Contributions to the Method of Paired Comparisons

Henry F. Kaiser and Ronald C. Serlin
University of California, Berkeley

A least-squares solution for the method of paired comparisons is given. The approach provokes a theorem regarding the amount of data necessary and sufficient for a solution to be obtained. This theorem establishes that it is possible to find a solution when there is a great deal of missing data. A measure of the internal consistency of the least-squares fit is developed. It is indicated that the method of paired comparisons need not be applied only to data obtained experimentally from the law of comparative judgment; indeed, an example (rating university football teams) involving observational data is worked out.

The method of paired comparisons may be described as follows. It is desired to scale n objects by attaching to each a number x_i, $i = 1, 2, \ldots, n$, with respect to some characteristic or quality. Empirical data, observed or experimentally determined, provide numerical "differences," d_{ij}, between objects i and j for some of the $n(n-1)/2$ possible pairs of objects. The metric for these differences is not specified—beyond requiring that $d_{ij} = -d_{ji}$. For example, if i is observed to be seven greater than j, then j is observed to be seven less than i.

Before embarking on a solution to the problem posed, three examples are offered to clarify the type of scientific questions to which this method may be applied. In the first example, n objects are to be scaled according to their apparent weight along a psychological scale into x_i. Pairs of these weights are compared repeatedly, and the proportion that weight i is judged heavier than weight j is observed. According to the law of comparative judgment, this proportion is converted into a normal deviate, which is the measure of difference d_{ij}. This, of course, is a classic application of the method of paired comparisons to psychophysical scaling. In the second example, n students are to be given grades x_i on an essay they each have written, and pairs of these essays are compared. The difference in quality (in some numerical sense) d_{ij} of two essays thus compared is recorded for all pairs of essays compared. In the third example, n football teams are to be rated (given x_i values), on the basis of $d_{ij} = +1, 0, -1$ (winning, tying, losing) in a game between teams i and j, for all games played among the n teams.

A Least-Squares Solution

Ideally, any two to-be-determined scale values x_i and x_j should be such that $(x_i - x_j) = d_{ij}$ for all pairs for which a d_{ij} is observed. (This obviously implies that the observed differences would have the

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 2, No. 3 Summer 1978 pp. 421–430

property that $d_{ij} + d_{jk} = d_{ik}$ for all $i, j,$ and k.) For this ideal case, a solution for the x_i would be trivially simple: Taking an arbitrary value for one of the x_i, the remaining $n - 1$ values could be given directly from the relationships $(x_i - x_j) = d_{ij}$—if enough data d_{ij} are available to establish all connections. However, for the typical problem occurring in practice, the perfect internal consistency of the ideal situation will seldom occur: $d_{ij} + d_{jk}$ does not in general equal d_{ik}. If Illinois beats Stanford, and Stanford beats Texas A. & M., this does not imply that Illinois has or could beat Texas A. & M.

In order to obtain a solution in the usual case, consider the error, in general different from zero,

$$d_{ij} - (x_i - x_j). \tag{1}$$

It would seem desirable to choose the x_i so that, in some sense, such errors for all available data will be minimized. This may be done most easily by appealing to the principle of least-squares. Consider then the function

$$f = \sum_{i \neq j}\sum [d_{ij} - (x_i - x_j)]^2 \tag{2}$$

to be minimized with respect to the x_i. The double sum goes over all elements of the skew-symmetric matrix $D = [d_{ij}]$ for the observations which have been made. (The elements of D for which d_{ij} has not been observed are considered as zero.)

Now

$$\frac{\partial f}{\partial x_i} = -2 \sum_{\substack{j \\ j \neq i}}^{n_i} [d_{ij} - (x_i - x_j)], \tag{3}$$

where the notation indicates that the sum goes over only those j for which there is a comparison with i. Thus n_i is the number of objects with which object i has been compared.

For convenience, let

$$d_i = \sum_{\substack{j \\ j \neq i}}^{n_i} d_{ij}, \tag{4}$$

the sum of the observed differences for object i. Upon setting the derivatives equal to zero, the system of simultaneous linear equations is found:

$$\sum_{\substack{j \\ j \neq i}}^{n_i} x_i - \sum_{\substack{j \\ j \neq i}}^{n_i} x_j = d_i$$

$$n_i x_i - \sum_{\substack{j \\ j \neq i}}^{n_i} x_j = d_i. \tag{5}$$

The symmetric matrix of coefficients of the x_i in Equation 5 is singular; this is most easily seen by observing that the sum of these coefficients for all rows and columns is zero. That this singularity should occur is obvious from inspecting Equation 2: it would be possible there to add any constant to all the x_i without altering the value of the function to be minimized.

