On the Method of Paired Comparisons

Author(s): M. G. Kendall and B. Babington Smith

Source: *Biometrika*, Mar., 1940, Vol. 31, No. 3/4 (Mar., 1940), pp. 324-345
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2332613

ON THE METHOD OF PAIRED COMPARISONS

By M. G. KENDALL and B. BABINGTON SMITH

Introduction

1. Suppose we have a number of objects A, B, C, etc. which are to be considered according to the different degrees in which they exhibit some common quality. If the quality is measurable in some objective way the objects will yield a number of variate values, in which case the problem is amenable to treatment by well-understood methods. It may, however, happen either for theoretical or for practical reasons that the quality is not measurable. We then have to rely for a discussion of the variation of the quality on judgments of a more or less subjective kind carried out after a comparison of the objects among themselves.

One of the methods of comparison which has been widely used in this connexion is that of ranking. An observer examines the objects and arranges them in the order in which he judges them to possess the quality under consideration. This arrangement is called a ranking, and when two or more observers provide rankings of the same set of objects there arise the familiar questions of the type: is there any significant resemblance between the judgments of observers? or, do the data furnish any evidence that the objects have a "real" objective ranking?

2. The ranking method suffers from a serious drawback when the quality considered is not known with certainty to be representable by a linear variable. We may, for instance, ask an observer to rank a number of individuals in order of intelligence, and he may comply with the request in the full belief that he is doing something within his powers; but if intelligence is not measurable on a linear scale this ranking may fail to give a real picture either of the observer's preference or of the variation of intelligence among the individuals. It is not impossible that the observer should judge A more intelligent than B, B than C, and C than A, if the individuals are presented for his consideration one pair at a time. The likelihood of this happening is obviously increased when we are dealing with tastes in music, eatables or film stars; and in practice the event is not uncommon. Such "inconsistent" preferences can never appear in ranking, for if A is preferred to B and B to C, then A must automatically be shown as preferred to C. The use of ranking thus destroys what may be valuable information about preferences.

3. In this paper we consider a more general method of investigating preferences. With n objects, we shall suppose that each of the $\binom{n}{2}$ possible pairs is presented to an observer and his preference of one member of the pair noted.

We assume that a choice between two objects can always be made.* With m observers the data then comprise $m\binom{n}{2}$ preferences. The questions to be discussed include:

(a) Is there any evidence that a particular observer is capable of forming a reliable judgment of the quality under investigation; and if not, is the fault his, or is it due to the fact that he has been asked to perform an impossible task?

(b) Is there any significant concordance of preferences between observers?

(c) Can the quality under discussion be represented by a linear variable?

4. The method of offering for judgment objects two at a time is known as the method of paired comparisons. Hitherto it has been used mainly in human psychology, but it has some interesting applications in animal experimentation. For instance, in feeding experiments it is impossible to get an animal to rank a number of foods in order of preference but it is not difficult to offer pairs of foods and to note which is taken first. Experiments of this kind have, of course, to be conducted with great care to ensure that conditions operating when the different pairs are offered are as constant as possible; but the difficulties are far from being insuperable and the method of paired comparisons offers a useful technique in cases where the more usual procedures cannot be applied. From the point of view of theoretical statistics perhaps the most interesting part of the present work is that it offers some lines of approach to the difficult question whether a given quality can be legitimately regarded as based on a linear variable, i.e. whether ranking or scoring methods are justifiable or not.

CONSISTENCE IN PREFERENCES

5. If the object A is preferred to B we write $A \rightarrow B$ or $B \leftarrow A$. The $\binom{n}{2}$ preferences of a single observer may be represented in tabular form as shown in Table I.

In this table, which is shown for the six objects A to F, an entry of unity in column Y and row X means $X \rightarrow Y$, and is thus accompanied by a complementary zero in row Y and column X. The diagonals are blocked out. For example, in Table I, $A \rightarrow B$, $A \rightarrow C$, $D \rightarrow A$, etc.

The arrangement of the objects A to F in the row and column headings is quite arbitrary. There are $(n!)^2$ ways of representing the same configuration of preferences in such a table according to the permutations of objects in row and

* That is, we exclude cases in which an observer cannot make up his mind which object he prefers, just as in the ranking case one excludes the possibility of split ranks. In practice it sometimes happens that an observer is genuinely unable to reach a decision. To allow for this fact in the theoretical discussions would introduce complications of a most intractable kind. When the effect becomes important in practice it can be allowed for by selecting the set of preferences which are most unfavourable to the hypothesis under test.

