PAIRED COMPARISONS

by

Ralph A. Bradley*

FSU Statistics Report No. M615
ONR Technical Report No. 157

May, 1982

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

*The work of the author is supported in part by the Office of Naval Research under Contract N00014-80-C-0093. Reproduction in whole or in part is permitted for any purpose of the United States Government.

PREFACE

This technical report is an invited chapter for the *Handbook of Statistics: Nonparametric Methods*, Volume 4 in a series edited by P. R. Krishnaiah and P. K. Sen and to be published by North-Holland Publishing Company, Amsterdam. Much of the material by the author used in the chapter was developed under ONR-sponsored research at the Florida State University and earlier at the Virginia Polytechnic Institute and State University. Some minor new generalizations of earlier work are included here.

Ralph A. Bradley

Paired Comparisons

by

Ralph A. Bradley*
Department of Statistics
Florida State University
Tallahassee, FL 32306

1. Introduction

Interest in paired comparisons in statistics and psychometrics has developed in the contexts of the design of experiments, nonparametric statistics, and scaling, including multidimensional scaling. Applications have arisen in many areas, but most notably in food technology, marketing research, and sports competition. An extensive bibliography on paired comparisons by Davidson and Farquhar (1976) contains some 400 references.

Paired comparisons have been considered in design of experiments as incomplete block designs with block size two by Clatworthy (1955) and others. Scheffé (1952) developed an analysis of variance for paired comparisons with consideration for possible order effects for the two treatments within blocks. When the usual parametric models of analysis of variance are imposed, the analysis of such designs follows standard methods and will not be discussed here.

The emphasis in this chapter will be on paired comparisons as a means of designing comparative experiments when no natural measuring scale is available. The author's interest in paired comparisons arose in consideration of statistical methods in sensory difference testing.

---

*The work of the author is supported in part by the Office of Naval Research under Contract N00014-80-C-0093.

When responses of individuals to items under comparison are subjective, and particularly when sensory responses to taste, odor, color or sound are involved, evaluation is easier when the number of items or samples to be considered at one time is small and the effects of sensory fatigue are minimized. Probabilistic models for paired comparisons may be devised to represent the experimental situation and permit appropriate data analysis. The models provide probabilities of possible choices of items or treatments from pairs of items and hence depend on orderings. The statistical methods devised are thus ranking methods and, while they are not literally non-parametric methods, they are often so classified.

The basic paired comparisons experiment has t treatments, $T_1, \ldots, T_t$, and $n_{ij} \geq 0$ comparisons of T_i with T_j, $n_{ji} = n_{ij}$, $i \neq j$, $i, j = 1, \ldots, t$. For each comparison, preference or order is designated by $a_{ij\alpha}$, $a_{ij\alpha} = 1$ if T_i is "preferred" to T_j in the α^{th} comparison of T_i and T_j, $a_{ij\alpha} = 0$ otherwise, $a_{ij\alpha} + a_{ji\alpha} = 1$. In further definition of notation, let

$$a_{ij} = \sum_{\alpha=1}^{n_{ij}} a_{ij\alpha} \text{ and } a_i = \sum_{\substack{j \\ j\neq i}} a_{ij}, \text{ the total number of preferences for } T_i.$$

In sensory evaluations, responses may be preferences or attribute order judgments on such characteristics as sweetness, smoothness, whiteness, etc. We shall loosely refer to preference judgments.

Dykstra (1960) provides typical data on a paired comparisons preference taste test involving four variations of the same product. The data are summarized in Table 1. Note that the experiment is not balanced: $n_{12} = 140$, $n_{13} = 54$, $n_{14} = 57$, $n_{23} = 63$, $n_{24} = 58$, $n_{34} = 0$; treatments T_3 and T_4 were not compared. Unbalanced experiments are permissible as long as the design is connected: it is not possible to select a subset of the treatments

Table 1

Summary of Results of a Taste Test

	T_1	T_2	T_3	T_4	a_i
T_1	--	28	15	23	66
T_2	112	--	46	47	205
T_3	39	17	--	--	56
T_4	34	11	--	--	45

such that no treatment in the subset is compared directly with a treatment in the complementary subset. Balanced experiments are more efficient when there is equal interest in all treatments and treatment comparisons.

We shall return to analysis of the data of Table 1, which gives values of a_{ij}, after discussion of models for paired comparisons and establishment of basic procedures.

This chapter is organized in such a way as to give initial attention to the analysis of basic paired comparisons data like those of Table 1. Then extensions of the method are developed for factorial treatment combinations and for multivariate responses, responses on several attributes for each paired comparison. The emphasis is on the methodology and applications, although properties of procedures are noted and references given. We conclude with comments on additional methods of analysis.

2. Models for Paired Comparisons

When $t = 2$, a paired comparisons experiment with treatments T_1 and T_2 might be modelled as $n_{12} > 0$ independent Bernoulli trials with probabilities of choices for T_1 and T_2 being π_1 and π_2, $\pi_i \geq 0$, $i = 1, 2$, $\pi_1 + \pi_2 = 1$. Then in some sense π_1 and π_2 are measures of "worth" of T_1 and T_2. Binomial theory applies and the sign test may be used to test the hypothesis, H_0: $\pi_1 = \pi_2$.

Bradley and Terry (1952a) proposed a basic model for paired comparisons, extended by Dykstra (1960) to include unequal values of the n_{ij}. The approach was a heuristic extension of the special binomial when $t = 2$. Treatment parameters, $\pi_1, \ldots, \pi_t$, $\pi_i \geq 0$, $i = 1, \ldots, t$, are associated with the t treatments, $T_1, \ldots, T_t$. It was postulated that these parameters represent relative selection probabilities for the treatments so that the probability of selection of T_i when compared with T_j is

$$P(T_i \to T_j) = \pi_i/(\pi_i+\pi_j), \; i \neq j, \; i, j = 1, \ldots, t. \tag{2.1}$$

Since the right-hand member of (2.1) is invariant under change of scale, specificity was obtained by the requirement that

$$\sum_{i=1}^{t} \pi_i = 1. \tag{2.2}$$

The model proposed imposes structure in that the most general model might postulate binomial parameters π_{ij} and $\pi_{ji} = 1 - \pi_{ij}$ for comparisons of T_i and T_j so that the totality of functionally independent parameters is $\binom{t}{2}$ rather than $(t-1)$ as specified in (2.1) and (2.2).

The basic model (2.1) for paired comparisons has been discovered and rediscovered by various authors. Zermelo (1929) seems to have proposed it first in consideration of chess competition. Ford (1957) proposed the model independently. Both Zermelo and Ford concentrated on solution of normal equations for parameter estimation and Ford proved convergence of the iterative procedure for solution.

The model arises as one of the special simple realizations of more general models developed from distributional or psychophysical approaches. Bradley (1976) has reviewed various model formulations and discussed them under categories -- linear models, the Lehmann model, psychophysical models, and models of choice and worth.

David (1963, Section 1.3) supposes that T_i has "merit" V_i, $i = 1, \ldots, t$, when judged on some characteristic, and that these merits may be represented on a merit scale. He defined "linear" models to be such that

$$P(T_i \to T_j) = H(V_i - V_j), \tag{2.3}$$

where H is a distribution function for a symmetric distribution, $H(-x) = 1 - H(x)$. Model (2.1) is a linear model since it may be written in the form,

$$P(T_i \to T_j) = \tfrac{1}{4}\int_{-(\log \pi_i - \log \pi_j)}^{\infty} \operatorname{sech}^2 y/2 \, dy = \pi_i/(\pi_i + \pi_j), \tag{2.4}$$

as described by Bradley (1953) using the logistic density function.

Thurstone (1927) proposed a model for paired comparisons, that is also a linear model, through the concept of a subjective continuum, an inherent sensation scale on which order, but not physical measurement, could be discerned. Mosteller (1951) provides a detailed formulation and an analysis of Thurstone's important Case V. With suitable scaling, each treatment has a location point on the continuum, say μ_i for T_i, $i = 1, \ldots, t$. An individual is assumed to receive a sensation X_i in response to T_i, with responses X_i normally distributed about μ_i. When an individual compares T_i and T_j, he in effect is assumed to report the order of sensations X_i and X_j which may be correlated; $X_i > X_j$ may be associated with $T_i \to T_j$. Case V takes all such correlations equal and the variances of all X_i equal. The probability of selection may be writton

$$P(T_i \to T_j) = P(X_i > X_j) = \frac{1}{\sqrt{2\pi}} \int_{-(\mu_i-\mu_j)} e^{-y^2/2} dy. \tag{2.5}$$

It is apparent from (2.4) and (2.5) that the two models are very similar. The choice between the models is much like the choice between logits and probits in biological assay. The use of $\log \pi_i$ as a measure of location for T_i in the first model is suggested.

Models (2.4) and (2.5) give very similar results in applications. Comparisons are made by Fleckenstein, Freund and Jackson (1958) with test data on comparisons of typewriter carbon papers. In general, more extensions of model (2.4) exist and we shall use that model in this chapter.

3. Basic Procedures

The general approach to analysis of paired comparisons based on the model (2.1) is through likelihood methods. On the assumption of independent responses for the n_{ij} comparisons of T_i and T_j, the binomial component of the likelihood function for this pair of treatments is

$$\left(\frac{\pi_i}{\pi_i+\pi_j}\right)^{a_{ij}}\left(\frac{\pi_j}{\pi_i+\pi_j}\right)^{a_{ji}} = \pi_i^{a_{ij}}\pi_j^{a_{ji}}/(\pi_i+\pi_j)^{n_{ij}},$$

ties or no preference judgments not being permitted. The complete likelihood function, on the assumption of independence of judgments between pairs of treatments, is

$$L = \prod_i \pi_i^{a_i} / \prod_{i<j} (\pi_i+\pi_j)^{n_{ij}}. \tag{3.1}$$

It is seen that $a_1, \ldots, a_t$ constitute a set of sufficient statistics for the estimation of $\pi_1, \ldots, \pi_t$ and that a_i is the total number of preferences or selections of T_i, $i = 1, \ldots, t$, for the entire experiment.

