ON CONSTRUCTING INTERVAL SCALES USING DATA RESULTING FROM CATEGORICAL JUDGMENTS

Glenn F. Lindsay
Naval Postgraduate School

September, 1981

CATEGORICAL JUDGEMENTS: THE METHOD

OF SUCCESSIVE INTERVALS

A frequently used means of obtaining ratings from judges is that of categorical judgement, wherein judges assign instances to ranked categories. For example, corporate bonds may be rated as A,AA, and so on; student opinion forms ask the student to rate an instructor as poor, fair, average, excellent, or outstanding; pollsters often ask people to check one of a set of categories described as strongly agree, agree, no opinion, disagree, and strongly disagree. When an instructor assigns a students letter grades, he may be viewed as making a categorical judgement in that the possible grades are the categories and the students are the instances. Other examples of ranked categories are found in such diverse applications as restaurant sanitation ratings, military officer fitness reports, and motion picture ratings (G, PG, R, X). Usually, there are descriptors associated with each category which serve to help the judge with his rating task.

The method described in this paper is a scaling method which uses categorical ratings provided by judges, and constructs an interval scale which includes not only the instances but also the bounds between the categories.[1,3,4,5] Thus descriptive benchmarks appear on the final scale. Typically, five categories are used.[2] No assumptions are made about the relative interval sizes for the categories. The categories are understood to be a mutually exclusive set of successive intervals which collectively exhaust the property continuum.

Data Assembly

A direct way to aggregate categorical ratings of instances by judges is through a frequency array, with a row for each of the n instances and a column for each of the m categories. Thus we would describe an entry in this raw frequency array as f_{ij}, denoting the number of judges who rated instance i in category j. Columns in the f_{ij} array should be arranged in ascending order of category value, so that the category representing the least amount of the property is Column 1, and the category representing the greatest amount of the property is Column m. It is not necessary for a judge to rate all instances.

Working with the f_{ij} array, we may cumulate values in each row rightward and divide by the row total to achieve a relative cumulative frequency array p_{ij}, where p_{ij} is the proportion of judges rating instance i who rated it in or below category j. Since the values in the right-hand column of the p_{ij} array will always be 1.0, this column may be omitted for computational purposes, yielding a p_{ij} array with n rows and m-1 columns.

In the example given below, 80 judges were asked to rate four political candidates in terms of their "potential effectiveness as President of the USA." The categories were Very Ineffective, Ineffective, Marginal, Effective, and Very Effective. Raw frequency data is shown in Table 1.

TABLE 1. Raw Frequency Array

Candidate	Potential Effectiveness				
f_{ij}	very ineffective	ineffective	marginal	effective	very effective
A	10	20	27	21	2
B	4	30	35	11	0
C	20	43	15	2	0
D	3	2	34	30	11

From the raw frequency array, the p_{ij} array may be constructed as in Table 2.

TABLE 2. Cumulated Relative Frequency Array

Candidate	Potential Effectiveness			
p_{ij}	very ineffective	ineffective	marginal	effective
A	0.1250	0.3750	0.7125	0.9750
B	0.0500	0.4250	0.8625	1.0000
C	0.2500	0.7825	0.975	1.0000
D	0.0375	0.0625	0.4875	0.8625

$\frac{\Sigma A_i}{\Sigma\Sigma A_{ij}}$

The results in Table 2 say, for example, that 71.25% of judges found A no better than marginal. (Another way of looking at the values in a p_{ij} array would view the columns as upper bounds on adjacent categories, and thus we would say that 71.25% of judges placed Candidate A below the upper bound of the marginal category) This is the interpretation we will use in the work to come. Note that category

"Effective" will have an upper bound but the highest category, "Very Effective", will not. Similarly, the lowest category will have no lower bound.

Relationship to Paired Comparisons (Optional Section)

It is possible to view the p_{ij} array as arising from a ranking procedure such as the method of paired comparisons. Suppose we wished to scale n instances by having judges perform a ranking task such as that of paired comparisons. To the set of n actual instances, we add m-1 reference points whose ranking is already known to both us and the judges. This yields an enlarged set of n+m-1 instances to be compared, and the problem could be represented as an (n+m-1) by (n+m-1) array of comparisons divided into blocks as shown in Figure 1.