TABLE I

	A	B	C	D	E	F
A	—	1	1	0	1	1
B	0	—	0	1	1	0
C	0	1	—	1	1	1
D	1	0	0	—	0	0
E	0	0	0	1	—	1
F	0	1	0	1	0	—

column; but in practice it is generally desirable to have the order in row and column the same, and even among the $n!$ possible arrangements so given there are often practical considerations which determine one order as more convenient than others.

6. Paired comparisons may also be represented geometrically by a method which can be illustrated for the case of the six objects as follows:

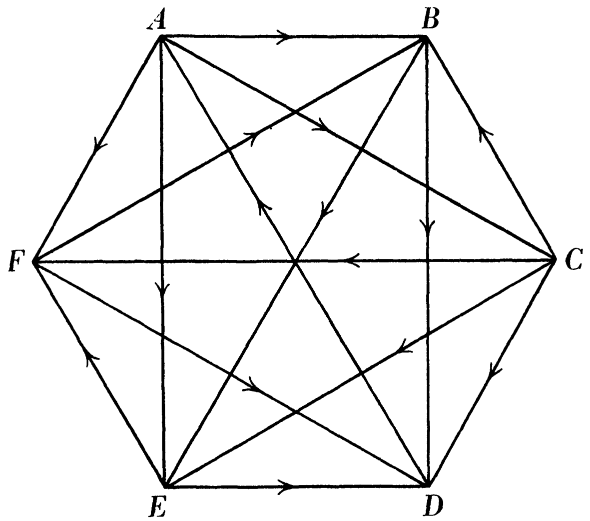

Fig. 1. Geometrical representations of the scheme of preferences of Table I.

We represent the six objects A to F by the six vertices of a regular hexagon and join the vertices in all possible ways by straight lines. If $A \rightarrow B$ we draw an arrow on the line AB pointing from A to B. The arrows shown on Fig. 1 correspond to the preferences shown in Table I.*

7. If an observer makes preferences of type $A \to B \to C \to A$ we say that the triad ABC is inconsistent. In the geometrical representation an inconsistent triad is shown by a triangle in which all the arrows go round in the same direction. We may thus speak of a "circular" triad of preferences. In Fig. 1 the triads ACD, BEF and three others are circular.

It is also possible to have inconsistent triads of greater extent; but any such circuit must contain at least two circular triads. Suppose, for instance, that $ABCD$ is circular, e.g. that $A \to B \to C \to D \to A$. Then either $A \to C$ or $C \to A$. In the first case ACD is circular, in the second ABC. Similarly either ABD or BCD is circular. Thus the circular tetrad must contain just two circular triads. On the other hand it is possible for a tetrad to contain circular triads without being itself circular.

Similarly, if $ABCDE$ is circular either ABC or $ACDE$ is circular and either BCD or $BDEA$ is circular. If the two tetrads are circular there must be at least three circular triads (not necessarily four, because ADE may be common to both). It is easy to see by an actual example based on this configuration that there need not be more than three circular triads; and it is clear that there must be at least three. For if the tetrads are not circular then ABC and BCD must be so and then either CDE is circular or $ABCE$ is so, adding at least one more.

Generally, it appears that a circular n-ad must contain at least $(n-2)$ circular triads; but it may contain more, and the fact that an n-ad contains $(n-2)$ circular triads does not mean that it is itself circular. In discussing inconsistences, therefore, it seems best to confine attention to circular triads, which, so to speak, constitute the inconsistent elements of the configuration, and to ignore the more ambiguous criteria associated with circular polyads of greater extent.

8. We now prove the following theorems:

(1) The maximum possible number of circular triads is $(n^3-n)/24$ if n is odd and $(n^3-4n)/24$ if n is even; and the minimum number is zero.

(2) These limits can always be attained by some configuration of preferences.

(3) For any integral number between the maximum and the minimum there exists at least one preference-configuration with that number of circular triads; and in general there will be more than one.

Consider a polygon of the type shown in Fig. 1 with n vertices. There will

* These preferences were obtained in an experiment on a dog, which was offered the following foods in pairs: meat, biscuit, chocolate, apple, pear and cheese. The members of a pair were cut to the same size and placed equidistantly from the dog, which was then released and allowed to choose. All the pieces of food were eaten avidly, it being that sort of dog, but there were considerable inconsistences in choice. We do not offer these data as more than an illustration of the method.

be $(n-1)$ lines emanating from each vertex. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the number of lines at the respective vertices on which the arrows *leave* the vertex.

Then
$$\underset{r=1}{\overset{n}{S}} (\alpha_r) = \binom{n}{2}$$
and the mean value of α_r is $(n-1)/2$.