3.1. Likelihood Estimation

ML estimators, p_i for π_i, $i = 1, \ldots, t$, are obtained through maximization of log L in (3.1) subject to the constraint (2.2). After minor simplifications, the resulting likelihood equations are

$$\frac{a_i}{p_i} - \sum_{\substack{j \\ j\neq i}} \frac{n_{ij}}{p_i+p_j} = 0, \quad i = 1, \ldots, t, \tag{3.2}$$

and

$$\sum_i p_i = 1. \tag{3.3}$$

Solution of equations (3.2) and (3.3) is done iteratively. If $p_i^{(k)}$ is the k^{th} approximation to p_i,

$$p_i^{(k)} = p_i^{*(k)}/\sum_i p_i^{*(k)},$$

where

$$p_i^{*(k)} = a_i/\sum_{\substack{j\\ j\neq i}} [n_{ij}/(p_i^{(k-1)}+p_j^{(k-1)})], \quad k = 1, 2, \ldots .$$

The iteration is started with initial specification of the $p_i^{(0)}$; one may take $p_i^{(0)} = 1/t$, $i = 1, \ldots, t$, and this is adequate although Dykstra (1956, 1960) has suggested better initial values.

We return to the example of Table 1. Values of a_i are given in the table and values of n_{ij} precede the table. Solution of equations (3.2) and (3.3) was begun with $p_i^{(0)} = 1/4$, $i = 1, \ldots, 4$. Results for initial iterations are summarized in Table 2 along with final values for p_i; typically approximately 10 iterations are sufficient for four-decimal accuracy in the final values. It is this iterative procedure that Ford (1957) has shown to converge. The procedure is easy to program on computers because of the symmetry of the equations to be solved. Bradley and Terry (1952a) and Bradley (1954a) have provided tables giving values of the p_i for equal values of the $n_{ij} = n$, $t = 3$, $n = 1, \ldots, 10$; $t = 4$, $n = 1, \ldots, 8$; $t = 5$, $n = 1, \ldots, 5$.

In small experiments, small values of the n_{ij}, perhaps with poorly selected treatments, the estimates p_i may define a point on a boundary of the parameter space. These situations may be recognized from tables like Table 1 and require special consideration. As an example, refer to Table 1 and suppose that T_2 and T_3 are always preferred to T_1 and T_4 and Table 1 is unchanged otherwise. Then $a_1 = 23$, $a_2 = 244$, $a_3 = 71$ and $a_4 = 34$. Treatments T_2 and T_3 dominate T_1 and T_4 and information on the

Table 2

Values of the Estimators in the Iterative Solution

T_i	$p_i^{(0)}$	$p_i^{(1)}$	$p_i^{(2)}$	$p_i^{(3)}$	$p_i^{(4)}$	$p_i^{(5)}$	p_i
1	.25	.1371	.1188	.1137	.1112	.1101	.1082
2	.25	.4094	.4656	.4918	.5049	.5131	.5193
3	.25	.2495	.2413	.2357	.2327	.2290	.2294
4	.25	.2040	.1743	.1588	.1512	.1478	.1431

relative values of T_2 and T_3 comes only from the direct comparisons of T_2 and T_3. It follows that $p_1 = 0$, $p_2 = 46/63 = .7302$, $p_3 = 17/63 = .2698$, and $p_4 = 0$. But there is also information on the relative values of π_1 and π_4. We find $p_1/p_4 = 23/34 = .4035/.5965$ and can write $p_1 = .4035\delta$ and $p_4 = .5965\delta$, δ infinitesimal. A formal analysis may be conducted through minimization of log L with respect to π_1^*, π_2, π_3, π_4^*, $\pi_2 + \pi_3 = 1$, $\pi_1^* + \pi_4^* = 1$, where $\pi_1 = \delta\pi_1^*$, $\pi_4 = \delta\pi_4^*$ and δ is small. Indeed, the maximum value of log L may be found in this way and it is needed in the computation of likelihood ratios as discussed below. Bradley (1954a) provides additional discussion of these special boundary problems. problems not usually encountered in applications.

3.2. Tests of Hypotheses

(i) The major test proposed by Bradley and Terry (1952) was that of treatment preference or selection equality. The null hypothesis is

$$H_0: \pi_1 = \pi_2 = \ldots = \pi_t = 1/t \tag{3.4}$$

and the general alternative hypothesis is

$$H_a: \pi_i \neq \pi_j \text{ for some } i, j, i \neq j, i, j = 1, \ldots, t. \tag{3.5}$$

If we designate the likelihood ratio as λ_1, it is easy to show that

$$-2 \log \lambda_1 = 2N \log 2 - 2B_1, \; N = \sum_{i<j} n_{ij},$$
$$B_1 = \sum_{i<j} n_{ij} \log(p_i+p_j) - \sum_i a_i \log p_i. \tag{3.6}$$

For large n_{ij}, $-2 \log \lambda_1$ has the central chi-square distribution with $(t-1)$ degrees of freedom under H_0. Values of B_1, together with exact significance levels, were provided with the cited tables* of estimators p_i. Comparison of significance levels for the large-sample test with small-sample exact significance levels in the tables suggests that the former may be used for modest values of the n_{ij}, a situation perhaps comparable to use of the normal approximation to the binomial.

For the values of the a_i of Table 1, the noted values of the n_{ij} above that table, $N = 372$, and the values of the p_i in Table 2, we have $B_1 = 206.3214$ and $-2 \log \lambda_1 = 103.06$ with 3 degrees of freedom. There is a clear indication that the π_i are not equal and that treatment preferences differ.

(ii) It is always encumbent on statisticians to check the validity of models used in statistical analyses when possible. We have noted above that a general "multi-binomial" model with $\binom{t}{2}$ functionally independent parameters π_{ij} may be posed that ignores the structure of paired comparisons in the sense that the same treatment is compared with more than

*Common logarithms were used to compute B_1 in these tables. In this paper, natural logarithms are used throughout.

one other treatment. The multi-binomial model fits the data of tables like Table 1 perfectly. This permits a test of the more restrictive model of (2.1).

The following likelihood ratio test was proposed by Bradley (1954b) and extended by Dykstra (1960). Consider the null hypothesis,

$$H_0\colon \pi_{ij} = \pi_i/(\pi_i+\pi_j),\ i \neq j,\ i,\ j = 1,\ \ldots,\ t, \tag{3.7}$$

and the alternative hypothesis,

$$H_a\colon \pi_{ij} \neq \pi_i/(\pi_i+\pi_j),\ \text{for some } i,\ j,\ i \neq j. \tag{3.8}$$

Under H_a, the likelihood estimator of π_{ij} is $p_{ij} = a_{ij}/n_{ij}$ when $n_{ij} > 0$ and the estimator is not needed when $n_{ij} = 0$. Under H_0, p_i is the estimator of π_i from equations (3.2) and (3.3). Designating λ_2 as the likelihood ratio statistic, we have

$$-2 \log \lambda_2 = 2\Big(\sum_{i\neq j} a_{ij} \log a_{ij} - \sum_{i<j} n_{ij} \log n_{ij} + B_1\Big). \tag{3.9}$$

For large n_{ij}, $-2 \log \lambda_2$ is taken to have the chi-square distribution with $\binom{t}{2} - (t-1) = \tfrac{1}{2}(t-1)(t-2)$ degrees of freedom under H_0. An alternative statistic, asymptotically equivalent to that of (3.9), is

$$\chi^2 = \sum_{i\neq j} (a_{ij}-a'_{ij})^2/a'_{ij}, \tag{3.10}$$

where $a'_{ij} = n_{ij}p_i/(p_i+p_j)$ and $a_{ij} = n_{ij}p_{ij}$. This alternate form may be rewritten,

$$\chi^2 = \sum_{i\neq j} n_{ij}\{p_{ij} - [p_i/(p_i+p_j)]\}^2/[p_i/(p_i+p_j)]. \tag{3.11}$$

Dykstra has noted that the test statistics may be distorted when some n_{ij} are small. Since there is no basis for pooling terms in this case, he suggested omitting terms in (3.11) with very small values of n_{ij} (and hence n_{ji}) and deleting one degree of freedom for each pair of terms so deleted.

For the data of Table 1, $n_{34} = 0$ and the tests for the fit of the model have $\frac{1}{2}(3)(2) - 1 = 2$ degrees of freedom. From (3.9), $-2 \log \lambda_2 = 2.02$ and there seems to be no reason to doubt the appropriateness of the model (2.1). The statistic in (3.10) is evaluated also for illustrative purposes. Values of the a'_{ij} are given in Table 3 and they may be compared directly with the values of a_{ij} in Table 1. Computation yields $\chi^2 = 2.00$; the close agreement of the two computations is typical.

Table 3
Estimated Frequencies for the Data of Table 1

	T_1	T_2	T_3	T_4	Row Sums
T_1	-	24.14	17.31	24.54	65.99
T_2	115.86	-	43.70	45.47	205.03
T_3	36.69	19.30	-	-	55.99
T_4	32.46	12.53	-	-	44.99

In the author's fairly extensive experience in fitting model (2.1) to data in food technology and consumer testing, the model is usually found to fit well. When the model does not fit, one or more treatments are often found to possess a characteristic not found in the others, possibly leading to preference judgments influenced by this attribute when such treatments are in a comparison.

(iii) In some uses of paired comparisons, responses may be obtained for several demographic groups, under different evaluation conditions, or other criterion for grouping responses. The possibility of group by treatment interaction or preference disagreement arises and this may be tested.

Let $u = 1, \ldots, g$ index groups of responses in paired comparisons, let π_i^u be the treatment parameter for T_i in group u, and suppose that sufficient comparisons are made within each group to obtain p_i^u, the estimator of π_i^u, $i = 1, \ldots, t$. Interest is in the hypotheses,

$$H_0\colon \pi_i^u = \pi_i,\ i = 1, \ldots, t;\ u = 1, \ldots, g, \tag{3.12}$$

and

$$H_a\colon \pi_i^u \neq \pi_i \text{ for some i and u.} \tag{3.13}$$

The likelihood ratio test depends on

$$-2 \log \lambda_3 = 2(B_1 - \sum_{u=1}^{g} B_{1u}),$$

where B_{1u} is computed from (3.6) for the data within group u and B_1 is computed similarly for the pooled data from all of the groups. For large values of the n_{iju}, the number of comparisons of T_i and T_j in group u, $-2 \log \lambda_3$ has the central chi-square distribution with $(g-1)(t-1)$ degrees of freedom under H_0 of (3.12).

