
DEPARTMENT OF STATISTICS

The University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT NO. 169

July 1968

ANALYZING PAIRED COMPARISON TESTS

By

Norman R. Draper
William G. Hunter
and
David E. Tierney

The University of Wisconsin at Madison

ANALYZING PAIRED COMPARISON TESTS

by

Norman R. Draper, William G. Hunter and David E. Tierney
The University of Wisconsin
Madison, Wisconsin 53706

SUMMARY

A new method for handling data of the "prefer A," "prefer B," or "no preference" type is discussed in this article. An informative graphical analysis is illustrated with four examples.

1. THE PROBLEM

Companies that manufacture consumer products often wish to compare the consumer appeal or the technical merit of two competing products, or two competing varieties of the same product, by a direct comparison. For example, the two competitors A and B are often given to a judge who is asked whether he prefers A, prefers B, or has no preference. Data of the following form, then, are obtained from a number of respondents:

Respondents N	Prefer A X	Prefer B Y	No Preference N-X-Y
100	64	16	20

What inferences can be made from data of this type? We note that A is preferred by more people in this sample and that the proportion of respondents with no preference is 0.20. What assurance can we have that, in the total population, more people will in fact prefer A to B?

How likely is it that A and B will be equally preferred, or even that B will be preferred to A, in the population as a whole? Let p, q be the probabilities that, in the population as a whole, A, B respectively are preferred. The observed preference ratios are $X/N = 64/100 = 0.64$ for A, and $Y/N = 16/100 = 0.16$ for B. What do these figures tell us about the probabilities p and q ? For example, is a value of p as low as 0.6 reasonable? Or 0.5? Is a value of p as high as 0.7 reasonable? Or 0.8?

A difficulty that arises in analyzing data of this type is the proper handling of the "no preference" response figures. Odesky (1967) recently discussed this type of difficulty in another context as follows:

"Which is the better method for paired comparison preference mail questionnaires: to force a choice by minimizing the no preference response opportunity or to allow an unrestricted vote? And how should no preference responses be handled in the analysis?

"These questions have always concerned product test researchers; because their answers depend somewhat on use of results (product versus market evaluation), several opinions exist (1).*

"Some researchers believe that an unqualified no preference response opportunity is a more realistic evaluation because test products are usually similar. On the other hand, a tester forced to make a random choice between two products when she actually has no preference is forced into a bias because other testers who do have a preference may not be divided on a 50-50 basis.

"Others feel that a no preference response opportunity is the easy way out for a tester and that forcing a preference is the only approach. Others say that an expression of degree

*Hansen and Hurwitz (1946).

of preference is the most meaningful response (2)** And still others recommend determining the direction of no preference responses--that is, does the respondent like both products the same or neither of them?

"If no preference responses are allocated among test results, the manner of allocation is important, particularly if most testers refuse a positive selection. This problem confronts the researcher when he tries to evaluate no preference responses and the total test score. He may decide to divide them equally on grounds that no preference means a purchaser has a 50 percent chance of buying either product if he prefers each equally. He may decide to disregard such responses altogether on the assumption that they mean what they say. Or he may decide to divide them proportionately between the preferences, rationalizing that no preference voters really have a preference proportionate to those stating a preference."

2. SOLUTION

The present authors have given a method for the proper handling of this type of data to take full account of the no preference results (see Draper, Hunter and Tierney, 1968). A Bayesian analysis was employed to get an exact joint posterior distribution for p = the probability that A is preferred and q = the probability that B is preferred. Drawing contours of this distribution is difficult. However, by first applying a transformation, then making a bivariate normal approximation, and then translating these contours back into the (p, q) space, an approximate but (we have found from experience) highly reliable representation of the true contours can be obtained.

** Eastlack (1964).

The results of this calculation for the data $(N, X, Y) = (100, 64, 16)$ are shown in Figure 1. The two closed contours are the 95% and 99% "translated approximate contours," the 95% contour being the inner one. We can say that, approximately, the points within the 95% contour have coordinates (p, q) which are considered reasonable in the light of the data at the 95% level of probability; similar remarks apply to the 99% contour. A "point estimate" of the pair (p, q) is given by the values

$$(\hat{p}, \hat{q}) = (X/N, Y/N)$$

if, before the sample is taken, there is no "prior information" that certain values of p and q are any more likely than other values, i. e., if the prior information is uniform over the possible range of values of p and q . This point $(p, q) = (0.64, 0.16)$ is marked with a star on the Figure.

The interpretation of Figure 1 is quite straightforward. To explain the interpretation, it is helpful to look at four diagrams based on the same sample size of $N = 100$ but with different X and Y values.

Specifically, we examine the following cases:

Case	N	X	Y	N-X-Y	Figure No.
1	100	64	16	20	1
2	100	32	8	60	2
3	100	40	40	20	3
4	100	20	20	60	4

Note that in each case we obtain $p = X/100$, $q = Y/100$ so that a mental two-place shift of a decimal point in X and Y gives p and q immediately.

The dashed 45 degree line on each figure divides the triangle into the "A-preferred" region (on the lower right) and the "B-preferred"

region (on the upper left). Thus, for example, if the whole of the posterior contours is entirely within one of these two regions, high preference is indicated accordingly. Thus Figures 1 and 2 indicate decisive A preference. In general, both the position of (p, q) and the position of the contours need be taken into account. Figures 3 and 4 indicate no decisive preference between A and B, since (p, q) lies on the dividing line and the contours straddle the line in each case.

