PSYCHOMETRIKA—VOL. 27, NO. 4
DECEMBER, 1962

OPTIMAL SCALING FOR ORDERED CATEGORIES*

R. A. BRADLEY AND S. K. KATTI

THE FLORIDA STATE UNIVERSITY

AND

IRMA J. COONS

RESEARCH CENTER, THE GENERAL FOODS CORPORATION

This paper deals with the determination of optimal weights for points on scoring scales for subjective comparative experiments. A scoring scale with a specific number of points is considered, and it is assumed that verbal or other indications imply an order to the scale points. The optimal spacing for the scale points is obtained in the sense that treatment or item differences are maximized relative to error or within-treatment variation. The method is presented in sufficiently generalized form to be used directly with any experimental design leading to the analysis of variance. An iterative procedure, suitable for computer use, yields the optimal differences among the ordered scale points. Properties of this procedure are discussed.

In many comparative experiments, it seems necessary to measure the responses of individuals on scoring scales. It is then also usually necessary to assign numerical values to points on these scales and to proceed with statistical analyses based on these numerical scores. Problems of this nature arise in many types of appraisal situations, and an important class of such problems arises in the consumer testing of food products.

Subjective scoring scales of various forms and numbers of scale points are used in food-product evaluations. These scales usually involve verbal descriptions for at least some of the points on the scales, but sometimes other representations of scale points are used. In any event, it is almost always implicitly assumed that the scale points are ordered, although there may be rare situations for which it is not clear that respondents understand the intended order. A common practice has been to assume that points on scoring scales are evenly spaced, implying an arithmetic scale, but such an assignment of scale values may not be optimum.

Bock [2] has considered the problem of optimal scaling and he states the problem succinctly as follows:

> The approach of optimal scaling is to assign numerical values to alternatives, or categories, so as to discriminate optimally among the objects . . . in some

*Research sponsored in part by the Statistics Branch, Office of Naval Research, and in part by the Research Center, the General Foods Corporation. Reproduction in whole or in part is permitted for any purpose of the United States Government.

> sense. Usually it is the least squares sense, and the values are chosen so that the variance between objects after scaling is a maximum with respect to that within objects.

Bock, in his paper, is largely concerned with the numerical method and with examples. Fisher ([3], pp. 285–298) seems to have been the first to use the approach. The problem is discussed in detail by Torgerson ([8], pp. 338–345), and his discussion is based on papers by Guttman [5] and more directly by Mosteller [7]. Essentially the criterion stated by Bock is used to obtain optimal discrimination among objects, and the concept of ordered categories or ordered scale points has not been considered by Torgerson or the earlier writers using this criterion. Torgerson includes his discussion under the heading "Deterministic Models for Categorical Data."

Other criteria are possible and have been used in considering ordered categories. Such a method has become known as the "method of successive intervals"; this is based on accumulated proportions for each treatment up to specified scale points. Proportions under normal curves are equated to these observed proportions, the corresponding normal deviates are obtained, and estimates of treatment means, standard errors, and scale values are found to minimize a pooled error variance. Basic references to the method of successive intervals are Horst [6], Gulliksen [4], Bock [1], and Torgerson ([8], Chap. 10). There is clearly no direct mathematical relationship between scale values obtained by the method of successive intervals and the procedures that we develop below; whether or not there is close agreement in scale values obtained, even though different criteria are used, must await extensive comparisons with good experimental data.

We have applied the scaling method for categorical data following Torgerson ([8], pp. 338–345) to the problem of determining scores for scale values in food testing. The results were rather meaningless in that the scores obtained did not at all match the order suggested by verbal, or otherwise indicated, orders on the scales. In view of this difficulty, the work of this paper was undertaken, and the principle of maximizing the variance between objects after scaling relative to within-objects variance was adopted, but with the difference that the maximization was carried out subject to imposed order restrictions on the scale points.

Formulation of the Problem

Suppose that a scoring scale has k points designated by $w_1, \cdots, w_k$. We assume an ordered scale so that $w_1 \leq w_2 \leq \cdots \leq w_k$. The origin of the scale is arbitrary (see Appendix A) and may be defined so that the scale values sum to zero, $\sum_{i=1}^{k} w_i = 0$. In a designed experiment, not necessarily a one-way classification, the frequencies of occurrences of $w_1, \cdots, w_k$ are recorded within each design division (treatments, blocks, replications, etc.) of the experiment. It is now possible to write algebraically the variance ratio F, for the test for treatment effects for the particular experimental design being used, in terms of the observed frequencies and $w_1, \cdots, w_k$.

The variance ratio will have the form

$$F = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} w_i w_j \Big/ \sum_{i=1}^{k} \sum_{j=1}^{k} b_{ij} w_i w_j . \tag{1}$$

The coefficients a_{ij} and b_{ij} in (1) are functions of the observed frequencies and of the design parameters of the experimental design employed. These values will be known for any given experiment. The optimal scale values for the ordered categories are obtained by maximizing F with respect to the w's, subject to the restrictions imposed on them. An alternative (and equivalent) procedure, for example for the one-way classification, would be to maximize R^2, the square of the multiple correlation coefficient as obtained from the ratio of treatment sum of squares to the total sum of squares. This alternative has some slight advantages through simplification of the coefficients in the denominator of R^2 corresponding to the b_{ij} in (1). Both F and R^2 are independent of the scale of the w's and, in numerical work, it may be advantageous to use the rule that the range of the w's, $w_k - w_1$, be unity although this is not necessary.

We proceed by transforming to new variables. Let

$$\begin{aligned} z_1 &= w_2 - w_1 , \\ z_2 &= w_3 - w_2 , \\ &\cdots \\ z_{k-1} &= w_k - w_{k-1} , \\ z_k &= w_1 + \cdots + w_k . \end{aligned} \tag{2}$$

In matrix notation,

$$z' = Tw', \tag{3}$$

with z' and w' being column vectors, and T being the matrix of coefficients indicated by (2). Now $w_{i+1} \geq w_i$ implies that $z_i \geq 0$, $i = 1, \cdots, k - 1$, and $\sum_{i=1}^{k} w_i = 0$ implies that $z_k = 0$. Substitution in F in (1) yields a new form,

$$F = \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} c_{ij} z_i z_j \Big/ \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} d_{ij} z_i z_j , \tag{4}$$

where the c_{ij} and d_{ij} are obtained directly from the a_{ij} and b_{ij}, respectively. The problem has been reduced now to the one of maximizing F in (4) with respect to the z's, subject to $z_i \geq 0$, $i = 1, \cdots, k - 1$. When values of $z_1, \cdots, z_{k-1}$ are obtained, values of $w_1, \cdots, w_k$ follow.

