

Testing the Equality of Scale Values and Discriminal Dispersions in Paired Comparisons

Mark L. Davison and Dennis P. McGuire, University of Minnesota

Tsuey-Hwa Chen, American Guidance Service

Ronald O. Anderson, University of Minnesota

General normal ogive and logistic multiple-group models for paired comparisons data are described. In these models, scale value and discriminial dispersion parameters are allowed to vary across stimuli and respondent populations. Submodels can be fit to choice proportions by nonlinearly regressing sample estimates of choice proportions onto a complex design matrix. By fitting various submodels and by appropriate coding of parameter effects, selected hypotheses about the equality of scale value and dispersion parameters across groups can be tested. Model fitting and hypothesis testing are illustrated using health care coverage data collected in two age groups. *Index terms:* Bradley-Terry-Luce Model, choice models, logistic regression, paired comparisons, probit regression, Thurstone's Law of Comparative Judgment.

McGuire & Davison (1991) described two related, unidimensional multiple-group paired comparisons models—one based on the normal ogive function and the second based on a logistic function. According to the normal ogive model

$$\pi_{jkg} = \left[1 / (2\pi)^{-1/2} \right] \int_0^{L_{jkg}} \exp\left(\frac{-t^2}{2}\right) dt, \quad (1)$$

where π_{jkg} is the probability that stimulus k ($k = 1, \dots, J$) is selected over stimulus j ($j = 1, \dots, J$) in population g ($g = 1, \dots, G$) and $L_{jkg} = x_{kg} - x_{jg}$, where x_{jg} and x_{kg} are scale values for stimuli j and k in population g . Equation 1 is a multiple-group extension of Thurstone's (1927) Law of Comparative Judgment Case V, and it is assumed that the discrim-

inal dispersions, σ_{jg} , are equal for all (j, g).

According to the McGuire & Davison (1991) logistic model,

$$\pi_{jkg} = \frac{\exp(L_{jkg})}{1 + \exp(L_{jkg})}, \quad (2)$$

where π_{jkg} and L_{jkg} are defined as in Equation 1. Equation 2 is a multiple-group extension of the Bradley-Terry-Luce model (BTLM; Bradley & Terry, 1952; Luce, 1959), which also assumes equal discriminial dispersions for all stimuli in all groups. McGuire & Davison (1991) then discussed schemes for dummy coding scale value effects that, when combined with logistic regression, probit analysis, or weighted least squares regression, lead to tests of hypotheses about the equality of scale values across groups.

The McGuire & Davison (1991) multiple-group model is extended here by relaxing the assumption of equal discriminial dispersions for all stimuli and groups. The examples use the logistic form of the model. However, the same ideas are readily extended to the normal ogive. The extension is accomplished by using a more general form of L_{jkg} , and this more general form serves to extend either model, because both models are phrased in terms of L_{jkg} .

By expanding McGuire & Davison's (1991) design matrix, scale value and discriminial dispersion parameters can be estimated by nonlinearly regressing choice proportions onto the design matrix. Fitting appropriate submodels and proper coding of scale values and/or dispersion effects leads to tests of various hypotheses about the equality of scale

value and/or dispersion parameters. Because McGuire and Davison described the coding of scale value effects, only the coding of discriminational dispersion effects and testing hypotheses about discriminational dispersions are discussed here.

The approach presented here leads to two types of hypothesis tests about dispersions. The first type consists of stimulus contrasts—comparisons of variability across stimuli within a respondent population. The second type consists of respondent contrasts—contrasts across respondent populations for a given stimulus.

Purpose

Two variations of the model are presented—logistic and normal ogive. Hierarchically embedded submodels are described using nonlinear regression, and procedures for implementing hypothesis tests about the equality of dispersions across respondent populations and stimuli are presented. Procedures for testing submodels also are described, because fitting all of the parameters in the full model simultaneously through nonlinear regression is impractical due to substantial multicollinearity among the design vectors for stimulus scale values and dispersion parameters. The general model is simply a framework for interrelating various submodels. Nonlinear regression permits fitting of submodels with low to moderate degrees of multicollinearity.

The Multiple-Group Model

Discriminal Dispersions

The methodology of paired comparisons involves repeated trials in which pairs of stimuli are presented, one pair at a time, and the respondent is asked which element of the pair is preferred. Here, a trial means the presentation of one stimulus pair to an individual. For a given stimulus pair, the data obtained consist of the proportion of trials in which one stimulus is selected over the other. The repeated presentations of a stimulus pair may be given to the same individual or to different individuals. Over trials, the perceived location of each stimulus is assumed to vary, and the discriminational dispersion is an index of that variability. If the discriminational dispersion equals 0.0, then the perceived location of

the stimulus is the same for every trial. Larger discriminational dispersions mean greater variability in perceptions of the given stimulus.