Define
$$T = \underset{r=1}{\overset{n}{S}} \left(\alpha_r - \frac{n-1}{2}\right)^2$$
$$= S(\alpha_r^2) - \frac{n(n-1)^2}{4}. \qquad \ldots\ldots(1)$$

We now show that if the direction of a preference is altered and the effect is to increase the number of circular triads by d, T is reduced by $2d$; and conversely. Consider the preference $A \to B$. The only triads affected by altering this to $B \to A$ are those containing the line AB. Suppose there are α preferences of type $A \to X$ (including $A \to B$) and β preferences of type $B \to X$. Then four possible types of triad arise:

$A \to X \leftarrow B$, say p in number
$A \leftarrow X \to B$,
$A \to X \to B$, which must number $\alpha - p - 1$
$A \leftarrow X \leftarrow B$, ,, ,, $\beta - p$.

When the preference $A \to B$ is reversed the first two remain non-circular. The third becomes circular, the fourth ceases to be so. The reduction in the value of T is
$$\alpha^2 - (\alpha-1)^2 + \beta^2 - (\beta+1)^2$$
$$= 2(\alpha - \beta - 1)$$
$$= 2d, \text{ say.}$$
The increase in the number of circular triads is
$$(\alpha - p - 1) - (\beta - p) = \alpha - \beta - 1$$
$$= d.$$

More generally, if as the result of reversing any number of preferences T is decreased by $2d$, then d must be an integer and the number of circular triads must be increased by d. This clearly follows from the previous results for the reversal of preferences can take place one at a time and the effect on T and the number of circular triads is cumulative.

We now investigate the maximum and minimum values of T. It is clear from the definition that T is greatest when the α's are the natural numbers $1, 2, \ldots, n$; and this is a possible case because it corresponds to ordinary ranking. Hence max. $(T) = (n^3 - n)/12$.

For the minimum value, consider the polygon $A_1, A_2, \ldots, A_n$. Set up the preferences $A_1 \to A_2 \to \ldots A_n \to A_1$. Clearly at any vertex this results in one arrow entering and one leaving the vertex, i.e. the contribution to α is unity at each vertex. Next set up the preferences $A_1 \rightarrow A_3 \rightarrow A_5 \rightarrow \ldots$. This circuit may either visit each vertex once, or not. In the latter case we proceed to an unvisited vertex and set up the preferences $A_r \rightarrow A_{r+2} \rightarrow A_{r+4} \rightarrow \ldots$ and so on. Again there will be a unit contribution to all the α's.

We then set up the preferences $A_1 \rightarrow A_4 \rightarrow A_7 \rightarrow$ etc. and so on; and in this way we shall ultimately complete the preference scheme.

If n is odd all the preferences described will consist of circular tours of the polygon, and thus the value of α for each vertex will be $(n-1)/2$. If n is even the last preference $A_1 \rightarrow A_{\frac{1}{2}n+1}$ will not be a tour but will consist of the single line joining one vertex with the symmetrically opposite vertex. Thus there will be $n/2$ vertices for which $\alpha = n/2$ and $n/2$ vertices for which $\alpha = (n-2)/2$. In this case $T = n/4$.

Now it is clear from the definition of T that it cannot be less than zero, or if n is even, be less than $n/4$. The configuration just given shows that these minima are, in fact, attainable.

Thus T can vary from a maximum of $(n^3-n)/12$ to a minimum of zero or $n/4$. Hence the maximum number of circular triads, being half the variation from maximum to minimum of T (the maximum of T corresponding to the ranking case in which there are no inconsistences) is $(n^3-4n)/24$ if n is even and $(n^3-n)/24$ if n is odd.

This establishes the first two results enunciated at the beginning of this section. To prove the third it is sufficient to give a systematic method of proceeding from the configuration of minimum to that of maximum inconsistence by steps decreasing T two at a time. Consider, for example, the case $n = 8$. For the minimum inconsistence the α's will have the values 0 to 7, which we set out thus:

A	B	C	D	E	F	G	H
0	1	2	3	4	5	6	7

We proceed by reversing the preferences between vertices whose α-values differ by two. This clearly reduces T by two.

Reversing the preferences between C and E we get

A	B	C	D	E	F	G	H
0	1	3	3	3	5	6	7

and between D and F we get

A	B	C	D	E	F	G	H
0	1	3	4	3	4	6	7

which we may rearrange as

A	B	C	E	D	F	G	H
0	1	3	3	4	4	6	7

Now reversing the preferences between B and E and between D and G and rearranging we have

A	B	E	C	F	D	G	H
0	2	2	3	4	5	5	7

and now interchanging A and B, G and H,

A	B	E	C	F	D	G	H
1	1	2	3	4	5	6	6

At this stage we have preserved the α-numbers 2, 3, 4 and 5 in the middle but reduced the extremes A and H. We can now carry out the process again, arriving at the α-numbers.