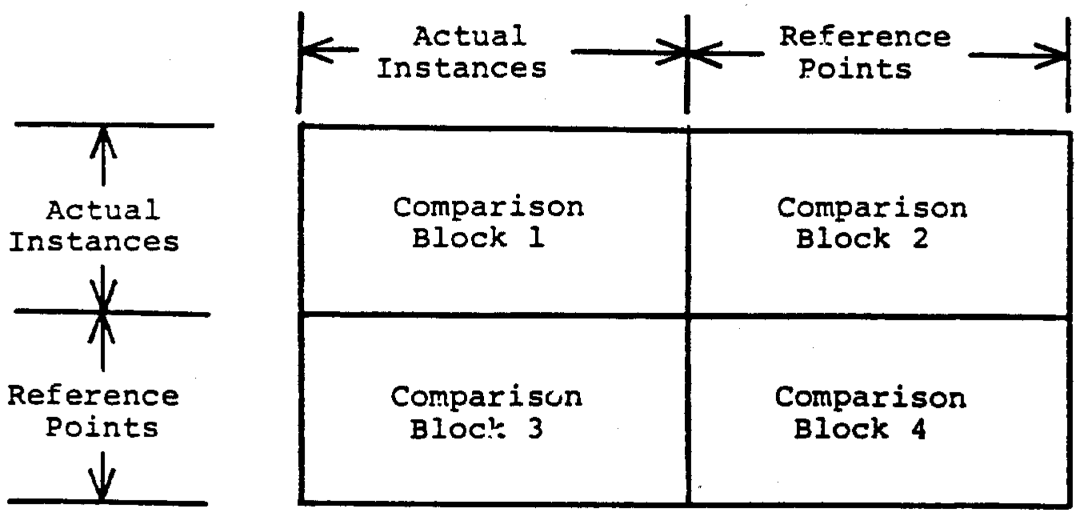


Figure 1. Subdivided Paired-Comparison Array

If judges ranked reference points along with instances, we could use the comparative scaling procedure to locate both the instances and the reference points on the same interval scale. However, since the rank order of the reference points is known, there is no need to ask judges to make the comparisons in Comparison Block 4. Block 3 is the complement of Block 2, and thus we would only need comparisons in Block 1 and Block 2.

Suppose further that in order to spare judges effort, we do not ask them to make comparisons in Block 1; that is, they will not compare actual instances against actual instances. This is legitimate since the paired-comparisons approach does not require that every instance be compared with every other instance.

We are then left with Block 2, which in the p_{ij} array of the paired comparisons approach will represent the proportion of judges who rated instance i as having less of the property than reference j. If the ranked reference points are interpreted as boundaries of adjacent, ranked categories, then the paired comparisons approach should yield a p_{ij} array similar to that from the categorical approach described earlier in this paper, assuming that the reference points are listed in rank order.

It would follow from the above that data from categorical judgements could be viewed as having come from rankings or paired comparisons, in the sense that when a judge places instance i in category j, he is saying that he ranks instance i below the upper bound of category j. Accordingly, we could use the paired comparisons procedure to scale both instances and category bounds. This approach has the disadvantage of not using all the information present, since the rank order of the reference points is already known.

What follows in this paper is a method which locates instances and category bounds on the same interval scale with fewer restrictive assumptions than the paired-comparisons approach. (End of Optional Section.)

Theory

We assume that a judge's "feelings" about the scale value of an instance i is a normally distributed random variable with mean S_i' and variance σ_i^2. We also assume that judges view the continuum of values for instances as being broken into successive intervals called categories, and that a judge's feelings about a category's upper bound is a normally distributed random variable so that for category j, the upper bound would be normally distributed with mean b_j' and variance v_j^2.

We want, for each instance i, an estimate S_i of its mean S_i'. To obtain these estimates, we will also have to obtain estimates b_j of the category upper bounds since the raw data will be sorted by category.

Since a judge's feelings about instance values and about category upper bounds are normally distributed random variables, the judge's feelings about the distance between an instance value and a category bound will also be a normally distributed random variable with mean $b_j' - S_i'$ and variance

$\sigma_i^2 + v_j^2 - 2\rho_{ij}\sigma_i v_j$. It is not unreasonable to assume that the value bound "feelings" are stochastically independent random variables, so that the correlation coefficient is zero for all pairs i and j. Besides $\rho_{ij} = 0$, we also assume that all category bounds have the same variance, so that for all j,*

$$v_j^2 = c\ .$$

Thus a judge's feelings about the distance from bound j to instance i can be viewed as a normally distributed random variable with mean $b_j' - S_i'$ and variance $\sigma_i^2 + c$. It follows that

$$\Pr(\text{Instance } i \text{ is rated below Bound } j) = \Pr\left(z < \frac{(b_j'-S_i') - 0}{\sqrt{\sigma_i^2+c}}\right),$$

where z is normally distributed with mean 0 and variance 1. From the frequency data from judges we obtain estimates p_{ij} of these probabilities, and then using the normal table we find the associated z_{ij} values. We now have $n(m-1)$ estimating equations of the form

$$z_{ij} = \frac{b_j - S_i}{\sqrt{\sigma_i^2+c}}\ , \quad i = 1, \ldots, n; \quad j = 1, \ldots, m-1\ . \tag{1}$$

For the present we will <u>assume</u> that all p_{ij} are such that $0.02 \le p_{ij} \le 0.98$, so that the <u>z_{ij} array is complete.</u> Ways of handling situations where this is not true are discussed later in this paper.

---

*An alternate approach would replace this by the assumption that the instances have the same variance.

Returning to our problem, we seek values for the estimates b_j and S_i, and the equations of the form of (1) represent a system of $n(m-1)$ equations in $2n+m-1$ unknowns.

There are a variety of ways to resolve this set of equations so as to yield values of the instance scale values. The method given here is reasonably quick and is considered to give satisfactory results.[1]

Our data is now summarized in the z_{ij} array, and we begin by examining the column sums of this array. Thus if we add estimates z_{ij} over instances i to obtain column sums, we have from (1)

$$\sum_{i=1}^{n} z_{ij} = b_j \left(\sum_{i=1}^{n} \left(\frac{1}{\sqrt{\sigma_i^2+c}} \right) \right) - \sum_{i=1}^{n} \left(\frac{S_i}{\sqrt{\sigma_i^2+c}} \right) , \qquad (2)$$

for $j = 1,2, \ldots, m-1$. Equation (2) shows a linear relationship between category upper bounds b_j and column sums of the z_{ij} array in that each column sum is a linear function of that category's upper bound. In the next section we will show how this linear relationship may be used to obtain values for the b_j.

Estimating Category Bounds b_j

Since we plan to locate both instance values and category bounds on the same interval scale, we have two degrees of freedom available to us since the unit and the origin of the scale are arbitrary. Equation (2) may be viewed as a linear transformation of an interval-scaled variable b_j, where the second term on the right-hand side establishes the origin of the scale. Using one of our degrees of freedom, we will set the origin of the scale so that

$$\sum_{i=1}^{n} \left(\frac{s_i}{\sqrt{\sigma_i^2 + c}} \right) = 0.$$

(Note that this does not imply that $\sum s_i = 0$.)