If “trial” refers to repeated presentations of stimulus pairs to one person, then the discriminational dispersion is a measure of the intraindividual variability in the perceived location of a stimulus. A dispersion of 0.0 means that the individual is completely consistent in his/her perception of the stimulus; larger dispersions indicate more inconsistency. If the repeated presentations are given to different individuals from the same population, then the discriminational dispersion reflects the consensus among individuals as to the perceived location of the stimulus. A discriminational dispersion of 0.0 means there is complete consensus; larger discriminational dispersions indicate less consensus.

In its most general form, Thurstone’s (1927) Law of Comparative Judgment explicitly allows unequal discriminational dispersions across stimuli. By contrast, no development of the BTLM allows unequal discriminational dispersions across stimuli (e.g., Yellott, 1969). Therefore, the developments here initially use the normal ogive framework. However, the logistic model will be discussed below.

The Normal Ogive Model

According to Thurstone (1927), when a stimulus is presented, it elicits a discriminational process corresponding to the location of the stimulus along the measured attribute. From trial to trial, the stimulus will give rise to varying locations. These varying locations for stimulus j are assumed to be normally distributed with mean x_j (the scale value) and variance σ_j^2 (the dispersion parameter). Thurstone’s concepts are extended to multiple groups by making the scale value and dispersion parameters for a given stimulus j specific to a population g . That is, in population g , the perceived location of stimulus j is assumed normally distributed with mean (or scale value) x_{jg} and variance σ_{jg}^2 (the dispersion parameter).

When two stimuli, j and k , are presented together, each elicits a discriminational process corresponding to two different locations along the judged dimension. In population g , if the two perceived locations for

stimuli j and k are independently and normally distributed, the difference in their perceived locations will be normally distributed with mean $x_{kg} - x_{jg}$ and variance $\sigma_{kg}^2 + \sigma_{jg}^2$. The probability that stimulus j is selected over stimulus k in population g , π_{jkg} , has the form shown in Equation 1 if L_{jkg} is redefined as follows:

$$L_{jkg} = \frac{(x_{kg} - x_{jg})}{(\sigma_{kg}^2 + \sigma_{jg}^2)^{1/2}}. \quad (3)$$

McGuire & Davison's (1991) normal ogive model incorporates a special case of Equation 3 in which the denominator is constrained to equal 1 for all (j, k, g) . This constraint is satisfied if $\sigma_{jg}^2 = .5$ for all (j, g) . Hereafter, Equation 1 together with Equation 3 (i.e., the new definition of L_{jkg}) will be referred to as the "normal ogive, multiple-group model" (NOMGM).

In the NOMGM, the parameters are determined up to an arbitrary choice of origin and unit in each population g . Following common convention, let the origin be defined so that the sum of scale values in each group equals 0.0:

$$\mu_g = \sum_j x_{jg} = 0.0 \text{ for all } g. \quad (4)$$

A corollary of Equation 4 is that the scale value for one of the stimuli in each group, for example Stimulus 1, can be expressed as the negative sum of the remaining stimulus scale values:

$$x_{1g} = -\sum_{j \neq 1} x_{jg}. \quad (5)$$

Let the unit be defined for each group so that the average dispersion parameter equals .5 in each group:

$$\sigma_g^2 = \left(\frac{1}{J}\right) \sum_j \sigma_{jg}^2 = .50 \text{ for all } g. \quad (6)$$

Equations 4 and 6 are the arbitrary constraints imposed without loss of generality to uniquely identify the parameters.

The Logistic Model

A corresponding logistic model of the form

shown in Equation 2 with L_{jkg} defined as in Equation 3 will be referred to as the "logistic, multiple-group model" (LMGM). Equation 2 was derived under the assumption that discriminational processes for individual stimuli have a squared hyperbolic secant distribution with equal discriminational dispersions for all stimuli (e.g., Yellott, 1969). (A third model could be generated from the assumption that differences in discriminational processes follow a double exponential distribution, but this will not be discussed here.) No justification has been found for a logistic model with varying discriminational dispersions across stimuli.

Although no justification has been found for the logistic model in its own right, it can be justified on the basis of its close approximation to the NOMGM. If the NOMGM is expressed in terms of parameter values $(x_{jg}, x_{kg}, \sigma_{jg}, \sigma_{kg})$ and the logistic model is expressed in terms of rescaled parameters $(1.7x_{jg}, 1.7x_{kg}, \sigma_{jg}, \sigma_{kg})$, then according to Haley (1952) the choice probabilities for the two models will differ by no more than .01 for any choice of $(x_{jg}, x_{kg}, \sigma_{jg}, \sigma_{kg})$. Therefore, if the NOMGM fits the data well so will the LMGM after rescaling the scale value parameters by 1.7. The logistic model of McGuire & Davison (1991) is a special case in which the denominator of Equation 3 is constrained equal to 1.0.