1	2	2	3	4	5	5	6

and twice again, giving

2	2	3	3	4	4	5	5

whence a final interchange gives

3	3	3	3	4	4	4	4

and this is the position of maximum inconsistence. It is readily verified by following the interchanges on a polygon diagram that the reversals are, in fact, legitimate.

Coefficient of consistence in paired comparisons

9. If d is the number of circular triads in an observed configuration of preferences we define

$$\left.\begin{aligned} \zeta &= 1-\frac{24d}{n^3-n}, \quad n \text{ odd} \\ &= 1-\frac{24d}{n^3-4n}, \quad n \text{ even} \end{aligned}\right\} \qquad \ldots\ldots(2)$$

and call ζ the coefficient of consistence. If and only if it is unity there are no inconsistences in the configuration, which may therefore be represented by a ranking. As ζ decreases to zero the inconsistence, as measured by the number of circular triads, increases.

For example, in the configuration of Fig. 1 there are five circular triads, ABD, ACD, AFD, AED and BEF. The maximum possible number is 8. Thus $\zeta = 0{\cdot}375$.

10. ζ can also be interpreted in the light of Table I. Suppose, in that table, we sum the rows. (The column sums are determined by the row sums and add no fresh information.) The sum of any row will be the α-number for that vertex in the polygon which corresponds to the object defining the row. T will then be the value of the sum of squares of deviations of row totals from the mean value $(n-1)/2$, that is to say, will be the variance of the row sums multiplied

by n. ζ is thus a linear function of this variance; but it cannot be tested in the χ^2 distribution as if Table I were a contingency table, for the border cells are not independent or linearly dependent.

11. If an individual observer produces a configuration of preferences which show inconsistence there are usually several explanations; he may be an incompetent judge, the objects may be so alike that consistent differentiation is not possible, or his attention may wander during the course of the experiment. We discuss these questions later. They are mentioned here to explain the motive for the next stage of the mathematics. With what probability can a value of ζ arise by chance if the observer allots his preferences at random with respect to the quality under consideration?

With n objects there are $2^{\binom{n}{2}}$ possible configurations of preferences. We proceed to investigate the distribution of d in this universe of $2^{\binom{n}{2}}$ different members. The method consists of proceeding from the distribution for n to that for $(n+1)$.

For $n = 3$ there are eight configurations, of which two give one circular triad and six no circular triads. Consider the effect of adding a new vertex D to the vertices ABC. Four cases arise:

(1) $D \rightarrow$ all A, B, C.
(2) $D \rightarrow$ two of A, B, C.
(3) $D \rightarrow$ one of A, B, C.
(4) $D \rightarrow$ none of A, B, C.

The last two are symmetrical with the first two and need not be separately considered.

Situation (1) arises in one way and clearly does not add any new circular triads other than those already existing in the configuration ABC. It therefore contributes six values $d = 0$ and two values $d = 1$. So does situation (4).

Situation (2) arises in three ways, according as $D \leftarrow A$, B, or C. The configurations so reached are similar and we may take any one, say $D \leftarrow C$, as the single preference. If $A \leftarrow C$ then DAC is not circular and if $B \leftarrow C$ the DBC is not circular. On the other hand $A \rightarrow C$ and $B \rightarrow C$ will each produce a circular triad. We then have the cases

	No. of circular triplets added
$A \leftarrow C \rightarrow B$	0
$A \rightarrow C \rightarrow B$	1
$A \leftarrow C \leftarrow B$	1
$A \rightarrow C \leftarrow B$	2

We now consider AB. In the first two cases just enumerated the direction of AB does not matter and no circular triads are added. With the third $A \rightarrow B$ gives no circular triad but $A \leftarrow B$ adds one. With the fourth $A \rightarrow B$ adds one and $A \leftarrow B$ adds none.

Thus the number of circular triads occurring for these four cases is found to be

No. of circular triplets	Frequency
0	2
1	2
2	4

We must multiply the frequency by three and by two to allow for similar symmetrical arrangements, and the final results are

No. of circular triplets	Frequency
0	24
1	16
2	24
Total	64

The principles of this method are clear enough and the work may be formalized by a number of conventions which we omit to save space. In common with many similar combinatorial problems, however, troubles arise from the sheer number of possibilities and the difficulty of ensuring that nothing is overlooked. Up to the present we have found the distribution of d for n up to and including 7. The frequencies and probabilities are given in Table II.

