values that complies with the constraints mentioned above. However, in order to prevent ending up with an estimation vector at a local log likelihood maximum, it is important to start with a parameter vector that presumably lies close to the wanted configuration. In ThurCatD, first order estimates of the set of parameters are calculated before the iteration starts. As an option, the user may apply his/her own start configuration. In that case, it will be read from an additional input file.

Calculation of start configuration. If no start configuration file is present, the initial parameter setting is automatically calculated by ThurCatD. First, arbitrary scale values v_k are assigned to each category with

$$v_k = k, \ \forall k = 1, \ldots, nc. \tag{23}$$

The unnormalized bound values are assumed to be the average scale value of the two categories it separates, or

$$\tilde{B}_k = \frac{1}{2}(v_{k-1} + v_k) = k - \frac{1}{2}. \tag{24}$$

Then, the unnormalized expected stimulus values are calculated by

$$\tilde{S}_i = \frac{1}{n_i}\sum_{k=1}^{nc} k \cdot f_{ik}, \forall i = 1, \ldots, ns. \tag{25}$$

The internal noise is best represented by the pooled *within* standard deviation:

$$SD_{\tilde{S}} = \sqrt{\frac{\sum_{i=1}^{ns}\left(\sum_{k=1}^{nc} f_{ik} \cdot k^2 - \frac{1}{n_i}\left(\sum_{k=1}^{nc} f_{ik} \cdot k\right)^2\right)}{\sum_{i=1}^{ns}(n_i - 1)}}. \tag{26}$$

According to the model, the perceived stimulus values have Gaussian distributions with noise spread σ. With σ arbitrarily set to 1, the normalized values, denoted by

$$\frac{1}{SD_{\tilde{S}}} \cdot \tilde{S}_i, \tag{27}$$

do have unit pooled standard deviation. The start values for the attributes are obtained by translating the scale values in order to comply with Equation 12:

$$\tilde{S}_i = \frac{1}{SD_{\tilde{S}}} \cdot \left(\tilde{S}_i - \frac{1}{ns}\sum_{m=1}^{ns} \tilde{S}_m\right), \forall i = 1, \ldots, ns. \tag{28}$$

Similarly, the start values for the category boundaries are calculated with

$$\hat{B}_k = \frac{1}{SD_{\tilde{S}}} \cdot \left(k - \frac{1}{2} - \frac{1}{ns}\sum_{m=1}^{ns} \tilde{S}_m\right), \forall k = 2, \ldots, nc. \tag{29}$$

Stopping Criteria

The iteration process of ThurCatD is equipped with two stopping criteria. First, iteration is stopped if the number of iteration steps reaches a previously defined maximum. This maximum number of steps is 5,000 by default or may be altered in the options settings. The second stopping criterion is defined by the machine precision of the computations. The iteration is stopped if the difference between the current and the previous value of $\ln L$ is less than or equal to the machine-dependent precision.

Cramér–Rao Bounds on Estimation Errors

After estimation of the parameters, ThurCatD calculates the Cramér–Rao bounds on estimation errors (e.g., van Trees, 1968). These measures, which determine the asymptotic values for the standard error of the estimates, are derived from the Fisher matrix **J**. This symmetrical square ($np \times np$) matrix is also known as the *information matrix*, of which the elements are defined by

$$J_{ij} = -E\left\{\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right\}_{\vec{\theta}_{ml}}, \forall i, j = 1, \ldots np, \tag{30}$$

where $\vec{\theta}_{ml}$ indicates that the elements are calculated at the maximum likelihood configuration. The pseudo inverse $\mathbf{J}^+$ of this Fisher matrix is the asymptotic variance–covariance matrix for the estimated parameters. The diagonal elements of $\mathbf{J}^+$ are the asymptotic parameter variances, and the off-diagonal elements are the covariances of parameter pairs. Hence, the asymptotic standard error of estimate θ_i is given by

