A COMPARISON OF TWO SCALING PROCEDURES IN PAIRED - COMPARISON EXPERIMENTS INVOLVING TIES

by

Shiu - Hon Wong

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

Statistics

May 1960

Blacksburg, Virginia

TABLE OF CONTENTS

I. INTRODUCTION

The paired - comparison experiment was introduced by Thurstone [6] for the purpose of estimating the relative strengths of treatment stimuli through subjective testing. The method was later refined and extended by Mosteller [4]. In the resulting Thurstone-Mosteller method it is assumed that sensations over a response continuum are jointly normally distributed with equal standard deviations and equal correlations between pairs. It is further assumed that all differences, however small, are perceptible, and tied observations are not permitted.

Recently Glenn and David [2] have proposed a modification of this method which makes possible the analysis of data involving ties. The basis of the modification is the assumption that whenever the difference between his responses to two stimuli under comparison lies below a certain threshold, the judge will declare a tie. This threshold and the mean stimulus responses are estimated by least squares.

In the estimation procedure, the no-preferences in every pair are added to the clear preferences expressed for each member of the pair. A scaling transformation is then applied to the corresponding proportions. The sums and differences of the resulting transforms for each pair are used as transformed data, to which the least squares estimation procedure is applied. When the response proportions are scaled with normal deviates, the sum and the difference of the transforms are correlated. This difficulty is overcome, at least in large samples, by postulating an angular response law for the response-scale differences. The angular law is such that an arc sine transformation is used in scaling the response proportions. As a result of the variance-stabilizing property of the arc sine transformation the sum and the difference are uncorrelated, at least for large samples.

The problem considered in this thesis is whether, in the estimation procedure described in [2], the results obtained by scaling with arc sine transforms are in closer agreement with the observations than those obtained by scaling with normal deviates. Otherwise stated, the question is whether the proportions predicted from the arc sine solution are closer to the observations than those given by scaling with normal deviates. This problem is motivated in part by a suggestion recently made by Mosteller [5]. In a general discussion of scales of measurement in psychometric work, he comments that it would be useful to explore the sensitivity of the method of paired comparisons to the shape of the curve used to scale the response proportions. He cites experience indicating that while the spacings obtained by the use of various scaling procedures are roughly invariant except for linear transformations, the reproduction of the original percentages is not the same. This experience is with paired comparisons not involving ties, of course, but the general question of the suitability of a scaling procedure has relevance also to situations in which ties are permitted.

The present study is carried out by applying both the arc sine and the normal scaling procedures to data obtained from two important fields of application of the paired-comparison technique. These fields are the testing of uniformity of differences in color scales and the sensory testing of food products. The closeness of agreement between observed and expected numbers is measured by means of a chi-square test of goodness of fit proposed in [2]. The results are presented in Chapter IV. In Chapter V results are given for certain hypothetical examples constructed by modifying some of the live data of Chapter IV, with a view to studying the possible effects of features of the data which might have bearing on the comparison of the two scaling procedures.

The computations in this study were made with the aid of an IBM-650. Programs giving the analysis by each of the two scaling procedures as well as the chi - square values have been written, and are described in the appendix. The program decks are on file in the Virginia Polytechnic Institute Computing Center.

II. MODEL

Consider a paired - comparison experiment involving t treatments, and let X_i and X_j be single responses of a judge to the i th and j th stimuli. Let S_i denote the true response to the i th stimulus $(i = 1,\ldots,t)$. Then under the Thurstone-Mosteller model as described in [4] the probability distribution of the difference $X_i - X_j$ $(i \neq j)$ is normal with mean $S_i - S_j$ and unit variance. Let us define

$$F(a) = \frac{1}{\sqrt{2\pi}} \int_{-a}^{\infty} e^{-\frac{1}{2}y^2}\, dy \,, \tag{2.1}$$

from which it is evident that

$$F(-a) = 1 - F(a) \,. \tag{2.2}$$

Denoting treatments i and j by T_i and T_j respectively, we may write the probability that T_i is preferred when T_i and T_j are compared as

$$\pi_{ij} = P(X_i > X_j) = F(S_i - S_j) \,. \tag{2.3}$$

It is evident from (2.2) and (2.3) that the Thurstone-Mosteller model makes no provision for tied observations. In the modification proposed in [2] the admission of ties is accomplished by postulating that there exists an interval of length 2τ, centered at the origin of the distribution of $X_i - X_j$, within which the judge cannot distinguish between

X_i and X_j, and will declare a tie. The probabilities that T_i, T_j, or neither are preferred in the comparison of T_i and T_j are defined respectively as

$$\pi_{i.ij} = P[(X_i - X_j) > \tau] = F(-\tau + S_i - S_j) ,$$

$$\pi_{j.ij} = P[(X_i - X_j) < -\tau] = 1 - F(\tau + S_i - S_j) ,$$

and

$$\pi_{o.ij} = P[|X_i - X_j| \leq \tau] = F(\tau + S_i - S_j) - F(-\tau + S_i - S_j) .$$

This set of relations replaces (2.3) and by virtue of (2.2) may be expressed in the form

$$\pi_{i.ij} + \pi_{o.ij} = F(\tau + S_i - S_j)$$

and (2.4)

$$\pi_{j.ij} + \pi_{o.ij} = F(\tau - S_i + S_j) .$$

Suppose n observations are made in each of the $\binom{t}{2}$ comparisons and let the data recorded be

$p_{i.ij} = n_{i.ij}/n$ = proportion of preferences for T_i,

$p_{j.ij} = n_{j.ij}/n$ = proportion of preferences for T_j,

and

$p_{o.ij} = n_{o.ij}/n$ = proportion of ties for each of

the combinations of T_i and T_j, where

$$n_{i.ij} + n_{j.ij} + n_{o.ij} = n$$

so that

$$p_{i.ij} + p_{j.ij} + p_{o.ij} = 1.$$

If we denote the experimental values of τ and S_i by $\tau'_{(ij)}$ and S_i', according to (2.4) we have

$$p_{i.ij} + p_{o.ij} = F(\tau'_{(ij)} + S_i' - S_j') = a_{ij}$$

and (2.5)

$$p_{j.ij} + p_{o.ij} = F(\tau'_{(ij)} - S_i' + S_j') = a_{ji},$$

where a_{ij} and a_{ji} are used for brevity. Hence, we may write

$$\tau'_{(ij)} + S_i' - S_j' = F^{-1}(a_{ij})$$

and

$$\tau'_{(ij)} - S_i' + S_j' = F^{-1}(a_{ji}),$$

where $F^{-1}(a_{ij})$ and $F^{-1}(a_{ji})$ are discrete variates which may be called the pseudo-normal deviates exceeded with probabilities $(1-a_{ij})$ and $(1-a_{ji})$ respectively. Solving for $\tau'_{(ij)}$ and $S_i' - S_j'$, we have

$$\tau'_{(ij)} = \tfrac{1}{2}[F^{-1}(a_{ij}) + F^{-1}(a_{ji})]$$

and (2.6)

$$S_i' - S_j' = \tfrac{1}{2}[F^{-1}(a_{ij}) - F^{-1}(a_{ji})].$$

When $F^{-1}(a_{ij})$ and $F^{-1}(a_{ji})$ represent normal deviates, the data $\tau'_{(ij)}$ and $(S_i'-S_j')$ are not independent. If we let

$\rho[\tau'_{(ij)},(S'_i-S'_j)]$ denote the correlation between $\tau'_{(ij)}$ and $(S'_i-S'_j)$, it can be shown that asymptotically

$$\rho[\tau'_{(ij)},(S'_i-S'_j)] = (V_1-V_2)/\sqrt{(V_1+V_2+2V_3)(V_1+V_2-2V_3)}, \quad (2.7)$$

where

$$V_1 = \mathrm{Var}[F^{-1}(a_{ij})] = (1-\pi_{j.ij})(\pi_{j.ij})/\{n[f(\tau+S_i-S_j)]^2\},$$

$$V_2 = \mathrm{Var}[F^{-1}(a_{ji})] = (1-\pi_{i.ij})(\pi_{i.ij})/\{n[f(\tau-S_i+S_j)]^2\},$$

$$V_3 = \mathrm{Cov}[F^{-1}(a_{ij}),F^{-1}(a_{ji})]$$

$$= -\pi_{i.ij}\pi_{j.ij}/\{n[f(\tau+S_i-S_j)][f(\tau-S_i+S_j)]\},$$

and $f(\tau+S_i-S_j)$, $f(\tau-S_i+S_j)$ denote the ordinates of the standard normal curve at abscissa values $(\tau+S_i-S_j)$ and $(\tau-S_i+S_j)$ respectively.