The least-squares solution developed thus far was first given by Gulliksen (1956), although Horst (1932) and Mosteller (1951) had derived it for the special case of an observed d_{ij} for all possible $n(n-1)/2$—the "complete data" case.

As suggested by $\sum\sum_{i \neq j} d_{ij} = \sum_i d_i = 0$, the "natural" condition

$$\sum_i x_i = 0 \qquad [6]$$

will be taken to establish the location of the x_i.

To apply the condition Equation 6, add Equations 5 and 6; a least-squares solution for the method of paired comparisons is given by solving the system of simultaneous linear equations

$$(n_i + 1)x_i + \sum_{\substack{j \\ j \neq i}}^{\text{not } n_i} x_j = d_i. \qquad [7]$$

The cumbersome notation of the second term of the left-hand member indicates summation over only those j for which there is *not* a comparison with i. It will be noted that the number of terms in this summation is $[(n-1) - n_i]$. Thus, the sum of the coefficients of the x_i for each of the Equations 7 is $(n_i + 1) + [(n-1) - n_i] = n$. Since the matrix of these coefficients is symmetric, this property holds also for the columns.

Designating the matrix of the coefficients of the x_i in the set of linear Equations 7 by **K**, it is seen that its elements have the following form:

$$k_{ij} = 1 \text{ or } 0 \text{ for all } i, j, i \neq j,$$

$$k_{ii} = (n_i + 1), \qquad [8]$$

$$\sum_i k_{ij} = \sum_j k_{ij} = n \text{ for all } i, j,$$

where $k_{ij} = 0$ when a comparison has been made (i.e., d_{ij} observed) and $k_{ij} = 1$ when there is no observation.

One extreme form of **K** occurs when every object has been compared with every other and thus $\mathbf{K} = n\mathbf{I}$, an easily inverted diagonal matrix which yields, by inspection, the solution $x_i = d_i/n$. At the other extreme, when there is no empirical data, every element of **K** equals one, and **K** is obviously singular. The concern of the next section is with the problem between these two extremes; that is, to state the necessary and sufficient condition for **K** to be nonsingular—answering the question of how much empirical data d_{ij} are needed to obtain a solution.

A Necessary and Sufficient Condition for a Solution

Intuition suggests that if a solution is to be obtained, that is, if **K** is to be nonsingular, then every object should be at least indirectly connected with every other; while the difference d_{ij} between objects

i and j need not be observed directly, there must be an empirical link, however remote, between them through other d_{ij}s. Another way of saying this is that there exists no self-contained group of m objects within the n objects, where no one of the m objects has any observed comparison d_{ij} with the remaining $(n-m)$ objects.

Theorem. This intuitive idea may be translated into mathematical terms as the theorem: **K** will be nonsingular if and only if after suitable symmetric) elementary transpositions, it is *not* possible to partition **K** such that

$$\left[\begin{array}{c:c} K_{11} & K_{12} \\ \hdashline K_{21} & K_{22} \end{array}\right], \qquad [9]$$

where $\mathbf{K}_{11}$ is symmetric and of order $m \times m$, $\mathbf{K}_{22}$ is symmetric and of order $(n-m)\times(n-m)$, and $\mathbf{K}_{12}$ and its transpose $\mathbf{K}_{21}$ consist entirely of ones.

Proof. From classical results of Frobenius (1908, 1912), the eigenvalue $\varkappa_1$ of maximum modulus is equal to n. The column eigenvector associated with it is $\{1\ 1\ 1\ \ldots\ 1\ 1\}$, which may be normalized to $n^{-1/2}\ \{1\ 1\ 1\ \ldots\ 1\ 1\} = q$. There may be other eigenvalues equal to n, but for that to be so, **K** must be reducible in the sense of Equation 13 below. If **K** is irreducible, there is only one maximal eigenvalue, and conversely. In any case, the spectral component associated with $\varkappa_1$ is

$$nqq' = [e_{ij}] = E, \text{ where } e_{ij} = 1. \qquad [10]$$

Let a matrix **L**, of the same type as **K**, but complementary to it, be introduced in the sense that $\ell_{ij} = 1 - k_{ij}$, $i \neq j$, row sums being equal to n as before. Evidently,

$$L = nI + E - K,\ K = nI + E - L, \qquad [11]$$

while **L** also has maximal eigenvalue $\lambda_1 = n$ and associated spectral component **E**.