$$\hat{\sigma}_{\theta_i} = \sqrt{J_{ii}^+}. \tag{31}$$

The asymptotic standard error is valid for infinite sample sizes. The real standard error is larger in general, and the underestimation depends on the sample size. Figure 3 illustrates the effect of increasing the sample size on the underestimation of the standard error. This figure is the result of a number of Monte Carlo simulations. The output model parameters for the data set of Example 1 (see the Appendix) were used to simulate data sets with various sample sizes. Samples were taken at random from Gaussian distributions that were determined by the scale values in the ThurCatD output for the data set of Example 1. These sample scale values were categorized afterward by applying the category bounds found in the output of the Example 1 data set. For each sample size, 200 data sets were generated and analyzed with ThurCatD. An example of a simulated data set with a sample size of 100 is given by Example 2 in the Appendix. The real standard deviations of the estimates were calculated over the 200 sets of output parameter configurations. These values were compared with the average as-

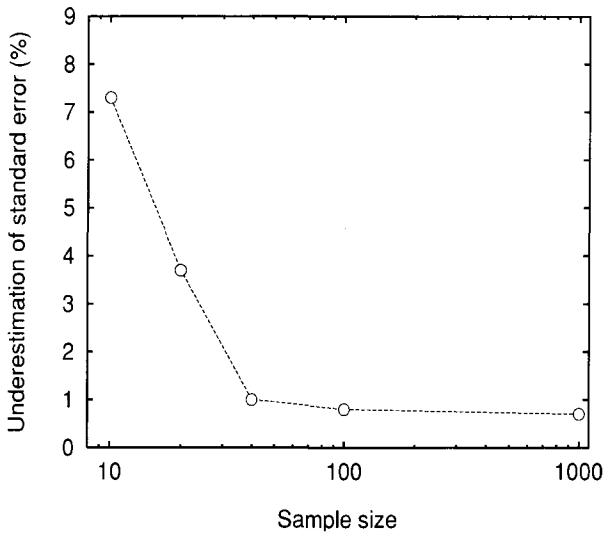

Figure 3. Underestimation of the asymptotic standard error as a function of the sample size. The underestimation is expressed as a percentage of the real standard deviation of estimated scale values as obtained over 200 Monte Carlo simulations.

ymptotic standard errors, which resulted in Figure 3. It shows that in the case of a sample size of 10—as in Example 1—the real standard errors on the estimated scale values are about 7% larger than the asymptotic values.

Disconnected Conditions

Occasionally, the frequency data for a subset of conditions do not have sufficient overlapping categories with the remaining conditions. In those cases, a range of categories (e.g., at the low or high end of the scale) are used exclusively by this subset of conditions, which are therefore indicated to be disconnected from the remaining set of conditions. The algorithm will fail to converge properly, and the output of ThurCatD will show high asymptotic standard errors, the standard error of the disconnected conditions being substantially larger than the others. The output will also show large negative correlations between the disconnected condition and the other conditions. The positions of the disconnected conditions on the Ψ-scale with respect to those of the remaining set can, therefore, not be determined with sufficient accuracy.

Following an arbitrary criterion, the asymptotic standard error $\hat{\sigma}_{S_j}$ of an estimated scale value S_j is considered to be an outlier if it deviates from the average more than three times the standard deviation of $\hat{\sigma}s_i$-values, or

$$\hat{\sigma}_{S_j} - \frac{1}{ns}\sum_{i=1}^{ns}\hat{\sigma}_{S_j} \geq 3\sqrt{\frac{\sum_{i=1}^{ns}\hat{\sigma}_{S_i}^2 - \frac{1}{ns}\left(\sum_{i=1}^{ns}\hat{\sigma}_{S_i}\right)^2}{ns-1}}. \qquad (32)$$

Conditions with an outlying asymptotic standard error are considered to be *suspicious and probably disconnected.* The user is warned about such conditions and is advised to remove them from the input and rerun the analysis.