In practice, computer programs have been developed for various experimental designs to obtain the c_{ij} and d_{ij} coefficients in (4); these are determined numerically through substitution of the required frequencies into the computer. Development of such programs is straightforward and will not be discussed here; it is sufficient to note that

$$c_{ij} = \sum_{i'=1}^{i} \sum_{j'=1}^{j} a_{i'j'} , \tag{5}$$

and that

$$d_{ij} = \sum_{i'=1}^{i} \sum_{j'=1}^{j} b_{i'j'} , \tag{6}$$

as is demonstrated in Appendix B.

Some difficulties may enter. Suppose an extreme scale point, say w_k , is never used. Then w_k does not appear in (1) and

$$a_{ik} = a_{kj} = b_{ik} = b_{kj} = 0 \qquad (i, j = 1, \cdots, k).$$

This results in

$$c_{i,k-1} = c_{k-1,j} = d_{i,k-1} = d_{k-1,j} = 0 \qquad (i, j = 1, \cdots, k-1),$$

as can be seen from (5) and (6) and the fact that

$$\sum_{i=1}^{k-1} a_{ij} = \sum_{j=1}^{k-1} a_{ij} = \sum_{i=1}^{k-1} b_{ij} = \sum_{j=1}^{k-1} b_{ij} = 0$$

resulting from the discussion of Appendix B. Now w_k and z_{k-1} are not estimable; a sensible rule is then to consider the original scale to be one with only $k - 1$ points and to proceed. Similarly, if w_1 is never used, the first row and first column of the C matrix and of the D matrix consist of zeros and again a $(k - 1)$-point scale must be considered. If more than one adjacent extreme scale point is not used, similar reduction of the number of original scale points must be made. If a central scale point is not used, say w_s , $1 < s < k$, the procedures outlined below may still operate. However, an element of indeterminancy enters, and the values of $z_1 , \cdots , z_{k-1}$ to maximize (4) are not unique with z_s and z_{s-1} being determined only to the extent of obtaining the optimal value for $z_s + z_{s-1}$.

Let us consider the C and D matrices in more detail. In the analysis of variance, the treatment and error sums of squares may be written in the forms yS_ty' and yS_ey', where y is the vector of the totality of N original observations, and S_t and S_e are nonnegative definite, symmetric, square matrices of order N and of ranks equal to the degrees of freedom for treatments and error, respectively. In the situation with a scoring scale of k points, there exists an N by k matrix Δ associating the observations y with the scale points w so that Δ contains one and only one nonzero element, unity, in each row; thus

$$\Delta w' = y', \qquad A = \Delta' S_t \Delta, \quad \text{and} \quad B = \Delta' S_e \Delta.$$

Transformation to the z's yields

$$C^* = \underline{(T^{-1})'}\ \Delta' S_t\ \Delta T^{-1} \quad \text{and} \quad D^* = (T^{-1})'\ \Delta' S_e\ \Delta T^{-1}$$

as defined in Appendix B with C and D being the $(k - 1)$-square, first principal minors of C^* and D^*, respectively. If all scale values are used at least once in the experiment, the rank of Δ is k. Since the rank of S_e or S_t is that of the degrees of freedom associated with it, it follows that the rank of C^* cannot exceed the smaller of k and the number of degrees of freedom for treatments, and the rank of D^* cannot exceed the smaller of k and the number of degrees of freedom for error. Similarly, the rank of C cannot exceed the smaller of $k - 1$ and the number of degrees of freedom treatments; the rank of D cannot exceed the smaller of $k - 1$ and the number of degrees of freedom for error. It follows that C is more likely to be singular than D because often the degrees of freedom for treatments will be less than $k - 1$, whereas usually the degrees of freedom for error will exceed $k - 1$.

To proceed, it would be easy if we could assert that D is positive definite (of rank $k - 1$) but such is not always the case. A clear exception occurs when a subset of adjacent extreme scores is used exclusively to evaluate all samples on one treatment. In this exception an infinite F can be obtained by setting all w's in the subset equal to one value and the remaining w's equal to another value. This may not be the only type of exception, and we have not been able to catalog all of the possibilities. It is clear that the difficulties arise in those configurations of score assignments that permit a zero error sum of squares. Thus, the difficulties are less likely to arise in larger experiments and in well-designed experiments without extreme treatments.

In applications of the method to be developed, we have encountered cases wherein C is singular, but we have not encountered cases where D is not positive definite. In the following work, we shall assume that D is positive definite and suggest that the nonsingularity of D be checked in applications before proceeding further. Several of our demonstrations will depend on D being positive definite, but the method of estimation may still be valid. The procedures will be valid still if D is singular provided the subspace over which the singularities exist does not intersect the principal quadrant of the z space. The procedures should also lead to a point in the principal quadrant of the z space for which an infinite F exists, but the proof of convergence given does not apply. It is very unlikely that a well-designed and comprehensive experiment for the calibration of a k-point scale will lead to D singular.

First Quadrant Maximization

To maximize F in (4), we require the point in, or on a boundary of, the principal quadrant of the z space that yields the supremum of F. While the procedures to be outlined may apply in somewhat more general situations, we shall limit consideration to the situation in which the denominator of (4) is a positive-definite quadratic form (D is positive definite). It does not follow that F has a zero-derivative maximum, and an iterative procedure has been developed for the maximization. It does follow that, if D is positive definite, we do have a finite maximum in, or on a boundary of, the principal quadrant and that F and its derivative are continuous. That F has an absolute maximum and that the procedure to be given yields that maximum, subject to D being positive definite, will be demonstrated in Appendix C.