12. For the values already obtained the moments are given by the following formulae:

$$\mu_1' \text{ (about 0)} = \frac{1}{4}\binom{n}{3}, \qquad \text{......(3)}$$

$$\mu_2 = \frac{3}{16}\binom{n}{3}, \qquad \text{......(4)}$$

$$\mu_3 = -\frac{3}{32}\binom{n}{3}(n-4), \qquad \text{......(5)}$$

$$\mu_4 = \frac{3}{256}\binom{n}{3}\left\{9\binom{n-3}{3}+39\binom{n-3}{2}+9\binom{n-3}{1}+7\right\}. \qquad \text{......(6)}$$

We have very little doubt that these results are true in general but can offer no rigorous proof. In so far, however, as the moments are in a sense symmetric sums it appears highly probable that they are given by polynomials

TABLE II

Frequency (f) of values of d and probability (P) that values will be attained or exceeded

Value of d	$n=2$		$n=3$		$n=4$		$n=5$		$n=6$		$n=7$	
	f	P	f	P	f	P	f	P	f	P	f	P
0	2	1·000	6	1·000	24	1·000	120	1·000	720	1·000	5,040	1·000
1			2	0·250	16	·625	120	·883	960	·978	8,400	·998
2					24	·375	240	·766	2,240	·949	21,840	·994
3							240	·531	2,880	·880	33,600	·983
4							280	·297	6,240	·792	75,600	·967
5							24	·023	3,648	·602	90,384	·931
6									8,640	·491	179,760	·888
7									4,800	·227	188,160	·802
8									2,640	·081	277,200	·713
9											280,560	·580
10											384,048	·447
11											244,160	·263
12											233,520	·147
13											72,240	·036
14											2,640	·001
Total	2	—	8	—	64	—	1,024	—	32,768	—	2,097,152	—

in n; and if this is so the values obtained are sufficient to establish polynomials of degree six or less.

It is also to be noted that from the above values of the moments

$$\beta_1 = \mu_3^2/\mu_2^3 \sim 8/n, \qquad \beta_2 = \mu_4/\mu_2^2 \sim 3 + 12/n,$$

from which it appears that a Type III distribution would fit the d-distribution fairly closely for moderate or large values of n. But as the distribution of d is of interest mainly for low values of n, which are all that occur in practice, it hardly seems worth while attempting to fit a curve.

AGREEMENT AMONG SEVERAL OBSERVERS

13. We now consider the investigation of similarities of judgments for m observers. Suppose that in a table of the form of Table I we enter a unit in the cell in row X and column Y whenever $X \to Y$ and count the units in each cell. A cell may then contain any number from 0 to m. If the observers are in complete agreement there will be $\binom{n}{2}$ cells containing the number m, the remaining $\binom{n}{2}$ cells being zero. The agreement may be complete even if there are inconsistences present.

Suppose that the cell in row X and column Y contains the number γ. Let

$$\Sigma = S\binom{\gamma}{2}, \qquad \text{......(7)}$$

the summation extending over the $n(n-1)$ cells of the table (the diagonal cells being ignored). Σ is then the sum of the number of agreements between pairs of judges. Put

$$u = \frac{2\Sigma}{\binom{m}{2}\binom{n}{2}} - 1. \qquad \text{......(8)}$$

The maximum number of agreements, occurring if $\binom{n}{2}$ cells each contain m, is $\binom{n}{2}\binom{m}{2}$ and thus in the case of complete agreement, and only in this case, $u = 1$. The further we go from this case, as measured by agreements between pairs of observers, the smaller u becomes. The minimum number of agreements occurs when each cell contains $m/2$ if m is even or $(m \pm 1)/2$ if m is odd. That is, if m is even, the minimum number of agreements is

$$2\binom{\frac{m}{2}}{2}\binom{n}{2} = \tfrac{1}{4}m(m-2)\binom{n}{2},$$

and in this case

$$u = -\frac{1}{m-1}. \qquad \text{......(9)}$$

When m is odd the minimum value of u is found to be

$$u = -\frac{1}{m}. \qquad \text{......(10)}$$

14. We propose to call u the coefficient of agreement. It is unity if and only if there is complete agreement in the comparisons. Its minimum value is not -1 except when $m = 2$. This, however, is to be expected in a measure of agreement for there can be no such thing as complete disagreement among three or more observers in paired comparisons. If observer P differs in certain comparisons from observers Q and R, the two latter must agree on those comparisons.

When $m = 2$, u reduces to

$$u = \frac{2\Sigma}{\binom{n}{2}} - 1 \qquad \text{......(11)}$$

and Σ becomes twice the number of cases in which the two observers agree about a comparison. u is thus a generalization of a coefficient τ proposed by Kendall (1938) to measure the correlation between two rankings. For general m, if the entries in the table were constrained to the ranking type, u would be the average intercorrelation τ between observers taken two at a time.