The aforementioned trivial conditions are special cases of disconnected conditions that are excluded from analysis before the iteration starts. Disconnected conditions are only discovered afterward by checking the statistics of the parameter estimates.

PROBABILITY STRESS AND GOODNESS OF FIT

In ThurCatD, two types of measures are calculated that give an indication about the quality of the model fit. The first measure, which equals the weighted average absolute discrepancy between observed cumulative proportions and model cumulative probabilities, is called the *probability stress.* It is defined by

$$PS = \frac{\sum_{i=1}^{ns}\sum_{k=1}^{nc}\left| fc_{ik} - n_i \cdot pc_{ik} \right|}{nc\sum_{i=1}^{ns} n_i}, \qquad (33)$$

where fc_{ij} denotes the observed cumulative frequencies and pc_{ik} denotes the model cumulative probabilities for cell (i,k). The contribution of each condition is weighted by its sample size n_i. In cases in which the sample size is equal for all conditions ($n_i = n, \forall i = 1 \ldots ns$), the stress measure of Equation 33 resembles the discrepancy measure proposed by Edwards and Thurstone (1952). If *PS* is *small*, it is concluded that the model predicts the experimental data properly. It is hard to define a generalized criterion for this stress measure that indicates the quality of the fit. However, from a number of model simulations, we found an empirical rule of thumb for the probability stress criterion. It depends on the average sample size

$$n_{av} = \frac{1}{ns}\sum_{i=1}^{ns} n_i.$$

It states that whenever

$$PS < \frac{0.15}{\sqrt{n_{av}}}, \qquad (34)$$

one may conclude that the model fits properly.

Mosteller (1951) introduced a χ^2-test for the deviation of the observed proportion and model probability matrices for paired comparison data. He used an inverse-sine transformation of probabilities and proportions in order to obtain a proper χ^2-test variable. Freeman and Tukey (1950) proposed a slightly different arcsin transformation for small samples. In ThurCatD, Mosteller's test variable, using the alternative transformation method, is applied to the individual cell probabilities of the scale categories. A working hypothesis is that these probabilities

are independent. For a sufficient number of categories, this is a reasonable assumption. The test variable is denoted by

$$\chi^2 = \sum_{i=1}^{ns} n_i \sum_{k=1}^{nc-1} \left(\arcsin \sqrt{\frac{fc_{ik}}{n_i + 1}} + \arcsin \sqrt{\frac{fc_{ik} + 1}{n_i + 1}} - 2 \arcsin \sqrt{pc_{ik}} \right)^2 \sim \chi^2[df], \quad (35)$$

where the angles are expressed in radians. The number of degrees of freedom,

$$df = ns(nc - 1) - (np - 1) = (ns - 1)(nc - 2), \quad (36)$$

is determined by the number of assumed independent probability cells [$= ns(nc - 1)$] minus the number of free parameters. The latter equals $(np - 1)$ as the sum of scale values is arbitrarily set to 0. The model fit is assumed to be adequate if the upper tail χ^2-probability > .05. In that case, the set of predicted probabilities do not significantly differ from the set of observed proportions.

SAMPLE RUNS OF ThurCatD

Example 1

In a numerical category scaling experiment performed by Boschman and Roufs (1997), subjects assessed the visual comfort of 19 visual display conditions presented to them. The subjects used a 10-point numerical scale, with Categories 1–10 to express their judgments. The results of one of the subjects are saved in the frequency file mcbtot.tcin, which is listed in Example 1 of the Appendix. This file, which was used as input for ThurCatD, shows for each condition the distribution of 10 ratings over the categories. The history information that ThurCatD produced during the run is also listed in the Appendix. It shows that the first three categories were ignored, since neither of the conditions were rated in these categories. It further shows that two of the conditions (Stimulus 5 and Stimulus 18) were left out from the analysis because they were trivial—that is, they always scored in one of the extreme categories. After removal of these stimuli, Category 10 became an unused category and was consequently also left out from the analysis. The estimated parameters and their asymptotic standard errors are printed to the ThurCatD output file mcbtot.tcout, which is also listed in the Appendix. The probability stress value indicates that the model cell probability deviates on average about 3% from the experimental cell proportions. The Mosteller χ^2 statistic shows that the fit is sufficient.