Since **E** itself has the single nonzero eigenvalue n, all others being zero, and since **nI** can be resolved into spectral components $nq_{(i}q'_{(i}$ based on *any* complete set of orthonormal vectors $q_{(i}$ each of n elements, it follows that **nI + E, K** and **L** can be simultaneously brought to diagonal canonical form by the *same* orthogonal transformation **H**′ ( )**H.** Hence since their respective sets of eigenvalues are

$$2n,\ n,\ n,\ \ldots,\ n;\ n,\ \kappa_2,\ \kappa_3,\ \ldots,\ \kappa_n;\ n,\ \lambda_2,\ \lambda_3,\ \ldots,\ \lambda_n, \qquad [12]$$

and since $2n$, n, n are respectively maximal in these, we must have, by Equation 11, $\lambda_i = n - \varkappa_i, > 1$. Further, all $\varkappa_i$ and λ_i must be non-negative.

Hence, matrices of type **K** are positive definite or non-negative definite. If **K** has p eigenvalues equal to zero, it can be said that it is non-negative definite of nullity p. In such a case, **L** will have $p+1$ eigenvalues equal to n; and by the results of Frobenius, this cannot be so unless by a similar permutation of rows and columns **L** can be brought to the shape

$$\begin{bmatrix} L_1 & & & & \\ & L_2 & & \text{(zeros)} & \\ & & \ddots & & \\ & \text{(zeros)} & & \ddots & \\ & & & & L_{p+1} \end{bmatrix}, \qquad [13]$$

where the diagonal submatrices $\mathbf{L}_i$ are not further reducible and where all non-diagonal blocks are null submatrices. But this again, by Equation 11, implies that by the same permutation of rows and columns, **K** has been brought to the shape

$$\begin{bmatrix} K_1 & & & & \\ & K_2 & & \text{(units)} & \\ & & \ddots & & \\ & \text{(units)} & & \ddots & \\ & & & & K_{p+1} \end{bmatrix}, \qquad [14]$$

where now all non-diagonal blocks are exclusively of unit elements.

This proves the theorem which stated that a necessary and sufficient condition for **K** to be singular was that it should reduce, by a similar permutation of rows and columns, to the shape

$$\begin{bmatrix} K_{11} & E_{12} \\ & \\ E_{21} & K_{22} \end{bmatrix}, \qquad [15]$$

where $\mathbf{E}_{12}$, $\mathbf{E}_{21}$ had exclusively unit elements.

With a little effort constructing examples of matrices of the form **K,** the implication of the theorem becomes obvious: It is possible to scale the weights, grade the essays, or rate the football teams with remarkably little data d_{ij}. The amount of data needed is much less than suggested by Schonemann (1970).

Internal Consistency

After obtaining the values x_i, any observed d_{ij} may be partitioned into two parts

$$d_{ij} = [x_i - x_j] + [d_{ij} - (x_i - x_j)]. \qquad [16]$$

The first part, $[x_i - x_j]$, may be considered the predicted d_{ij}, while $[d_{ij} - (x_i - x_j)]$ is the residual or error part. In this section, a measure of the efficacy of the least-squares fit will be developed.

As suggested by other measures of this sort, it seems reasonable to take for the desired measure r^2 the ratio of the predicted sum of squares to the total sum of squares:

$$r^2 = \frac{\text{SS (predicted)}}{\text{SS (total)}} = \frac{\sum\sum_{i \neq j} (x_i - x_j)^2}{\sum\sum_{i \neq j} d_{ij}^2} \qquad [17]$$

where the double summations in both numerator and denominator go over only those pairs for which observations d_{ij} have been made. (There is no correction for means in Equation 17, because both components have mean zero.)

Alternatively, a seemingly appropriate measure would be given by one minus the ratio of the error sum of squares to the total sum of squares:

$$r^2 = 1 - \frac{\text{SS (error)}}{\text{SS (total)}} = 1 - \frac{\sum\sum_{i \neq j} [d_{ij} - (x_i - x_j)]^2}{\sum\sum_{i \neq j} d_{ij}^2} . \qquad [18]$$

It can be shown that Equations 17 and 18 are equivalent (as might be expected from the use of least squares).