15. In discussing the significance of u it is desirable to know whether the set of preferences which give rise to it could have arisen by chance if the preferences had been assigned at random with respect to the quality under consideration. The procedure which first suggests itself is a generalization of the method used for the case of m rankings (Kendall & Babington Smith, 1939). That is to say, we sum the entries in the rows of the table and consider the variance of these entries. If the preferences are allotted at random we expect to find about equal numbers given to each object, and the variance will be low; in other cases it will be higher.

The difficulty about this suggestion is that it has not been found possible to ascertain the distribution of the variance in the $2^{m\binom{n}{2}}$ possible sets of preferences. The case $m = 1$, corresponding to the distribution of d for inconsistences, is difficult enough to solve. For higher values of m we have failed to find any distributions except in trivial cases.

We can, however, offer a test based on the distribution of u (or Σ). The comparative simplicity of the distributions in this case is in accordance with the remark made by Kendall in the paper under reference that the distribution of τ is much simpler and much more regular than the distribution of the Spearman correlation coefficient ρ.

16. Consider one cell in the table in row X and column Y and let it contain the number γ. Then the corresponding cell in row Y and column X will contain $m-\gamma$. Thus these two contribute to Σ the amount $\binom{\gamma}{2}+\binom{m-\gamma}{2}$.

Now, of the total ways in which the units can be distributed in the first cell there will be $\binom{m}{\gamma}$ in which γ units occur. Consequently the distribution of Σ in the cell and the corresponding cell is given by the expression

$$f = t^{\binom{m}{2}} + \binom{m}{1} t^{\binom{m-1}{2}} + \binom{m}{2} t^{\binom{m-2}{2}+\binom{2}{2}} + \ldots + \binom{m}{\gamma} t^{\binom{m-\gamma}{2}+\binom{\gamma}{2}} + \ldots + t^{\binom{m}{2}}, \quad \ldots\ldots(12)$$

and since the distribution in other pairs of cells is independent if the preferences are allotted at random the distribution of Σ for the whole table is given by

$$D(\Sigma) = f^N, \quad \ldots\ldots(13)$$

where $N = \binom{n}{2}$.

17. The distributions have been worked out for the following values of m and n: $m = 3$, $n = 2$ to 8; $m = 4$, $n = 2$ to 6; $m = 5$, $n = 2$ to 5; $m = 6$, $n = 2$ to 4. Tables III to VI give the probabilities based on these distributions, i.e. the probabilities that a given value of Σ will be attain d or exceeded.

For constant n the distribution tends to the Type III form as m tends to infinity. In fact, for a single pair of related cells the variate value corresponding

22-2

TABLE III

The probability P that a value of Σ will be attained or exceeded, for $m = 3$, $n = 2$ to 8

$n=2$		$n=3$		$n=4$		$n=5$		$n=6$		$n=7$		$n=8$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
1	1·000	3	1·000	6	1·000	10	1·000	15	1·000	21	1·000	28	1·000
3	·250	5	·578	8	·822	12	·944	17	·987	23	·998	30	1·000
		7	·156	10	·466	14	·756	19	·920	25	·981	32	·997
		9	·016	12	·169	16	·474	21	·764	27	·925	34	·983
				14	·038	18	·224	23	·539	29	·808	36	·945
				16	·0046	20	·078	25	·314	31	·633	38	·865
				18	$\cdot 0^{3}24$	22	·020	27	·148	33	·433	40	·736
						24	·0035	29	·057	35	·256	42	·572
						26	$\cdot 0^{3}42$	31	·017	37	·130	44	·400
						28	$\cdot 0^{4}30$	33	·0042	39	·056	46	·250
						30	$\cdot 0^{6}95$	35	$\cdot 0^{3}79$	41	·021	48	·138
								37	$\cdot 0^{3}12$	43	·0064	50	·068
								39	$\cdot 0^{4}12$	45	·0017	52	·029
								41	$\cdot 0^{6}92$	47	$\cdot 0^{3}37$	54	·011
								43	$\cdot 0^{7}43$	49	$\cdot 0^{4}68$	56	·0038
								45	$\cdot 0^{9}93$	51	$\cdot 0^{4}10$	58	·0011
										53	$\cdot 0^{5}12$	60	$\cdot 0^{3}29$
										55	$\cdot 0^{6}12$	62	$\cdot 0^{4}66$
										57	$\cdot 0^{8}86$	64	$\cdot 0^{4}13$
										59	$\cdot 0^{9}44$	66	$\cdot 0^{5}22$
										61	$\cdot 0^{10}15$	68	$\cdot 0^{6}32$
										63	$\cdot 0^{12}23$	70	$\cdot 0^{7}40$
												72	$\cdot 0^{8}42$
												74	$\cdot 0^{9}36$
												76	$\cdot 0^{10}24$
												78	$\cdot 0^{11}13$
												80	$\cdot 0^{13}48$
												82	$\cdot 0^{14}12$
												84	$\cdot 0^{16}14$