An alternative and simpler model for interpreting numerical category scaling results is to assume that numerical category labels (in the example, numbers from 1 to 10) can be interpreted as interval data. If this model was adequate, the ThurCatD results should be linearly related to the average scale values. Figure 4 shows a typical relation between average numerical category ratings and ThurCatD scale values. It shows a moderate deviation from linearity at the upper and lower ends of the scale.

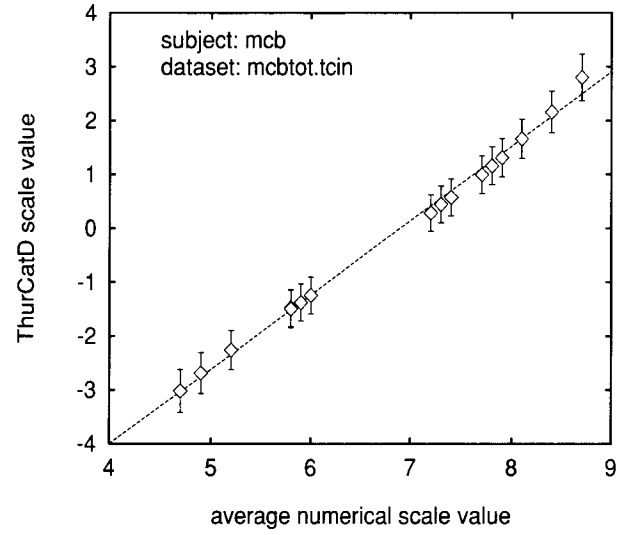

Figure 4. ThurCatD scale values versus average numerical scale values for the 17 remaining conditions of the sample data set listed in the Appendix. In two occasions, symbols (nearly) coincide with those of other conditions. The dotted line is the regression line.

Example 2

The sample data set of Example 2 in the Appendix is a simulated set with a sample size of 100, using the model parameters of Example 1. The output of the ThurCatD analysis shows that the estimated scale values very much resemble those of Example 1, which is of course not surprising, since the data set of Example 2 assumes the same set of values for the model parameters. The asymptotic standard errors, however, are considerably smaller. This is due to the larger sample size.

AVAILABILITY

ThurCatD is an in-house program developed by the author. The Windows 95/98 version of ThurCatD is available for noncommercial use. It can be downloaded (from http://www.ipo.tue.nl/homepages/mboschma/tools.htm).

REFERENCES

BOSCHMAN, M. C., & ROUFS, J. A. J. (1997). Text quality metrics for visual display units: II. An experimental survey. *Displays*, **18**, 45-64.

EDWARDS, A. L., & THURSTONE, L. L. (1952). An internal consistency check for scale values by the method of successive intervals. *Psychometrika*, **17**, 169-180.

FREEMAN, M. F., & TUKEY, J. W. (1950). Transformations related to the angular and the square root. *Annals of Mathematical Statistics*, **21**, 607-611.

GULLIKSEN, H. (1954). A least squares solution for successive intervals assuming unequal standard deviations. *Psychometrika*, **19**, 117-139.

KRUSKAL, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, **29**, 115-129.

MOSTELLER, F. (1951). Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, **16**, 207-218.

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

TORGERSON, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

TUCKER, L. R. (1952). A level of proficiency scale for a unidimensional skill [Abstract]. *American Psychologist*, **7**, 408.

VAN TREES, H. L. (1968). *Detection, estimation and modulation theory: Pt. I. Detection and linear modulation theory*. New York: Wiley.

APPENDIX
Input and Output of ThurCatD

In this appendix, the ThurCatD analysis of a realistic experimental data set (Example 1) and a simulated data set (Example 2) will be demonstrated, and their input and output will be listed.

