

UNCLASSIFIED

AD

401 405

*Reproduced
by the*

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

63-3-2

PRL-TDR-63-1

CATALOGED BY ASTIA
AS AD No. 401405

Chance Distribution of Inconsistent Response Patterns in Paired Comparison and Multiple Ranking Designs

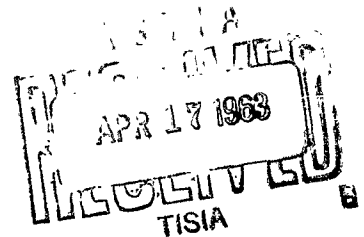
By

Charles E. Zimmer

401 405

Technical Documentary Report PRL-TDR-63-1

January 1963



6570TH PERSONNEL RESEARCH LABORATORY
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
Lackland Air Force Base, Texas

Project 7719, Task 771901

N O T I C E

Copies of this document may be purchased from the Office of Technical Services, US Department of Commerce.

Qualified requestors may obtain copies from ASTIA. Orders will be expedited if placed through the librarian or other person designated to request documents from ASTIA.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

PRL-TDR-63-1

**CHANCE DISTRIBUTION OF INCONSISTENT RESPONSE PATTERNS
IN PAIRED COMPARISON AND MULTIPLE RANKING DESIGNS**

**By
Charles E. Zimmer**

**Technical Documentary Report PRL-TDR-63-1
January 1963**

**6570TH PERSONNEL RESEARCH LABORATORY
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
Lackland Air Force Base, Texas**

Project 7719, Task 771901

ABSTRACT

Comparative judgments are used in developing scales for various personnel and occupational criteria. In scaling data from paired comparisons, frequency of inconsistent responses is crucial. To determine whether information from the simpler and more economical multiple ranking design can be evaluated by the same techniques as for a complete paired comparison design, computer programs were adapted whereby the full population of possible response patterns could be randomly sampled to determine the chance distribution of inconsistent responses for both designs. Results for the 1000 randomly selected patterns showed that the multiple rank order design restricts the possible number of response patterns and reduces the frequency of inconsistent patterns. The distributions were so different that techniques devised for testing significance of extreme frequencies for data from the classic paired comparison design are inappropriate for evaluating extreme occurrences in multiple ranking data. Since the multiple ranking distribution approximates the normal distribution, it would be suitable to evaluate empirical data by comparison with the parameters here determined for the random sample of the full population of response patterns.

This report has been reviewed and is approved.

Fred E. Holdrege, Col, USAF
Commander

A. Carp
Technical Director

Hq 6570th Personnel Research Laboratory

CHANCE DISTRIBUTION OF INCONSISTENT RESPONSE PATTERNS IN PAIRED COMPARISON AND MULTIPLE RANKING DESIGNS

1. INTRODUCTION

The method of paired comparisons has long enjoyed an honored reputation in the field of psychological measurement. It is particularly useful in developing scales for subjective observations where direct quantitative measurements are not available. The method has been widely used in investigating sensory discrimination and establishing preference scales. In personnel research it has proved useful in developing interest and activity inventories and in evaluating occupations and job components. Furthermore, fairly precise techniques are available for evaluating the significance of paired comparison data in relation to chance expectation. The method suffers from the serious difficulty of becoming completely unwieldy for the subject when dealing with more than a dozen or so stimulus objects. A comparable unwieldiness for the experimenter occurs in the attempt to evaluate judgments based on more than a minimal number of stimulus objects. Treatment of 31 stimulus objects (a convenient number for multiple ranking designs) requires that the subject perform a total of 465 judgments, a rather large demand for the average experimental situation. The number of unique response patterns that may arise from these 465 pairs is in excess of 9.5×10^{139} ; evaluation of the significance of any given response configuration presents a task of no mean complexity.

The multiple rank order designs proposed by Gulliksen & Tucker (1961) provide a convenient means of presenting larger numbers of stimulus objects in a format that considerably eases the subject's task. The problem still remains, however, that as the number of stimulus objects is increased, the task of evaluating a specific response pattern becomes alarmingly more complex.