TABLE IV

The probability P that a value of Σ will be attained or exceeded, for $m = 4$ and $n = 2$ to 6
(for $n = 6$ only values beyond the 1 % point are given)

$n=2$		$n=3$		$n=4$		$n=5$		$n=5$		$n=6$		$n=6$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
2	1·000	6	1·000	12	1·000	20	1·000	42	·0048	57	·014	79	$\cdot 0^{8}42$
3	·625	7	·947	13	·997	21	1·000	43	·0030	58	·0092	80	$\cdot 0^{8}28$
6	·125	8	·736	14	·975	22	·999	44	·0017	59	·0058	81	$\cdot 0^{9}98$
		9	·455	15	·901	23	·995	45	$\cdot 0^{3}73$	60	·0037	82	$\cdot 0^{9}15$
		10	·330	16	·769	24	·979	46	$\cdot 0^{3}41$	61	·0022	83	$\cdot 0^{9}12$
		11	·277	17	·632	25	·942	47	$\cdot 0^{3}24$	62	·0013	84	$\cdot 0^{10}51$
		12	·137	18	·524	26	·882	48	$\cdot 0^{4}90$	63	$\cdot 0^{3}76$	86	$\cdot 0^{11}30$
		14	·043	19	·410	27	·805	49	$\cdot 0^{4}37$	64	$\cdot 0^{3}44$	87	$\cdot 0^{11}17$
		15	·025	20	·278	28	·719	50	$\cdot 0^{4}25$	65	$\cdot 0^{3}23$	90	$\cdot 0^{13}28$
		18	·0020	21	·185	29	·621	51	$\cdot 0^{5}93$	66	$\cdot 0^{3}13$		
				22	·137	30	·514	52	$\cdot 0^{5}21$	67	$\cdot 0^{4}72$		
				23	·088	31	·413	53	$\cdot 0^{5}17$	68	$\cdot 0^{4}36$		
				24	·044	32	·327	54	$\cdot 0^{6}74$	69	$\cdot 0^{4}18$		
				25	·027	33	·249	56	$\cdot 0^{7}66$	70	$\cdot 0^{5}97$		
				26	·019	34	·179	57	$\cdot 0^{7}38$	71	$\cdot 0^{5}47$		
				27	·0079	35	·127	60	$\cdot 0^{9}93$	72	$\cdot 0^{5}20$		
				28	·0030	36	·090			73	$\cdot 0^{5}10$		
				29	·0025	37	·060			74	$\cdot 0^{6}51$		
				30	·0011	38	·038			75	$\cdot 0^{6}18$		
				32	$\cdot 0^{3}16$	39	·024			76	$\cdot 0^{7}78$		
				33	$\cdot 0^{4}95$	40	·016			77	$\cdot 0^{7}44$		
				36	$\cdot 0^{5}38$	41	·0088			78	$\cdot 0^{7}15$		

TABLE V

The probability P that a value of Σ will be attained or exceeded, for m = 5 and n = 2 to 5

$n=2$		$n=3$		$n=4$		$n=5$		$n=5$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
4	1·000	12	1·000	24	1·000	40	1·000	76	$\cdot 0^{4}50$
6	·375	14	·756	26	·940	42	·991	78	$\cdot 0^{4}16$
10	·063	16	·390	28	·762	44	·945	80	$\cdot 0^{5}50$
		18	·207	30	·538	46	·843	82	$\cdot 0^{5}15$
		20	·103	32	·353	48	·698	84	$\cdot 0^{6}39$
		22	·030	34	·208	50	·537	86	$\cdot 0^{6}10$
		24	·011	36	·107	52	·384	88	$\cdot 0^{7}23$
		26	·0039	38	·053	54	·254	90	$\cdot 0^{8}53$
		30	$\cdot 0^{3}24$	40	·024	56	·158	92	$\cdot 0^{8}12$
				42	·0093	58	·092	94	$\cdot 0^{9}14$
				44	·0036	60	·050	96	$\cdot 0^{10}46$
				46	·0012	62	·026	100	$\cdot 0^{12}91$
				48	$\cdot 0^{3}36$	64	·012		
				50	$\cdot 0^{3}12$	66	·0057		
				52	$\cdot 0^{4}28$	68	·0025		
				54	$\cdot 0^{5}54$	70	·0010		
				56	$\cdot 0^{5}18$	72	$\cdot 0^{3}39$		
				60	$\cdot 0^{7}60$	74	$\cdot 0^{3}14$		

TABLE VI

The probability P that a value of Σ will be attained or exceeded, for m = 6 and n = 2 to 4