In order to overcome the difficulty presented by lack of independence, at least in large samples, the arc sine transformation is proposed in [2]. This is accomplished by defining

$$F(a) = \tfrac{1}{2}\int_{-a}^{\pi/2} \cos y \, dy = \tfrac{1}{2}(1+\sin a), \quad (2.8)$$

where a represents an angle measured in radians, such that

$$-\tfrac{1}{2}\pi \le a \le \tfrac{1}{2}\pi.$$

From (2.8) we have

$$F^{-1}(a_{ij}) = \sin^{-1}(2a_{ij}-1)$$

and

$$F^{-1}(a_{ji}) = \mathrm{Sin}^{-1}(2a_{ji} - 1).$$

Thus we may write

$$\tau'_{(ij)} = \tfrac{1}{2}[\mathrm{Sin}^{-1}(2a_{ij} - 1) + \mathrm{Sin}^{-1}(2a_{ji} - 1)]$$

and (2.9)

$$S'_i - S'_j = \tfrac{1}{2}[\mathrm{Sin}^{-1}(2a_{ij} - 1) - \mathrm{Sin}^{-1}(2a_{ji} - 1)].$$

Since it may be shown that approximately, for large samples,

$$\mathrm{Var}[\mathrm{Sin}^{-1}(2a_{ij} - 1)] = \mathrm{Var}[\mathrm{Sin}^{-1}(2a_{ji} - 1)] = 1/n,$$

it is evident that

$$\mathrm{Cov}[\tau'_{(ij)}, (S'_i - S'_j)] = 0.$$

Thus for large samples, $\tau'_{(ij)}$ and $(S'_i - S'_j)$ as given by (2.9) are uncorrelated.

In this thesis we are concerned with comparing the results obtained by using the model (2.6), in which F^{-1} denotes a pseudo-normal deviate, with those obtained by using (2.9), in which the pseudo-normal deviate is replaced by the arc sine function as indicated. The criterion of comparison is the goodness of fit between the observed and expected numbers in each category. Thus in effect we are inquiring as to whether the arc sine transformation, which overcomes the problem of correlated data, at least for large samples, gives results in better agreement with the observations than those given by ignoring the correlations and scaling with the normal curve.

The method described in [2] includes a weighted analysis and also the analysis of non-balanced experiments. In this study, with the interest centering on the comparison of the two scaling procedures, we consider only the unweighted analysis of balanced experiments. This is not unreasonable for the purpose at hand, since it has been indicated in [2] that the unweighted analysis yields results very close to the final results of a weighted solution carried out iteratively. In addition, the present writer has also found this to be true in a number of examples.

III. COMPUTATIONAL METHOD

3.1 Estimation of Parameters

Consider a paired-comparison experiment involving t treatments, in which n observations are made on each of the $\binom{t}{2}$ pairs. For the purpose of the present study we use the unweighted least squares estimates of the parameters τ and S_i $(i = 1,\ldots,t)$ as obtained in [2], Section 3. For brevity we define

$$\tau'_{(ij)} = \tfrac{1}{2}[F^{-1}(a_{ij}) + F^{-1}(a_{ji})] = g_{ij} \qquad (3.1)$$

and

$$S'_i - S'_j = \tfrac{1}{2}[F^{-1}(a_{ij}) - F^{-1}(a_{ji})] = h_{ij} \qquad (3.2)$$

From (3.1) we obtain the least squares estimate for τ as τ^* such that

$$Q_1 = \sum_{i<j}^{t}(\tau - g_{ij})^2$$

is a minimum for $\tau = \tau^*$, where $\sum_{i<j}^{t}$ denotes the sum over all $\binom{t}{2}$ pairs for which $i < j$. Setting

$$\frac{\partial Q_1}{\partial \tau} = 0$$

we have $\tau^* = \dfrac{2}{t(t-1)} \sum_{i<j}^{t} g_{ij}$. (3.3)

From (3.2) we obtain the least squares estimates of S_i as S_i^* $(i = 2,\ldots,t)$, S_1 being taken as origin, such that

$$Q_2 = \sum_{i<j}^{t}(S_i - S_j - h_{ij})^2$$

is a minimum for $S_i = S_i^*$ $(i = 2,\ldots,t)$, and $S_1^* = 0$. For convenience we express Q_2 in matrix form. Let $\underline{y}$ be a $\binom{t}{2}\times(1)$ column vector with elements h_{ij}, i.e.,

$$\underline{y}' = [h_{12},\ldots,h_{ij},\ldots,h_{t-1,t}],$$

and let $\underline{b}$ be a $(t-1)\times(1)$ column vector with elements S_i $(i = 2,\ldots,t)$, i.e.,

$$\underline{b}' = [S_2,\ldots,S_t].$$

Let X be a $\binom{t}{2}\times(t-1)$ matrix consisting of 1's, -1's and 0's such that for each element in $\underline{y}$ there is a row in X and for each column in X there is an element in $\underline{b}$. The row corresponding to h_{ij} has +1 in the column corresponding to S_i and -1 in the column corresponding to S_j. All other elements are zero. Now we may write

$$Q_2 = (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b}). \tag{3.4}$$

After setting

$$\frac{\partial Q_2}{\partial \underline{b}} = \underline{0}$$

we have

$$\underline{b}^* = (X'X)^{-1}X'\underline{y}. \tag{3.5}$$

It is evident that

$$X'X = tI - \underline{j}\underline{j}', \tag{3.6}$$

where I, of order $(t-1)$, is a unit matrix, and $\underline{j}$ is a column vector of order $(t-1)\times(1)$ with elements equal to unity.

From (3.6), the inverse of X'X is found to be

$$(X'X)^{-1} = \frac{1}{t} I + \frac{1}{t} \underline{j}\underline{j}' . \tag{3.7}$$

Let us denote by H_1 a (t)x(t) matrix such that its elements above the main diagonal are h_{ij}, those below the main diagonal are $h_{ji} = -h_{ij}$, and those on the main diagonal are zero. If we delete the first row in H_1 and call the resulting matrix H, (t-1)x(t), it is found that

$$X'\underline{y} = H\underline{j} , \tag{3.8}$$

which is a (t-1)x(1) column vector, the elements of which are the row sums of H.

Finally, substituting from (3.7) and (3.8) into (3.5), one obtains

$$\underline{b}^* = \frac{1}{t}(I + \underline{j}\underline{j}')H\underline{j} = \frac{1}{t}(H\underline{j} + \underline{j}\underline{j}' H\underline{j}) , \tag{3.9}$$

the elements of which may be written as

$$S_i^* = \frac{1}{t}\left(\sum_{j=1}^{t} h_{ij} + \sum_{i=2}^{t} \sum_{j=1}^{t} h_{ij}\right) ,$$

where $h_{ji} = -h_{ij}$, $h_{ii} = 0$, and $i = 2,\dots,t$.