Consider the partial derivatives

$$(7) \qquad \frac{\partial F}{\partial z_i} = \frac{2[(\sum_j c_{ij}z_j)(\sum_i \sum_j d_{ij}z_iz_j) - (\sum_j d_{ij}z_j)(\sum_i \sum_j c_{ij}z_iz_j)]}{(\sum_i \sum_j d_{ij}z_iz_j)^2}$$

$$= K(z)[(\sum_j c_{ij}z_j)(\sum_i \sum_j d_{ij}z_iz_j) - (\sum_j d_{ij}z_j)(\sum_i \sum_j c_{ij}z_iz_j)],$$

wherein $K(z) \geq 0$ and the summations with respect to i and j extend over values $1, \cdots, k-1$. The procedure is a simple one. Take $i = 1$ in (7), substitute simple initial values of $z_2, \cdots, z_{k-1}$ in (7), and then obtain the value of z_1 that maximizes F in (4) under the required conditions and on the line described by the initial values. When z_1 has been so obtained, designate it by $z_1^{(1)}$. If the initial values were $z_2^{(0)}, \cdots, z_{k-1}^{(0)}$, repeat setting $i = 2$ in (7), and substitute $z_1^{(1)}, z_3^{(0)}, \cdots, z_{k-1}^{(0)}$ to obtain $z_2^{(1)}$ and so on. This procedure always converges since the values of F corresponding to each stage of the iterations form a bounded increasing sequence. We now consider how (7) is used in this process.

To be specific, consider $i = 1$ in (7) and substitute $z_2^{(0)}, \cdots, z_{k-1}^{(0)}$ with these numerical values less than unity without loss of generality. It is clear that the sign of $\partial F/\partial z_i$ depends on

$$(8) \qquad Q_i = (\sum_j c_{ij}z_j)(\sum_i \sum_j d_{ij}z_iz_j) - (\sum_j d_{ij}z_j)(\sum_i \sum_j c_{ij}z_iz_j).$$

When $i = 1$ and the initial values have been substituted, Q_1 is a polynomial in z_1. This polynomial is at most of degree three, but it is immediately evident that the coefficient of z_1^3 is always zero. The coefficient of z_1^2 is

$$(9) \qquad c_{11}\left(\sum_{j=2}^{k-1} d_{1j}z_j^{(0)}\right) - d_{11}\left(\sum_{j=2}^{k-1} c_{1j}z_j^{(0)}\right),$$

the coefficient of z_1 is

$$(10) \qquad c_{11}\left(\sum_{i=2}^{k-1} \sum_{j=2}^{k-1} d_{ij}z_i^{(0)}z_j^{(0)}\right) - d_{11}\left(\sum_{i=2}^{k-1} \sum_{j=2}^{k-1} c_{ij}z_i^{(0)}z_j^{(0)}\right),$$

and the constant term is

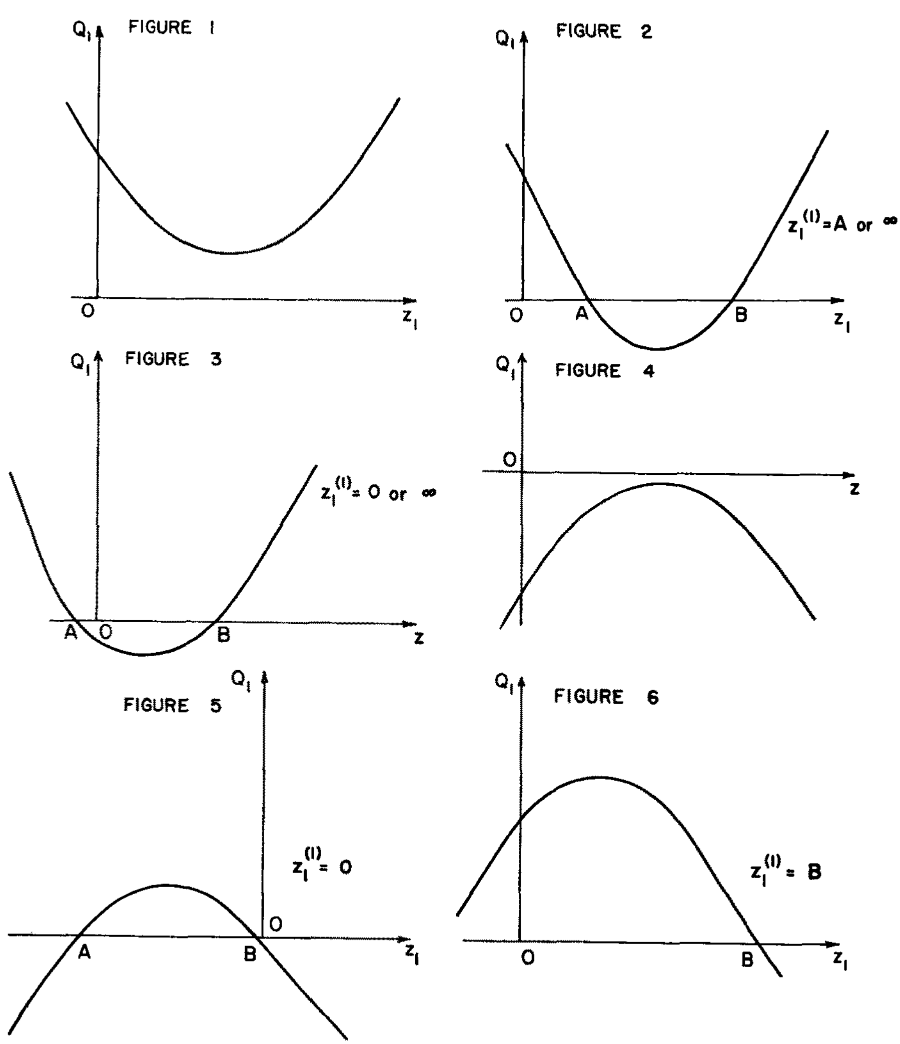
$$(11) \qquad \left(\sum_{j=2}^{k-1} c_{1j}z_j^{(0)}\right)\left(\sum_{i=2}^{k-1} \sum_{j=2}^{k-1} d_{ij}z_i^{(0)}z_j^{(0)}\right) - \left(\sum_{j=2}^{k-1} d_{1j}z_j^{(0)}\right)\left(\sum_{i=2}^{k-1} \sum_{j=2}^{k-1} c_{ij}z_i^{(0)}z_j^{(0)}\right).$$

We shall consider two main cases separately, (i) Q_1 quadratic and (ii) Q_1 not quadratic.