An omnibus test of treatment equality may be described:

$$H_0\colon \pi_i^u = 1/t,\ i = 1,\ \ldots,\ t;\ u = 1,\ \ldots,\ g,$$

$$H_a\colon \pi_i^u \neq 1/t \text{ for some } i \text{ and } u,$$

$$-2 \log \lambda_4 = 2N \log 2 - 2 \sum_{u=1}^{g} B_{1u},\ N = \sum_{u=1}^{g} N_u = \sum_{u} \sum_{i<j} n_{iju}.$$

The test statistic is taken to have the chi-square distribution with $g(t-1)$ degrees of freedom under H_0. An analysis of chi-square table may be formed: $-2 \log \lambda_4 = -2 \log \lambda_3 - 2 \log \lambda_1$, where $-2 \log \lambda_1$ is the test statistic of (3.6) based on the pooled data.

Bradley and Terry (1952a) gave a small example for two tasters evaluating pork roasts from hogs with differing diets, $t = 3$, $g = 2$, $n_{iju} = 5$ for all i, j, u, $i \neq j$. The data are summarized in Table 4 and Table 5 is the analysis of chi-square table. The large total treatment effect is seen to be due to disagreement of the two judges on preferences.

Table 4

Roast Pork Preference Data for Two Judges

Diet	Judge 1		Judge 2		Pooled Data	
T_i	$a_i^{(1)}$	$p_i^{(1)}$	$a_i^{(2)}$	$p_i^{(2)}$	a_i	p_i
1	1	.0526	7	.5324	8	.2479
2	7	.4737	5	.2993	12	.4268
3	7	.4737	3	.1683	10	.3253
	$B_{11} = 6.7166$		$B_{12} = 9.2895$		$B_1 = 20.2565$	

Table 5

Analysis of Chi Square, Roast Pork Data

Test	Statistic	d.f.	χ^2
Treatments, given agreement	$-2 \log \lambda_1$	2	1.07
Judge by Treatment Interaction	$-2 \log \lambda_3$	2	8.50
Treatments	$-2 \log \lambda_4$	4	9.58

(iv) Tests for specified treatment contrasts, contrasts on the log π_i, may be made by the method of Section 5.

Bradley and Terry (1952a) proposed one additional test. It was assumed that the treatments fell into two groups, say $T_1, \ldots, T_s$ and $T_{s+1}, \ldots, T_t$, with $\pi_1 = \ldots = \pi_s = \pi$ and $\pi_{s+1} = \ldots = \pi_t = (1-s\pi)/(t-s)$. The test is of the equality of π and $(1-s\pi)/(t-s)$, or equivalently of $\pi_i = 1/t$, $i = 1, \ldots, t$, against the two-group alternative of the assumption. The reader is referred to the reference for details.

3.3. Confidence Regions

Large-sample theory may be used to obtain variances and covariances for the estimators $p_1, \ldots, p_t$ or their logarithms in paired comparisons. Bradley (1955) considered this theory with each $n_{ij} = n$ and Davidson and Bradley (1970), considering the multivariate model discussed in Section 6 obtained results for general n_{ij} as a special case.

Let $\mu_{ij} = n_{ij}/N$. Then $\sqrt{N}(p_1-\pi_1), \ldots, \sqrt{N}(p_t-\pi_t)$ have the singular multivariate normal distribution of dimensionality $(t-1)$ in a space of t dimensions with zero mean vector and dispersion matrix $\underset{\sim}{\Sigma} = [\sigma_{ij}]$ such that

$$\sigma_{ij} = \text{cofactor of } \lambda_{ij} \text{ in } \begin{bmatrix} \underset{\sim}{\Lambda} & \underset{\sim}{1} \\ \underset{\sim}{1}' & 0 \end{bmatrix} \Big/ \begin{vmatrix} \underset{\sim}{\Lambda} & \underset{\sim}{1} \\ \underset{\sim}{1}' & 0 \end{vmatrix}, \tag{3.14}$$

where $\underset{\sim}{\Lambda} = [\lambda_{ij}]$, $\underset{\sim}{1}'$ is the t-dimensional unit row vector, and

$$\lambda_{ii} = \frac{1}{\pi_i} \sum_{\substack{j \\ j \neq i}} \mu_{ij}\pi_j/(\pi_i+\pi_j)^2, \ i = 1, \ldots, t,$$

and (3.15)

$$\lambda_{ij} = -\mu_{ij}/(\pi_i+\pi_j)^2, \ i \neq j, \ i, j = 1, \ldots, t.$$

In order to use these results in applications, σ_{ij} must be estimated; this is done through substitution of p_i for π_i in (3.15) to obtain the $\hat{\lambda}_{ij}$, and subsequent substitution in (3.14) yields the $\hat{\sigma}_{ij}$'s.

For the data of Table 1, values of $p_1, \ldots, p_4$ in Table 2 are used to obtain

$$\hat{\underset{\sim}{\Lambda}} = \begin{bmatrix} 10.4963 & -.9558 & -1.2740 & -2.4259 \\ -.9558 & .4304 & -.3022 & -.3553 \\ -1.2740 & -.3022 & .7441 & 0 \\ -2.4259 & -.3553 & 0 & 3.1237 \end{bmatrix}$$

from whence

$$\hat{\Sigma} = \begin{bmatrix} .0800 & -.0695 & -.0314 & .0208 \\ -.0695 & .6644 & -.4689 & -.1260 \\ -.0314 & -.4689 & .6784 & -.1781 \\ .0208 & -.1260 & -.1781 & .2833 \end{bmatrix}. \qquad (3.16)$$

Note that $\hat{\Sigma}$ is singular, the row and column sums being zero.

Approximate confidence regions may be obtained. The confidence interval on π_i is developed from the fact that $\sqrt{N}(p_i-\pi_i)/\sqrt{\hat{\sigma}_{ii}}$ is standard normal for large N. In the example, the .95-confidence interval for π_1 is (.0795, .1369). Let π^* be a vector containing any subset of t* distinct parameters of the set, t* < t. The (1-α)-confidence region for these t* parameters is that ellipsoidal region of the parameter subspace for which

$$N(\pi^*-p^*)'\hat{\Sigma}^{*-1}(\pi^*-p^*) \leq \chi^2_{\alpha,t^*}. \qquad (3.17)$$

In (3.17), p^* is the vector of estimates corresponding to π^*, $\hat{\Sigma}^*$ is the dispersion matrix for $\sqrt{N}(p^*-\pi^*)$ obtainable from (3.16), and χ^2_{α,t^*} is the (1-α)-percentage point of the central chi-square distribution with t* degrees of freedom. As an example, let $\pi^* = (\pi_1, \pi_2)'$ and then $p^* = (.1082, .5193)$,

$$\hat{\Sigma}^* = \begin{bmatrix} .0800 & -.0695 \\ -.0695 & .6644 \end{bmatrix} \text{ and } \hat{\Sigma}^{*-1} = \begin{bmatrix} 13.7441 & 1.4372 \\ 1.4372 & 1.6553 \end{bmatrix},$$

with $\alpha = .01$, $t^* = 2$, $\chi^2_{.01,2} = 9.210$, it may be verified that (3.17) yields the .99-confidence region,

$$13.7441(\pi_1-.1082)^2 + 1.6553(\pi_2-.5193)^2 + 2.8744(\pi_1-.1082)(\pi_2-.5193) \leq .0248.$$

Since it may be appropriate to regard $\log \pi_i$ as the location parameter for T_i, $i = 1, \ldots, t$, in view of (2.4) and (2.5), confidence intervals or regions on the $\log \pi_i$ may be desired. It follows that $\sqrt{N}(\log p_1 - \log \pi_1), \ldots, \sqrt{N}(\log p_t - \log \pi_t)$ have the singular multivariate normal distribution with zero mean vector and dispersion matrix $\underline{D}\underline{\Sigma}\underline{D}$, where $\underline{D}$ is the diagonal matrix with typical element $1/\pi_i$. Estimated variances and covariances are as follows: est. var.$(\sqrt{N} \log p_i) = \hat{\sigma}_{ii}/p_i^2$, est. covar.$(\sqrt{N} \log p_i, \sqrt{N} \log p_j) = \hat{\sigma}_{ij}/p_i p_j$, $i \neq j$. Confidence intervals or regions on the $\log \pi_i$ may be obtained analogously to those shown above for the π_i. If a method of multiple comparisons is to be used, the necessary variances and covariances may be obtained from the information given.

In the very special case when each $n_{ij} = n$, approximate variances and covariances may be obtained if the treatments are not too disparate. Then, on the assumption that $\pi_i = 1/t$, $i = 1, \ldots, t$, $\sigma_{ii} = 2(t-1)^2/t^3$ and $\sigma_{ij} = -2(t-1)/t^3$, $i \neq j$, while $N = n\binom{t}{2}$. Like the binomial with its stable variance for its parameter in a middle range, so are the variances and covariances stable in paired comparisons when the π_i are near $1/t$ and the $n_{ij} = n$. This can reduce computational effort for balanced experiments.

3.4. Asymptotic Relative Efficiency

It is well known that the asymptotic relative efficiency of the sign test to the Student test is $2/\pi$ when assumptions for the latter apply and appropriate data could be obtained. Bradley (1955) showed that, under similar conditions, the asymptotic relative efficiency of paired comparisons relative to a randomized complete block design with the same number of treatment replications is $t/\pi(t-1)$, when each $n_{ij} = n$. This result may be adjusted to show that the relative efficiency of paired comparisons relative to the analysis of variance for the similar balanced incomplete block design is $2/\pi$ by the methods of Raghavarao (1971, Sections 4.3 and 4.5).

While tho asymptotic relative efficiency factor of $2/\pi$ suggests loss of efficiency through use of the ranking or proference designations of paired comparisons, the method is usually used because measurement scales are not available for sensory or judgment evaluations.

4. Extensions of the Basic Model

4.1. Adjustments for Ties

The basic paired comparisons experiment forces decision on the part of the respondent and data like those of Table 1 result. Nevertheless, ties or "non-selection" judgments often arise, for example, in consumer testing.