For scaling with normal deviates we take $F^{-1}(a_{ij})$ and $F^{-1}(a_{ji})$ in (3.1) and (3.2) as defined for (2.6); whereas for scaling with arc sine transforms we replace (2.6) with (2.9). The solutions are otherwise formally the same. IBM-650 programs for performing the actual calculations are described in the appendix.

3.2 Chi - Square Test of Goodness of Fit

As a criterion of comparison of the two scaling procedures we use the goodness of fit between observed and expected numbers as measured by a chi-square test proposed in [2]. Having obtained the estimates τ^* and S_i^* $(i = 2,\ldots,t)$, $S_1^* = 0$, we first determine the expected values of a_{ij} and a_{ji}, denoted by a_{ij}^* and a_{ji}^* respectively. For scaling with normal deviates these are given by

$$a_{ij}^* = F(\tau^* + S_i^* - S_j^*) \tag{3.11}$$

and

$$a_{ji}^* = F(\tau^* - S_i^* + S_j^*) \ , \tag{3.12}$$

in accord with (2.5), in which F stands for the cumulative form of the standard normal function as defined in (2.1). For scaling with arc sine transforms we have

$$a_{ij}^* = \tfrac{1}{2}[1 + \sin(\tau^* + S_i^* - S_j^*)] \tag{3.13}$$

and

$$a_{ji}^* = \tfrac{1}{2}[1 + \sin(\tau^* - S_i^* + S_j^*)] \ , \tag{3.14}$$

arising from the definition (2.8).

Let the expected numbers be denoted by

$n_{i.ij}^*$ = expected number of preferences for T_i,

$n_{j.ij}^*$ = expected number of preferences for T_j, and

$$n^*_{o.ij} = \text{expected number of ties,}$$

when T_i and T_j are compared, where

$$n^*_{i.ij} + n^*_{j.ij} + n^*_{o.ij} = n.$$

From the definitions of a^*_{ij} and a^*_{ji} it follows that

$$n^*_{i.ij} + n^*_{o.ij} = na^*_{ij}$$

and

$$n^*_{j.ij} + n^*_{o.ij} = na^*_{ji} ,$$

from which we obtain

$$n^*_{j.ij} = n(1 - a^*_{ij}) , \qquad (3.15)$$

$$n^*_{i.ij} = n(1 - a^*_{ji}) , \qquad (3.16)$$

and

$$n^*_{o.ij} = n(a^*_{ij} + a^*_{ji} - 1) . \qquad (3.17)$$

For the test of goodness of fit proposed in [2] the chi-square statistic takes the form

$$X^2 = \sum_{i<j}^{t}[(n_{i.ij} - n^*_{i.ij})^2/n^*_{i.ij} + (n_{j.ij} - n^*_{j.ij})^2/n^*_{j.ij} + (n_{o.ij} - n^*_{o.ij})^2/n^*_{o.ij}].$$

For sufficiently large values of the expected numbers, X^2 is distributed approximately as chi-square with degrees of freedom determined as follows. There are $\binom{t}{2}$ pairs which yield two independent observations each, or $2\binom{t}{2} = t(t-1)$ independent observations in all. From the data we have estimated the parameters τ and S_i $(i = 2,\ldots,t)$, or a total of t parameters. Thus the degrees of freedom for X^2 are

$$t(t-1) - t = t(t-2).$$

The calculation of X^2 is included in the IBM-650 programs described in the appendix.

IV. EXAMPLES: LIVE DATA

4.1 Sources of Data

The data used in this study come from two important fields of application of the paired-comparison technique. These are: (a) research for the establishment of uniform color scales and (b) consumer-preference testing of food products. They are described briefly below, the former at somewhat more length than the latter, which is the better-known application. All of the cases considered are balanced paired-comparison experiments involving ties. That is, for each of the possible pairs data are recorded in the trinomial form consisting of preferences for one member or the other and no-preferences, with the same number of observations on all pairs in a given experiment.

4.1.1 Color Scaling Experiments

In determining physically reproducible standards for colors, the Munsell system of color specification is widely used. In this system it is attempted to arrange conceptual colors in a three-dimensional space according to their psychological attributes, in such a way that they are equally spaced visually in all directions. The problem of relating differences in the system specifications to visual differences is the problem of establishing a color metric for the system. In testing the metric, experimenters must determine whether the differences between stimuli from equally spaced points on the scale do in fact correspond to equal visual differences.

Series of experiments in this field are being conducted by the OSA Committee on Uniform Color Scales [3], which organization supplied the first series of data considered in this chapter. The basic materials used in these experiments are sets of colored tiles, each set corresponding to a Munsell-value level or plane in the three-dimensional color space. From the tiles at a given level pairs are made up corresponding to certain hue differences. These hue differences are combined with certain masking arrangements when viewed by the observers. A particular hue difference and mask combination is referred to as a chromaticness difference. One type of problem considered is the following: We are given a particular chromaticness difference at each of t Munsell-value levels, and it is desired to obtain a scaling of the Munsell-value levels for that chromaticness difference. This leads to a paired-comparison experiment involving t treatments, viz. the t Munsell-value levels.

Since it often occurs that an observer is unable to distinguish between the items of a pair in these experiments, it is considered desirable to permit ties.

The first series of data considered in this chapter is of the type described above. Six hue differences were associated with two masking arrangements in all possible combinations for a total of 12 chromaticness differences. For each of these 12 chromaticness differences three Munsell-value levels were compared. We shall designate the three Munsell-value levels as A, B, and C, and number the chromaticness differences 1-12. We thus have 12 paired-comparison experiments involving three treatments each. These are designated by CS 1 to CS 12 (CS for "color scaling"). The data, involving 97 observations on each pair, are given in Table 4.1, which includes estimates of $\rho[\tau'_{(ij)}, (S'_i - S'_j)]$ in normal scaling for the separate pairs. These estimates are denoted by $r[\tau'_{(ij)}, (S'_i - S'_j)]$, and their calculation is described in the appendix. The estimates τ^* and S^*_i, $i = 2, 3$, together with the X^2 values for each of the two scaling procedures are given in Table 4.2.

4.1.2 Food Research Experiments

The paired-comparison technique has been widely used in consumer-preference testing, principally of such items as various types of food products. In many cases ties have not been permitted, while in other cases tied observations have been discarded before analyzing the data. These practices have tended to waste information contained in a no-preference class. The method proposed in [2] was motivated in part by the feeling that in the estimation of response-scale values, tied observations should be taken into account.

The second series of examples considered in this chapter is based on data supplied by General Foods Research Center [1]. The data pertain to four different consumer-preference studies in which the treatments represent variations of a consumer product, and in each case are coded A,B,C,---. The numbers of treatments involved in these studies are 3,4,4, and 6 respectively, and the respective numbers of observations on each pair are 98, 80, 150, and 34 for the four studies. These are designated by FR 1, FR 2, FR 3, and FR 4 (FR for "food research"). The data and the estimates of $\rho[\tau'_{(ij)},(S'_i-S'_j)]$ in normal scaling for the separate pairs are given in Table 4.3. The estimates τ^* and S^*_i, together with the X^2 values for each of the two scaling procedures are given in Table 4.4.

4.2 Results of Analysis

The data of Tables 4.1 and 4.3 have been analyzed using each of the two scaling procedures by the method described in Chapter III. The estimates τ^* and S_i^*, $i = 2,\ldots,t$, as well as the chi-square statistics X^2 for each example under each scaling procedure are given in Tables 4.2 and 4.4. X_n^2 and X_s^2 are used to denote the X^2 values under normal curve and arc sine scaling, respectively.