Then, using our other degree of freedom, we will set the unit for the scale so that the transformation slope is n, or

$$\sum_{i=1}^{n} \left(\frac{1}{\sqrt{\sigma_i^2 + c}} \right) = n.$$

With unit and origin thus established, Equation (2) now becomes simply

$$\sum_{i=1}^{n} z_{ij} = nb_j, \qquad j = 1,2, \ldots, m-1,$$

$$b_j = \frac{\sum_{i=1}^{n} z_{ij}}{n}, \qquad j = 1,2, \ldots, m-1. \tag{3}$$

Equation (3) says that our estimates of the category upper bounds may be obtained simply by taking the column averages of the z_{ij} array.

This solves part of our problem since we have now established a way to locate the category upper bounds on an interval scale. The next task is to locate the n instance values on the same scale, and unfortunately, we have already exhausted our usual tricks of judiciously selecting scale unit and origin.

Estimating Instance Values S_i'

Our basic estimating equations (1) were

$$z_{ij} = \frac{b_j - S_i}{\sqrt{\sigma_i^2 + c}} .$$

We have estimates of the z_{ij} from the judges, and values of the b_j from the column averages of the z_{ij} array. We seek values for the S_j and we would prefer to acquire these values without restrictive assumptions about the instance variances.

Having had some success with column averages, we will now look at the row averages of the z_{ij} array. The row averages are

$$\bar{z}_i = \frac{\sum_{j=1}^{m-1} z_{ij}}{m-1}$$

$$= \frac{\sum_{j=1}^{m-1} \frac{b_j - S_i}{\sqrt{\sigma_i^2 + c}}}{m-1} ,$$

which become

$$\bar{z}_i = \left(\frac{1}{\sqrt{\sigma_i^2 + c}}\right)\left(\frac{\sum_{j=1}^{m-1} b_j}{m-1} - \frac{\sum_{j=1}^{m-1} S_i}{m-1}\right) ,$$

or

$$\bar{z}_i = \left(\frac{1}{\sqrt{\sigma_i^2 + c}}\right)(\bar{b} - S_i) , \quad i = 1,2, \ldots, n. \tag{4}$$

In (4), $\bar{b}$ is the average of the column averages, and thus also the grand average of the z_{ij} array.

Now we shall solve for the variance terms. Subtracting Equation (4) from Equation (1), we may write

$$(z_{ij} - \bar{z}_i) = \frac{(b_j - S_i)}{\sqrt{\sigma_i^2+c}} - \frac{(\bar{b} - S_i)}{\sqrt{\sigma_i^2+c}}$$

or

$$(z_{ij} - \bar{z}_i) = \frac{b_j - \bar{b}}{\sqrt{\sigma_i^2+c}} .$$

We square both sides of this equation, and sum over categories j, and we have

$$\sum_{j=1}^{m-1} (z_{ij} - \bar{z}_i)^2 = \left[\frac{1}{\sigma_i^2+c}\right] \sum_{j=1}^{m-1} (b_j - \bar{b})^2 ,$$

from which we obtain the variance estimates

$$\sigma_i^2+c = \frac{\sum_{j=1}^{m-1} (b_j - \bar{b})^2}{\sum_{j=1}^{m-1} (z_{ij} - \bar{z}_i)^2} , \quad i = 1,2, \ldots, n. \tag{5}$$

These variance estimates are the final components we need in our basic estimating equation (1)

$$z_{ij} = \frac{b_j - S_i}{\sqrt{\sigma_i^2+c}} .$$

We solve this for S_i,

$$S_i = b_j - z_{ij}\sqrt{\sigma_i^2+c} ,$$

and enter the variance es.'mate (5) to obtain

$$S_i = b_j - z_{ij}\sqrt{\frac{\sum_{j=1}^{m-1}(b_j - \bar{b})^2}{\sum_{j=1}^{m-1}(z_{ij} - \bar{z}_i)^2}} \; .$$

We sum both sides of this equation over categories j:

$$\sum_{j=1}^{m-1} S_i = \sum_{j=1}^{m-1} b_j - \sum_{j=1}^{m-1} z_{ij}\sqrt{\frac{\sum_{j=1}^{m-1}(b_j - \bar{b})^2}{\sum_{j=1}^{m-1}(z_{ij} - \bar{z}_i)^2}} \; .$$

The left-hand side of this expression is equal to $(m-1)S_i$, and so we divide through the equation by (m-1) to obtain, finally,

$$S_i = \bar{b} - \bar{z}_i\sqrt{\frac{\sum_{j=1}^{m-1}(b_j - \bar{b})^2}{\sum_{j=1}^{m-1}(z_{ij} - \bar{z}_i)^2}} \quad , \quad i = 1,2, \ldots, n. \quad (6)$$

Equation (6) is our final computing form for the instance values S_i. Note that within this expression, $\bar{b}$ is the grand average of the values in the z_{ij} array, $\bar{z}_i$ is the average value of the ith row of that array, the numerator within the radical is the sum of squares of the column averages, and the denominator within the radical is the sum of squares of the values in the ith row.

With the mathematical work completed, it remains to establish a suitable computational framework by which numerical values for the instance values and the category bounds (as well as the variances) may be readily obtained. This is given in the following section, which begins on the next page.

Step by Step Procedure for Obtaining Scale Values, z_{ij} Array Complete.