The treatment of ties in the sign test has received considerable attention. Hemelrijk (1952) demonstrated that the most powerful test of significance was obtained by omission of ties and use of a conditional binomial test on the sample results so reduced. But the treatment of ties must depend on experimental objectives, see Gridgeman (1959), and estimation of potential share of a consumer market surely must require other considerations. Decisions for paired comparisons must be similar to those for the sign test. Two formal methods for the treatment of ties in paired comparisons are available.

Rao and Kupper (1967) introduced a parameter $\theta \geq 1$ and adjusted probabilities associated with the comparison of T_i and T_j to obtain

$$P(T_i \to T_j) = \pi_i/(\pi_i+\theta\pi_j) = \tfrac{1}{4} \int_{-(\log \pi_i - \log \pi_j)+\eta}^{\infty} \operatorname{sech}^2 y/2 \, dy,$$

and

$$P(T_i = T_j) = (\theta^2-1)\pi_i\pi_j/(\pi_i+\theta\pi_j)(\theta\pi_i+\pi_j) \tag{4.1}$$

$$= \int_{-(\log \pi_i - \log \pi_j)-\eta}^{-(\log \pi_i - \log \pi_j)+\eta} \operatorname{sech}^2 y/2 \, dy, \; i \neq j,$$

where $\eta = \log \theta$. It is seen that the model extends the linear model of (2.4) and that $\log \theta$ is, in a sense, a threshold parameter associated with discriminatory ability.

Rao and Kupper extended the theory in parallel with that given above. Unfortunately, they assumed that $n_{ij} = n$, but the work is easily extended. We summarize only the results leading to the test of treatment equality, although they provide other asymptotic results including variances and covariances for their estimators. We use our notation. Let $N = \sum_{i<j} n_{ij}$ and b_{ij} be the sum of the number of ties and the number of preferences for T_i in the n_{ij} comparisons of T_i and T_j. Let $b_i = \sum_{j, j \neq i} b_{ij}$ and let b_0 be the total number of ties in the experiment. The likelihood equations are:

$$\frac{b_i}{p_i} - \sum_{\substack{j\\ j\neq i}} \frac{b_{ij}}{p_i+\hat\theta p_j} - \sum_{\substack{j\\ j\neq i}} \frac{b_{ij}}{\hat\theta p_i+p_j} = 0, \; i = 1, \ldots, t,$$

$$\sum_i p_i = 1, \tag{4.2}$$

$$\frac{b_0\hat\theta}{\hat\theta^2-1} - \tfrac{1}{2} \sum_{i\neq j} \frac{b_{ij}p_j}{p_i+\hat\theta p_j} = 0,$$

where p_i is the estimator of π_i and $\hat{\theta}$ of θ. The likelihood ratio test of H_0: $\pi_i = 1/t$, $i = 1, \ldots, t$, versus H_a: $\pi_i \neq 1/t$ for some i, leads to the statistic,

$$-2 \log \lambda_1^* = 2N \log 2N - 2b_0 \log 2b_0 - 2(N-b_0)\log(N-b_0) - 2B_1^*, \quad (4.3)$$

where

$$B_1^* = \sum_{i \neq j} b_{ij} \log(p_i + \hat{\theta} p_j) - \sum_i b_i \log p_i - b_0 \log(\hat{\theta}^2 - 1). \quad (4.4)$$

Again, for large N and under H_0, $-2 \log \lambda_1^*$ has the central chi-square distribution with (t-1) degrees of freedom. An iterative solution of equations (4.2) is suggested by Rao and Kupper. They provided also a test of the hypothesis, $\theta = \theta_0$, against the alternative, $\theta \neq \theta_0$.

Davidson (1970) proposed probabilities corresponding to those of (4.1) as

$$P(T_i \rightarrow T_j) = \pi_i / (\pi_i + \pi_j + \nu\sqrt{\pi_i \pi_j})$$

and (4.5)

$$P(T_i = T_j) = \nu\sqrt{\pi_i \pi_j} / (\pi_i + \pi_j + \nu\sqrt{\pi_i \pi_j}),$$

$\nu \geq 0$. This model preserves the odds ratio, $P(T_i \rightarrow T_j)/P(T_j \rightarrow T_i) = \pi_i/\pi_j$, consistent with the Luce (1959) choice axiom. In addition, the probability of a tie is a maximum when $\pi_i = \pi_j$ and diminishes as π_i and π_j differ, an intuitively desirable effect.

Let b_{ij}^* be the sum of the number of ties and twice the number of preferences for T_i in the n_{ij} comparisons of T_i and T_j and let $b_i^* = \sum_{\substack{j \\ j \neq i}} b_{ij}$.

Davidson's likelihood equations are

$$\frac{b_i^*}{p_i} - \sum_{\substack{j \\ j\neq i}} n_{ij}(2+\hat{\nu}\sqrt{p_j/p_i})/(p_i+p_j+\hat{\nu}\sqrt{p_ip_j}) = 0,\ i = 1, \ldots, t,$$

$$\sum_i p_i = 1, \tag{4.6}$$

$$\frac{b_0}{\hat{\nu}} - \sum_{i<j} n_{ij}\sqrt{p_ip_j}/(p_i+p_j+\hat{\nu}\sqrt{p_ip_j}) = 0,$$

where p_i is the estimator of π_i and $\hat{\nu}$ of ν. The likelihood ratio statistic corresponding to (4.3) is of the same form with B_1^* replaced by

$$B_1^{**} = \sum_{i<j} n_{ij} \log(p_i+p_j+\hat{\nu}\sqrt{p_ip_j}) - \tfrac{1}{2}\sum_i b_i^* \log p_i - b_0 \log \hat{\nu}. \tag{4.7}$$

Davidson also proposed an iterative solution for the equations (4.6) and examined large-sample theory. He showed that the Rao-Kupper test and the Davidson test for treatment equality are asymptotically equivalent.

The choice between the two methods for extending the basic paired comparisons model to a model allowing for ties seems to be a matter of intuitive appeal. Both give very similar results in applications.

4.2. Adjustments for Order

In paired comparisons, there is often concern for the effects of order of presentation of the two items in a pair. Experiments are often conducted so that, for each pair of treatments, each order of presentation is used equally frequently in an effort to "balance out" the effects of order. Scheffé (1952) addressed this problem in the analysis of variance. Beaver and Gokhale (1975) extended our basic model to allow for order effects.

Davidson and Beaver in an undated manuscript describe the Beaver-Gokhale model as having additive order effects and discuss also a model with multiplicative order effects suggested by Beaver (1976). For the ordered pair (T_i, T_j), Beaver and Gokhale defined

$$P_{ij}(T_i \to T_j) = \frac{\pi_i + \delta_{ij}}{\pi_i + \pi_j}, \quad P_{ij}(T_j \to T_i) = \frac{\pi_j - \delta_{ij}}{\pi_i + \pi_j} \tag{4.8}$$

and, for the ordered pair (T_j, T_i),

$$P_{ji}(T_i \to T_j) = \frac{\pi_i - \delta_{ij}}{\pi_i + \pi_j}, \quad P_{ji}(T_j \to T_i) = \frac{\pi_j + \delta_{ij}}{\pi_i + \pi_j}. \tag{4.9}$$

The corresponding probabilities for the model with multiplicative order effects are

$$\begin{aligned} P_{ij}(T_i \to T_j) &= \frac{\theta_{ij}\pi_i}{\theta_{ij}\pi_i + \pi_j}, \quad P_{ij}(T_j \to T_i) = \frac{\pi_j}{\theta_{ij}\pi_i + \pi_j}, \\ P_{ji}(T_i \to T_j) &= \frac{\pi_i}{\pi_i + \theta_{ij}\pi_j}, \quad P_{ji}(T_j \to T_i) = \frac{\theta_{ij}\pi_j}{\pi_i + \theta_{ij}\pi_j}. \end{aligned} \tag{4.10}$$

The model given by (4.8) and (4.9) requires that $|\delta_{ij}| \le \max(\pi_i, \pi_j)$, an awkward feature, while the model (4.10) only requires that $\theta_{ij} > 0$. Advantages of the multiplicative model (4.10) are:

(i) Preference probabilities depend on the worth parameters π_i and π_j only through the ratio π_i/π_j.

(ii) Model (4.10) admits a sufficient statistic whose dimension is that of the parameter space.

(iii) Model (4.10) is a linear model and, for example,

$$P_{ij}(T_i \to T_j) = \tfrac{1}{4} \int_{-(\log \pi_i - \log \pi_j) - \log \theta_{ij}}^{\infty} \operatorname{sech}^2 y/2 \, dy.$$

For these reasons, we limit further discussion to (4.10).

Explicit methodology for model (4.10) and its special cases does not appear in the statistical literature, although it is implied by Davidson and Beaver. Various likelihood ratio tests and associated estimation procedures can be developed easily when needed. We consider only the special case when $\theta_{ij} = \Theta$ for all $i \neq j$. Then the likelihood equations are

$$\frac{a_i}{p_i^*} - \sum_{\substack{j \\ j \neq i}} \frac{n_{ij}\hat{\theta}}{(\hat{\theta}p_i^* + p_j^*)} - \sum_{\substack{j \\ j \neq i}} \frac{n_{ji}}{(p_i^* + \hat{\theta}p_j^*)} = 0, \quad i = 1, \ldots, t,$$

$$\sum_i p_i^* = 1, \tag{4.11}$$

$$\frac{f}{\hat{\theta}} - \sum_{i \neq j} \frac{n_{ij}p_i^*}{(\hat{\theta}p_i^* + p_j^*)} = 0,$$

where f is the total number of preferences for the first presented item of a pair, p_i^* is the estimator of π_i and $\hat{\theta}$ of θ, while n_{ij} is the number of judgments on the ordered pair (T_i, T_j) and n_{ji} is the number of judgments on the ordered pair (T_j, T_i). The likelihood ratio statistic for H_0: $\pi_i = 1/t$, $i = 1, \ldots, t$, versus H_a: $\pi_i \neq 1/t$ for some i in the presence of an order effect is

$$-2 \log \lambda_1^+ = 2N \log N - 2f \log f - 2(N-f)\log(N-f) - 2B_1^+, \tag{4.12}$$

where

$$B_1^+ = \sum_{i \neq j} n_{ij} \log(\hat{\theta} p_i^* + p_j^*) - \sum_i a_i \log p_i^* - f \log \hat{\theta}. \tag{4.13}$$

Again, under H_0, $-2 \log \lambda_1^+$ has the central chi-square distribution with $(t-1)$ degrees of freedom. A test for the presence of a common order effect, H_0: $\theta = 1$ versus H_a: $\theta \neq 1$, follows immediately. For this test,

$$-2 \log \lambda_4 = 2(B_1 - B_1^+) \tag{4.14}$$

has the central chi-square distribution with 1 degree of freedom when $\theta = 1$. In (4.14), B_1 is taken from (3.6).