Upon comparing the X_n^2 and X_s^2 values in Table 4.2, one finds that $X_s^2 < X_n^2$ in only two cases (CS 1 and CS 8). In the other 10 color scaling examples $X_n^2 < X_s^2$, indicating that in these cases normal curve scaling gives results in somewhat better agreement with the observations than does arc sine scaling. It is evident from Table 4.1 that the $r[\tau'_{(ij)},(S_i'-S_j')]$ are, on the average over the three pairs, at least as large for some of the other examples as they are for CS 1 and CS 8. However, the two largest single values of $r[\tau'_{(ij)},(S_i'-S_j')]$ are associated with the latter examples. This tends to provide a small indication that preference for the arc sine procedure may be associated with relatively large values of this correlation.

Looking at the results in Table 4.2 from another standpoint, let us postulate that the agreement between observed and expected numbers is satisfactory unless X^2 exceeds, say, the 90th percentile of the chi-square distribution with 3 degrees of freedom. That is, unless X^2 exceeds 6.25 we regard the agreement as satisfactory. On this basis, there are six examples (CS 4, 7, 8, 10, 11, and 12) in which neither scaling procedure gives satisfactory results. In four examples (CS 1, 2, 3, 5) both procedures are satisfactory, while in the remaining two (CS 6 and CS 9) the results of normal curve scaling are acceptable while those of arc sine scaling are not. The latter conclusion is of minor importance, of course, since the 90th percentile was an arbitrarily chosen point. The conclusions in such borderline cases will necessarily depend on the level of significance adopted in making the chi-square test.

Applying the criterion of the above paragraph to the results for the FR examples (Table 4.4), we find the conclusions to be very similar to those for the CS examples. In FR 1 both procedures yield results in satisfactory agreement with the data. The same is true of FR 2, the comparison value being 13.4, the 90th percentile of the chi-square distribution with 8 degrees of freedom. The same critical value applies in FR 3, and it is concluded that normal curve scaling gives satisfactory results, while arc sine scaling does not. This distinction is, however, rather artificial in view of the closeness of the two X^2 values. Finally, in FR 4, neither scaling procedure is satisfactory when judged by comparing the X^2 values with 33.2, the 90th percentile of the chi-square distribution with 24 degrees of freedom. As in the CS examples, we find also in the FR examples only very slight evidence that the relative merits of the two scaling procedures are affected by $r[\tau'_{(ij)}, (S'_i - S'_j)]$, the values of which are included in Table 4.3. Some relatively large values of $r[\tau'_{(ij)}, (S'_i - S'_j)]$ are found in FR 4, in which X_s^2 is slightly less than X_n^2. This fact is of minor importance, however, since both X_s^2 and X_n^2 are so large.

In general, one concludes from these examples that for data such that the general model is satisfactory, both scaling procedures will give satisfactory results. Similarly, if the general model is unsuitable for the given data, both procedures will give approximately equally unsatisfactory results. There is a small indication that the preference, if any, is slightly in favor of the normal curve procedure.

Table 4.1

Data and Correlation Estimates

for Color Scaling Examples

Example number	pair i,j	$n_{j.ij}$	$n_{o.ij}$	$n_{i.ij}$	$r[\tau'_{(ij)}, s'_i - s'_j)]$
CS 1	A,B	84	9	4	-.3926
	A,C	79	10	8	-.2721
	B,C	42	16	39	-.0086
CS 2	A,B	83	8	6	-.3013
	A,C	62	21	14	-.2194
	B,C	18	11	68	.1415
CS 3	A,B	89	5	3	-.3717
	A,C	78	12	7	-.3156
	B,C	24	17	56	.1049
CS 4	A,B	84	7	6	-.2869
	A,C	70	12	15	-.1763
	B,C	30	28	39	.0408
CS 5	A,B	65	16	16	-.1801
	A,C	47	28	22	-.1206
	B,C	26	20	51	.0899
CS 6	A,B	84	7	6	-.2869
	A,C	78	9	10	-.2223
	B,C	29	25	43	.0580
CS 7	A,B	76	15	6	-.3702
	A,C	85	5	7	-.2273
	B,C	56	27	14	-.2273
CS 8	A,B	69	23	5	-.4519
	A,C	56	30	11	-.2861
	B,C	43	25	29	-.0580
CS 9	A,B	73	17	7	-.3508
	A,C	82	9	6	-.3144
	B,C	53	25	19	-.1571
CS 10	A,B	78	13	6	-.3549
	A,C	71	16	10	-.2702
	B,C	54	20	23	-.1154
CS 11	A,B	35	45	17	-.1467
	A,C	65	24	8	-.3495
	B,C	46	30	21	-.1292
CS 12	A,B	74	17	6	-.3831
	A,C	82	8	7	-.2738
	B,C	65	18	14	-.2119

Table 4.2

Estimates and Chi-Square Values
for Color Scaling Examples

Example number	Normal Scaling τ^*	Normal Scaling s_i^*	Normal Scaling X_n^2	Arc Sine Scaling τ^*	Arc Sine Scaling s_i^*	Arc Sine Scaling X_s^2
CS 1	.2567	1.3153 1.2482	2.7720	.1634	.9286 .8971	1.4891
CS 2	.2583	1.3400 .6689	4.1623	.1754	.9811 .4854	6.0446
CS 3	.2616	1.6176 1.1685	0.5032	.1627	1.1294 .8137	1.9314
CS 4	.2678	1.1914 .9343	8.0891	.1898	.8658 .6898	11.2515
CS 5	.3127	.7038 .3587	2.7333	.2378	.5407 .2736	3.2934
CS 6	.2516	1.3000 1.0839	3.2574	.1729	.9416 .7885	7.2269
CS 7	.3208	1.0012 1.4688	9.4490	.2127	.7123 1.0625	22.0325
CS 8	.4594	.8989 .8966	12.5095	.3221	.6395 .6546	9.3106
CS 9	.3400	.9780 1.3713	2.1906	.2273	.6994 .9941	6.7675
CS 10	.3169	.9694 1.1700	9.7215	.2148	.6960 .8647	10.2894
CS 11	.5145	.3773 .8257	7.0253	.3808	.2723 .6078	8.0295
CS 12	.3146	.9142 1.4516	8.0688	.2039	.6490 1.0565	16.1646

Table 4.3

Data and Correlation Estimates

for Food Research Examples

Example numbers	pair i,j	$n_{j.ij}$	$n_{o.ij}$	$n_{i.ij}$	$r[\tau'_{(ij)},(s'_i-s'_j)]$
	A,B	39	14	45	.0155
FR 1	A,C	28	12	58	.0758
	B,C	40	11	47	.0147
	A,B	39	4	37	-.0034
	A,C	25	16	39	.0577
FR 2	A,D	30	7	43	.0307
	B,C	32	13	35	.0103
	B,D	29	11	40	.0343
	C,D	30	14	36	.0220
	A,B	61	23	66	.0088
	A,C	50	11	89	.0456
FR 3	A,D	30	17	103	.1296
	B,C	51	11	88	.0437
	B,D	35	6	109	.0707
	C,D	46	19	85	.0639
	A,B	18	4	12	-.0402
	A,C	19	0	15	.0000
	A,D	19	1	14	-.0141
	A,E	17	0	17	.0000
	A,F	6	1	27	.0854
	B,C	7	3	24	.1147
	B,D	5	1	28	.1020
FR 4	B,E	6	0	28	.0000
	B,F	1	1	32	.3074
	C,D	15	3	16	.0053
	C,E	1	2	31	.3993
	C,F	6	0	28	.0000
	D,E	13	1	20	.0226
	D,F	6	1	27	.0854
	E,F	14	5	15	.0071

Table 4.4

Estimates and Chi-Square Values

for Food Research Examples

	Normal Scaling			Arc Sine Scaling		
Example number	τ^*	s_i^*	χ_n^2	τ^*	s_i^*	χ_s^2
FR 1	.1630	-.1549 -.3223	1.6238	.1285	-.1219 -.2538	1.6077
FR 2	.1731	-.0371 -.1460 -.2206	11.8895	.1371	-.0290 -.1150 -.1745	11.9251
FR 3	.1327	-.0254 -.3381 -.6765	12.9703	.1023	-.0195 -.2647 -.5264	13.5783
FR 4	.0842	.7060 .2896 -.0034 -.5148 -.7932	46.5881	.0556	.5140 .1999 .0044 -.3667 -.5736	45.7421

V. EXAMPLES: HYPOTHETICAL DATA

The color scaling examples of Chapter IV involve only three treatments, and hence there are observations on only three pairs in each case. Because of this small number, it is relatively easy to introduce modifications in the data with a view to studying the effects of such modifications on X_n^2 and X_s^2. In particular, we are interested in determining what features of the data, if any, have bearing on the relative suitability of the normal curve and arc sine scaling procedures.