Of particular interest are the internally inconsistent response patterns and the resultant circular triads, i.e., intransitive loops involving three stimulus objects wherein stimulus A is judged to be greater than B, stimulus B is judged greater than C, but C is judged greater than A. In theory, inconsistent judgments may arise when three or more stimulus objects are perceived as being indetical in respect to the quality under investigation. Neither the classic paired comparison method nor the multiple ranking variations permit direct expressions of "equal." Inconsistencies are thus the only means available to the subject for expressing conditions of equality. (Note that without replication of pairings, the method is insensitive to situations wherein two stimuli are regarded as equal; such a condition can normally be detected only if it holds for a high percentage of the population.) In practice, inconsistent judgments may also be regarded either as an indication that the subject is operating without benefit of a well-defined criterion, or that he is just plain careless. The two phenomena are not necessarily independent.

Exact probabilities for the chance expectancy of circular triads arising in the complete paired comparison method have been determined for problems involving as many as seven stimulus objects (Kendall, 1948). Slater (1961) has similarly tabled the exact probabilities attached to inconsistent responses for cases involving from two to eight stimuli. It would be difficult to justify the expenditure of time and effort that would be required to extend such tables to include even a few additional stimulus objects. The chance expectancy of circular triads arising from situations involving eight or more stimuli can, however, be estimated from a chi-square approximation given by Kendall.

It should be noted that these evaluating schemes have been derived from the complete paired comparison method. Inherent in the multiple rank ordering designs are certain response restrictions which cast considerable cloudiness on the picture. A multiple ranking design involving 31 stimulus objects (a balanced block scheme of 31 items of 6 stimuli each) permits but 3.8×10^{88} possible unique response patterns, several million trillions less than are possible when the 31 stimulus

objects are presented in all possible pairings. One obvious difference between the two designs is that it is impossible to give inconsistent responses within a rank ordering presentation, and thus there can be no triadic loops within the combinations of stimuli which are grouped in a single item. It appears that the restrictive nature of the balanced block design tends to lessen the chance formation of circular triads. In consequence, correspondingly greater significance must be attached to the event.

No adequate investigation has been made as to the extent by which the multiple ranking design differs from the complete paired comparison method in regard to the chance expectation of triadic loops. Development of a suitable chi-square formula for characterizing the chance distribution of circular triads has been hampered by the complexities induced by the response restrictions, a situation made more difficult by the fact that the nature of the constraints is a function of the particular idiosyncracies of a given design. Since the advantages of the multiple ranking designs accrue only when a dozen or more stimulus objects are involved, a precise definition of the occurrence of triadic loops by means of an analysis of all possible response patterns is prohibitive. In view of the successes that have been realized through application of Monte Carlo techniques to other probabilistic situations of comparable complexity, it appears that a random sampling of response configurations might shed considerable light on the relationship between the complete paired comparison method and the multiple ranking designs.

2. PROCEDURE

Only certain numbers of stimuli lend themselves to the balanced designs appropriate for multiple rank ordering procedures. This investigation concerns a design involving 31 stimulus objects; the multiple ranking format consists of 31 items of 6 stimuli each, balanced so that each stimulus object occurs once and only once with every other stimulus object. The specific design was chosen because of the availability of an IBM 650 Tape RAMAC program for scoring and computing summary information on "6-31" data. A similar program developed by Gulliksen & Tucker for handling the 6-31 design is available from the IBM Program Library. The popularity of this specific design arises from the fact that it is the largest of the multiple ranking designs that may conveniently be handled by present-day medium-sized computers.

Because of the forced nature of the judgments in the full paired comparison model, the response to a given pair of stimulus objects may be considered as a binary decision; in the multiple ranking format, the response to an item or block consists of a permutation of the digits 1 through n , where n represents the number of stimuli in the item or block.

Artificial random data for the 6-31 design were obtained through random selection from the 720 permutations of the digits 1 through 6. Each permutation group corresponds to a rank ordering of the six stimuli which comprise a single item. Thirty-one such permutation groups selected at random comprise a complete response configuration for one "subject." Random response patterns were generated for 1000 "cases"; these were processed by the program normally used to handle data obtained from conventional testing situations.

Data for the classic paired comparison method consisted of generating a string of 465 random digits reduced to a binary pattern by converting the digits 5 through 9 to 1 and the remaining digits to zero. The random binary string thus corresponded to 465 decisions involved in all possible pairings of 31 stimulus objects. One thousand such "cases" were generated and scored by a modified version of the program for scoring the incomplete balanced block design.¹

¹ The author is indebted to Dr. John Merck and Miss Kathleen Davis for providing the basic program for generating pseudo-random numbers. The pseudo-random number generating routine has a known periodicity; the output requirements for these two samples were considerably less than one full cycle. Although the generating scheme has been subjected to all the standard tests for randomization, additional checks were performed on the specific output for these two samples. No evidence was found for questioning the adequacy of randomization.