(i) Q_1 *Quadratic*

Consider the graph of Q_1 plotted against z_1. Ordinarily, a quadratic form can have one of the six shapes shown in Figs. 1, $\cdots$, 6. We observe first that Figs. 1 and 4 are impossible in the present situation. The proof is given in Appendix D. We discuss the remaining figures under two subcases.

GRAPHS OF Q_1 DETERMINING
THE SIGNS OF dF/dz_1



(a) *The Coefficient of* $z_1^2 > 0$

When the coefficient of z_1^2 in (9) exceeds zero, Fig. 2 or 3 applies. If the graph is as in Fig. 2, the maximum of F occurs either at $z_1 = A$, the smaller root of the quadratic, or at $z_1 = \infty$. This follows since $\partial F/\partial z_1 > 0$ for $0 \leq z_1 < A$ and for $z_1 > B$, the larger root of the quadratic, indicating that F is increasing in these regions. F is evaluated for $z_1 = A \quad (z_2 = z_2^{(0)}, \cdots, z_{k-1} = z_{k-1}^{(0)})$ and as $z_1 \to \infty$. If the larger value of F occurs when $z_1 = A$, we take $z_1^{(1)} = A$. If the larger value occurs when $z_1 = \infty$, we take $z_1^{(1)}$ large, but finite, relative to $z_2^{(0)}, \cdots, z_{k-1}^{(0)}$ and proceed. In numerical work we have simply adopted the rule of taking $z_1^{(1)}$ to be 20 times the larger of $z_2^{(0)}, \cdots, z_{k-1}^{(0)}$ and proceeding.

If the graph is as in Fig. 3, $z_1^{(1)} = 0$ or $z_1^{(1)} = \infty$. As for Fig. 2, the decision between these alternatives is made by calculating F for each and taking the alternative leading to the larger F. If $B < 0$, it is clear that $z_1^{(0)} = \infty$.

(b) *The Coefficient of* $z_1^2 < 0$

When the coefficient of z_1^2 in (9) is negative, Fig. 5 or 6 applies. If Fig. 5 is appropriate, $z_1^{(1)} = 0$. Similarly, if Fig. 6 prevails, $z_1^{(1)} = B$.

Now the general procedure is to fix z_1 at $z_1^{(1)}$, to fix $z_3, \cdots, z_{k-1}$ at $z_3^{(0)}, \cdots, z_{k-1}^{(0)}$, and to repeat the procedure outlined for $z_1^{(1)}$ to obtain $z_2^{(1)}$ and so on. The iterative procedure is continued until successive sets of values of $z_1, \cdots, z_{k-1}$ are nearly equal, implying convergence of the process. (In programming this problem for the IBM 650 computer, we repeated the iterations until successive values of the maximum of F were in agreement to a desired accuracy—we chose the change in F to be less than 10^{-7}.)

The details of recognizing which one of the configurations of Figs. 2, 3, 5, and 6 applies may be set forth. The first step is to evaluate (9) to decide whether (a) or (b) applies. One next considers the roots of the equation $Q_1 = 0$. With the smaller root designated by A and the larger by B, for (a), the graph is as in Fig. 2 if $A \geq 0$ and as in Fig. 3 if $A \leq 0$. Similarly, for (b), the graph is as in Fig. 5 if $B \leq 0$ and as in Fig. 6 if $B \geq 0$. The figures for successive z_i are, of course, of the same form and all of the arguments above apply at all stages of each cycle of iterations.

(ii) Q_1 *Not Quadratic*

When Q_1 is not quadratic, several situations may again prevail. The most obvious one is that for which the first row and the first column of both C and D matrices are entirely zeros. Then all coefficients of Q_1 vanish, and $\partial F/\partial z_1 \equiv 0$. Thus z_1 is indeterminate, and we have the situation previously discussed where we must effectively consider not a k-point scale but at most a $(k - 1)$-point scale.

When the first row and first column of the C matrix consist of zeros but the D matrix does not have this form, we may proceed. The coefficient of z_1^2 in (9) is now zero but the coefficient of z_1 and the constant term are not zero. We may rewrite (10) as

$$(12) \qquad -d_{11}\left(\sum_{i=1}^{k-1}\sum_{j=1}^{k-1} c_{ij}z_i^{(0)}z_j^{(0)}\right)$$

and (11) as

$$(13) \qquad -\left(\sum_{j=2}^{k-1} d_{ij}z_j^{(0)}\right)\left(\sum_{i=1}^{k-1}\sum_{j=1}^{k-1} c_{ij}z_i^{(0)}z_j^{(0)}\right),$$

and hence Q_1 and $\partial F/\partial z_1$ are zero when

$$(14) \qquad z_1 = -\left(\sum_{j=2}^{k-1} d_{ij}z_j^{(0)}\right)\Big/ d_{11} .$$

Thus, since (12) is negative because the C and D matrices are nonnegative definite, it follows that the line Q_1 has negative slope with z_1 intercept as given in (14). Hence to maximize F, take z_1 as in (14) if the right-hand member of (14) is positive and take $z_1 = 0$ otherwise.

We have not encountered other situations where (9) may be zero. It is clear that values of $z_2^{(0)}, \cdots, z_{k-1}^{(0)}$ may be chosen which by chance will make (9) zero; if this occurs, it should be possible to take new trial values and to proceed. If other special configurations of scale-point frequencies in special experimental designs lead to difficulties through the vanishing of (9), they will require special consideration.

The remarks of this subsection apply to all Q_i as the iterations proceed.

Details on computer programming will not be outlined. We have an IBM 650 program that is initiated with the input of the coefficients c_{ij} and d_{ij} of (4) and (8), and trial values $z_2^{(0)}, \cdots, z_{k-1}^{(0)}$. Output shows successive values of the z's and corresponding values of F as the iterations proceed. This program has also been adapted for use, with some refinements added, with the IBM 1620. It is perhaps sufficient to state that this program is quite long, due to the various decisions needed in regard to Figs. 1 to 6, but the principles employed are as outlined in this section. When final values of $z_1, \cdots, z_{k-1}$ have been obtained, values of $w_1, \cdots, w_k$, suitably scaled, follow from (2).