In a paired-comparison experiment involving three treatments A, B, and C, the data on the first two pairs (A B and A C) imply a certain ordering of A, B, and C on the response scale. For example, if both B and C are preferred to A, but the preference for C over A is greater than that for B over A, the data on these pairs imply the order ABC. Now if the observations on the pair B C imply that C is preferred to B, we shall say that the data are internally consistent. Otherwise, a degree of inconsistency exists.

In the color scaling data, three examples (CS 1,8, and 10) contain this type of inconsistency, while the others are internally consistent in this respect. Upon interchanging

$n_{B.BC}$ and $n_{C.BC}$ in each of the 12 sets of data, we have a series of 12 hypothetical examples, three of which are internally consistent, while the other nine are not. The analysis of these examples shows a considerable reduction in both X_n^2 and X_s^2 for the three (CS 1, 8, and 10), and the opposite for the other nine, relative to the original X^2 values. However, X_n^2 and X_s^2 seem to be approximately equally affected by these changes and the conclusions regarding the relative merits of normal curve and arc sine scaling are in effect identical to those of Section 4.2. That is, when the general model is satisfactory for the data, both procedures work well; otherwise neither is satisfactory.

Another important feature of paired-comparison data involving ties might be termed its conformity to the general model postulated in [2]. In that paper it is demonstrated that the probability of a tie is inversely related to the difference of the true stimulus responses of the treatments being compared. Thus when the observations indicate a relatively strong preference for one of the treatments in a pair, we should expect fewer ties than when the two treatments seem to be about equally preferable. That is, for data which conform to the model in this respect, we should expect more ties when $n_{j.ij}$ and $n_{i.ij}$ are nearly equal than when they differ considerably.

One approach to the study of the effects of conformity to the model, or lack of it, has been tried on the data of the color scaling examples. This approach consisted of varying $n_{o.BC}$, the number of recorded ties in the third pair, while retaining the original data on the first two pairs. It was accomplished by adding to (or subtracting from) $n_{o.BC}$ even amounts in increments of four while subtracting from (or adding to) each of $n_{B.BC}$ and $n_{C.BC}$ one half the corresponding amount. Each set of data was analyzed by both scaling procedures and the variations were continued until a point was found for which X_n^2 was least, and similarly a point for which X_s^2 was least. Let us denote these optimum values of $n_{o.BC}$ by $n_{o.BC(n)}$ and $n_{o.BC(s)}$ respectively. They are optimum in the sense that for the given data on the first two pairs, these values of $n_{o.BC}$ will yield the least values of X_n^2 and X_s^2 in the pattern of variation described above. In general, $n_{o.BC(n)}$ and $n_{o.BC(s)}$ did not coincide exactly, although they tended to be fairly close together.

The modified data involving $n_{o.BC(n)}$ and that involving $n_{o.BC(s)}$ for each of the 12 examples were subjected to further study by varying $n_{B.BC}$ and $n_{C.BC}$. This was accomplished by adding to (or subtracting from) $n_{B.BC}$ amounts in increments of four while at the same time subtracting from (or adding to) $n_{C.BC}$ the compensating amounts. The two series of resulting data were analyzed by both scaling procedures, and the variations were carried out on either side of the optimum point in each case until X^2 values in excess of 6.25 resulted. This corresponds to choosing the 90th percentile of the chi-square distribution with three degrees of freedom as the critical point. In all, this study included over 180 variations of the data, analyzed by each of the scaling procedures.

The X^2 values for the variations of the color scaling data about $n_{o.BC(n)}$ are depicted graphically in Figure 5.1 (Examples 1-6) and Figure 5.2 (Examples 7-12). The corresponding values for the variations about $n_{o.BC(s)}$ are presented in Figures 5.3 and 5.4. For convenience the origin on the horizontal scale has been taken at $n_{o.BC(n)}$ in Figures 5.1 and 5.2 and at $n_{o.BC(s)}$ in Figures 5.3 and 5.4. The modifications, obtained in the manner described above, are designated by 1, 2, 3,--- on one side of the optimum point and by -1, -2, -3,--- on the other. The graphs of X^2_n are depicted with solid lines, while those of X^2_s are represented with broken lines.

From the graphs of Figures 5.1 and 5.2 we observe that in these examples normal curve scaling yields generally better results than does arc sine scaling, since the graph of X_n^2 lies below that of X_s^2 in the principal part of each range. In Figures 5.3 and 5.4, we note that, despite of the advantage given to the arc sine procedure in these cases, it is superior to the normal curve procedure in only half the cases considered. We note, too, that when the data contains serious lack of internal consistency or lack of conformity to the general model, neither procedure may be expected to give good results. Otherwise, there is some evidence to the effect that, for the cases considered, the normal curve procedure seems to be more likely to provide the better agreement between observed and expected numbers.

The graphs in Figures 5.1 through 5.4 also indicate that the range over which the normal curve procedure gives good results is in general somewhat longer than the corresponding range for the arc sine procedure.