Model Fitting

Both the NOMGM and LMGM can be expressed in terms of a complex design matrix and the unknown scale value and dispersion parameters. McGuire & Davison (1991) described the design matrix elements for the special case in which there is only one group and there are no dispersion parameters in the model, and the special case in which there are multiple groups and no dispersion parameters in the model. They then demonstrated how their models can be fit to data using either probit regression (for Thurstone's normal ogive model) or logistic regression (for the BTLM) of the sample choice proportions onto the design matrix. Two practical problems arise in extending the logistic or probit regression approaches from the McGuire and Davison submodels to the more general NOMGM and LMGM.

First, most logistic or probit regression algo-

rithms assume that L_{jkg} is a linear function of parameters. Obviously, L_{jkg} in Equation 3 is a nonlinear function of the unknown scale value and dispersion parameters. Fortunately, nonlinear regression algorithms have been developed for fitting logistic or probit functions in which L_{jkg} is a nonlinear function of parameters.

Second, there can be substantial multicollinearity between the design matrix elements corresponding to scale value parameters and those corresponding to dispersion parameters. As a result, the algorithm may not converge; or if it does, the parameter estimates may have very large standard errors (SEs). Consequently, regressing choice proportions onto design matrix elements using logistic or probit regression appears practical only as a means of fitting submodels in which the multicollinearity is slight to moderate. Large SEs for parameter estimates are an indication that the degree of multicollinearity exceeds tolerable limits.

Although a researcher may simply wish to estimate parameters in a submodel, more often the researcher wants to test various hypotheses about the parameters. This involves three steps: (1) reparameterizing the submodel in terms of contrasts corresponding to the hypotheses, (2) constructing the corresponding design matrix, and (3) regressing the sample choice proportions onto that design matrix. After completing these three steps, the researcher can test the hypotheses using fit statistics, the estimates of contrasts, and the SEs on those contrasts.

Examples

Because the variety of possible submodels and hypotheses is limitless, this process is illustrated using two examples. One example illustrates testing a hypothesis about the equality of dispersion parameters across populations; the other illustrates the testing of a hypothesis about the equality of dispersion parameters across stimuli.

Data

The two populations were the youngest age group, ages 25–34, and the oldest age group, ages 55 and above, from McGuire & Davison (1991). Respondents compared all possible pairs of seven

health care coverage types in terms of importance. The seven types of coverage were preventive, routine, chronic, major nonemergency, major emergency, extraordinary, and custodial/long-term. For the youngest age group, the number of respondents judging each pair ranged from 746 to 760; for the oldest age group, the number judging each pair ranged from 845 to 861. The choice proportions, number of respondents judging each pair, and study details are given in McGuire and Davison. The SYSTAT (Wilkinson, 1988) NONLIN routine was used to obtain maximum likelihood estimates of logistic model parameters.

Example 1: The Proportionality and Equality Submodels

The proportionality submodel. McGuire & Davison (1991) focused on the null hypothesis that scale values were equal in two or more groups. In some cases, the researcher will be more interested in the null hypothesis that the scale values are perfectly correlated in two or more groups; that is, scale values are proportional when expressed as deviations around the population mean scale value.

Because scale values and dispersions have a multiplicative relationship, the proportionality submodel can be viewed in one of two essentially equivalent ways—either discriminial dispersions are equal and scale values are proportional across groups or scale values are equal and discriminial dispersions are proportional across groups. In the proportionality submodel, the scale values in any two groups, x_{jg} and $x_{jg'}$, are related by a constant of proportionality $\sigma_g/\sigma_{g'}$ such that $x_{jg'} = (\sigma_g/\sigma_{g'})x_{jg}$; here σ_g and $\sigma_{g'}$ refer to the discriminial dispersions common to all stimuli in groups g and g' , respectively. If the proportionality hypothesis holds, but not strict equality of scale values, the rank order of stimuli is the same in all groups, but for any stimulus pair, j and k , the strength of the preference for k over j varies across populations.

In the proportionality submodel, because scale values for a given stimulus are proportional across all populations and because dispersions are equal across stimuli within a population, L_{jkg} may be written as

$$L_{jkg} = \frac{(x_k - x_j)}{(2\sigma_g^2)^{1/2}} \tag{7}$$

To define the unit, let the dispersion in the average group equal .5 [i.e., $\sigma^2 = (1/G)\sum_g \sigma_g^2 = .5$]. Then the dispersion in any particular group can be written as $\sigma_g^2 = .5 + \sigma_g^{*2}$; that is, written in terms of a contrast between the dispersion in group g and the dispersion in the average group, $\sigma_g^{*2} = \sigma_g^2 - .5$. Because the dispersion in the average group equals .5, the dispersion in one of the populations—say Population 1—can be written in terms of the remaining dispersion contrasts as $.5 + \sum_{g \neq 1} \sigma_g^{*2}$.