Examples

Optimal scale points have been obtained for a number of scoring scales and a number of product groups. The examples below all relate to one-way classifications, and the D matrices are for the total sum of squares rather than the error sum of squares. The first example is given in some detail whereas only data and results are reported for the remaining.

Table 1 shows the data for a three-treatment experiment with a five-point scoring scale. The C and D matrices are given in the bottom of the table. Note that the C matrix has rank 2 whereas the D matrix has rank 4.

To obtain the C matrix, the A matrix must first be calculated and is found from the coefficients of the quadratic form,

$$[(w_1 + 2w_2 + 18w_3 + 12w_4 + 11w_5)^2/44] + [(5w_1 + 7w_2 + 12w_3 + 14w_4 + 5w_5)^2/43] + [(27w_1 + 7w_2 + 9w_3)^2/43] - [(33w_1 + 16w_2 + 39w_3 + 26w_4 + 16w_5)^2/130].$$

To obtain the D matrix (for the total sum of squares), the B matrix is calculated from

$$33w_1^2 + 16w_2^2 + 39w_3^2 + 26w_4^2 + 16w_5^2 - [(33w_1 + 16w_2 + 39w_3 + 26w_4 + 16w_5)^2/130].$$

The C and D matrices follow from A and B matrices through (5) and (6).

Iteration was begun with $z_1 = z_2 = z_3 = z_4 = 1$ for initial trial values. The corresponding value of R^2 (R^2, not F, since we are maximizing the ratio

TABLE 1

Frequencies of Scale Values and C and D Matrices for Example 1

Scale values	Treatments I	II	III	Totals
w_5 Excellent	11	5	0	16
w_4 Good	12	14	0	26
w_3 Fair	18	12	9	39
w_2 Poor	2	7	7	16
w_1 Terrible	1	5	27	33
Totals	44	43	43	130
		C Matrix		
9.18068	10.37390	7.92950	3.23014	
	11.96787	8.96026	3.88542	
		6.84884	2.79007	
			1.36216	
		D Matrix		
24.62307	20.56153	10.66153	4.06153	
	30.53076	15.83076	6.03076	
		28.43076	10.83076	
			14.03076	

TABLE 2

Successive Cycles of Iterations for Example 1

Cycle no.	Score Differences z_1	z_2	z_3	z_4	R^2
1	1	1	1	1	0.4439719
2	1.100	0.774	0.682	0.036	0.4691211
3	1.085	0.797	0.628	0.135	0.4696671
4	1.084	0.795	0.618	0.159	0.4696947
5	1.084	0.795	0.616	0.163	0.4696956
6	1.084	0.795	0.615	0.164	0.4696956

of treatment to total sums of squares) is 0.44397. The results of complete cycles of iterations are given in Table 2.

Successive scale values may be obtained by accumulating the z's. Thus $w_1 = 0, w_2 = 1.084, w_3 = 1.879, w_4 = 2.494$, and $w_5 = 2.658$. But for standardization, we adjust these to sum to zero and to have unit range: $w_1 = -0.611$, $w_2 = -0.203$, $w_3 = 0.096$, $w_4 = 0.327$, and $w_5 = 0.389$.

Four additional experiments with the same scoring scale and other variants of the same product type have been completed. We summarize these briefly by showing the data in Table 3, values of the z's and R^2 in Table 4 and values of the scale points in Table 5.

It is to be noted in Table 2 that values of R^2 do not change greatly as we move towards the optimal scale values. Also, in Table 5, it is seen that the standardized optimal scale values are similar for all five examples. In Table 6, we show values of R^2 obtained as the scales are interchanged among examples. Optimal values of R^2 show on the principal diagonal and, for example, the second R^2 in the first row is that obtained using the scale of Example 1 on the data of Example 2. It is seen that scales may be interchanged without great reduction in values of R^2 and that equal spacing gives an intermediate value of R^2 for all but the data of Example 3.

Discussion

We have provided a method for obtaining optimal weights for scoring scales while retaining the intended order on the scale. Some of the properties of the procedure have been investigated but others have not. Further work might be undertaken on the reliability of scale point estimates obtained and on the resulting distributions of the optimal F or R^2 derived. While these further investigations would be helpful, we believe that the applicability of

TABLE 3

Frequencies of Scale Values for Four Additional Examples

No.	Scale values	Treatments					
		I	II	III	IV	V	
2	w_5	10	5	11	0	0	
	w_4	8	24	10	3	4	
	w_3	10	11	15	9	5	
	w_2	4	3	5	16	7	
	w_1	1	3	3	11	16	
		I	II	III	IV	V	
3	w_5	4	4	0	8	2	
	w_4	13	20	7	5	30	
	w_3	9	10	6	3	10	
	w_2	5	3	13	15	2	
	w_1	9	7	14	11	0	
		I	II	III			
4	w_5	9	4	0			
	w_4	9	14	1			
	w_3	16	17	3			
	w_2	5	4	27			
	w_1	2	5	12			
		I	II	III	IV	V	VI
5	w_5	2.5*	7	4	5	6	1
	w_4	19	21	15	19	11	5
	w_3	7.5*	13	12	11	16	15
	w_2	11	1	3	4	10	14
	w_1	6	3	10	3	2	10

*Due to attempts to score between scale points.

TABLE 4

Score Differences and R^2 for Four Additional Examples

Example	z_1	z_2	z_3	z_4	R^2
2	0.690	1.568	0.589	0.606	0.347
3	0	41.563	7.652	0	0.244
4	0	2.563	2.260	0.515	0.124
5	0	8.981	0.513	0.870	0.474

TABLE 5

Values of Scale Points for the Five Examples

Example	w_1	w_2	w_3	w_4	w_5
1	-0.611	-0.203	0.096	0.327	0.389
2	-0.536	-0.336	0.118	0.288	0.464
3	-0.569	-0.569	0.276	0.431	0.431
4	-0.557	-0.557	0.310	0.359	0.443
5	-0.477	-0.477	0.003	0.427	0.523

TABLE 6

Values of R^2 Obtained on Interchanging Optimal Scales

Data of Example	Optimal Scale for Example					Equal Spacing
	1	2	3	4	5	
1	0.470	0.455	0.414	0.408	0.408	0.444
2	0.333	0.347	0.325	0.327	0.322	0.333
3	0.195	0.205	0.244	0.215	0.234	0.174
4	0.103	0.113	0.110	0.124	0.106	0.111
5	0.340	0.419	0.472	0.406	0.474	0.343

the method will be in the conduct of methodological studies to ascertain scale values for subsequent routine use of the scale with the same product type.