1. Arrange the raw frequency data in a table where the rows are instances and the columns the categories. Columns should be in rank order, with Column 1 representing the least favorable category, etc.

2. Compute relative cumulative frequencies for each row, and record these in a new table. The last column of this new table will consist of unit values, and may be omitted.

3. Treating these values as leftward areas under a Normal (0,1) curve, go to a table of the normal distribution and find the z values for these areas. Record these in a new n by $(m-1)$ table. This is the z_{ij} array for the computations which follow.

4. For each row i in the z_{ij} array, compute the row average, $\overline{z_i}$.

5. For each column j in the z_{ij} array, compute the column average. Call these column averages b_j, and note that b_j is the value of the upper bound of category j on our scale.

6. Compute a grand average of all the values in the z_{ij} array. This is readily done by simply averaging the column averages. Call the grand average $\bar{b}$.

7. Compute*

$$B = \sum_{j=1}^{m-1} (b_j - \bar{b})^2 .$$

*Sum-of-squares computational shortcuts may save time and effort here.

8. For each row, compute*

$$A_i = \sum_{j=1}^{m-1} (z_{ij} - \overline{z_i})^2 .$$

9. For each row, compute $\sqrt{B/A_i}$. This is an estimate of $\sqrt{\sigma_i^2+c}$.

10. Finally, for each row (instance) compute

$$s_i = \bar{b} - \overline{z_i} \sqrt{B/A_i} .$$

These are the scale values of the instances, and they are on the same interval scale as the category bounds b_j. We now have the desired scale, and may perform any linear transformation $y = \alpha + \beta x$, $\beta>0$, to move the scale where we want it. Remember to use the same transformation to move both instance values and the category bounds.

*Sum-of-squares computational shortcuts may save time and effort here.

Example 1

We will continue the example started on Page 2, with p_{ij} array on Page 3. Because of the high p_{ij} values in the last column, we will have to pool the Effective and Very Effective categories. Steps 1 and 2 have already been completed, and the z_{ij} array below with associated calculations represents Steps 3,4,5, and 6.

.125 for 1-F(·) cic pg 579 [-]   .7125 for F(·) [+]

Candidate z_{ij}		Potential Effectiveness: Very Ineffective	Ineffective	Marginal	row total	row average: $\bar{z}_i$
A	1	-1.15	-0.32	0.56	-0.91	-0.303
B	2	-1.64	-0.19	1.09	-0.74	-0.247
C	3	-0.67	0.78	1.96	2.07	0.690
D	4	-1.78	-1.53	-0.03	-3.34	-1.113
Column totals		-5.24	-1.26	3.58	-2.92 = Grand Total	
Column averages: b_j		-1.310	-0.315	0.895	-0.243 = Grand Average: $\bar{b}$	

Step 7. $B = (-1.310-(-0.243))^2 + (-0.315-(-0.243))^2 + (0.895-(-0.243))^2 = \underline{2.439}$

Step 8. $A_1 = (-1.15-(-0.303))^2 + (-0.32-(-0.303))^2 + (0.56-(-0.303))^2 = \underline{1.462}$

Similarly . . . . $A_2 = \underline{3.731}$

$A_3 = \underline{3.471}$

$A_4 = \underline{1.792}$

Step 9. $\sqrt{B/A_1} = 1.292$, $\sqrt{B/A_2} = 0.809$, $\sqrt{B/A_3} = 0.838$, $\sqrt{B/A_4} = 1.167$.

Step 10.

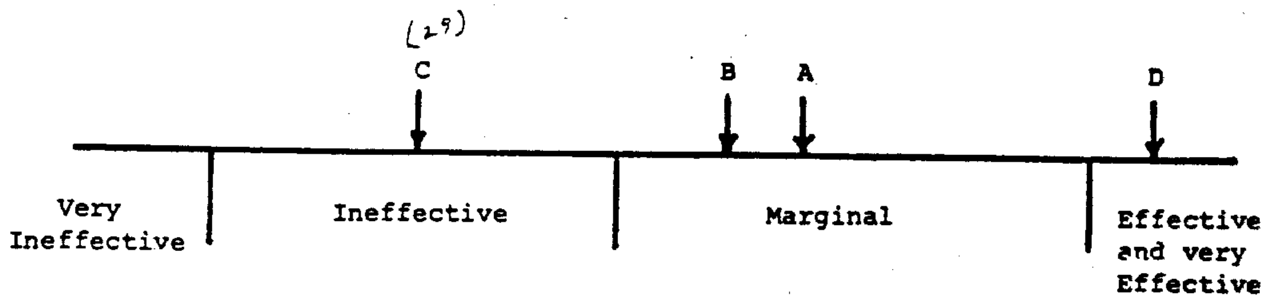
$$S_1 = -0.243 - (-0.303)(1.292) \quad = \underline{0.148}$$
$$S_2 = -0.243 - (-0.247)(0.809) \quad = -0.043$$
$$S_3 = -0.243 - (0.690)(0.838) \quad = \underline{-0.821}$$
$$S_4 = -0.243 - (-1.113)(1.167) \quad = \underline{1.055}$$

Also, Upper Bound on the Very Ineffective Category is -1.310

Upper Bound on the Ineffective Category is -0.315

Upper Bound on the Marginal Category is 0.895

Graphical Presentation of Results



Incomplete z_{ij} Arrays

As is the case in the method of paired comparisons, z_{ij} array entries corresponding to $p_{ij} > 0.98$ and $p_{ij} < 0.02$ should be omitted to avoid undue influence by a small number of judges. With the categorical method, however, the computational procedure should not be applied to an incomplete z_{ij} array. Accordingly, in order to use the method described in this paper one must modify the specific scaling problem at hand in such a way that complete z_{ij} arrays are obtained.