Equation 7 then can be written in terms of unknown stimulus scale values x_i ($i = 2, \dots, J$), known design elements corresponding to stimulus scale values d_i ($i = 2, \dots, J$), unknown dispersion contrasts σ_h^{*2} ($h = 2, \dots, G$), and known design elements corresponding to dispersion contrasts c_h ($h = 2, \dots, G$):

$$L_{jkg} = \frac{\left(\sum_{i \neq 1} x_i d_i \right)}{\left(1 + \sum_{h \neq 1} c_h \sigma_h^{*2} \right)^{1/2}} \tag{8}$$

McGuire & Davison (1991) explained the coding of the scale value design elements d_i and the relationship between terms in the numerators of Equations 7 and 8. Design elements corresponding to population dispersion parameters are defined as

$$\begin{aligned} c_h &= 2 \text{ if } h = g \text{ and } g \neq 1 \\ &= -2 \text{ if } g = 1 \\ &= 0 \text{ otherwise.} \end{aligned} \tag{9}$$

Table 1 shows the design elements for the two groups and the seven stimuli in this example. Because there were only two groups, and hence one dispersion contrast, the subscript was dropped from the design element variable c and the dispersion contrast σ^* . Each row of Table 1 corresponds to a choice proportion, π_{jkg} . The columns labeled j , k , and g indicate the stimuli and group to which a row corresponds. The design elements for scale values are shown in the columns labeled $d_2 - d_7$.

Each stimulus can be represented by a vector of design elements ($d_2, d_3, d_4, d_5, d_6, d_7$). The elements $d_2 - d_7$ in a row corresponding to stimuli j and k are constructed from the design vectors for stimuli j and k by subtracting the design vector for j from the design vector for k . Because the scale value for Stimulus 1 can be expressed as the negative sum of scale values for the other stimuli, $x_1 = -\sum_{i \neq 1} x_i$, the vector for Stimulus 1 is a vector of -1 s ($-1 -1 -1 -1 -1 -1$). For all other stimuli, $j \neq 1$, the vector consists of 0s for all elements except $d_j = 1$; for example, for Stimulus 2 the vector is $(1 \ 0 \ 0 \ 0 \ 0 \ 0)$.

For a row corresponding to stimuli j and k , for which $L_{jkg} = x_{kg} - x_{jg}$, the design elements are constructed by subtracting the vector for stimulus j from the vector for k . For example, the first row in which $j = 2$ and $k = 1$ was constructed by subtracting the row vector for $j = 2$ $(1 \ 0 \ 0 \ 0 \ 0 \ 0)$ from the vector for $k = 1$ $(-1 -1 -1 -1 -1 -1)$ to yield $(-2 -1 -1 -1 -1 -1)$. Similarly, the row for which $j = 7$ and $k = 6$ was constructed by subtracting the vector for Stimulus 7 $(0 \ 0 \ 0 \ 0 \ 0 \ 1)$ from the row vector for Stimulus 6 $(0 \ 0 \ 0 \ 0 \ 1 \ 0)$ to yield $(0 \ 0 \ 0 \ 0 \ 1 -1)$.

The design element c corresponds to the unknown dispersion contrast σ^{*2} . More precisely, c equals the coefficient on σ^{*2} in the denominator of Equation 8. Specifically, the element $c = -2$ for Population 1 because the denominator in Equation 8 has the form $(1 - 2\sigma^{*2})^{1/2}$ for Population 1. For Population 2, the denominator in Equation 8 has the form $(1 + 2\sigma^{*2})^{1/2}$, and hence $c = 2$ for Population 2.

Table 2 shows the results obtained by regressing the 21 sample choice proportions shown in the appendix of McGuire & Davison (1991, p. 182) for their youngest and oldest age groups (Populations 1 and 2) onto the design elements in Table 1. The “weights” corresponding to Stimuli 2–7 are the regression weights corresponding to predictors $d_2 - d_7$ and are maximum likelihood estimates of corresponding stimulus scale values. According to Equation 5, $x_1 = -\sum_{i \neq 1} x_i$; therefore, the scale value for Stimulus 1 ($-.23$) was computed as the negative sum of the other scale value estimates. The value listed in the “weight” column for σ^{*2} is the regression weight on the predictor c and is a