Upon rewriting Equation 7 as

$$Kx = d \qquad [19]$$

and verifying that

$$\sum\sum_{i \neq j} (x_i - x_j)^2 = 2x'Kx , \qquad [20]$$

a simple computing formula for this measure r^2 is given by noting from Equations 19 and 20 that

$$\sum\sum_{i \neq j} (x_i - x_j)^2 = 2x'Kx = 2x'd = 2\sum_i x_i d_i \qquad [21]$$

—a variant of the usual least-squares property that the sum of squares accounted for is given by the inner product of the solution of the normal equations and the right-hand members of the normal equations. Substituting Equation 21 in Equation 17 gives

$$r^2 = \frac{2\sum_i x_i d_i}{\sum\sum_{i \neq j} d_{ij}^2} , \qquad [22]$$

or even more simply

$$r^2 = \frac{\sum_i x_i d_i}{\sum\sum_{i<j} d_{ij}^2} = \frac{\sum_i x_i d_i}{\sum\sum_{i>j} d_{ij}^2} . \qquad [23]$$

The coefficient r^2 cannot be considered a measure of reliability in the sense of stability. Indeed, consider a situation in which there were just barely enough data to obtain a solution; in this case, it would more easily be possible to find the x_i such that $x_i - x_j = d_{ij}$ for some d_{ij}—thus tending to increase r^2—but not a situation which could be expected to yield more stable x_i (in that further observations would tend not to affect the x_i). As the extreme example, let there be only the absolute minimum $n - 1$ observations $d_{i,i+1}$, $i = 1, 2, \ldots, n - 1$. Here, according to the theorem, **K** is nonsingular, but the direct connections between the objects are limited to one on either side of all but the first and last, which have only one connection. It is readily seen that it is possible in this extreme case of minimum data and thus probably minimum stability to find the x_i such that $x_i - x_j = d_{ij}$ for all d_{ij} observed, and consequently $r^2 = 1$.

Table 1
The Skew-Symmetric Data Matrix D of d_{ij} for the Games of 12 Major Independent University Football Teams, 1976 Season

	Air Force	Army	Boston College	Colgate	Georgia Tech	Holy Cross	Miami (Florida)	Navy	Notre Dame	Penn State	Pittsburgh	Rutgers	Sum, d_i
Air Force	0	−1	0	0	0	0	0	+1	0	0	0	0	0
Army	+1	0	−1	+1	0	+1	0	−1	0	−1	−1	0	−1
Boston College	0	+1	0	0	0	+1	−1	+1	0	0	0	0	+2
Colgate	0	−1	0	0	0	+1	0	0	0	0	0	−1	−1
Georgia Tech	0	0	0	0	0	0	0	−1	+1	0	−1	0	−1
Holy Cross	0	−1	−1	−1	0	0	0	0	0	0	0	0	−3
Miami (Florida)	0	0	+1	0	0	0	0	0	−1	−1	−1	0	−2
Navy	−1	+1	−1	0	+1	0	0	0	−1	0	−1	−1	−3
Notre Dame	0	0	0	0	−1	0	+1	+1	0	0	−1	0	0
Penn State	0	+1	0	0	0	0	+1	0	0	0	−1	0	+1
Pittsburgh	0	+1	0	0	+1	0	+1	+1	+1	+1	0	0	+6
Rutgers	0	0	0	+1	0	0	0	+1	0	0	0	0	+2
Sum, $-d_i$	0	+1	−2	+1	+1	+3	+2	+3	0	−1	−6	−2	0

However, r^2 is surely a measure of reliability in the sense of internal consistency. For it is simply a standard measure of the internal closeness of the least-squares fit of the x_i to the data available—regardless of whether these x_i will bear any relationship to the x_i for the same problem with different data. It is in this sense of internal consistency that r^2 should be interpreted.

Example

In Table 1 are given the differences d_{ij} for games of the 1976 season between 12 major independent university football teams. These differences are assigned the value +1 if team i beat team j, 0 if the two teams tied (which does not occur here), or if the teams did not play and −1 if team i lost to team j. For example, Georgia Tech beat Notre Dame (and, of course, Notre Dame lost to Georgia Tech). Table 1, then, is the skew-symmetric data matrix of observed d_{ij}, together with the row sums d_i of the difference between games won and games lost.

The matrix of coefficients **K** for this problem is given in Table 2. Thus, from the third column, Boston College played Army, Holy Cross, Miami (Florida), and Navy. And the diagonal element, $k_{33} = 5$ $(= n_3 + 1)$, indicates that Boston College played $n_3 = 4$ games with the teams under consideration. By counting the number of zeros on one side of the diagonal K, it is seen that these teams played 24 games with each other during the 1976 season; thus about 36% of the possible 66 $(= n(n-1)/2)$ comparisons were made.

The vector **d** of row sums of **D**, the right-hand sides of the normal equations, is repeated in Table 2, and the vector **x** of solutions to these equations is given alongside. These values x_i may be considered ratings for the teams.