Use of the method for various scales and product types suggests the following conclusions.

(i) When many scale points are used, optimal scale values induce a collapsing of the scale and the effective number of points is smaller than the original number of points or categories.

(ii) When real and appreciable product acceptability differences occur, increased stability in scale values has been observed, and the scale values that are optimal tend to approach equally spaced scale values.

(iii) The results obtained generally add assurance in the continued use of equally spaced scales with approximately five points.

These conclusions should be regarded as tentative; they are, however, based on fairly extensive studies. Verification in other types of applications of scoring scales should be undertaken. These conclusions are based on extensive applications of the method of this paper to problems in food testing. Between fifty and one hundred sets of data have been considered. We hope to summarize these applications in a subsequent paper but this has not yet been done; such extensive summaries would not be appropriate here.

One concluding note is needed. It is clear that the method for ordered categories will yield the same result as the similar method (based on the same criterion) for unordered categories when the unrestricted maximum occurs in the principal quadrant of the z space. A reviewer has noted that this is true for Examples 1 and 2. This reviewer has also noticed that the remaining examples are disordered only in neighboring categories and that the method gives such categories identical scale values; the same results would have been obtained if one had decided a priori to group these categories and apply the method for unordered categories. From these examples, one might believe that perturbations in the data would suggest appropriate initial groupings of adjacent ordered categories. But we do have many sets of data where such grouping would not be obvious a priori and where the method of unordered categories would scramble scale values to orders where at least an extreme amount of grouping would be necessary. The important point of this paper is that such a priori judgment and insight is not required. One might argue that, where the intended natural ordering is not achieved, reconsideration of the scale used and descriptive terminology related to it should be reconsidered.

Appendices

A. *Arbitrary Origin and Scale*

The scoring scale or w scale has been discussed, and it has been noted that, for convenience, we can set $\sum_{i=1}^{k} w_i = 0$ and take $w_k - w_1 = 1$. It is perhaps intuitively obvious that this may be done but a further note may be helpful.

In the analysis of variance with N total observations, there are $N - 1$ degrees of freedom associated with the analysis after estimation of the over-all mean. These $N - 1$ degrees of freedom are associated with $N - 1$ observation contrasts, some assignable to the measurement of error, some of treatment effects, some of blocks, and so on. The sums of squares and mean squares for error, treatments, and blocks are either simple sums of squares of these contrasts or perhaps quadratic forms in them. If observations are denoted by $y_{ij\cdots}$, a contrast is a linear function $\sum a_{ij\cdots} y_{ij\cdots}$ of the observations, with $\sum a_{ij\cdots} = 0$. This implies that a contrast may always be written in terms of a linear function of differences $y_{ij\cdots} - y_{i'j'\cdots}$. But in our situation each $y_{ij\cdots}$ is equal to one of $w_1, \cdots, w_k$ and $y_{ij\cdots} - y_{i'j'\cdots} = w_h - w_{h'}$, a difference in the w's. Hence, if each $w_i = a + w_i'$, this relationship constituting a change in origin, the contrasts in terms of the w''s are identical to those in terms of the w's. The choice of a, or the choice of origin, is thus arbitrary and we have made this choice by taking $\sum_{i=1}^{k} w_i = 0$.

Once the origin has been determined, the choice of scale or units of measurement is also arbitrary. The effect of a change in scale is to multiply each contrast by the same constant, to multiply each sum of squares or mean square by the square of that constant, and to leave ratios of mean squares, such as F or R^2, unchanged.

B. *Transformation to Difference Variates*

The transformation (2) permits expression of F in (1) in terms of $z_1, \cdots, z_{k-1}$ in (4). The new coefficients c_{ij} and d_{ij} in (5) and (6) have been expressed in terms of the original coefficients a_{ij} and b_{ij}. This requires proof.

The inverse of T in (3) is

$$\text{(B1)} \qquad T^{-1} = -\frac{1}{k}\begin{bmatrix} k-1 & k-2 & k-3 & \cdots & 1 & -1 \\ -1 & k-2 & k-3 & \cdots & 1 & -1 \\ -1 & -2 & k-3 & \cdots & 1 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & -2 & -3 & \cdots & 1 & -1 \\ -1 & -2 & -3 & \cdots & -(k-1) & -1 \end{bmatrix}$$

and the pertinent quadratic forms, wAw' and wBw', transform to $z(T^{-1})'AT^{-1}z'$ and $z(T^{-1})'BT^{-1}z'$. Thus $C^* = (T^{-1})'AT^{-1}$ and $D^* = (T^{-1})'BT^{-1}$. Direct matrix multiplication yields

$$\text{(B2)} \qquad c_{ij} = \sum_{i'=1}^{i} \sum_{j'=1}^{j} a_{i'j'} - \frac{i}{k} \sum_{i'=1}^{k} \sum_{j'=1}^{j} a_{i'j'} - \frac{j}{k} \sum_{i'=1}^{i} \sum_{j'=1}^{k} a_{i'j'} + \frac{ij}{k^2} \sum_{i'=1}^{k} \sum_{j'=1}^{k} a_{i'j'} ,$$

$$\text{(B3)} \qquad d_{ij} = \sum_{i'=1}^{i} \sum_{j'=1}^{j} b_{i'j'} - \frac{i}{k} \sum_{i'=1}^{k} \sum_{j'=1}^{j} b_{i'j'} - \frac{j}{k} \sum_{i'=1}^{i} \sum_{j'=1}^{k} b_{i'j'} + \frac{ij}{k^2} \sum_{i'=1}^{k} \sum_{j'=1}^{k} b_{i'j'} ,$$

for $i, j = 1, \cdots, k - 1$. Since $z_k = 0$ by definition, we are only interested in the principal $(k - 1)$-square minors, C and D, of C^* and D^* and their elements are those of (B2) and (B3).