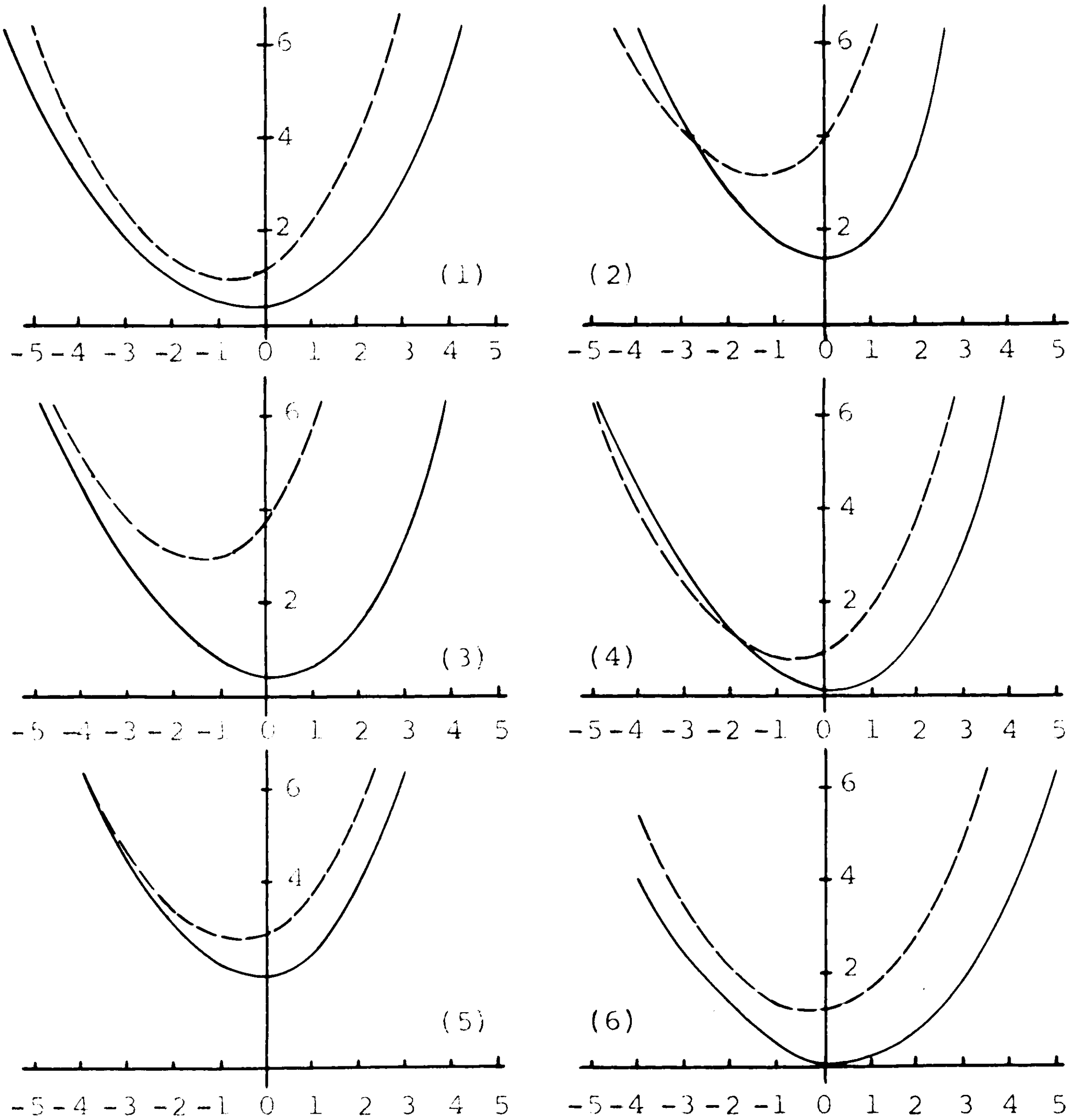

Figure 5.1

Variation of X^2 for Modifications of Color Scaling Data about $n_{o.BC(n)}$. Examples 1-6.

Solid line represents X_n^2. Broken line represents X_s^2.

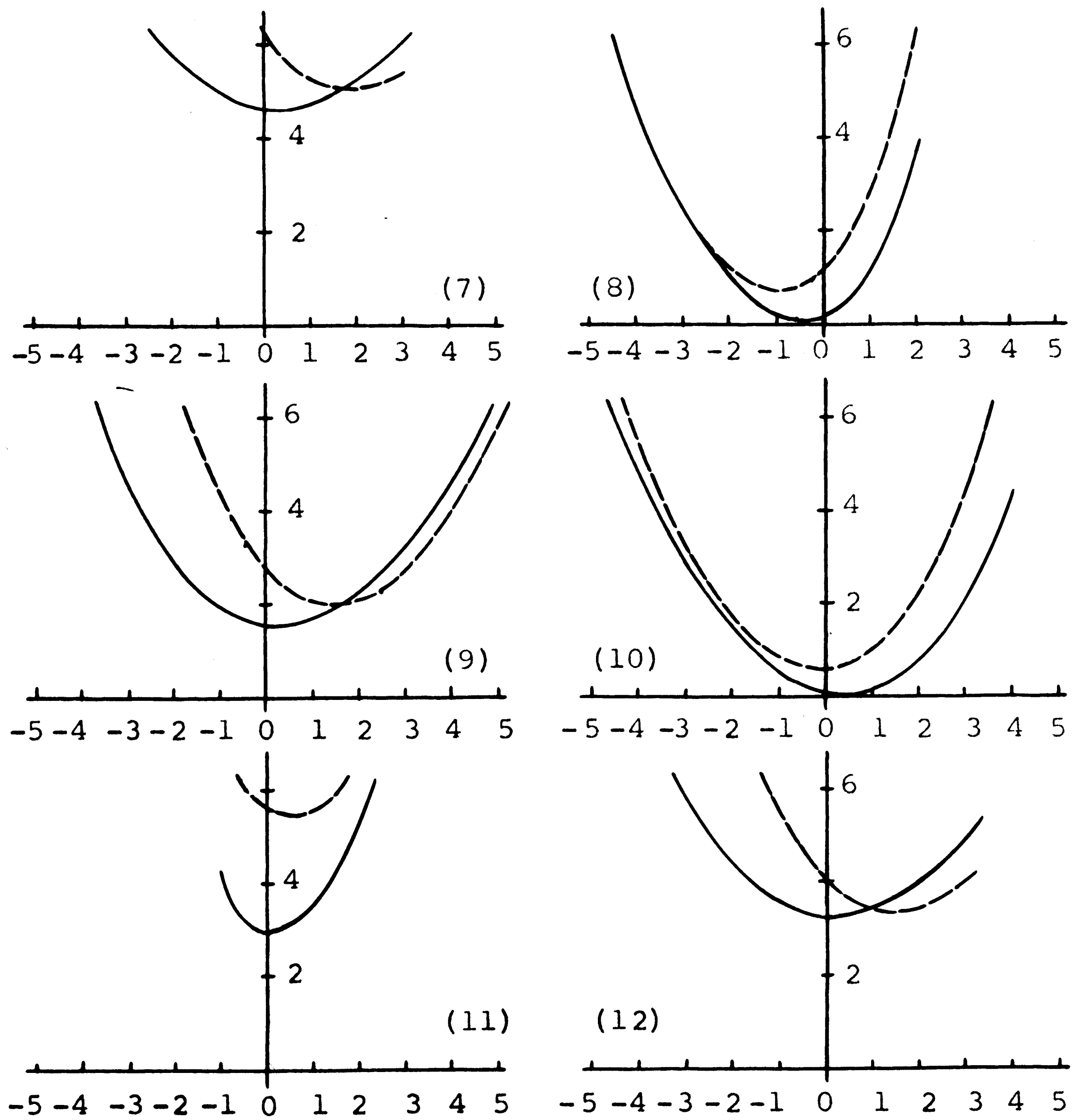

Figure 5.2

Variation of X^2 for Modifications of Color Scaling Data about $n_{o.BC(n)}$. Examples 7-12.

Solid line represents X_n^2. Broken line represents X_s^2.