The measure of internal consistency r^2 for this problem is given, from Equation 23, by finding the inner product of the vectors **d** and **x** in Table 2, and dividing by the sum of squares of elements below (say) the diagonal of **D** in Table 1, that is,

$$r^2 = \frac{13.5982}{24} = .5667. \qquad [24]$$

To illustrate that r^2 is not a measure of reliability in the sense of stability, consider the effect of adding a 13th team, USC, to the problem. Of the teams listed, USC played only Notre Dame, and won. Obviously, then, USC would have a rating exactly one greater than Notre Dame's—for this comparison $x_i - x_j = d_{ij}$—and the numerator and denominator in Equation 24 would both be increased by one, increasing r^2 to .5839. Yet, based on only one game, USC could not be regarded as having received a stable rating.

For this example the d_{ij} is considered to be +1, 0, −1 (winning, tying, losing). This surely seems to be "what really counts" in a football game. But the d_{ij}s as the "point spreads" in the games could have been observed. For a given problem, what constitutes the appropriate d_{ij} may be a difficult question and is beyond the scope of this paper.

A Linear Transformation on the Scale Values

The origin of the scale values x_i is zero, which is arbitrary. For the example, the scale is uninteresting. For the latter problem, let a constant, c, be chosen with which to multiply each of the x_i to minimize

$$g = \sum_{i \neq j}\sum [p_{ij} - c(x_i - x_j)]^2, \qquad [25]$$

where p_{ij} is the "point spread" in the game between team i and team j and the x_i are as given in Table

Table 2
The Symmetric Matrix K, the vector d, and the Vector
$x = K^{-1}d$ of Ratings for 12 Major Independent University Football Teams, 1976 Season

	Air Force	Army	Boston College	Colgate	Georgia Tech	Holy Cross	Miami (Florida)	Navy	Notre Dame	Penn State	Pittsburgh	Rutgers	d_i	x_i
Air Force	3	0	1	1	1	1	1	0	1	1	1	1	0	-.2194
Army	0	8	0	0	1	0	1	0	1	0	0	1	-1	-.2262
Boston College	1	0	5	1	1	0	0	0	1	1	1	1	+2	+.0724
Colgate	1	0	1	4	1	0	1	1	1	1	1	0	-1	-.6386
Georgia Tech	1	1	1	1	4	1	1	0	0	1	0	1	-1	+.0244
Holy Cross	1	0	0	0	1	4	1	1	1	1	1	1	-3	-1.2641
Miami (Florida)	1	1	0	1	1	1	5	1	0	0	0	1	-2	-.0076
Navy	0	0	0	1	0	1	1	8	0	1	0	0	-3	-.2126
Notre Dame	1	1	1	1	0	1	0	0	5	1	0	1	0	+.2180
Penn State	1	0	1	1	1	1	0	1	1	4	0	1	+1	+.6114
Pittsburgh	1	0	1	1	0	1	0	0	0	0	7	1	+6	+1.0679
Rutgers	1	1	1	0	1	1	1	0	1	1	1	3	+2	+.5744
Sum	12	12	12	12	12	12	12	12	12	12	12	12	0	0.0000

2. Solving the elementary calculus problem for the example, $c = 22.70$ is found. To remove the negative signs 100 is added to the cx_i, yielding the more convenient and perhaps more interesting scale values $100 + cx_i$:

Pittsburgh	124.2	Miami (Florida)	99.8
Penn State	113.9	Navy	95.2
Rutgers	113.0	Air Force	95.0
Notre Dame	105.0	Army	94.9
Boston College	101.6	Colgate	85.5
Georgia Tech	100.6	Holy Cross	71.3

References

Frobenius, G. Ueber Matrizen aus positiven Elementen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, 1908, 471–476; 1909; 514–518.

Frobenius, G. Ueber Matrizen aus nicht-negativen Elementen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, 1912, 456–477.

Gulliksen, H. A least squares solution for paired comparisons with incomplete data. *Psychometrika*, 1956, *21*, 125–134.

Horst, P. A method for determining the absolute value of a series of stimulus situations. *Journal of Educational Psychology*, 1932, *23*, 418–440.

Mosteller, F. Remarks on the method of paired comparisons. I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 1951, *16*, 3–11.

Schonemann, P. A note on Gulliksen's least squares solution. *British Journal of Mathematical and Statistical Psychology*, 1970, *23*, 69–71.

Acknowledgments

We are deeply indebted to the late Professor A. C. Aitken of the University of Edinburgh who proved the sufficiency part of our theorem; the proof of the theorem, as it appears above, is due to Aitken. The research reported in this paper was supported in part by the Program in Applied Mathematics and Statistics, National Science Foundation, and by the National Institute of Education, Department of Health, Education, and Welfare.

Author's Address

Henry F. Kaiser, School of Education, University of California, Berkeley, CA 94720