Because they correspond to cumulative relative frequencies, missing z_{ij} values are always found in one or more outermost columns of the z_{ij} array, involving the highest ranked categories, the lowest ranked categories, or both. Causes could include most judges rating an instance in the low categories, most judges rating an instance in the high categories, or judges rating in a middle category with close agreement (low σ_i^2) among judges.

Because of the variety of situations that can occur, it is probably best not to attempt to provide here specific instructions on how to cope with an incomplete z_{ij} array. We will suggest three tactics, any or all of which might be applicable to a specific problem. All involve a cost, leaving one with such penalties as an unscaled instance, an unscaled category bound, or a bound or instance scaled using less information than was anticipated. Accordingly, how one proceeds is up to the analyst and his goals for study results.

Tactic 1. One may delete those rows with missing z_{ij} values to obtain a smaller but complete z_{ij} array, and then apply the method given in this paper. This means, of course, that instances represented by those deleted rows will not be scaled directly. One either discards these instances, or "pieces" then onto the scale in some way that will hopefully be defendable (but will seldom be altogether satisfactory).

Tactic 2. One may pool extreme categories to obtain a z_{ij} array void of missing values. For example, if Column 1 has missing z_{ij} values and Column 2 is complete, we combine Categories 1 and 2 into a single category, and use the z_{ij} values of Column 2. As another example, if the last column (column $m-1$) has missing z_{ij} values and the next to last column is complete, we combine the last two categories that the judges used ($m-1$ and m) and delete the last column in the z_{ij} array. This means that some category bounds will not be on the scale. In many cases, this approach is preferred to Tactic 1 described previously.

Tactic 3. A third approach is to break the z_{ij} array down into smaller arrays, applying the previously described tactics so that one has several complete but smaller z_{ij} arrays. These are scaled separately. If one has been clever in dividing the original array, the resulting set of scales will have exactly the proper amount of "two points in common" so that appropriate linear transformations will place all instances and bounds on the same scale without redundancy. This approach involves a number of arbitrary decisions in its execution, but will often provide a resulting scale that is complete in the sense that all n instances and $m-1$ category bounds are present.

Tactic 4 Add instances sent out questionnaires again [illegible]

In order to illustrate how these tactics may be employed, we shall give an example showing how, in one real situation, the problem of an incomplete z_{ij} array was resolved.

Example 2

In a study done as a classroom scaling exercise, nineteen dog breeds (instances) were to be rated in terms of their "deterrent effectiveness against intruders and trespassers". Five ranked categories were used; the complete questionnaire is shown in Figure 2.

About 65 judges responded by completing the questionnaire, meaning that an eventual missing z_{ij} value in column j will correspond to $f_{ij} \leq 1$ in the lower categories. In the upper categories, an $f_{ij} \leq 1$ will yield a missing z_{ij} value in the (j-1)st column. Raw frequencies f_{ij} are shown in Table 3. In this case, missing z_{ij} values were handled at the f_{ij} array level.

A glance at Table 3 shows that not all dogs can be directly scaled together with all category bounds, since a complete z_{ij} array would have many missing values. Since it was desired that all dog breeds be scaled, Tactic 1 (eliminate instances) was not employed. Direct application of Tactic 2 (pool categories) would in this case actually destroy the category distinctions. Accordingly, it was decided to employ Tactic 3 by subdividing the array.

A useful first step was to rearrange the instances so that rows which would be similar in missing z_{ij} values were

DETERRENT EFFECTIVENESS OF WATCHDOGS

Please rate the following dog breeds according to your opini of their deterrent effectiveness against intruders and trespasser

Deterrent Effectiveness Against Intruders and Trespassers.

	Ineffective	Slighuly Effective	Effective	Highly Effective	Extremely Effective	Don't Know
Boston Terrier	( )	( )	( )	( )	( )	( )
St. Bernard	( )	( )	( )	( )	( )	( )
English Bulldog	( )	( )	( )	( )	( )	( )
Dalmation	( )	( )	( )	( )	( )	( )
Boxer	( )	( )	( )	( )	( )	( )
Chihuahua	( )	( )	( )	( )	( )	( )
Lhasa Apso	( )	( )	( )	( )	( )	( )
Labrador Retriever	( )	( )	( )	( )	( )	( )
English Setter	( )	( )	( )	( )	( )	( )
Cocker Spaniel	( )	( )	( )	( )	( )	( )
Irish Setter	( )	( )	( )	( )	( )	( )
German Shephard	( )	( )	( )	( )	( )	( )
Great Dane	( )	( )	( )	( )	( )	( )
Dachshund	( )	( )	( )	( )	( )	( )
Basset Hound	( )	( )	( )	( )	( )	( )
Beagle	( )	( )	( )	( )	( )	( )
Standard Poodle	( )	( )	( )	( )	( )	( )
Miniature Poodle	( )	( )	( )	( )	( )	( )
Fox Terrier	( )	( )	( )	( )	( )	( )

Thank you for your help. For a copy of the results of this stud please list your SMC or mail code number ___________.