The preponderance of "no preference" responses is indicated by the distance of the contours from the diagonal boundary. Thus we see in Figures 1 and 3 (both 20% "no preference" results) that the contours are close to the boundary while in Figures 2 and 4 (both 60% "no preference" results) the contours are far from the boundary.

Thus we see, by our analysis, how the "no preference" results can be taken into account in a natural way and allowed to throw light on the questions at issue. For example, in Section 1 we asked what values of p might be reasonable on the basis of the first set of data. It is immediately obvious by examining the contours in Figure 1 that, at the 95% probability level, values of p between 0.52 and 0.73, approximately, are reasonable. Corresponding reasonable values of q can be read from the figure for each value of p in the range given.

4. THE TWO MARGINAL DISTRIBUTIONS

We have now finished describing our solution to the "no-preference" problem. It consists of calculating, plotting, and examining the joint posterior distribution for p and q . The main part of the paper is thus complete. Additional insight can be gained by a study of two

marginal distributions. In all four figures, the two marginal distributions are shown erected on the diagonal border of the (p, q) region, and the scale on that border applies to them.

The solid curve is obtained by integrating the (exact) joint distribution of p and q , given by equation (2.3) of our previous paper (1968), over the lines $p - q = \text{constant}$.

We recall that one of the suggested methods for handling the $N - X - Y$ "no preference" results was to pretend that half of them had instead been "prefer A" and that half of them had instead been "prefer B." The solid distribution provides us with the nearest equivalent of this procedure in our formulation.

The dashed curve is obtained by integrating the (exact) joint distribution of p and q , given by equation (2.3) of our previous paper (1968), over the lines $p/q = \text{constant}$.

We recall that another of the suggested methods for handling the $N - X - Y$ "no preference" results was to pretend that a proportion $X/(X + Y)$ of them had instead been "prefer A" and a proportion $Y/(X + Y)$ of them had instead been "prefer B." The dashed distribution provides us with the nearest equivalent of this procedure in our formulation.

Note that both marginals give more information than their equivalent ad hoc procedures, since we can examine a complete distribution and not simply a point estimate. Thus we obtain immediately a clear idea of the precision of our information on p . (It might be argued that the ad hoc procedures can also provide a measure of precision since a confidence interval for p , whose length depends on $1/\sqrt{N}$, can be calculated.

This, however, raises an additional drawback of the ad hoc methods. If the "no preference" results were ignored, an appropriate confidence interval for p would have a length proportional to $1/\sqrt{X + Y}$ instead of $1/\sqrt{N}$. Thus the more "no preferences" we obtain, the greater is the apparent precision. Of course, this is completely opposite to what common sense would suggest. Our method, on the other hand, gives answers which agree with common sense.)

In considering the marginal distributions, we see that the inferences to be made from them depend greatly both on the proportion of "no preference" results and the method of averaging out over them. Clearly, if the X, Y values already indicate a high preference for A , dividing the no preference results in the ratio $X:Y$ will even further accentuate the choice of A , compared with a 50:50 division. This is shown in Figs. 1 and 2. Note that the more "no preference" responses there are, the greater will be the difference between the locations of the two marginal distributions. The relative spreads are also influenced, as we can see in Figures 1 and 2. Comparison of Figures 3 and 4 brings this out even more strongly; here we see that the two marginals have the same mean $p = 0.50$, but that the more "no preference" responses exist, the more is the "X:Y marginal" spread out in relation to the "50:50 marginal." As remarked earlier, this behavior agrees with common sense.

A third alternative mentioned by Odesky (1967), and quoted earlier, is to disregard the "no preferences" altogether. The equivalent procedure, in our Bayesian framework, is to look at only a slice (defined by $p + q = (X+Y)/N$) of the joint distribution, rather than the whole of it. We do not subscribe to such a procedure, however, because it clearly throws away information, i. e., the conditional distribution provides us with less information than the joint distribution.

5. RECOMMENDATION

Our recommendation for the analysis of data of the "prefer A," "prefer B," "no preference" type is thus as follows: Plot the translated approximate Bayesian contours, see where they lie, and make inferences accordingly. If desired, one can also plot the two marginal distributions we have described and carefully examine the differences in them.

If the contours or the marginals cover extremely wide ranges, this is a sign that the data do not provide a sufficiently precise basis for decision. (In general, the larger the value of N , the tighter will the distributions be.) Therefore, in some cases, the appearance of the various plots may suggest that the collection of additional data is warranted.

6. THE COMPUTER PROGRAM

Figures 1 through 4 were all obtained by means of a Fortran 63 computer program written by the authors. Use of the program is extremely simple, requiring only the addition of a single punched card, containing the values of N , X , and Y , for each set of data. The average execution time for the figures given here was 2.2 minutes per figure on a CDC 1604 computer.

REFERENCES

1. Draper, N. R., Hunter, W. G., and Tierney, D. E. (1968). Which product is better? University of Wisconsin Statistics Dept. Technical Report No. 161.
2. Eastlack, J. O., Jr. (1964). Consumer flavor preference factors in food product design. Journal of Marketing Research, 1, (February), 38-42.

3. Hansen, M. H., and Hurwitz, W. N. (1946). The problem of non-response in sample surveys, Journal of the American Statistical Association, 41, (December), 517-29.
4. Odesky, S. H. (1967). Handling the neutral vote in paired comparison product testing, Journal of Marketing Research, 4, (May), 199-201.

Figure 1. Posterior densities for
 $N=100$, $X=64$, and $Y=16$.