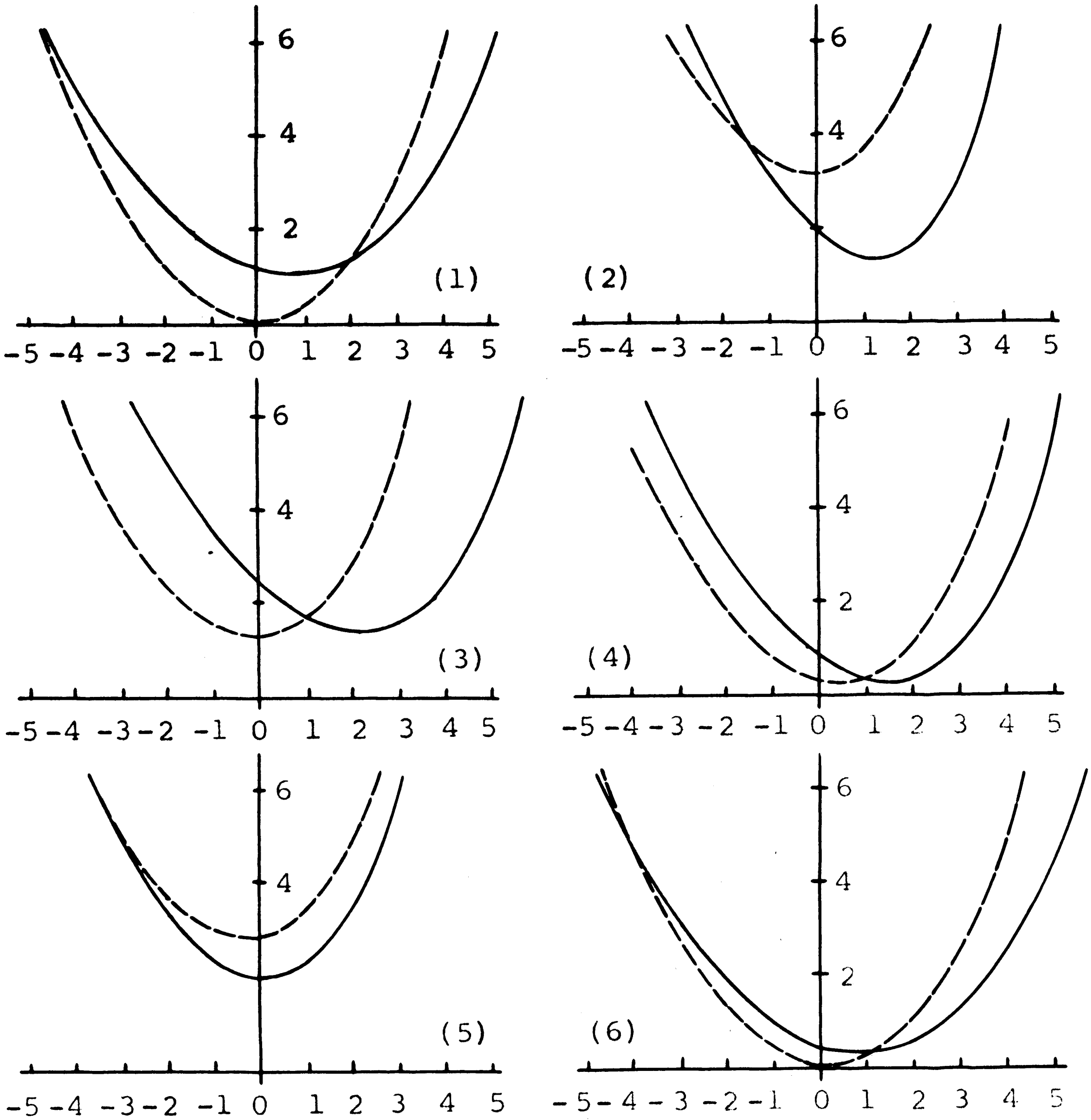

Figure 5.3

Variation of X^2 for Modifications of Color Scaling Data about $n_{o.BC(s)}$. Examples 1-6.

Solid line represents X_n^2. Broken line represents X_s^2.

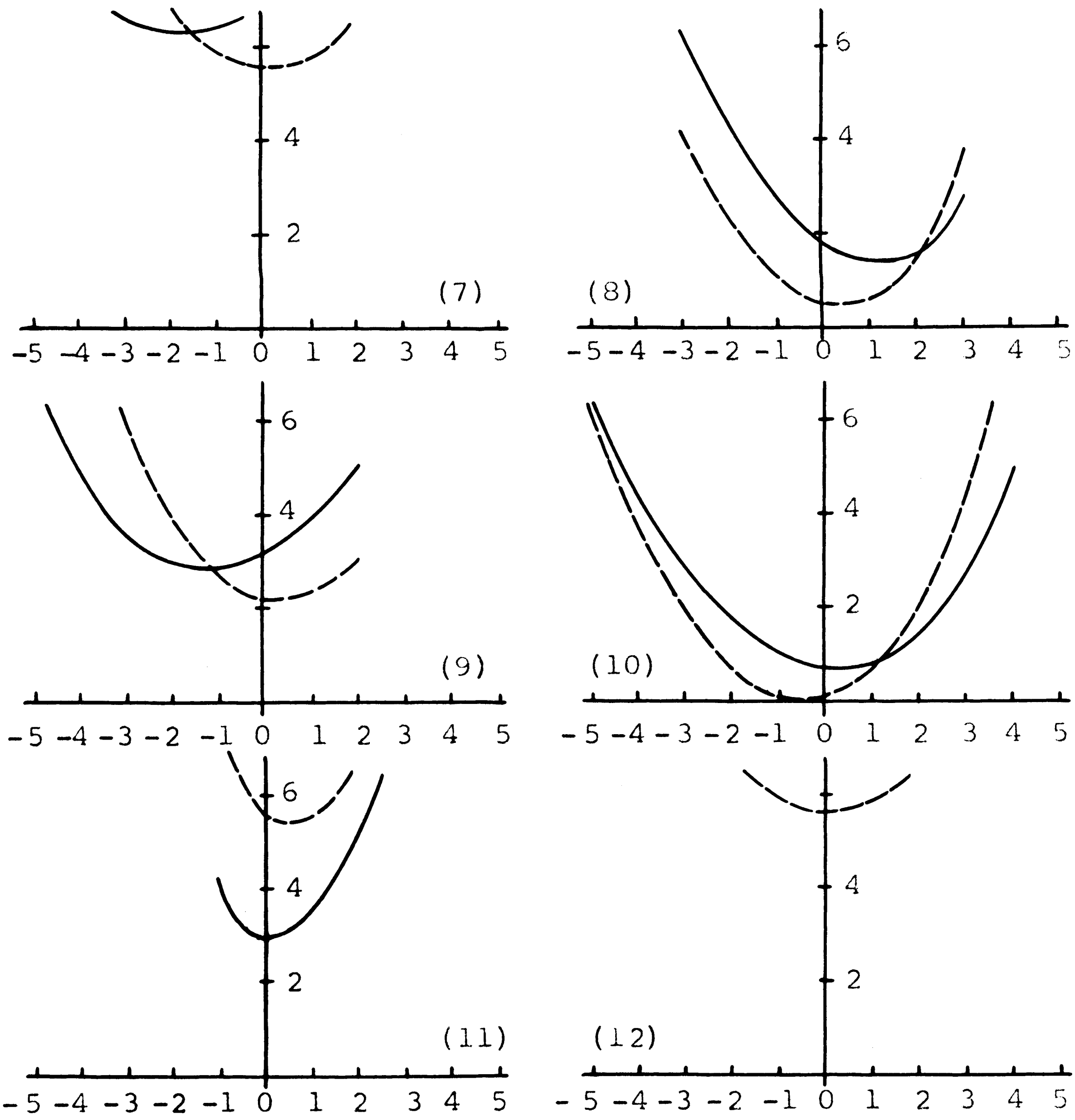

Figure 5.4

Variation of X^2 for Modifications of Color Scaling Data about $n_{o.BC(s)}$. Examples 7-12.

Solid line represents X_n^2. Broken line represents X_s^2.

VI. SUMMARY AND DISCUSSION

A recent modification [2] extends the Thurstone-Mosteller method of paired comparisons to cover cases in which ties are admitted. The modification is effected by postulating that whenever the difference between the judge's responses to two stimuli under comparison lies below a certain threshold, a tie will be declared. The threshold and the mean stimulus responses are estimated by least squares.

The pertinent data in the estimation procedure have the form of a sum and a difference of scaled response proportions. When the scaling is based on the normal curve as postulated in the Thurstone-Mosteller model, the data are correlated. This difficulty is overcome in [2], at least for large samples, by postulating an angular response law for the response-scale differences. The law is such as to lead to the scaling of the response proportions by means of arc sine transforms. For large samples these transforms have approximately a stable variance, so that the covariance of the sum and difference of two such transforms is asymptotically zero.

The research worker who uses the paired comparison technique is usually interested in determining a set of response-scale values in the best possible agreement with his observed data. Accordingly, the question naturally arises as to which of two (or more) available procedures will lead to the best agreement between the observed and expected numbers. The purpose of this paper is to compare the results of scaling with arc sine transforms as in [2] with those obtained by ignoring the correlations and scaling with normal deviates. The rationale for such a comparison is that factors other than the correlations (such, for example, as the shape in the tail areas of the response curve) may have more important effects on the agreement between observed and expected numbers than do the correlations.