Table 3. Response Precuencies, Example 2

	Deterrent Effectiveness Against Intruders and Trespassers				
	Ineffective	Slightly Effective	Effective	Highly Effective	Extremely Effective
Boston Terrier	0	27	18	7	1
St. Bernard	5	6	26	21	6
English Bulldog	0	6	29	19	9
Dalmation	2	14	30	15	2
Boxer	0	1	20	22	20
Chihuahua	34	14	8	6	2
Lhasa Apso	20	12	9	1	1
Labrador Retriever	2	8	32	15	7
English Setter	3	22	29	11	0
Cocker Spaniel	10	32	20	3	1
Irish Setter	3	17	25	14	5
German Shephard	0	0	2	6	55
Great Dane	0	1	3	20	38
Dachshund	20	28	13	2	0
Basset Hound	16	31	15	1	0
Beagle	7	35	18	5	0
Standard Poodle	9	27	21	6	2
Miniature Poodle	37	24	7	2	0
Fox Terrier	1	23	14	6	0

grouped. Results of this are shown in Table 4, and necessary category pooling is shown in Table 5. Here we see that if we treat the center group of instances (St. Bernard through Standard Poodle) as a separate problem, it will be possible to scale these dog breeds together with all category bounds. After grouping categories, the first three dog breeds in the list could be scaled as a three-category problem, and since the resulting scale would have two points (category bounds) in common with the scale from the center group of dog breeds, the two interval scales could be merged via linear transformation. The same could be said for the Boston Terrier and Fox Terrier group. These observations provided a plan of attack, with only a few details to be worked out.

If categories are pooled as shown, the English Bulldog would be rated in four categories, whereas three-category ratings are needed if ultimate scales are to merge with the center group. It was decided to group the English Bulldog with the Boston and Fox Terriers, since adding an instance to that group will give better category bound estimates. Also, rather than leave the Basset Hound stranded, it was decided to add the remaining instances (Llasa Apso through Miniature Poodle) to his group.

The results of these decisions are shown in Table 6, and the final resolution into four scaling problems is portrayed in Table 7, where we have numbered the problems as Problems 1,2,3, and 4. The vertical lines in Table 7 indicate the

Table 4. Response Frequencies with Rearranged Instances

	Ineffective	Slightly Effective	Effective	Highly Effective	Extremely Effective
German Shephard	0	0	2	6	55
Great Dane	0	1	3	20	38
Boxer	0	1	20	22	20
English Bulldog	0	6	29	19	9
Boston Terrier	0	27	18	7	1
Fox Terrier	1	23	14	6	0
St. Bernard	5	6	26	21	6
Dalmation	2	14	30	15	2
Chihuahua	34	14	8	6	2
Labrador Retriever	2	8	32	15	7
Irish Setter	3	17	25	14	5
Standard Poodle	9	27	21	6	2
Lhasa Apso	20	12	9	1	1
English Setter	3	22	29	11	0
Cocker Spaniel	10	32	20	3	1
Dachshund	20	28	13	2	0
Beagle	7	35	18	5	0
Miniature Poodle	37	24	7	2	0
Basset Hound	16	31	15	1	0

Table 5. Necessary Category Groupings

	Ineffective	Slightly Effective	Effective	Highly Effective	Extremely Effective
German Shephard	0	0	2	6	55
Great Dane	0	1	3	20	38
Boxer	0	1	20	22	20
English Bulldog	0	6	29	19	9
Boston Terrier	0	27	18	7	1
Fox Terrier	1	23	14	6	0
St. Bernard	5	6	26	21	6
Dalmation	2	14	30	15	2
Chihuahua	34	14	8	6	2
Labrador Retriever	2	8	32	15	7
Irish Setter	3	17	25	14	5
Standard Poodle	9	27	21	6	2
Lhasa Apso	20	12	9	1	1
English Setter	3	22	29	11	0
Cocker Spaniel	10	32	20	3	1
Dachshund	20	28	13	2	0
Beagle	7	35	18	5	0
Miniature Poodle	37	24	7	2	0
Basset Hound	16	31	15	1	0

Table 6. Grouped Categories for Four Scaling Problems

	Ineffective	Slightly Effective	Effective	Highly Effective	Extremely Effective
German Shephard	0	0	2	6	55
Great Dane	0	1	3	20	38
Boxer	0	1	20	22	20
English Bulldog	0	6	29	19	9
Boston Terrier	0	27	18	7	1
Fox Terrier	1	23	14	6	0
St. Bernard	5	6	26	21	6
Dalmation	2	14	30	15	2
Chihuahua	34	14	8	6	2
Labrador Retriever	2	8	32	15	7
Irish Setter	3	17	25	14	5
Standard Poodle	9	27	21	6	2
Lhasa Apso	20	12	9	1	1
English Setter	3	22	29	11	0
Cocker Spaniel	10	32	20	3	1
Dachshund	20	28	13	2	0
Beagle	7	35	18	5	0
Miniature Poodle	37	24	7	2	0
Basset Hound	16	31	15	1	0

Table 7. Reduction to Four Scaling Problems

	Ineffective	Slightly Effective	Effective	Highly Effective	Extremely Effective	
German Shephard		2		6	55	
Great Dane		4		20	38	Problem 1
Boxer		21		22	20	
English Bulldog	6		29	28		
Boston Terrier	27		18	8		Problem 2
Fox Terrier	24		14	6		
St. Bernard	5	6	26	21	6	
Dalmation	2	14	30	15	2	
Chihuahua	34	14	8	6	2	
Labrador Retriever	2	8	32	15	7	Problem 3
Irish Setter	3	17	25	14	5	
Standard Poodle	9	27	21	6	2	
Lhasa Apso	20	12		11		
English Setter	3	22		40		
Cocker Spaniel	10	32		24		
Dachshund	20	28		15		Problem 4
Beagle	7	35		23		
Miniature Poodle	37	24		9		
Basset Hound	16	31		16		

category bounds that will be obtained when each of the four scales is constructed, and we can see that by using an appropriate linear transformation for each of the scales from Problems 1,2, and 4, these scales can be merged with the scale from Problem 3.