Other tests could be developed. One of interest is the test for a common order effect: H_0: $\theta_{ij} = \theta$ for all $i \neq j$, H_a: $\theta_{ij} \neq \theta$ for some i, j, $i \neq j$. Such a test could be described as a test of order by treatment pair interaction.

Note that neither model for order effects suggests that an effort to balance out the effects of order is exactly right. Note also that both order effects and ties could be important and this is the situation addressed by Davidson and Beaver in their unpublished manuscript.

4.3. A Bayesian Approach

Davidson and Solomon (1973) considered a Bayesian approach to the estimation of the worth parameters $\pi_1, \ldots, \pi_t$ of paired comparisons. Let $\underset{\sim}{a}^o = [a_{ij}^o]$ and $\underset{\sim}{n}^o = [n_{ij}^o]$, $n_{ii}^o = a_{ii}^o = 0$, $n_{ij}^o = n_{ji}^o$. They formulated a conjugate prior distribution for the parameters,

$$\phi(\underset{\sim}{\pi}) = A(\underset{\sim}{a}^o, \underset{\sim}{n}^o) \prod_{i<j} \pi_i^{a^o_{ij}} \pi_j^{a^o_{ji}} / (\pi_i + \pi_j)^{n^o_{ij}}, \ \underset{\sim}{\pi} \in \Omega, \tag{4.15}$$

$$= A(\underset{\sim}{a}^o, \underset{\sim}{n}^o) \prod_i \pi_i^{a^o_i} / \prod_{i<j} (\pi_i + \pi_j)^{n^o_{ij}},$$

where $\Omega = \{\underset{\sim}{\pi}: \pi_i \geq 0,\ i = 1, \ldots, t,\ \sum_i \pi_i = 1\}$. They restricted attention to densities (4.15) for which $a^o_{ij} \geq 0$ and $a^o_{ij} + a^o_{ji} = n^o_{ij}$. They noted that, even with these restrictions, each $(\underset{\sim}{a}^o, \underset{\sim}{n}^o)$ determines a distinct prior distribution and that the family of priors can represent a wide spectrum of prior beliefs. Davidson and Solomon suggested that the experimenter think of his prior beliefs in terms of a conceptual experiment with n^o_{ij} responses to the pair (T_i, T_j) with a^o_{ij} of them being preferences for T_i. Choice of n^o_{ij} is to be made as a measure of the strength of the experimenter's beliefs on the pair (T_i, T_j).

It is noted that the selection of an estimator for the vector of worth parameters $\underset{\sim}{\pi}$ is of central interest. This is to be done on the basis of the prior distribution (4.15) and the results of experimentation summarized in the likelihood function conditioned on $\underset{\sim}{\pi}$,

$$\ell(\underset{\sim}{a}|\underset{\sim}{\pi}) = \prod_i \pi_i^{a_i} \prod_{i<j} \binom{n_{ij}}{a_{ij}} (\pi_i + \pi_j)^{-n_{ij}}. \tag{4.16}$$

The estimator of $\underset{\sim}{\pi}$ can be used to estimate pairwise preference probabilities or to provide a ranking of the items or treatments in the experiment.

One estimator of $\underset{\sim}{\pi}$ is the mode $\underset{\sim}{p}^*$ of the posterior distribution of $\underset{\sim}{\pi}$. This mode is shown to be the solution of the set of equations,

$$\frac{a_i'}{p_i^*} - \sum_{\substack{j \\ j \neq i}} \frac{n_{ij}'}{(p_i^* + p_j^*)} = 0,\ i = 1, \ldots, t,$$

$$\sum_i p_i^* = 1, \qquad (4.17)$$

where $n_{ij}' = n_{ij}^o + n_{ij}$ and $a_i' = a_i^o + a_i$, $i < j$, $i, j = 1, \ldots, t$. It is seen that the choice of prior distribution led to a natural combination of prior and experimental information as seen from the definitions of n_{ij}' and a_i'. Further, equations (4.17) have the form of equations (3.2) and (3.3).

Davidson and Solomon considered also the Bayes estimator of $\underset{\sim}{\pi}$ under a quadratic loss function, namely $\underset{\sim}{\bar{p}}$, the mean of the posterior distribution of $\underset{\sim}{\pi}$. While they did not obtain a closed expression for $\underset{\sim}{\bar{p}}$, they did show that, if $n_{ij}' = n'$ for all $i < j$, the rankings determined by $\underset{\sim}{p}^*$ and $\underset{\sim}{\bar{p}}$ are identical with the Bayes ranking determined by the posterior score $\underset{\sim}{a}'$.

4.4. Triple Comparisons

The basic model for paired comparisons can be extended to triple comparisons in at least two ways. Bradley and Terry (1952b) proposed the model,

$$P(T_i \to T_j \to T_k) = \pi_i \pi_j / (\pi_i + \pi_j + \pi_k)(\pi_j + \pi_k) \qquad (4.18)$$

for comparison of T_i, T_j and T_k in a triplet, $i \neq j \neq k$, $i, j, k = 1, \ldots, t$. Pendergrass and Bradley (1960) proposed the model,

$$P(T_i \to T_j \to T_k) = \pi_i^2 \pi_j / [\pi_i^2(\pi_j + \pi_k) + \pi_j^2(\pi_i + \pi_k) + \pi_k^2(\pi_i + \pi_j)]. \qquad (4.19)$$

In both models, the π's may again be regarded as worth parameters with $\sum_i \pi_i = 1$. Both models have some desirable properties as discussed in the second reference. Model (4.18) is consistent with the Luce choice axiom and can be written as a Lehmann model (see Bradley (1976)). Model (4.19) has the property that the set of treatment rank sums constitutes a set of sufficient statistics for the estimation of $\pi_1, \ldots, \pi_t$. Basic methodology for the second model is well developed including estimation procedures, tests of hypotheses including goodness of fit, and asymptotic theory.

We show only the estimating equations and the basic test for model (4.19). If $p_1, \ldots, p_t$ are the estimators of $\pi_1, \ldots, \pi_t$, they result from solution of the equations,

$$\frac{a_i}{p_i} - \sum_{\substack{j<k \\ j,k\neq i}} \frac{n_{ijk}[2p_i(p_j+p_k)+p_j^2+p_k^2]}{D_{ijk}(\underset{\sim}{p})} = 0, \; i = 1, \ldots, t,$$

$$\sum_i p_i = 1, \tag{4.20}$$

where

$$D_{ijk}(p) = p_i^2(p_j+p_k) + p_j^2(p_i+p_k) + p_k^2(p_i+p_j) \tag{4.21}$$

and n_{ijk} is the number of repetitions or rankings on the triplet (T_i, T_j, T_k), $i < j < k$. The quantity a_i in (4.20) is such that $a_i = 3 \sum_{\substack{j<k \\ j,k\neq i}} n_{ijk} - R_i$, where R_i is the total sum of ranks for T_i in the experiment. Pendergrass and Bradley suggest iterative means of solution of the equations (4.20) although they held each $n_{ijk} = n$ for all $i < j < k$.

The likelihood ratio test of H_0: $\pi_i = 1/t$, $i = 1, \ldots, t$, versus H_a: $\pi_i \neq 1/t$ for some i, is based on

$$-2 \log \lambda_5 = 2N \log 6 + 2 \sum_i a_i \log p_i - 2 \sum_{i<j<k} n_{ijk} \log D_{ijk}(\underset{\sim}{p}), \tag{4.22}$$

where $N = \sum_{i<j<k} n_{ijk}$. Under H_0, $-2 \log \lambda_5$ has the central chi-square distribution with (t-1) degrees of freedom for large N.

Park (1961) applied the Pendergrass-Bradley procedures to experimental data and compared the results with those from companion experiments using paired comparisons. He found good model fits and estimator agreement.

5. Treatment Contrasts and Factorials

It became apparent very early in applications of paired comparisons to sensory experimentation that there was need for special analyses when the treatments represented factorial treatment combinations. Abelson and Bradley (1954) attempted to address this need with very limited success and it remained an open problem until solved by Bradley and El-Helbawy (1976). They considered factorial treatment combinations in the more general framework of specified treatment contrasts. This simplified both notation and theory.

In Table 6, we show paired comparisons data for treatments representing a 2^3 factorial set of treatment combinations. The data are taken from Bradley and El-Helbawy (1976) and arise from a consumer preference taste test on coffees, where the factors are brew strength, roast color and coffee brand, each at two levels. Twenty-six preference judgments were obtained on each of the 28 distinct treatment comparisons. Note that it is convenient to replace the typical treatment T_i by $T_{\alpha_1\alpha_2\alpha_3}$, $\alpha_i = 1$ or 0, $i = 1, 2, 3$, so that the subscripts indicate the chosen levels of the factors. We shall return to these data to illustrate use of the general method explained below with factorials.