Comparisons are made by applying both scaling procedures to data obtained from two important fields of application of the paired-comparison technique. The criterion of comparison is the goodness of fit between the observations and the expected numbers determined from the solution, as measured by a chi-square statistic proposed in [2]. Computations of the parameter estimates and of the chi-square statistics are made with the aid of an IBM-650, for which the necessary programs have been written.

The results of the analysis of 16 actual experiments are given in Chapter IV, in each case for both the arc sine and the normal curve scaling procedures, including the chi-square statistics. In general it appears that, for the cases considered, if the data are such that the general model fits well, both procedures yield satisfactory results; otherwise neither procedure is satisfactory. Estimates of the correlation of the sum and difference in the data for each pair under normal curve scaling are listed. These indicate that there is only slight evidence that when the correlations are large the arc sine procedure may be preferable. To a greater extent there is evidence that more generally the preference, if any, would be in favor of the normal curve procedure.

In Chapter V a report is given on certain series of studies of hypothetical data obtained by modifying the live data of Chapter IV. The objective in these series is to determine if possible whether any features of the data can be said to have a bearing on the relative merits of the two scaling procedures. The principal properties of the data which are considered are a) internal consistency of the observations on the various pairs and b) conformity to the general model proposed in [2]. The results of these studies support the general conclusions of Chapter IV. They also provide some additional evidence to the effect that the normal curve procedure is the more likely to give good agreement between the observed and expected numbers.

VII. ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Dr. William A. Glenn for his guidance and advice. This paper is made possible only with his assistance.

The author wishes also to thank Dr. Rolf E. Bargmann for his invaluable IBM-650 subroutine for the normal curve and Dr. Herbert A. David for his helpful criticism and suggestions. Thanks are due also to who prepared the typewritten copies.

VIII. BIBLIOGRAPHY

[1] Dykstra, Otto, Jr., Associate Technologist, General Foods Research Center. (1959). Private communication.

[2] Glenn, W. A. and David, H. A. (1960). Ties in paired-comparison experiments using a modified Thurstone-Mosteller model. Biometrics 16, 86.

[3] Judd, Deane B., Chairman, OSA Committee on Uniform Color Scales, National Bureau of Standards. (1959). Private communication.

[4] 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.

[5] Mosteller, F. (1958). The mystery of the missing corpus. Psychometrika 23, 279.

[6] Thurstone, L. L. (1927). Psychophysical analysis. Amer. J. Psychol. 38, 368.

IX. APPENDIX

IBM-650 Programs

IBM-650 programs have been written to obtain the estimates τ^* and S_i^* for the parameters τ and S_i using each of the normal and arc sine scaling procedures. In these programs the range for t, the number of treatments to be compared, is $3 \leq t \leq 17$; and n, the number of observations, is the same for all pairs. Two separate program decks in seven-per-card form are on file at the Virginia Polytechnic Institute Computing Center under File Number 6.1.002.1 (VPI). Normal transformation is used in the N-deck and arc sine transformation in the AS-deck.

Input and output are in the same form in both cases, except that the correlation between $\tau'_{(ij)}$ and $(S_i'-S_j')$, as given by (2.7), is estimated for each pair in the normal deviates case. The estimate for each pair, denoted by $r[\tau'_{(ij)},(S_i'-S_j')]$, is obtained from (2.7) by replacing $\pi_{j.ij}$ by $1 - a_{ij}$, $\pi_{i.ij}$ by $1 - a_{ji}$, $f(\tau + S_i - S_j)$ by the ordinate at $F^{-1}(a_{ij})$ and $f(\tau - S_i + S_j)$ by the ordinate at $F^{-1}(a_{ji})$. There is, of course, no counterpart of this for the arc sine case, since the large-sample covariance terms are zero.

After the parameters have been estimated, the test of goodness of fit described in Section 3.2 is performed by each of the programs.

Use one data card for each of the $\binom{t}{2}$ pairs, the pairs (i,j) being arranged in such a way that $i < j$, i.e., $i = 1,\ldots,t-1$ and $j = 2,\ldots,t$.

Input Data

word 1 $n_{j.ij}$

word 2 $n_{o.ij}$

word 3 $n_{i.ij}$

word 4 $n = n_{j.ij} + n_{o.ij} + n_{i.ij}$

(words 1-4 use 00 0000 nnnn form)

word 5 zeros if the last treatment is involved in the pair.

non-zero otherwise.

word 6 00 00xx 0000, where xx = t.

zeros for the last pair.

word 7 identification

word 8 zeros

Output

All words are in xxxx.xxxxxx form, except that the n's are in input form. Words not mentioned are left blank.

<u>stage 1</u> -- $\binom{t}{2}$ cards

word 1 $n_{j.ij}$

word 2 $n_{o.ij}$

word 3 $n_{i.ij}$

word 4 n

word 5 identification

word 6 $r[\tau'_{(ij)},(S'_i-S'_j)]$. (*)

<u>stage 2</u> -- (t-1) cards

word 1 S^*_i, $i = 2,\ldots,t$.

word 5 identification

<u>stage 3</u> -- 1 card

word 1 τ^*

word 5 identification

<u>stage 4</u> -- $\binom{t}{2}$ cards

word 1 $X^2_{j.ij}$

word 2 $X^2_{o.ij}$

word 3 $X^2_{i.ij}$

word 4 $X^2_{j.ij}+X^2_{o.ij}+X^2_{i.ij}$

word 5 identification

<u>stage 5</u> -- 1 card

word 1 $\sum_j X^2_{j.ij}$

(*) for N-deck only; zeros for AS-deck.

word 2 $\sum_{o} X^2_{o.ij}$

word 3 $\sum_{i} X^2_{i.ij}$

word 4 $\sum_{j} X^2_{j.ij} + \sum_{o} X^2_{o.ij} + \sum_{i} X^2_{i.ij}$

word 5 identification

<u>Control Panel</u>

Use 80-80 board.

<u>Input Deck</u>

Place the input data after the program deck as follows:

<u>program deck</u>

<u>input data 1</u>

<u>input data 2</u>

(*) ---

<u>Operating Notes</u>

(1) Set 70 1951 9000 in storage entry switches.

(2) Set switch to RUN for:

(a) half-cycle

(b) control

(3) Set switch to STOP for:

(a) programmed

(b) overflow

(c) error

(*) The number of problems is unlimited.

(4) Set display switch to DISTRIBUTOR.

(5) Place input deck into 533.

(6) Place blank cards into punch unit, (t^2+1) for each problem.

(7) Press the computer reset key.

(8) Press program start key, read start key and punch start key.

(9) When read unit empties, press end of file key.

If more problems are to be run, reset 00 0000 1350 in storage entry switches and start as before.

The vita has been removed from the scanned document

ABSTRACT

A recently-proposed modification of the Thurstone-Mosteller method of paired comparisons makes possible the analysis of data involving tied observations. The modification includes the postulating of an angular response law such that the response proportions are scaled with arc sine transforms instead of with normal deviates.

In this paper a comparison is made of the arc sine and normal curve scaling procedures in paired comparisons involving ties. This is done by applying both methods to data from two important fields of application. Comparisons are also made on several series of hypothetical data. The criterion of comparison is the goodness of fit between the observations and the expected numbers computed from the solution, as measured by means of a chi-square statistic. Computations of parameter estimates and chi-square statistics are made with the aid of an IBM-650, for which the necessary programs have been written.

It is concluded that for data conforming well to the model as proposed, both scaling procedures tend to give results in satisfactory agreement with the observations. There is some evidence that, for the cases considered, the preference, if any, is for the normal curve procedure.