Further simplification follows from the nature of A and B. We consider A and note that the treatment mean square in the analysis of variance is a quadratic form, say $t\Lambda t'$, in the treatment contrasts of the vector t. (Often $\Lambda = KI$, i.e., a multiple of the identity matrix.) The treatment contrasts are linear functions of $w_1, \cdots, w_k$, say $t' = \Gamma w'$, and Γ has elements γ_{ij} such that $\sum_{i=1}^{k} \gamma_{ij} = 0$. It follows that $A = \Gamma' \Lambda \Gamma$ and that $\sum_{i=1}^{k} a_{ij} = \sum_{j=1}^{k} a_{ij} = 0$. Then (5) follows, as does (6) by a similar argument.

C. *The Attainment of an Absolute Maximum*

First we shall eliminate the indeterminate situations for which a scale of less than k points should be considered. Thus, we shall take C to be nonnegative definite and D to be positive definite (making the same restriction on D that was made in discussing the iterative procedure). We show that the iterative procedure that we have employed reaches a unique absolute maximum for F. This maximum is unique in the sense that points $z_1, \cdots, z_{k-1}$ and $Kz_1, \cdots, Kz_{k-1}$, where $K > 0$, are taken to be identical and yield the same optimal scaling.

We refer to (8) and note that $\partial F/\partial z_i = 0$ at no more than two points. Thus, in reference to Figs. 1 to 6, it is clear that F has at most one maximum at a finite point on a line parallel to an axis in the z space. Furthermore, since C and D are nonnegative definite and D has nonzero diagonal elements,

$$\lim_{z_i \to \infty} F = \lim_{z_i \to -\infty} F = c_{ii}/d_{ii} .$$

Hence, for z_j fixed, $j \neq i$, the graph of F plotted against z_i must have

> (a) one maximum, one minimum, and equal horizontal asymptotes (Fig. 7), (b) one maximum, no minimum, and equal horizontal asymptotes (Fig. 8), (c) no maximum, one minimum, and equal horizontal asymptotes (Fig. 9), or (d) zero slope everywhere on the line.

That F simply cannot have constant but nonzero slope follows from (9), (10), and (11) for the expression in (11) must be zero if those of (9) and (10) are zero. It is to be noted that the maxima and minima in Figs. 7, 8, and 9 may occur for any values of z_i, positive, negative, or zero.

Consider a one-to-one linear transformation on the z_i, say from z's to t's. Now in matrix notation $F = TUT'/TVT'$, wherein U and V are again nonnegative definite. The arguments above apply again, and hence we may conclude that the plot of F along any straight line in the z space is either constant or as in one of Figs. 7, 8, or 9. In particular note that F is constant on any straight line through the origin but that points on such lines are identical in regard to the maximization of F.

We shall now prove the uniqueness of an absolute maximum in the positive quadrant of the z space by reductio ad absurdum and through elimination of possible alternatives. Several cases arise and they will be treated separately.

(i) *There cannot be more than one local maximum of F in the interior of the principal quadrant of the z space and not collinear with the origin.*

Suppose that A and B are two points in the principal quadrant of the z space yielding local maxima for F and that A and B are not collinear with the origin. F has zero derivatives at both A and B points on the line AB. Since A and B both yield maxima, there must be a minimum between A and B and one with a zero derivative. Thus the existence of A and B implies that F has zero derivatives at three points (but not all points) on the line contrary to the possible graphs of F in Figs. 7, 8, or 9. This contradition establishes the proposition (i). It also follows that A and B, not collinear with the origin, cannot exist anywhere in the z space, both with zero derivatives on AB.

(ii) *There cannot be both a local maximum for F in the interior and on a boundary of the principal quadrant of the z space.*

Suppose that one or more local maxima of F are on the boundaries of the principal quadrant of the z space, and let A be the point yielding the largest of these. Let B be the point in the principal quadrant yielding a local maximum for F. Consider the line AB and note that $\partial F/\partial s < 0$ at A when s is the direction A to B; this follows since A would not be a local maximum if $\partial F/\partial s > 0$ and $\partial F/\partial s \neq 0$ by (i). The graph of F on AB must be as in Fig. 7 with a minimum between A and B and a maximum at B. If the line AB is not perpendicular to the boundary containing A, the extension of AB intersects another section of the boundary at A'. Furthermore, since the figure is as described, F_A (F at point A) is less than and $F_{A'}$ is greater than the horizontal asymptote. Then A' is a boundary point yielding a larger F than A, and we have a contradiction. If the line AB is perpendicular to the boundary at A, $F_A < F_{AB\infty}$, where $F_{AB\infty}$ is the height of the asymptote on AB. But now $F_{AB\infty}$ is achieved as a limit of F as the nonzero coordinates of A are held fixed and as the other coordinates become large in fixed ratios (described by these coordinates at B). An equivalent value of F is obtained when those values of the z's that approached infinity in constant ratios are held finite in the proper ratios and the remaining z's are taken to be zero. Hence, there is a point on a boundary of the principal quadrant yielding