Table 6

Preference Data in Coffee Testing

Treatment preferred, $T_{\underset{\sim}{\alpha}}$		Treatment not preferred, $T_{\underset{\sim}{\beta}}$ $\underset{\sim}{\beta}$								
		000	001	010	011	100	101	110	111	$a_{\underset{\sim}{\alpha}}$
	000	--	15	15	16	19	14	19	16	114
	001	11	--	10	15	15	14	15	12	92
	010	11	16	--	15	15	14	18	15	104
	011	10	11	11	--	14	11	15	13	85
$\underset{\sim}{\alpha}$	100	7	11	11	12	--	9	14	13	77
	101	12	12	12	15	17	--	16	18	102
	110	7	11	8	11	12	10	--	12	71
	111	10	14	11	13	13	8	14	--	83

$$a_i/p_i - \phi_i(\underset{\sim}{p}) = 0,\ i = 1, \ldots, t, \qquad \sum_i \log p_i = 0, \tag{5.4}$$

where $\underset{\sim}{p} = (p_1, \ldots, p_t)$,

$$\phi_i(\underset{\sim}{p}) = \sum_{\substack{j \\ j\neq i}} \frac{n_{ij}}{p_i+p_j} - \frac{1}{p_i} \sum_{\substack{j \\ j\neq i}} E_j(\underset{\sim}{p}) \frac{D_{ij}}{D_{ii}}, \tag{5.5}$$

$$E_i(\underset{\sim}{p}) = a_i - \sum_{\substack{j \\ j\neq i}} n_{ij}p_i/(p_i+p_j), \tag{5.6}$$

$i = 1, \ldots, t$, and D_{ij} is the typical element of

$$\underset{\sim}{D} = \underset{\sim}{I}_t - \underset{\sim}{B}_m'\underset{\sim}{B}_m, \tag{5.7}$$

$D_{ii} > 0$, $\underset{\sim}{I}_t$, the t-square identity matrix. Note the similar forms in (5.6) and (3.2). If $m = 0$, the estimation process involves solution of (3.2) replaced by (5.3).

Iterative solution of equations (5.4) is discussed briefly by Bradley and El-Helbawy (1976) and in detail by El-Helbawy and Bradley (1977). In the latter reference, it is shown that the proposed iterative procedure converges and yields a maximum of the likelihood function over the parameter space $\{\underset{\sim}{\pi}: \pi_i > 0,\ i = 1, \ldots, t,\ \sum_i \log \pi_i = 0,\ \underset{\sim}{B}_m \log \underset{\sim}{\pi} = \underset{\sim}{0}_m\}$.

A class of likelihood ratio tests may be developed. Let $\underset{\sim}{B}_{m_a}$, $\underset{\sim}{B}_{m_1}$, and $\underset{\sim}{B}_{m_0} = \begin{bmatrix} \underset{\sim}{B}_{m_a} \\ \underset{\sim}{B}_{m_1} \end{bmatrix}$ be matrices like $\underset{\sim}{B}_m$, $0 \le m_a, m_1 \le m_0 \le (t-1)$, $m_0 = m_a + m_1$. With the condition that $\sum_i \log \pi_i = 0$, we test

$$H_0: \underset{\sim}{B}_{m_0} \log \underset{\sim}{\pi} = 0 \tag{5.8}$$

against

$$H_a: \underset{\sim}{B}_{m_a} \log \underset{\sim}{\pi} = 0. \tag{5.9}$$

The test statistic is

$$-2 \log \lambda_{m_0, m_a} = 2[B_1(\underset{\sim}{p}_0) - B_1(\underset{\sim}{p}_a)], \tag{5.10}$$

where B_1 is defined in (3.6), and, for large $N = \sum_{i<j} n_{ij}$ and under H_0 in (5.8), the statistic has the central chi-square distribution with m_1 degrees of freedom. In (5.10), $\underset{\sim}{p}_0$ is the solution of (5.4) where $\underset{\sim}{B}_m = \underset{\sim}{B}_{m_0}$ and $\underset{\sim}{p}_a$, the solution when $\underset{\sim}{B}_m = \underset{\sim}{B}_{m_a}$. Basically, the test involves the assumption that

$$\begin{bmatrix} \underset{\sim}{\lambda}_t' \\ \underset{\sim}{B}_{m_a} \end{bmatrix} \log \underset{\sim}{\pi} = \underset{\sim}{0}_{m_a+1}$$

and a test of the additional constraints,

$$\underset{\sim}{B}_{m_1} \log \underset{\sim}{\pi} = \underset{\sim}{0}_{m_1},$$

$\underset{\sim}{B}_{m_1}$ consisting of m_1 orthonormal rows orthogonal to those of $\underset{\sim}{B}_{m_a}$.

The test procedure is illustrated with the data of Table 6. Treatments T_i have subscripts in the lexicographic order of $T_{\underset{\sim}{a}}$ in the table. Suppose that we wish to test the hypothesis that there are no two-factor interactions on the assumption that there is no three-factor interaction. Then $t = 8$, $m_a = 1$, $m_1 = 3$, $m_0 = 4$ with

$$\underset{\sim}{B}_{m_a} = \frac{1}{\sqrt{8}} (1, -1, -1, 1, -1, 1, 1, -1)$$

and

$$\underset{\sim}{B}_{m_1} = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{bmatrix}.$$

Necessary calculations yield:

$$\underset{\sim}{p}_0 = (1.300, 1.275, 1.060, 1.040, 0.962, 0.944, 0.784, 0.769),$$

$$\underset{\sim}{p}_a = (1.515, 1.060, 1.342, 0.855, 0.790, 1.193, 0.647, 0.890),$$

$$B_1(\underset{\sim}{p}_0) = 497.81, \; B_1(\underset{\sim}{p}_a) = 490.14,$$

$$-2 \log \lambda_{m_0, m_a} = 2(497.81 - 490.14) = 15.34.$$

The statistic, $-2 \log \lambda_{m_0, m_a}$ has the central chi-square distribution with 3 degrees of freedom and is large. It is possible also to partition this chi-square into three chi-squares, each with 1 degree of freedom, as is done in Table 7.

The general test procedure for hypothesis (5.8) versus (5.9) based on the statistic (5.10) may be used repeatedly to produce an analysis of chi-square table. Two such analyses are given in Tables 7 and 8 for the data of Table 6. Rows in these tables correspond to rows of the usual analysis of variance table for a 2^3 factorial and similar descriptive terms have been used. In order to preserve orthogonality of the various chi-squares, they must be sequenced properly; each row requires that certain conditions be assumed, equivalent to the specification of $\underline{B}_{m_a}$. Both Tables 7 and 8 are shown to illustrate two different sequencings of the rows and to suggest that the choice of sequencing does not have substantial effects on the inferences that may be made. Additional details on computations for Tables 7 and 8 are given by Bradley and El-Helbawy (1976).

The analyses below were done through recognition of the factorial structure of the treatments. Factorial parameters may be introduced formally, although it is not necessary to do so. We illustrate with the 2^3 factorial. Let $\pi_{\underline{\alpha}}$ replace π_i for the treatment $T_{\underline{\alpha}} \equiv T_i$, where $\underline{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$, $\alpha_r = 0$ or 1, $r = 1, 2, 3$. We reparameterize by writing

$$\pi_{\underline{\alpha}} = \prod_{r=1}^{3} \pi_{\alpha_r}^{(r)} \cdot \prod_{r<s} \pi_{\alpha_r\alpha_s}^{(rs)} \cdot \pi_{\alpha_1\alpha_2\alpha_3}^{(123)}. \tag{5.11}$$

The parameters on the right-hand side of (5.11) are new factorial parameters. The transformation is linear if logarithms are taken; the logarithms of the

Table 7

An Analysis of Chi-square for the Coffee Data

Hypothesis tested*	Conditions assumed	Degrees of freedom	Chi-square
No F_1 effect	No F_2 F_3, no interactions	1	9.28
No F_2 effect	No F_3, no interactions	1	4.29
No F_3 effect	No interactions	1	0.04
No F_1F_2, F_1F_3, F_2F_3 interactions	No $F_1F_2F_3$ interaction	3	15.34
No F_2F_3 interaction	No $F_1F_2F_3$ interaction	1	0.22
No F_1F_3 interaction	No F_2F_3, $F_1F_2F_3$ interactions	1	14.96
No F_1F_2 interaction	No F_1F_3, F_2F_3, $F_1F_2F_3$ interactions	1	0.15
No $F_1F_2F_3$ interaction	None	1	0.63
No treatment effects	None	7	29.58

*F_1 is brew strength, F_2 is roast colour, F_3 is brand.

Table 8

An Alternative Analysis of Chi-square for the Coffee Data

Hypothesis tested*	Conditions assumed	Degrees of freedom	Chi-square
No F_1 effect	None	1	9.47
No F_2 effect	No F_1 effect	1	4.33
No F_3 effect	No F_1, F_2 effects	1	0.04
No F_1F_2, F_1F_3, F_2F_3 interactions	No main effects	3	15.12
No F_1F_2 interaction	No main effects	1	0.16
No F_1F_3 interaction	No main effects, no F_1F_2 interaction	1	14.73
No F_2F_3 interaction	No main effects, no F_1F_2, F_1F_3 interactions	1	0.24
No $F_1F_2F_3$ interaction	No main effects, no two-factor interactions	1	0.62
No treatment effects	None	7	29.58

*F_1 is brew strength, F_2 is roast colour, F_3 is brand.

new factorial parameters are subject to the usual linear constraints for factorial parameters in the analysis of variance in order to make the transformation one-to-one. Estimators of the factorial parameters are functions of the estimators p_{α}. A full explanation of these procedures is given by El-Helbawy and Bradley (1976).

Special treatment contrasts may be of interest in paired comparisons. Suppose that, in a coffee taste test experiment with $t = 4$, T_4 represents an experimental coffee produced by a new process while the other treatments came from a standard process. One may wish to compare T_4 with the other three treatments. Two approaches are possible. The first assumes nothing, $m_a = 0$, and takes

$$\underset{\sim}{B}_{m_1} = \frac{1}{\sqrt{12}} (1, 1, 1, -3).$$

The second approach assumes that $\pi_1 = \pi_2 = \pi_3$, $m_a = 2$,

$$\underset{\sim}{B}_{m_a} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \end{bmatrix}$$

and retains the same $\underset{\sim}{B}_{m_1}$. With these matrices defined, the general test procedure of this section is used.

We have presented a method for the examination of specified treatment contrasts and the analysis of factorial paired comparison experiments together with examples. These methods provide much new flexibility.

6. Multivariate Paired Comparisons

Multivariate responses to paired comparisons are often obtained. For example, this happens in consumer testing where, on paired samples, preferences on a number of characteristics are solicited.