The method given in this paper was used to construct the four scales. Raw results are shown graphically in Figure 4, with numerical values given in Table 8.

To provide general results with some numerical meaning, it was decided to assign a value of zero to the category bound between the Ineffective and Slightly Effective categories, and a value of 100 to the bound between the Highly Effective and Extremely Effective categories. Thus, dog breeds rated ineffective would have negative scores, and the extremely effective dogs would have scores greater than 100. Raw results from Problem 3 were transformed to this scale, and then raw results from the other three problems were transformed to the new Problem 3 scale. Final scale values are shown in the right-hand column of Table 8. Project results are summarized in Table 9.

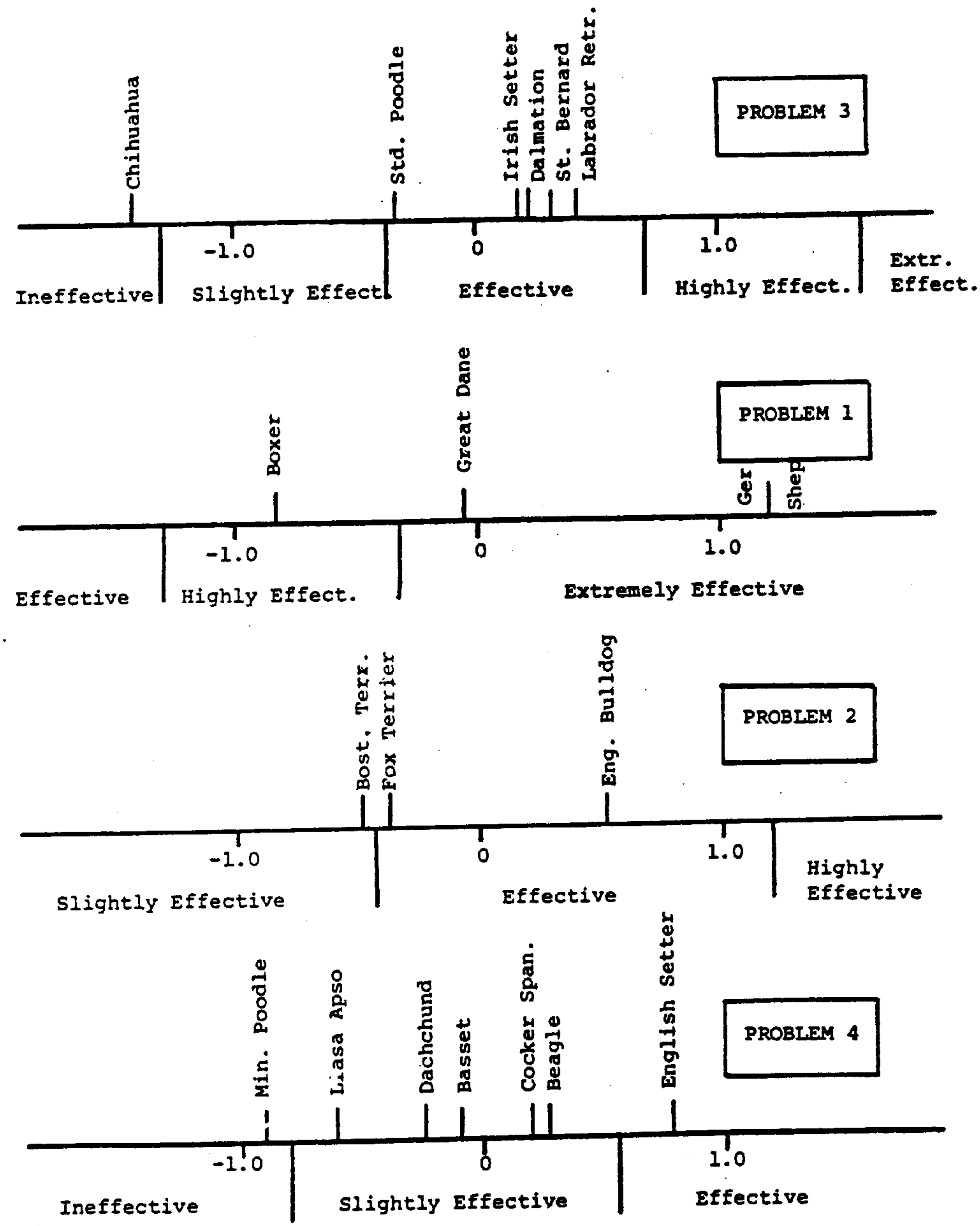


Figure 4. Raw Results for the Four Scaling Problems

Table 8. Solutions to the Four Scaling Problems

Problem 3 Results	Raw Results	Transformed
Upper bound, Highly Effective category	1.5933	100.0
Upper bound, Effective category	0.6763	68.3
Upper bound, Slightly Effective category	-0.37	32.2
Upper bound, Ineffective category	-1.304	0.0
St. Bernard	0.3666	57.7
Dalmation	0.2169	52.5
Chihuahua	-1.4213	-4.0
Labrador Retriever	0.4144	59.3
Irish Setter	0.2	51.9
Standard Poodle	-0.3578	32.7
Problem 1 Results		**Transformed to Prob. 3 Scale**
Upper bound, Highly Effective cateory	-0.318	100.00
Upper bound, Effective category	-1.268	68.3
German Shephard	1.198	150.1
Great Dane	-0.097	107.4
Boxer	-0.816	83.4
Problem 2 Results		
Upper bound, Effective category	1.151	68.3
Upper bound, Slightly Effective category	-.445	32.2
English Bulldog	0.525	64.83
Boston Terrier	-0.4715	31.3
Fox Terrier	-0.3734	34.6
Problem 4 Results		
Upper bound, Slightly Effective category	0.5136	32.2
Upper bound, Ineffective category	-0.7293	0.0
Lhasa Apso	-0.5796	3.9
English Setter	0.7728	39.0
Cocker Spaniel	0.1983	24.1
Dachshund	-0.2311	12.9
Beagle	0.2250	24.7
Miniature Poodle	-0.8114	-2.1
Basset Hound	-0.1079	16.1