GRAPHS OF F SHOWING MAXIMA, MINIMA, AND HORIZONTAL ASYMPTOTES

FIGURE 7

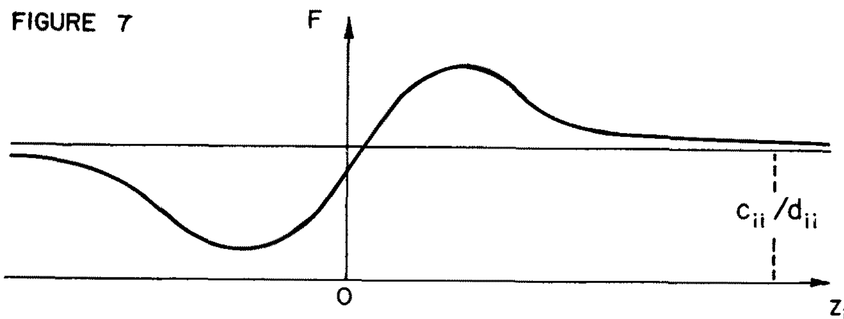


FIGURE 8

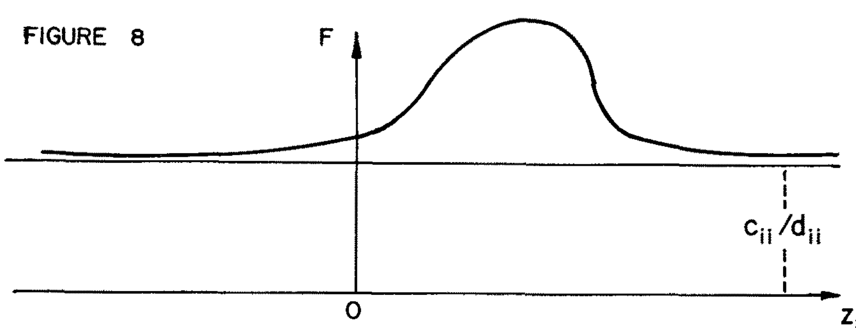
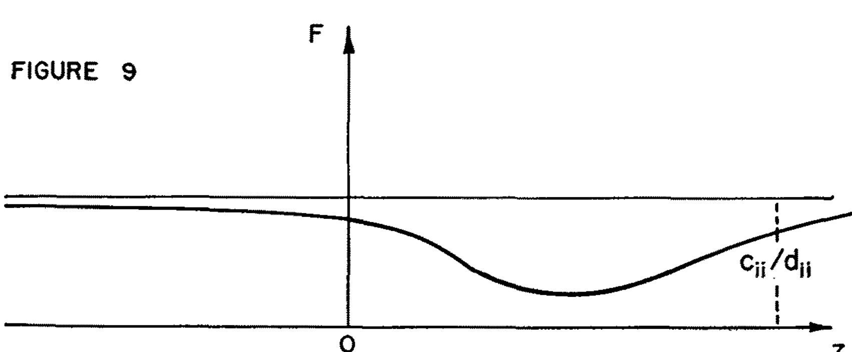


FIGURE 9



$F_{AB_\infty} > F_A$ and again we have a contradiction. Proposition (ii) now follows.

(iii) *There cannot be two local maxima for F, not collinear with the origin, on the boundary of the principal quadrant of the z space.*

Let A and B be two points, not collinear with the origin, on the boundary of the principal quadrant of the z space yielding local maxima for F, and let one of A and B be the point yielding the largest local maximum. Consider two cases: (a) A and B in the same boundary subspace and (b) A and B in boundary subspaces orthogonal to each other. If (a) prevails, the argument of (i) applies in the subspace and yields the contradiction. If (b) applies, join AB and, since the derivatives at A and B on AB must be of opposite, nonzero signs, there must be a minimum between A and B. Since one of A and B is the largest local maximum on a boundary and since there cannot be a local maximum in the principal quadrant as well, there cannot be a maximum between A and B. Then F_A and F_B are both less than F_{AB_∞} and this implies that there is a point in or on the boundary of the principal quadrant with a value of F exceeding F_A and F_B, yielding a contradiction. Thus, this third proposition follows.

We have established that there is at most one local maximum in or on the boundary of the principal quadrant of the z space. F has one maximum since F is finite for all $z_1, \cdots, z_{k-1}$. The iterative procedure will yield increasingly larger values of F as the path of iteration moves in the principal quadrant of the z space, and larger values of F can be found until the path reaches the point yielding the maximum. Hence we have shown that F has a unique absolute maximum and that it is attained by the iterative process described.

D. *Figures* 1 *and* 4 *are Impossible*

Consider the maximization of F with respect to z_1 (when $z_2 = z_2^{(0)}, \cdots, z_{k-1} = z_{k-1}^{(0)}$) and note that we may write

$$F = \frac{a_1 z_1^2 + a_2 z_1 + a_3}{b_1 z_1^2 + b_2 z_1 + b_3}. \tag{D1}$$

Since the numerator and denominator in (D1) are nonnegative definite, we have

$$a_2^2 \leq 4a_1 a_3 , \tag{D2}$$

$$b_2^2 \leq 4b_1 b_3 . \tag{D3}$$

We observe first that we can take $a_2 = 0$ without loss of generality for, if $a_1 \neq 0$, this involves only a translation and if $a_1 = 0$, (D2) implies that $a_2 = 0$. With $a_2 = 0$, $\partial F/\partial z_1$ leads to

$$Q_1 = z_1^2(a_1 b_2) + 2z_1(a_1 b_3 - a_3 b_1) - a_3 b_2 = 0. \tag{D4}$$

The discriminant of Q_1 is $(a_1b_3 - a_3b_1)^2 + 4a_1a_3b_2^2$, and this is obviously nonnegative in view of (D2). Hence Figs. 1 and 4, depending on negative discriminants, are ruled out.

When the discriminant is positive, the roots of Q_1 are unequal and the graphs in Figs. 2, 3, 5, 6 follow. When the roots of Q_1 are equal, this implies that the maximum of F is the same as the minimum of F and indeed F is constant independent of z_1 . An algebraic proof has been obtained to show directly that this is so but it is not included here for brevity.

Acknowledgments

This work was undertaken after discussions with a number of workers at the General Foods Corporation who also followed the progress of the work and provided data and helpful comment. These include Mavis B. Carroll, Donald Gardner, Otto Dykstra, George Ferris, Barbara Balcom, and John Coleman. We are pleased to acknowledge their assistance.

REFERENCES

[1] Bock, R. D. Note on the least squares solution for the method of successive categories. *Psychometrika*, 1957, **22**, 231–240.
[2] Bock, R. D. Methods and applications of optimal scaling. Report No. 25, Psychometric Laboratory, Univ. of North Carolina, 1960.
[3] Fisher, R. A. *Statistical methods for research workers* (10th ed.). Edinburgh: Oliver and Boyd, 1946.
[4] Gulliksen, H. A least squares solution for successive intervals assuming unequal standard deviations. *Psychometrika*, 1954, **19**, 117–139.
[5] Guttman, L. The quantification of a class of attributes: A theory and method of scale construction. In P. Horst *et al.*, *The prediction of personal adjustment.* New York: Social Science Research Council, 1941. Pp. 319–348.
[6] Horst, P. *The prediction of personal adjustment.* New York: Social Science Research Council, 1941.
[7] Mosteller, F. A theory of scalogram analysis, using noncumulative types of items: A new approach to Thurstone's method of scaling attitudes. Report No. 9, Laboratory of Social Relations, Harvard Univ., 1949.
[8] Torgerson, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

Manuscript received 9/14/61
Revised manuscript received 3/13/62