Davidson and Bradley (1969) extended the paired comparisons model to the multivariate case. Let $\underset{\sim}{s} = (s_1, \ldots, s_p)$, $s_\alpha = i$ or j, be the response vector on attributes $\alpha = 1, \ldots, p$ for the treatment pair (T_i, T_j), $s_\alpha = i$ indicating preference for T_i on attribute α. The probability of response $\underset{\sim}{s}$ on (T_i, T_j) is

$$P(\underset{\sim}{s}|i,j) = p^{(1)}(\underset{\sim}{s}|i,j)h(\underset{\sim}{s}|i,j), \tag{6.1}$$

where

$$p^{(1)}(\underset{\sim}{s}|i,j) = \prod_{\alpha=1}^{p} \pi_{\alpha s_\alpha}/(\pi_{\alpha_i}+\pi_{\alpha_j}) \tag{6.2}$$

and

$$h(\underset{\sim}{s}|i,j) = 1 + \sum_{\alpha<\beta} \delta(s_\alpha \cdot s_\beta)\rho_{\alpha\beta}(\pi_{\alpha i}/\pi_{\alpha j})^{-\delta(i,s_\alpha)/2}(\pi_{\beta i}/\pi_{\beta j})^{-\delta(i,s_\beta)/2}, \tag{6.3}$$

for all $\underset{\sim}{s}$, $i < j$, $i, j = 1, \ldots, t$. Notation is as follows: $\pi_{\alpha i}$ is the worth parameter for T_i on attribute α, $\sum_i \pi_{\alpha i} = 1$, $\rho_{\alpha\beta}$ is a "correlation" parameter for attributes α and β assumed constant for all treatment pairs, and $\delta(s_\alpha, s_\beta) = 1$ or -1 as the two arguments of the indicator function agree or disagree. Note that $\underset{\sim}{\rho} = \underset{\sim}{0}$ implies independence of responses on attributes; $\underset{\sim}{\rho}$ has typical element $\rho_{\alpha\beta}$. It is necessary to restrict the parameter space so that $\pi_{\alpha i} \geq 0$, $\alpha = 1, \ldots, p$, $i = 1, \ldots, t$, and $h(\underset{\sim}{s}|i,j) \geq 0$ for each of the 2^p cells associated with each of the $\binom{t}{2}$ treatment pairs.

Let

$$B(\underset{\sim}{\pi}) = -\sum_{\alpha=1}^{p} B_1(\underset{\sim}{\pi}_\alpha) \tag{6.4}$$

and

$$C(\underset{\sim}{\pi}, \underset{\sim}{\rho}) = \sum_{i<j} \sum_{\underset{\sim}{s}} f(\underset{\sim}{s}|i,j) \log h(\underset{\sim}{s}|i,j), \tag{6.5}$$

where $\underset{\sim}{\pi}$ has typical element $\pi_{\alpha i}$ and $\underset{\sim}{\pi}_\alpha$ is the α^{th} row of $\underset{\sim}{\pi}$. The quantity $B_1(\underset{\sim}{\pi}_\alpha)$ is the function B_1 of (3.6) with p_i there replaced by $\pi_{\alpha i}$ and a_i replaced by $a_{\alpha i}$, the total number of preferences for T_i on attribute α. In addition, $f(\underset{\sim}{s}|i,j)$ is the number of times the preference vector $\underset{\sim}{s}$ occurs among the n_{ij} responses to the pair (T_i, T_j). We may express the logarithm of the likelihood function as

$$\log L = C(\underset{\sim}{\pi}, \underset{\sim}{\rho}) + B(\underset{\sim}{\pi}). \tag{6.6}$$

Consider first a test for independence: H_0: $\underset{\sim}{\rho} = \underset{\sim}{0}$ versus H_a: $\rho_{\alpha\beta} \neq 0$ for some $\alpha < \beta$, $\alpha, \beta = 1, \ldots, p$. Under H_0, the likelihood equations reduce to equations (3.2) and (3.3) for each $\alpha = 1, \ldots, p$. If $\underset{\sim}{p}_\alpha^0$ is the solution for the α^{th} set of equations and becomes the α^{th} row of $\underset{\sim}{p}^0$, $\underset{\sim}{p}^0$ estimates $\underset{\sim}{\pi}$ under H_0. Under H_a, the equations to be solved are:

$$\sum_{i<j} f(\underset{\sim}{s}|i,j)h^{-1}(\underset{\sim}{s}|i,j)\delta(s_\alpha,s_\beta)(\pi_{\alpha i}/\pi_{\alpha j})^{-\delta(i,s_\alpha)/2}(\pi_{\beta i}/\pi_{\beta j})^{-\delta(i,s_\beta)/2}\Bigg|_{\substack{\underset{\sim}{\pi}=\underset{\sim}{p} \\ \underset{\sim}{\rho}=\hat{\underset{\sim}{\rho}}}} = 0,$$

$\alpha < \beta$, α, $\beta = 1, \ldots, p$,

(6.7)

$$\frac{a_{\alpha i}+R_{\alpha i}}{p_{\alpha i}} - \sum_{\substack{j \\ j\neq i}} \frac{n_{ij}}{p_{\alpha i}+p_{\alpha j}} = 0,$$

$i = 1, \ldots, t$, $\alpha = 1, \ldots, p$,

$$\sum_i p_{\alpha i} = 1, \ \alpha = 1, \ldots, p,$$

where

$$R_{\alpha i} = -\tfrac{1}{4} \sum_{\substack{j \\ j \neq i}}^{t} \sum_{\underset{\sim}{s}} f(\underset{\sim}{s}|i,j) h^{-1}(\underset{\sim}{s}|i,j) \times$$

$$(\pi_{\alpha i}/\pi_{\alpha j})^{-\delta(i,s_\alpha)/2} \sum_{\substack{\beta \\ \beta \neq \alpha}} \delta(i,s_\beta)\rho_{\alpha\beta}(\pi_{\beta i}/\pi_{\beta j})^{-\delta(i,s_\beta)/2} . \tag{6.8}$$

Solutions of equations (6.7) is discussed by Davidson and Bradley (1969). If we let $\underset{\sim}{p}$ and $\hat{\underset{\sim}{\rho}}$ be the estimators of $\underset{\sim}{\pi}$ and $\underset{\sim}{\rho}$ from equations (6.7), the likelihood ratio test statistic is

$$-2 \log \lambda_6 = 2\{B(\underset{\sim}{p}) - B(\underset{\sim}{p}^0) + C(\underset{\sim}{p},\hat{\underset{\sim}{\rho}})\} \tag{6.9}$$

and, under H_0, it has the central chi-square distribution with $\tfrac{1}{2}p(p-1)$ degrees of freedom.

If it is assumed that $\underset{\sim}{\rho} = \underset{\sim}{0}$, tests on the parameters $\underset{\sim}{\pi}_\alpha$ may be made separately as in the univariate case for each $\alpha = 1, \ldots, p$.

An overall test of no treatment preferences may be made in the presence of correlations. Then we have H_0: $\underset{\sim}{\pi} = [1/t]$ and H_a: $\pi_{\alpha i} \neq 1/t$ for some α and i. Under H_a, the estimators from equations (6.7) are again $\underset{\sim}{p}$ and $\hat{\underset{\sim}{\rho}}$. Under H_0, the estimators of $\underset{\sim}{\pi}$ and $\underset{\sim}{\rho}$ are $[1/t]$ and $\hat{\underset{\sim}{\rho}}_0$, the latter obtained from solution of (6.7) with $\underset{\sim}{p} = [1/t]$. The test statistic is

$$-2 \log \lambda_7 = 2\{B(\underset{\sim}{p}) + C(\underset{\sim}{p},\hat{\underset{\sim}{\rho}}) + pN \log 2 - C(1/t,\hat{\rho}_0)\} \tag{6.10}$$

with the central chi-square distribution with p(t-1) degrees of freedom under H_0.

A likelihood ratio test of the fit of the model (6.1) is given by Davidson and Bradley. An alternative test may be based on

$$\chi^2 = \sum_{i<j} \sum_{\underset{\sim}{s}} \{f(\underset{\sim}{s}|i,j)-\hat{f}(\underset{\sim}{s}|i,j)\}^2/\hat{f}(\underset{\sim}{s}|i,j) \tag{6.11}$$

and, under the model, has the central chi-square distribution for large N with $\{(2^p-1)\binom{t}{2}-p(t-1)-\binom{p}{2}\}$ degrees of freedom. The estimators $\underset{\sim}{p}$ and $\hat{\underset{\sim}{\rho}}$ are substituted in (6.1) to obtain expected cell frequencies $\hat{f}(\underset{\sim}{s}|i,j) = n_{ij}\,\hat{P}(\underset{\sim}{s}|i,j)$.

Davidson and Bradley (1970) examine large-sample properties of procedures discussed above. Davidson and Bradley (1971) examine regression relationships among the characteristics in the multivariate problem.

We conclude this section with one of the examples given by Davidson and Bradley (1969). Table 9 shows the observed and expected cell frequencies, the latter in parentheses, for a chocolate pudding test with t = 3, p = 3, the treatments being brands, and the attributes being taste, color and texture.

Table 9

Observed and Expected Cell Frequencies
for a Chocolate Pudding Test

Treatment Pair	Cell Frequencies $f(\underset{\sim}{s}\|i,j)$								Frequency
i, j	Cells $\underset{\sim}{s}$								n_{ij}
	(iii)	(jii)	(iji)	(jji)	(iij)	(jij)	(ijj)	(jjj)	
1, 2	8 (7.93)	1 (1.09)	1 (1.15)	1 (1.69)	0 (0.76)	2 (0.97)	0 (0.37)	9 (8.03)	22
1, 3	6 (6.25)	0 (0.60)	1 (1.24)	1 (0.92)	1 (1.12)	0 (0.62)	1 (0.64)	9 (7.61)	19
2, 3	7 (6.92)	1 (0.37)	1 (1.26)	1 (0.60)	3 (1.70)	1 (0.75)	1 (1.10)	6 (8.31)	21

Details on calculations are not given. However, as a possible check on computer programming, the solution of (6.7) is as follows:

$$\underset{\sim}{p} = \begin{bmatrix} 0.312 & 0.360 & 0.328 \\ 0.307 & 0.321 & 0.372 \\ 0.338 & 0.288 & 0.374 \end{bmatrix}, \quad \begin{aligned} \hat{\rho}_{12} &= 0.675 \\ \hat{\rho}_{13} &= 0.654 \\ \hat{\rho}_{23} &= 0.588\,. \end{aligned}$$

Tests are summarized in Table 10. It is seen that the major effects are the high correlations among responses on attributes.