TABLE 1. Distributions of Circular Triads Obtained From Generating Random Response Patterns for 31 Stimulus Objects

	Complete paired comparison design	"6-31" multiple ranking design
N	1000	1000
M	1123.03	969.12
σ	29.27	60.33
Range	1018-1202	716-1106

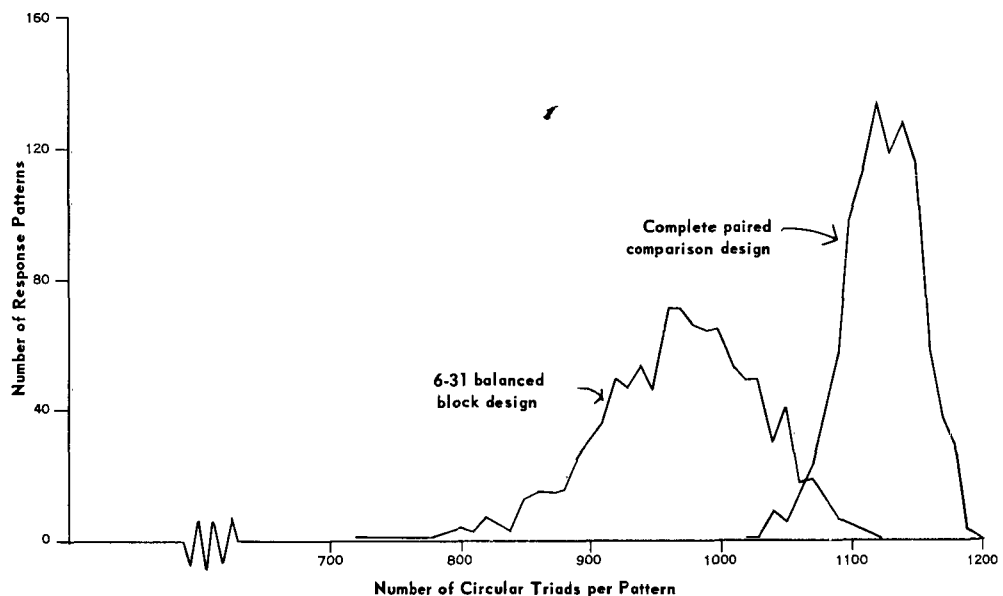


Fig. 1. Distributions of the occurrence of circular triads for multiple rank order and complete paired comparison designs.

3. RESULTS

Samples of 1000 represent a truly infinitesimal fraction of the total population of unique response configurations. Nevertheless the tabled summary data of Table 1 and the accompanying graph (Fig. 1) give indication that reasonably well-defined distribution functions have been obtained. The two distributions differ significantly in respect to their means as well as their standard deviations (well beyond the .001 level). The mean and sigma for the complete paired comparison sample

are in close agreement with the values we derived from a "backward" application of Kendall's chi-square approximation ($M = 1125.94$, $\sigma = 28.83$).

Maximum inconsistency in the judgments involving 31 stimuli will result in the formation of 1240 triadic loops; this is true for both the classic paired comparison method and the balanced incomplete block design. A response pattern that is entirely consistent within itself will, of course, yield no circular triads. In the case of the complete paired comparison method, it can be shown that the distribution of circular triads is continuous throughout the entire range of 0 through 1240.²

The distribution representing the multiple rank ordering design displays a certain amount of unevenness that cannot readily be explained at this time. Insufficient sampling is, of course, the most obvious explanation. In view of the fact that a comparable unevenness is not apparent in the complete paired comparison sample, it does not appear entirely justified to explain away the irregularities in the multiple ranking model in terms of inadequate sampling. There is no reason to assume that the "true" curve is necessarily smooth and regular in shape. As mentioned above the response restrictions imposed by the multiple ranking design bar the formation of triadic loops among the combinations of stimuli that occur together in a single item. The net effect is an obviously significant decrease in the overall expectancy of such loops. It is not unreasonable to postulate that these restrictions might also tend to inhibit the chance occurrence of certain numbers of circular triads, thus creating troughs in the curve.

Nonetheless the distribution curve obtained from the multiple ranking data appears sufficiently well defined to permit considerable generalization. In view of the observed highly significant differences between the two methods, it would seem inappropriate to apply Kendall's chi-square test to the evaluation of extreme cases arising from the multiple ranking design. Rather, it is suggested that a far more realistic evaluation can be obtained by use of the mean and sigma obtained from this random distribution.