Table 1
 Design Matrix for the Proportionality Submodel

<i>j</i>	<i>k</i>	<i>g</i>	Design Elements							Disper- sion (<i>c</i>)
			Scale Values							
			<i>d</i> ₂	<i>d</i> ₃	<i>d</i> ₄	<i>d</i> ₅	<i>d</i> ₆	<i>d</i> ₇		
2	1	1	-2	-1	-1	-1	-1	-1	-2	
3	1	1	-1	-2	-1	-1	-1	-1	-2	
4	1	1	-1	-1	-2	-1	-1	-1	-2	
5	1	1	-1	-1	-1	-2	-1	-1	-2	
6	1	1	-1	-1	-1	-1	-2	-1	-2	
7	1	1	-1	-1	-1	-1	-1	-2	-2	
3	2	1	1	-1	0	0	0	0	-2	
4	2	1	1	0	-1	0	0	0	-2	
5	2	1	1	0	0	-1	0	0	-2	
6	2	1	1	0	0	0	-1	0	-2	
7	2	1	1	0	0	0	0	-1	-2	
4	3	1	0	1	-1	0	0	0	-2	
5	3	1	0	1	0	-1	0	0	-2	
6	3	1	0	1	0	0	-1	0	-2	
7	3	1	0	1	0	0	0	-1	-2	
5	4	1	0	0	1	-1	0	0	-2	
6	4	1	0	0	1	0	-1	0	-2	
7	4	1	0	0	1	0	0	-1	-2	
6	5	1	0	0	0	1	-1	0	-2	
7	5	1	0	0	0	1	0	-1	-2	
7	6	1	0	0	0	0	1	-1	-2	
2	1	2	-2	-1	-1	1	-1	-1	2	
3	1	2	-1	-2	-1	-1	-1	-1	2	
4	1	2	-1	-1	-2	-1	-1	-1	2	
5	1	2	-1	-1	-1	-2	-1	-1	2	
6	1	2	-1	-1	-1	-1	-2	-1	2	
7	1	2	-1	-1	-1	-1	-1	-2	2	
3	2	2	1	-1	0	0	0	0	2	
4	2	2	1	0	-1	0	0	0	2	
5	2	2	1	0	0	-1	0	0	2	
6	2	2	1	0	0	0	-1	0	2	
7	2	2	1	0	0	0	0	-1	2	
4	3	2	0	1	-1	0	0	0	2	
5	3	2	0	1	0	-1	0	0	2	
6	3	2	0	1	0	0	-1	0	2	
7	3	2	0	1	0	0	0	-1	2	
5	4	2	0	0	1	-1	0	0	2	
6	4	2	0	0	1	0	-1	0	2	
7	4	2	0	0	1	0	0	-1	2	
6	5	2	0	0	0	1	-1	0	2	
7	5	2	0	0	0	1	0	-1	2	
7	6	2	0	0	0	0	1	-1	2	

maximum likelihood estimate of the dispersion contrast parameter σ^{*2} .

The hypothesis that the dispersion parameters are equal in the two groups is equivalent to the hypothesis that the dispersion contrast parameter

equals 0.0 ($\sigma^{*2} = 0.0$). This can be tested with a *t* statistic obtained by dividing the estimate of σ^{*2} by its asymptotic SE (ASE): $t = .004/.026 = .15$, which would not be significant at any conventional significance level. Thus, the hypothesis that the dis-

Table 2
 Regression Weights, ASEs, and Scale Values for the
 Seven Stimuli and Dispersion Parameter Estimates
 From Fitting the Proportionality Model

Stimulus and Dispersion	Weight	ASE	Scale Value
Stimulus			
1			-.23
2	.10	.02	.10
3	-.18	.02	-.18
4	-.18	.02	-.18
5	1.36	.02	1.36
6	-.68	.02	-.68
7	-.20	.02	-.20
Dispersion			
σ^{*2}	-.004	.026	
σ_1^2			.496
σ_2^2			.504

dispersion parameters are equal in the two groups is retained.

The equality submodel. As a further test of the hypothesis, a submodel was fit in which the dispersion contrast parameter was constrained equal to 0.0 by regressing the choice proportions onto design variables $d_2 - d_7$ only (i.e., dropping design variable c). This model, the equality submodel, is a hierarchically embedded submodel of the proportionality submodel and corresponds to the hypothesis that scale values are equal across the two groups. For the proportionality and equality submodels, the difference in the likelihood ratio fit statistics, which is approximately distributed as χ^2 with 1 degree of freedom, was $\chi^2 = .02$, which would not be significant at any conventional level. Like the t statistic for the hypothesis $\sigma^{*2} = 0.0$, this χ^2 statistic would not lead to rejection of the hypothesis that the dispersion parameters are equal in the two groups in favor of the model that scale values are proportional.

Whereas the equality submodel is hierarchically embedded in the proportionality submodel, the proportionality submodel is itself hierarchically embedded in the model of Equation 2 in which L_{jkg} is defined as below Equation 1. In Equation 2, scale values are allowed to vary freely across groups rather than being constrained to be proportional

across groups. The difference in the likelihood ratio fit statistics for the model of Equation 2 and the proportionality submodel was $\chi^2(5) = 1,088$. This statistic would lead to rejection of the hypothesis that scale values are proportional (perfectly correlated) across groups in favor of the more general multiple-group model in Equation 2 in which scale values vary freely across groups.