Table 10

Test Statistics for Hypotheses for the Chocolate Pudding Data

Test	Statistic	Ref. No.	Value	d.f.
Test of Independence	$-2 \log \lambda_6$	(6.9)	62.665	3
Test of Equal Inferences	$-2 \log \lambda_7$	(6.10)	2.362	6
Test of Model Fit	χ^2	(6.11)	7.557	12

As a final comment on the example, cell frequencies are small and asymptotic theory must be regarded only as approximate. The tests do, however, seem to work well and be adequately indicative.

7. Other Methods of Paired Comparisons

Our efforts in this chapter have concentrated on one method of paired comparisons and its extensions. This was done because it has been most fully developed and has been found to work well in applications. Even so, it has been necessary to be brief and applications require computer programs that are easily developed after review of pertinent references for additional detail.

We have seen that the Thurstone model is very similar to the one used here. It has had less attention. However, three papers do extend the Thurstone model: Harris (1957) generalized the model to allow for possible order effects, Glenn and David (1960) allowed for ties, and Sadasivan (1982) permitted unequal numbers of judgments on pairs.

Other approaches to the analysis of paired comparisons exist. Kendall and Babington Smith (1940) considered the count of circular triads as a measure of consistency of judgments and also developed a coefficient of concordance as a measure of agreement of judgments by several judges. Guttman (1946) developed a method of scaling treatments in paired comparisons, the objective of Zermello. Saaty (1977) proposed a consensus method through evaluation by group discussion to provide treatment or item scores on a ratio scale. Bliss, Greenwood and White (1956) used "rankits" in the analysis of paired comparisons. Mehra (1964) and Puri and Sen (1969) extended the idea of signed ranks to paired comparisons. Wei (1952) and Kendall (1955) have proposed an iterative scoring system that takes into account not only direct comparisons but also roundabout comparisons involving other items.

No attention has been given here to the design of tournaments. There is an extensive literature on this subject included in the Davidson-Farquhar bibliography.

REFERENCES

Abelson, R. M. and Bradley, R. A. (1954). A 2×2 factorial with paired comparisons. *Biometrics 10*, 487-502.

Beaver, R. J. (1976). Discussion: Science, statistics and paired comparisons. *Biometrics 32*, 233-235.

Beaver, R. J. and Gokhale, D. V. (1975). A model to incorporate within-pair order effects in paired comparisons. *Comm. Statist. A4*, 923-939.

Bliss, C. I., Greenwood, M. L. and White, E. S. (1956). A rankit analysis of paired comparisons for measuring the effects of sprays on flavor. *Biometrics 12*, 381-403.

Bradley, R. A. (1953). Some statistical methods in taste testing and quality evaluation. *Biometrics 9*, 22-38.

Bradley, R. A. (1954a). The rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika 41*, 502-537.

Bradley, R. A. (1954b). Incomplete block rank analysis: On the appropriateness of the model for a method of paired comparisons. *Biometrics 10*, 375-390.

Bradley, R. A. (1955). Rank analysis of incomplete block designs. III. Some large-sample results on estimation and power for a method of paired comparisons. *Biometrika 42*, 450-470.

Bradley, R. A. (1976). Science, statistics and paired comparisons. *Biometrics 32*, 213-232.

Bradley, R. A. and El-Helbawy, A. T. (1976). Treatment contrasts in paired comparisons: Basic procedures with application to factorials. *Biometrika 63*, 255-262.

Bradley, R. A. and Terry, M. E. (1952a). The rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika 39*, 324-345.

Bradley, R. A. and Terry, M. E. (1952b). *Statistical Methods for Sensory Difference Tests of Food Quality, Appendix A*. Biannual Rpt. No. 4, Virginia Agric. Exp. Sta., Blacksburg, Va.

Clatworthy, W. H. (1955). Partially balanced incomplete block designs with two associate classes and two treatments per block. *J. Res. Nat. Bur. Stand. 54*, 177-190.

David, H. A. (1963). *The Method of Paired Comparisons*. Griffin, London.

Davidson, R. R. (1970). On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *J. Amer. Statist. Assoc. 65*, 317-328.

Davidson, R. R. and Beaver, R. J. (undated). On extending the Bradley-Terry model to incorporato within-pair order effects. Unpublished manuscript.

Davidson, R. R. and Bradley, R. A. (1969). Multivariate paired comparisons: The extension of a univariate model and associated estimation and test procedures. *Biometrika 56*, 81-95.

Davidson, R. R. and Bradley, R. A. (1970). Multivariate paired comparisons: Some large sample results on estimation and tests of equality of preference. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.), Cambridge University Press, 111-125.

Davidson, R. R. and Bradley, R. A. (1971). A regression relationship for multivariate paired comparisons. *Biometrika 58*, 555-560.

Davidson, R. R. and Farquhar (1976). A bibliography on the method of paired comparisons. *Biometrics 32*, 241-252.

Davidson, R. R. and Solomon, D. L. (1973). A Bayesian approach to paired comparison experimentation. *Biometrika 60*, 477-487.

Dykstra, O. (1956). A note on the rank analysis of incomplete block designs -- applications beyond the scope of existing tables. *Biometrics 12*, 301-306.

Dykstra, O. (1960). Rank analysis of incomplete block designs: A method of paired comparisons employing unequal repetitions on pairs. *Biometrics 16*, 176-188.

El-Helbawy, A. T. and Bradley, R. A. (1976). Factorial treatment combinations in paired comparisons. *Proc. Int. Conf. on Statistics, Computer Science and Social Research 1*, 121.1-144.1, Cairo University Press, Cairo.

El-Helbawy, A. T. and Bradley, R. A. (1977). Treatment contrasts in paired comparisons: Convergence of a basic iterative scheme for estimation. *Commun. Statist. A6*, 197-207.

El-Helbawy, A. T. and Bradley, R. A. (1978). Treatment contrasts in paired comparisons: Large-sample results, applications and some optimal designs. *J. Amer. Statist. Assoc. 73*, 831-839.

Fleckenstein, M., Freund, R. A., and Jackson, J. E. (1958). A paired comparison test of typewriter carbon papers. *Tappi 41*, 128-130.

Ford, L. R. Jr. (1957). Solution of a ranking problem from binary comparisons. *Amer. Math. Monthly 64(8)*, 28-33.

Glenn, W. A. and David, H. A. (1960). Ties in paired-comparison experiments using a modified Thurstone-Mosteller method. *Biometrics 16*, 86-109.

Gridgeman, N. T. (1959). Pair comparison, with and without ties. *Biometrics 15*, 382-388.

Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. *Ann. Math. Statist. 17*, 144-163.

Harris, W. P. (1957). A revised law of comparative judgment. *Psychometrika 22*, 189-198.

Hemelrijk, J. (1952). A theorem on the sign test when ties are present. *Indag. Math. 14*, 322-326.

Kendall, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics 11*, 43-62.

Kendall, M. G. and Babington Smith, B. (1940). On the method of paired comparisons. *Biometrika 31*, 324-345.

Luce, R. D. (1959). *Individual Choice Behavior*. John Wiley & Sons, Inc., New York.

Mehra, K. L. (1964). Rank tests for paired-comparison experiments involving several treatments. *Ann. Math. Statist. 35*, 122-137.

Mosteller, F. (1951). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika 16*, 3-9.

Park, G. T. (1961). Sensory testing by triple comparisons. *Biometrics 17*, 251-260.

Pendergrass, R. N. and Bradley, R. A. (1960). Ranking in triple comparisons. *Contributions to Probability and Statistics*, (I. Olkin *et al.*, ed.). Stanford Univ. Press, 331-351.

Puri, M. L. and Sen, P. K. (1969). On the asymptotic theory of rank order tests for experiments involving paired comparisons. *Ann. Inst. Statist. Math. 21*, 163-173.

Raghavarao, D. (1971). *Construction and Combinatorial Problems in Design of Experiments*. John Wiley & Sons, Inc., New York.

Rao, P. V. and Kupper, L. L. (1967). Ties in paired-comparison experiments: A generalization of the Bradley-Terry model. *J. Amer. Statist. Assoc. 62*, 194-204, Corregenda *63*, 1550.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *J. Math. Psych. 15*, 234-280.

Sadasivan, G. (1982). A Thurstone-type model for paired comparisons with unequal numbers of repetitions. *Commun. Statist.-Theor. Math. 11*, 821-833.

Scheffé, H. (1952). An analysis of variance for paired comparisons. *J. Amer. Statist. Assoc.* *47*, 381-400.

Thurstone, L. L. (1927). Psychophysical analysis. *Amer. J. Psych.* *38*, 368-389.

Wei, T. H. (1952). *The Algebraic Foundations of Ranking Theory.* Unpublished thesis, Cambridge University.

Zermelo, E. (1929). Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung. *Math. Zeit.* *29*, 436-460.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU Report No M615 ONR Report No 157	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and subtitle) Paired Comparisons		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER FSU Report No. M615
7. AUTHOR(s) Ralph A. Bradley		8. CONTRACT OR GRANT NUMBER(s) ONR Contract No. N00014-80-C-00
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Florida State University Tallahassee, FL 32306		10. PROGRAM ELEMENT, PROJECT, TASK & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Arlington, VA 22217		12. REPORT DATE May, 1982 13. NUMBER OF PAGES 46
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from re

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Paired comparisons, factorials, categorical data, multivariate discrete methods, goodness of fit.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This technical report is an invited chapter for the *Handbook of Statistics: Nonparametric Methods*, Volume 4 in a series edited by P. R. Krishnaiah and P. K. Sen and to be published by North-Holland Publishing Company, Amsterdam. Much of the material by the author used in the chapter was developed under ONR-sponsored research at the Florida State University and earlier at the Virginia Polytechnic Institute and State University. Some minor new generalizations of earlier work are included here.