Random sampling of this type cannot resolve the question as to whether the two designs yield stimulus scale values of comparable magnitude. By the very nature of randomization, summations over all cases serves to balance out the individual response restrictions inherent in the multiple ranking design. No significant difference was found between the two samples in respect to the distributions of normal deviate scale values. This is not to say, however, that the two designs would necessarily yield comparable results in a meaningful testing situation. The extent to which the multiple ranking design differs from the complete paired comparison method is undoubtedly a function of the specific stimuli involved, or more precisely, of the particular grouping of stimuli into blocks for ranking.

4. CONCLUSIONS

Random samples of 1000 response patterns were generated in order to compare certain aspects of the complete paired comparison method with the multiple ranking design. The response restrictions inherent in the multiple ranking design were found to impose a significant reduction in the chance occurrence of circular triads; a highly significant increase in variance was also found. It thus appears most inappropriate to evaluate the significance of triadic loops arising from the multiple rank order testing situation by use of Kendall's chi-square and related techniques derived from the complete paired comparison method. The distribution of circular triads for the

²The writer believes that this situation likewise holds for balanced incomplete block designs but has thus far been unable to provide a proof.

"6-31" balanced block design does not appear to be appreciably influenced by the imposition of an end-point on the continuum. Approximation of the normal curve is sufficiently close to suggest that extreme occurrences of circular triads may be evaluated against a pure chance distribution with a mean of 969 and a sigma of 60. Although derived from a comparatively small sample of the total population of response patterns, these values are believed to approximate the "true" characteristics of the distribution sufficiently well to serve as a meaningful framework within which the occurrence of triadic events may be evaluated. Application of an appreciably larger random sample would enable more precise definition of these values.

REFERENCES

- Gulliksen, H. & Tucker, L.R. A general procedure for obtaining paired comparisons from multiple rank orders. *Psychometrika*, 1961, **26**, 173-184.
- Gulliksen, H. & Tucker, L.R. Paired comparisons from balanced incomplete blocks. IBM 650 Program Library, File Number 6.0.038, undated.
- Kendall, M.G. *Rank correlation methods*. London: Charles Griffin, 1948.
- Slater, P. Inconsistencies in a schedule of paired comparisons. *Biometrika*, 1961, **48**, 303-312.

6570th Personnel Research Laboratory (AMD). Lackland AF Base, Tex. Rpt No. PRL-TDR-63-1. CHANCE DISTRIBUTION OF INCONSISTENT RESPONSE PATTERNS IN PAIRED COMPARISON AND MULTIPLE RANKING DESIGNS. Jan 63, iii + 5 p. incl illus. Unclassified Report Comparative judgments are used in developing scales for various personnel and occupational criteria. In scaling data from paired comparisons, frequency of inconsistent responses is crucial. To determine whether information from the simpler and more economical multiple ranking design can be evaluated by the same techniques as for a complete paired comparison design, computer programs were adapted whereby	1 Psychometrics 2 Statistical functions 3 Mathematical computer data 4 Probability 5 Test methods 6 Programming I AFSC Project(Task) 7719(01) II C.E. Zimmer III Avd fr OTS IV In ASTIA collection
the full population of possible response patterns could be randomly sampled to determine the chance distribution of inconsistent responses for both designs. Results for the 1000 randomly selected patterns showed that the multiple rank order design restricts the possible number of response patterns and reduces the frequency of inconsistent patterns. The distributions were so different that techniques devised for testing significance of extreme frequencies for data from the classic paired comparison design are inappropriate for evaluating extreme occurrences in multiple ranking data. Since the multiple ranking distribution approximates the normal distribution, it would be suitable to evaluate empirical data by comparison with the parameters here determined for the random sample of the full population of response patterns.	the full population of possible response patterns could be randomly sampled to determine the chance distribution of inconsistent responses for both designs. Results for the 1000 randomly selected patterns showed that the multiple rank order design restricts the possible number of response patterns and reduces the frequency of inconsistent patterns. The distributions were so different that techniques devised for testing significance of extreme frequencies for data from the classic paired comparison design are inappropriate for evaluating extreme occurrences in multiple ranking data. Since the multiple ranking distribution approximates the normal distribution, it would be suitable to evaluate empirical data by comparison with the parameters here determined for the random sample of the full population of response patterns.