The model of Equation 2 also fit significantly better than the equality submodel [$\chi^2(6) = 1088.02$, $p < .01$]. Thus, after examining the fit of the equality submodel, the proportionality submodel, and the model of Equation 2, it can be concluded that the scale values differ across groups but the form of the differences is not captured by the proportionality submodel. The data, therefore, suggest that scale values are neither equal nor perfectly correlated (proportional) across the two groups.

Example 2: The Equality of Dispersions Across Stimuli

The model. McGuire & Davison (1991) found highly significant differences in scale values across the five age groups. The scale values for Stimuli 1, 3, 4, and 7 varied more across age groups than did the scale values for Stimuli 2, 5, and 6. This finding led to the following research question. Because the perceptions of Stimuli 1, 3, 4, and 7 varied more across groups, did they also vary more within groups? That is, were the dispersions for Stimuli 1, 3, 4, and 7 equal to those for Stimuli 2, 5, and 6? Thus, a model in which L_{jkg} has the following form was posited:

$$L_{jkg} = \frac{(x_{kg} - x_{jg})}{(\sigma_k^2 + \sigma_j^2)^{1/2}} \tag{10}$$

This model assumes there were two distinct dispersion parameters, one for Stimuli 1, 3, 4, and 7, and a second dispersion parameter for Stimuli 2, 5, and 6. Also, for each stimulus, this model allows the scale value in Population 1 to differ from its scale value in Population 2.

The design matrix for this model is shown in Table 3. The coding of the scale value design elements was explained in detail by McGuire &

Table 3
 Design Matrix for the Model Allowing Variability in Dispersions Across Stimuli

<i>j</i>	<i>k</i>	<i>g</i>	Design Elements													Dispersion (<i>c</i>)
			Scale Values													
			<i>d</i> ₂₁	<i>d</i> ₃₁	<i>d</i> ₄₁	<i>d</i> ₅₁	<i>d</i> ₆₁	<i>d</i> ₇₁	<i>d</i> ₂₂	<i>d</i> ₃₂	<i>d</i> ₄₂	<i>d</i> ₅₂	<i>d</i> ₆₂	<i>d</i> ₇₂		
2	1	1	-2	-1	-1	-1	-1	-1	0	0	0	0	0	0	.25	
3	1	1	-1	-2	-1	-1	-1	-1	0	0	0	0	0	0	-1.50	
4	1	1	-1	-1	-2	-1	-1	-1	0	0	0	0	0	0	-1.50	
5	1	1	-1	-1	-1	-2	-1	-1	0	0	0	0	0	0	.25	
6	1	1	-1	-1	-1	-1	-2	-1	0	0	0	0	0	0	.25	
7	1	1	-1	-1	-1	-1	-1	-2	0	0	0	0	0	0	-1.50	
3	2	1	1	-1	0	0	0	0	0	0	0	0	0	0	.25	
4	2	1	1	0	-1	0	0	0	0	0	0	0	0	0	.25	
5	2	1	1	0	0	-1	0	0	0	0	0	0	0	0	2.00	
6	2	1	1	0	0	0	0	-1	0	0	0	0	0	0	2.00	
7	2	1	1	0	0	0	0	0	-1	0	0	0	0	0	.25	
4	3	1	0	1	-1	0	0	0	0	0	0	0	0	0	-1.50	
5	3	1	0	1	0	0	-1	0	0	0	0	0	0	0	.25	
6	3	1	0	1	0	0	0	-1	0	0	0	0	0	0	.25	
7	3	1	0	1	0	0	0	0	-1	0	0	0	0	0	-1.50	
5	4	1	0	0	1	-1	0	0	0	0	0	0	0	0	.25	
6	4	1	0	0	1	0	0	-1	0	0	0	0	0	0	.25	
7	4	1	0	0	1	0	0	0	-1	0	0	0	0	0	-1.50	
6	5	1	0	0	0	1	-1	0	0	0	0	0	0	0	2.00	
7	5	1	0	0	0	1	0	-1	0	0	0	0	0	0	.25	
7	6	1	0	0	0	0	1	-1	0	0	0	0	0	0	.25	
2	1	2	-2	-1	-1	-1	-1	-1	-2	-1	-1	-1	-1	-1	.25	
3	1	2	-1	-2	-1	-1	-1	-1	-1	-2	-1	-1	-1	-1	-1.50	
4	1	2	-1	-1	-2	-1	-1	-1	-1	-1	-2	-1	-1	-1	-1.50	
5	1	2	-1	-1	-1	-2	-1	-1	-1	-1	-1	-2	-1	-1	.25	
6	1	2	-1	-1	-1	-1	-2	-1	-1	-1	-1	-1	-2	-1	.25	
7	1	2	-1	-1	-1	-1	-1	-2	-1	-1	-1	-1	-1	-2	-1.50	
3	2	2	1	-1	0	0	0	0	1	-1	0	0	0	0	.25	
4	2	2	1	0	-1	0	0	0	1	0	-1	0	0	0	.25	
5	2	2	1	0	0	-1	0	0	1	0	0	-1	0	0	2.00	
6	2	2	1	0	0	0	0	-1	1	0	0	0	-1	0	2.00	
7	2	2	1	0	0	0	0	-1	1	0	0	0	0	-1	.25	
4	3	2	0	1	-1	0	0	0	0	1	-1	0	0	0	-1.50	
5	3	2	0	1	0	-1	0	0	0	1	0	-1	0	0	.25	
6	3	2	0	1	0	0	0	-1	0	1	0	0	-1	0	.25	
7	3	2	0	1	0	0	0	0	0	1	0	0	0	-1	-1.50	
5	4	2	0	0	1	-1	0	0	0	0	1	-1	0	0	.25	
6	4	2	0	0	1	0	0	-1	0	0	1	0	-1	0	.25	
7	4	2	0	0	1	0	0	-1	0	0	1	0	0	-1	-1.50	
6	5	2	0	0	0	1	-1	0	0	0	0	1	-1	0	2.00	
7	5	2	0	0	0	1	0	-1	0	0	0	1	0	-1	.25	
7	6	2	0	0	0	0	1	-1	0	0	0	0	1	-1	.25	