Example 1

The frequency input file. This file describes the observed data to be analyzed. It consists of a matrix with frequency data indicating how often a condition was rated on a certain category. The file is arranged as follows:

first line: *ns nc*

ns = the number of stimulus conditions
nc = the number of scale categories

following *ns* lines: frequencies (separated by spaces) for each of the *nc* categories.

In this example for 19 display conditions, the visual comfort was rated on a 10-category scale (1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The frequency input file mcbtot.tcin is listed below:

```
19 10
  0 0 0 1 3 2 4 0 0 0
  0 0 0 0 1 1 2 6 0 0
  0 0 0 2 7 1 0 0 0 0
  0 0 0 0 0 0 1 4 5 0
  0 0 0 0 0 0 0 0 0 0
  0 0 0 0 0 1 6 3 0 0
  0 0 0 0 0 1 0 8 1 0
  0 0 0 0 0 0 7 2 1 0
  0 0 0 0 4 4 2 0 0 0
  0 0 0 0 0 0 4 4 2 0
  0 0 0 0 0 1 1 8 0 0
  0 0 0 1 2 5 2 0 0 0
  0 0 0 4 5 1 0 0 0 0
  0 0 0 1 6 3 0 0 0 0
  0 0 0 0 0 0 1 4 5 0
  0 0 0 0 0 1 1 4 4 0
  0 0 0 1 0 7 2 0 0 0
  0 0 0 0 0 0 0 0 0 10
  0 0 0 0 0 0 0 3 7 0
```

History output. The following history information was displayed in a separate window during the run of the ThurCatD analysis of mcbtot.tcin:

```
History for C:\WINDOWS\Desktop\thurcatd\mcbtot.tcin generated by ThurCatD 2.2 on 4/17/00 at 14:44:27.
ThurCatD 2.2 will analyse C:\WINDOWS\Desktop\thurcatd\mcbtot.tcin using up to 5000 iterations.
a0=0.0100, pmin=0.0000010000, extended output=disabled, simple output=disabled
ThurCatD read preference file C:\WINDOWS\Desktop\thurcatd\mcbtot.tcin
Interval1 is empty and will be ignored.
Interval2 is empty and will be ignored.
Interval3 is empty and will be ignored.
Warning: Insufficient statistics in data of stimulus 5
         It will therefore be left out from the analysis!
Warning: Insufficient statistics in data of stimulus 18
         It will therefore be left out from the analysis!
Interval10 is empty and will be ignored.
Likelihood kernel=-1.842804397432E+0002 after nit=100
Stop criterion reached (likelihood step <= precision) after 168 iterations
Likelihood kernel=-1.84280439719198E+0002
ThurCatD created results file C:\WINDOWS\Desktop\thurcatd\mcbtot.tcout
ThurCatD finished successfully!!
```

APPENDIX (Continued)