Table 9. Study Results

DETERRENT EFFECTIVENESS OF WATCHDOGS

As a final project in a course on scaling methods, a categorical questionnaire on the deterrent effectiveness of watchdogs was prepared and circulated. Seventy people responded, rating each of 19 dog breeds in categories ranging from ineffective to extremely effective against intruders and trespassers.

On the scale of deterrent effectiveness developed from this data, dogs scoring more than 100 points are in the extremely effective category, while dogs with negative scores are in the ineffective category. Results are shown below.

Extremely Effective: 100+

150.1 German Shephard
107.3 Great Dane

Highly Effective: 68.3 - 100

83.4 Boxer

Effective: 32.2 - 68.3

64.8 English Bulldog
59.3 Labrador Retriever
57.7 St. Bernard
52.5 Dalmation
51.9 Irish Setter
39.0 English Setter
34.6 Fox Terrier
32.7 Standard Poodle

Slightly Effective: 0 - 32.2

31.5 Boston Terrier
24.7 Beagle
24.1 Cocker Spaniel
16.1 Basset Hound
12.9 Dachshund
3.9 Lhasa Apso

Ineffective:

-2.1 Miniature Poodle
-4.0 Chihuahua

It is interesting to note that correlation coefficient between deterrent effectiveness and dog weight was computed to be 0.61.

References

1. Burros, R. H. "The Estimation of the Discriminal Dispersion in the Method of Successive Intervals" Psychometrika 20, 299-305.

2. Dyer, R. F., Matthews, J. J., Wright, C. E., and K. L. Yudowitch. Questionnaire Construction Manual. Ft. Hood Field Unit, U.S. Army Research Institute for the Behavioral Sciences, September, 1976.

3. Rimoldi, H. J. A., and M. Hormaeche. "The Law of Comparative Judgement in the Successive Intervals and Graphic Rating Scaling Method." Psychometrika, 20, 307-318.

4. Saffir, M. "A Comparative Study of Scales Constructed by Three Psychophysical Methods". Psychometrika 2, 179-198.

5. Torgerson, W. S., Theory and Methods of Scaling, Wiley, 1958.

EXERCISES

1. What is the minimum number of categories needed to scale a set of ten instances using the method given in this paper? Assume complete z_{ij} arrays.

2. Using the method of this paper, what would be the result if you had five categories, one instance, and no missing z_{ij} values.

3. Raw frequencies are as follows:

	Extremely Poor	Poor	So-so	Good	Extremely Good
Coors	3	10	17	52	18
Olympia	6	20	45	24	5
Budweiser	4	8	25	48	15
Millers	5	15	38	31	11
Henry Weinhard	3	6	21	50	20

   a. Scale the instances and bounds.

   b. Transform the scale so that the upper bound of Good is 100 and the lower bound of So-so is zero.

4. Scale the SAM Systems using the f_{ij} data below. Set the upper bound on "Effective" at 100, and the upper bound on "Ineffective" at zero.

SAM System	Ineffective	Somewhat Ineffective	Marginally Effective	Effective	Highly Effective
A	1	15	38	40	9
B	0	3	21	36	33
C	12	43	27	4	0
D	0	0	3	33	60
E	9	27	25	18	11
F	17	44	24	2	0

5. Suppose the example given in this paper had included twenty dog breeds instead of nineteen, with the additional row in Table 3 (and thereafter) as follows:

	Ineffective	Slightly Effective	Effective	Highly Effective	Extremely Effective
Doberman Pinscher	0	0	0	6	36

Discuss how the presence of this data would affect the scale, and then, add this dog breed to the scale.

6. A Student Opinion Form (SOF) asks students at the end of the course to rate course, text, and exams each as <u>O</u> (Outstanding), <u>E</u> (Excellent), <u>A</u> (Average), <u>F</u> (Fair), or <u>P</u> (Poor). The scoring system awards five points for <u>O</u>, four points for <u>E</u>, three points for <u>A</u>, and so on, and uses these values to compute an average score. Some SOF data for various past courses is shown below. Using this data, what can you say about how well the scoring system described above agrees with the student's perception of category bounds?

		1	2	3	4	5
		P	F	A	E	O

How would you rate these courses?

		1 P	2 F	3 A	4 E	5 O
Course	A	1	2	18	14	2
	B	0	0	2	12	13
	C	2	2	14	15	4
	D	0	0	9	11	12
	E	0	0	2	8	18
	F	1	3	17	5	0

How would you rate the textbook?

	1 P	2 F	3 A	4 E	5 O
B	0	1	2	10	14
C	11	9	9	7	1
D	0	0	11	11	3
E	0	0	3	13	12
F	10	10	5	1	0

How would you rate the exams?

	1 P	2 F	3 A	4 E	5 O
A	3	3	11	15	4
B	0	1	2	12	12
C	4	3	13	11	6
D	0	1	9	9	8
E	0	0	2	12	13
F	1	1	10	14	0