Davison (1991) as Form I coding. The scale value design elements, *d*, can be divided into four submatrices, as shown in Table 3. The submatrix in the upper right corner consists of all 0s. The remaining three submatrices are all equal to each other and to

the scale value design elements, *d*, in Table 1.

In Table 3, the design element *c* is a coefficient on a re-expression of the denominator in Equation 10. That denominator can be written as $(1 + c\sigma^{*2})^{1/2}$ where σ^{*2} is a contrast parameter proportional to the

difference in the dispersion parameter for Stimuli 1, 3, 4, and 7 and that for Stimuli 2, 5, and 6. Because the average dispersion parameter is assumed to be .5, the dispersion for Stimuli 1, 3, 4, and 7 can be written as $.5 - .75\sigma^{*2}$ (the multiplier .75 results from there being four stimuli in one set and three in the other), and the dispersion for Stimuli 2, 5, and 6 can be written as $.5 + \sigma^{*2}$. Because the denominator in Equation 10 is the square root of the sum of dispersion parameters for stimuli j and k , the denominator can be written as $(1 - 1.5\sigma^{*2})^{1/2}$ if both j and k are contained in stimulus set {1, 3, 4, 7}, as $(1 + 2\sigma^{*2})^{1/2}$ if j and k are contained in stimulus set {2, 5, 6}, and as $(1 + .25\sigma^{*2})^{1/2}$ if one member of the pair j and k comes from the set {1, 3, 4, 7} and the other from the set {2, 5, 6}. Hence, $c = -1.5, 2$, or $.25$ depending on whether j and k both come from set {1, 3, 4, 7}, both come from set {2, 5, 6}, or one comes from set {1, 3, 4, 7} and the other from set {2, 5, 6}.

Results. Table 4 contains the results from fitting the model to the data. The weights in Group 1 are the regression weights on the parameters d_{j1} and are maximum likelihood estimates of scale values for Stimuli 2-7 in Group 1. Because $x_{i1} = -\sum_{i \neq 1} x_i$, the scale value for Stimulus 1 in Group 1 was estimated to be the negative sum of scale values for Stimuli 2-7 in Group 1. Given the coding of d in Group 2 [see McGuire & Davison (1991) for more details], the weights in Group 2 are regression weights on the parameters d_{j2} , and are maximum likelihood estimates of the contrast $x_j^* = x_{j2} - x_{j1}$; in other words, a maximum likelihood estimate of the difference in the scale value of stimulus j in Groups 1 and 2. The ASEs associated with the weights in Group 2 can be used to form a t ratio for testing the null hypothesis that the contrast $x_j^* = 0.0$; that is, the null hypothesis that the scale value for stimulus j is equal in the two populations.

The dispersion weight is a maximum likelihood estimate of the dispersion contrast parameter σ^{*2} (-.05). The null hypothesis $\sigma^{*2} = 0.0$ can be tested by taking the estimate of σ^{*2} and dividing it by its ASE: $t = -.05 / .03 = -1.67$ ($p > .05$). This t ratio did not lead to rejection of the null hypothesis that the dispersion parameters are equal for the two stimu-

Table 4
 Regressions Weights, ASEs, and Scale Values
 for the Seven Stimuli and Dispersion Parameter
 Estimates From Fitting the Model
 With Varying Dispersions Across Stimuli