ThurCatD output. The generated main output file mcbtot.tcout is listed below. The first three intervals (Categories 1, 2, and 3) are found empty and are therefore automatically removed from the analysis. Furthermore, Conditions 5 and 18 are automatically removed from the analysis, since they have trivial frequency distributions. As a result of removing Conditions 5 and 18, Category 10 became empty also and was therefore automatically discarded by ThurCatD. The output shows the estimated parameters and their asymptotic standard errors. It further summarizes the goodness-of-fit analysis by listing the *PS* value and the Mosteller χ^2 test results.

```
Output for C:\WINDOWS\Desktop\thurcatd\mcbtot.tcin generated by ThurCatD 2.2 on 4/17/00 at 14:44:27.

Configuration after 168 iterations:
Estimated parameters (Noise spread parameter arbitrary set to 1):
Scale value parameters:
  stimulus:  1 scale value: -1.3761; S_estimate: 0.3439
  stimulus:  2 scale value:  0.4445; S_estimate: 0.3400
  stimulus:  3 scale value: -2.6858; S_estimate: 0.3809
  stimulus:  4 scale value:  2.1582; S_estimate: 0.3856
  stimulus:  6 scale value:  0.2862; S_estimate: 0.3382
  stimulus:  7 scale value:  1.3110; S_estimate: 0.3553
  stimulus:  8 scale value:  0.5743; S_estimate: 0.3417
  stimulus:  9 scale value: -1.4764; S_estimate: 0.3456
  stimulus: 10 scale value:  1.1623; S_estimate: 0.3519
  stimulus: 11 scale value:  0.9952; S_estimate: 0.3485
  stimulus: 12 scale value: -1.4940; S_estimate: 0.3460
  stimulus: 13 scale value: -3.0145; S_estimate: 0.3973
  stimulus: 14 scale value: -2.2590; S_estimate: 0.3646
  stimulus: 15 scale value:  2.1582; S_estimate: 0.3856
  stimulus: 16 scale value:  1.6610; S_estimate: 0.3652
  stimulus: 17 scale value: -1.2438; S_estimate: 0.3419
  stimulus: 19 scale value:  2.7989; S_estimate: 0.4314

Interval bound parameters:
  Lower bound of interval 5: -3.2187; S_estimate: 0.2573
  Lower bound of interval 6: -1.8084; S_estimate: 0.1828
  Lower bound of interval 7: -0.6285; S_estimate: 0.1482
  Lower bound of interval 8: 0.6406; S_estimate: 0.1491
  Lower bound of interval 9: 2.2349; S_estimate: 0.1963

Intervals:
  interval 4: -infinity ... -3.2187
  interval 5: -3.2187 ... -1.8084
  interval 6: -1.8084 ... -0.6285
  interval 7: -0.6285 ... 0.6406
  interval 8:  0.6406 ... 2.2349
  interval 9:  2.2349 ... +infinity

log likelihood kernel=-1.84280439719198E+0002

Model fit:
  Probability stress=0.033996 -> stress<0.047434
    Model fit is OK according to rule of thumb for probability stress.

  Mosteller Chi-square= 53.0227, 64 Degrees of freedom.
    Upper tail P-value=0.8345 Model fit is OK: (P>0.05).
```

Example 2

In this example the output model parameters from Example 1 were used to simulate scaling data with a sample size of 100 for the remaining 17 conditions in the output of Example 1. The samples were taken from the Gaussian distributions that were determined by the scale values for the data of Example 1 and a standard deviation equal to 1. The samples were categorized afterward by applying the interval bounds of Example 1. In order to enable comparison with the previous example, the two trivial stimuli and the three empty categories were added.

The input file. The resulting frequency input file mc.tcin is listed below.

```
19 10
  0 0 0  1 32 41 24  2  0  0
  0 0 0  0  4 16 39 38  3  0
  0 0 0 30 53 15  2  0  0  0
```

APPENDIX (Continued)

```
  0 0 0   0   0   0   7 48 45   0
  0 0 0   0   0   0   0   0   0 100
  0 0 0   1   2 17 47 29   4   0
  0 0 0   0   0   5 29 54 12   0
  0 0 0   0   0 11 34 45 10   0
  0 0 0   6 36 30 25   3   0   0
  0 0 0   0   0   2 22 61 15   0
  0 0 0   0   0   6 22 56 16   0
  0 0 0   2 35 45 16   2   0   0
  0 0 0 43 47 10   0   0   0   0
  0 0 0 19 48 30   3   0   0   0
  0 0 0   0   0   0   5 50 45   0
  0 0 0   0   0   0 13 55 32   0
  0 0 0   4 28 43 23   2   0   0
  0 0 0   0   0   0   0   0   0 100
  0 0 0   0   0   0   2 28 70   0
```