Group, Stimulus, and Dispersion	Weight	ASE	Scale Value
Group 1			
Stimulus			
1			.27
2	.37	.03	.37
3	-.63	.03	-.63
4	.24	.03	.24
5	1.31	.04	1.31
6	-.72	.03	-.72
7	-.84	.03	-.84
Group 2			
Stimulus			
1			-.64
2	-.50	.04	-.13
3	.82	.04	.19
4	-.78	.04	-.54
5	.14	.05	1.46
6	.05	.04	-.67
7	1.17	.04	.33
Dispersion			
σ^{*2}	-.05	.03	
$\sigma^2(1, 3, 4, 7)$.46
$\sigma^2(2, 5, 6)$.55

lus sets {1, 3, 4, 7} and {2, 5, 6}. As a further test of the hypothesis, the model of Equation 2 (in which the dispersion contrast parameter is constrained equal to 0.0) also was fit by regressing the choice proportions onto the predictors d_{21}, \dots, d_{72} (i.e., deleting predictor c). The difference in the likelihood ratio fit statistics for models with and without c was $\chi^2(1) = 3.256$ ($p > .05$), leading to acceptance of the null hypothesis that the models with and without the constraint that $\sigma^{*2} = 0.0$ fit the data equally well. The analysis failed to support the conjecture that the stimuli with the most variation in perceived location across the age groups also had more variation in their perceived locations within groups.

Discussion

Although there are advantages to a probit/logistic regression approach, one limitation concerns dependencies potentially arising in some paired com-

parisons designs. When each respondent judges more than one pair and each pair is judged by several individuals, there can be dependencies across pairs. That is, errors on the judgment of pair (i, j) may not be independent of errors on pair (i, j') by that same individual. The dependencies may result from sequential judgment effects, systematic individual differences in perceived stimulus locations, or other factors. The logistic/probit regression approach requires independent trials. Takane (1989) has extended Thurstone's normal ogive model to develop an analysis of covariance structures for paired comparisons data from a single population, an analysis that relaxes the assumption of independent trials. The model itself is readily extended to more than one population, although the feasibility of an analysis based on the model remains to be explored, particularly when the model is extended to more than one population. Furthermore, some methods of fitting the model require distributional assumptions that have yet to be examined.

The BTLM and Thurstone Case V model for paired comparisons data are special cases of very general logistic and probit regression models in statistics. Although this fact has long been appreciated by statisticians (e.g., Cox & Snell, 1989), it has been little recognized by psychometricians. This is rather ironic, because psychophysical experiments of the 1800s led to both Thurstone's Law of Comparative Judgment and early probit analyses. Once the connection between paired comparisons models and logistic/probit regression is recognized, the recent advances in logistic/probit analyses can be used on paired comparisons data.

The advances are important in several respects. First, the advances mean that many paired comparisons models can be fit to data with readily available software. Most of the major statistical packages contain some form of logistic/probit regression. Some readily available, nonlinear regression algorithms also are capable of fitting logistic/probit models. These nonlinear regression algorithms are particularly important in the fitting of paired comparisons models with dispersion parameters, because in such models L_{jkg} is a nonlinear function of parameters.

Second, probit and logistic regression contain powerful hypothesis testing techniques. Because of these techniques, paired comparisons data are useful for far more than simply scaling stimuli in a given population. They can be used to test hypotheses about the equality of scale values across groups (McGuire & Davison, 1991) and selected hypotheses about the equality of dispersion parameters across stimuli and across populations. Probit and logistic regression include tests of model fit.

Taking advantage of the connection between paired comparisons models and logistic or probit regression is not without its problems. Fitting paired comparisons models with logistic or probit regression involves design matrices that are cumbersome to construct, as illustrated in Tables 1 and 3. The number of predictors is often too large for many maximum likelihood logistic or probit algorithms; however, this often can be overcome by using a weighted least squares approximation (McGuire & Davison, 1991), which can be implemented with many major multiple regression programs. Multicollinearity appears to be a potential problem in fitting models with dispersion parameters. Despite these problems, however, recognizing the link between paired comparisons models and logistic or probit regression greatly expands the research potential of paired comparisons data.

References

- Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39, 324-345.
- Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). New York: Chapman and Hall.
- Haley, D. C. (1952). *Estimates of the dosage mortality relationship when the dose is subject to error* (Technical Rep. No. 15). Stanford CA: Stanford University Applied Mathematics and Statistics Laboratory.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- McGuire, D. P., & Davison, M. L. (1991). Testing group differences in paired comparisons data. *Psychological Bulletin*, 110, 171-182.
- Takane, Y. (1989). Analysis of covariance structures and probabilistic binary choice data. In G. de Soete, H. Feger, & K. C. Klauer (Eds.), *New developments in psychological choice modeling* (pp. 139-160). New York: North-Holland.

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

Wilkinson, L. (1988). *SYSTAT: The system for statistics*. Evanston IL: SYSTAT, Inc.

Yellott, R. (1969). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution.

Journal of Mathematical Psychology, 15, 109–144.

Author's Address

Send requests for reprints or further information to Mark L. Davison, Department of Educational Psychology, University of Minnesota, 178 Pillsbury Dr. S.E., Minneapolis MN 55455, U.S.A. Internet: mld@vx.cis.umn.edu.