History output. The following history file was generated during the run of the ThurCatD analysis of mc.tcin.

```
History for C:\WINDOWS\Desktop\thurcatd\mc.tcin generated by ThurCatD 2.2 on 4/17/00 at 14:39:15.
ThurCatD 2.2 will analyse C:\WINDOWS\Desktop\thurcatd\mc.tcin using up to 5000 iterations.
a0=0.0100, pmin=0.0000010000, extended output=disabled, simple output=disabled
ThurCatD read preference file C:\WINDOWS\Desktop\thurcatd\mc.tcin
Interval1 is empty and will be ignored.
Interval2 is empty and will be ignored.
Interval3 is empty and will be ignored.
Warning: Insufficient statistics in data of stimulus 5
         It will therefore be left out from the analysis!
Warning: Insufficient statistics in data of stimulus 18
         It will therefore be left out from the analysis!
Interval10 is empty and will be ignored.
Likelihood kernel=-1.865060957657E+0003 after nit=100
Stop criterion reached (likelihood step <= precision) after 162 iterations
Likelihood kernel=-1.86506095758860E+0003
ThurCatD created results file C:\WINDOWS\Desktop\thurcatd\mc.tcout
ThurCatD finished successfully!!
```

ThurCatD output. The generated main output file mc.tcout is listed below.

```
Output for C:\WINDOWS\Desktop\thurcatd\mc.tcin generated by ThurCatD 2.2 on 4/17/00 at 14:39:15.

Configuration after 162 iterations:
Estimated parameters (Noise spread parameter arbitrary set to 1):
Scale value parameters:
  stimulus:  1 scale value: -1.2550; S_estimate: 0.1077
  stimulus:  2 scale value:  0.2902; S_estimate: 0.1065
  stimulus:  3 scale value: -2.6503; S_estimate: 0.1202
  stimulus:  4 scale value:  2.0668; S_estimate: 0.1207
  stimulus:  6 scale value:  0.1968; S_estimate: 0.1062
  stimulus:  7 scale value:  1.0152; S_estimate: 0.1099
  stimulus:  8 scale value:  0.7516; S_estimate: 0.1084
  stimulus:  9 scale value: -1.4021; S_estimate: 0.1084
  stimulus: 10 scale value:  1.2493; S_estimate: 0.1116
  stimulus: 11 scale value:  1.1545; S_estimate: 0.1109
  stimulus: 12 scale value: -1.4114; S_estimate: 0.1085
  stimulus: 13 scale value: -2.9977; S_estimate: 0.1259
  stimulus: 14 scale value: -2.2469; S_estimate: 0.1151
  stimulus: 15 scale value:  2.1017; S_estimate: 0.1212
  stimulus: 16 scale value:  1.7320; S_estimate: 0.1161
  stimulus: 17 scale value: -1.2978; S_estimate: 0.1079
  stimulus: 19 scale value:  2.7031; S_estimate: 0.1344

Interval bound parameters:
  Lower bound of interval 5: -3.1555; S_estimate: 0.0803
  Lower bound of interval 6: -1.7212; S_estimate: 0.0560
  Lower bound of interval 7: -0.6107; S_estimate: 0.0460
  Lower bound of interval 8:  0.5943; S_estimate: 0.0460
  Lower bound of interval 9:  2.1841; S_estimate: 0.0607
```

APPENDIX (Continued)

```
Intervals:
  interval 4: -infinity ... -3.1555
  interval 5: -3.1555 ... -1.7212
  interval 6: -1.7212 ... -0.6107
  interval 7: -0.6107 ... 0.5943
  interval 8:  0.5943 ... 2.1841
  interval 9:  2.1841 ... +infinity

log likelihood kernel=-1.86506095758860E+0003

Model fit:
  Probability stress=0.007139 -> stress<0.015000
    Model fit is OK according to rule of thumb for probability stress.

  Mosteller Chi-square= 40.8381, 64 Degrees of freedom.
    Upper tail P-value=0.9894 Model fit is OK: (P>0.05).
```

(Manuscript received August 17, 1999;
revision accepted for publication May 23, 2000.)