$n=2$		$n=3$		$n=4$		$n=4$		$n=4$	
Σ	P	Σ	P	Σ	P	Σ	P	Σ	P
6	1·000	18	1·000	36	1·000	55	·043	74	$\cdot 0^{4}12$
7	·688	19	·969	37	·999	56	·029	75	$\cdot 0^{5}89$
10	·219	20	·832	38	·991	57	·020	76	$\cdot 0^{5}49$
15	·031	21	·626	39	·959	58	·016	77	$\cdot 0^{5}32$
		22	·523	40	·896	59	·011	80	$\cdot 0^{6}68$
		23	·468	41	·822	60	·0072	81	$\cdot 0^{6}17$
		24	·303	42	·755	61	·0049	82	$\cdot 0^{6}12$
		26	·180	43	·669	62	·0034	85	$\cdot 0^{7}34$
		27	·147	44	·556	63	·0025	90	$\cdot 0^{8}93$
		28	·088	45	·466	64	·0016		
		29	·061	46	·409	65	$\cdot 0^{3}83$		
		30	·040	47	·337	66	$\cdot 0^{3}66$		
		31	·034	48	·257	67	$\cdot 0^{3}48$		
		32	·023	49	·209	68	$\cdot 0^{3}26$		
		35	·0062	50	·175	69	$\cdot 0^{3}16$		
		36	·0029	51	·133	70	$\cdot 0^{4}86$		
		37	·0020	52	·097	71	$\cdot 0^{4}68$		
		40	$\cdot 0^{3}58$	53	·073	72	$\cdot 0^{4}48$		
		45	$\cdot 0^{4}31$	54	·057	73	$\cdot 0^{4}16$		

to a frequency $\binom{m}{\gamma}$ is $\binom{m-\gamma}{2}+\binom{\gamma}{2}$, which is a quadratic in γ. Were the variate value a linear function of γ the distribution for the single cell would tend to normality in accordance with the well-known property of the binomial. The case of the quadratic value corresponds to a transformation of the variate of the type $x^2 = y$ and the transform of the normal form $\exp(-x^2)\,dx$ becomes the Type III form $\exp(-y)\,y^{-\frac{1}{2}}\,dy$. Since the N cells are independent and the sum of variates in the same Type III form is also distributed in that form, it follows that Σ is in the limit distributed as $\exp(-\Sigma)\,\Sigma^{\frac{N}{2}-1}\,d\Sigma$ except perhaps for some constants. Thus Σ or some multiple of it is distributed as χ^2.

For constant m the distribution tends to normality with increasing n.

18. The first of these results suggests that the Type III distribution will provide an approximation to the distribution (13) when m is moderately large. We proceed to find the first four moments of (13).

It is sufficient to find the first four moments of (12), those of (13) being obtainable therefrom in virtue of the relationships which connect seminvariants of independent distributions.

The rth moment of (12) about the origin is given by

$$2^m \mu_r' = \left[\left(t\frac{\partial}{\partial t}\right)^r f\right]_{t=1}, \qquad \ldots\ldots(14)$$

since 2^m is the total frequency. Thus we have

$$2^m \mu_1' = \mathop{S}_{r=0}^{m} \binom{m}{r}\left(r^2 - mr + \frac{m^2-m}{2}\right) = 2^m\binom{m}{2} + S\binom{m}{r}(r^2 - mr). \qquad \ldots\ldots(15)$$

Sums such as $S\binom{m}{r}r^p$ can be obtained by operating on the binomial $(1+x)^m$ p times by $x\frac{\partial}{\partial x}$, e.g. we find

$$S\left\{\binom{m}{r}r\right\} = 2^m\frac{m}{2},$$

$$S\binom{m}{r}r^2 = 2^m\left\{\frac{m}{2}+\frac{1}{2}\binom{m}{2}\right\}$$

and hence, substituting in (15),

$$\mu_1' = \frac{1}{2}\binom{m}{2}.$$

Thus the mean of the distribution (13) is given by

$$\mu_1' = \tfrac{1}{2}N\binom{m}{2}. \qquad \ldots\ldots(16)$$

In a similar way we find

$$\mu_2 = \tfrac{1}{4}N\binom{m}{2}, \qquad \text{......(17)}$$

$$\mu_3 = \tfrac{3}{4}N\binom{m}{3}, \qquad \text{......(18)}$$

$$\mu_4 = N\binom{m}{2}\left\{\frac{3m^2-15m+17}{8} + \tfrac{3}{32}N(m^2-m)\right\}. \qquad \text{......(19)}$$

These are the moments of Σ. Those of u are obtained by dividing by an appropriate power of $N\binom{m}{2}$ and it may be noted in particular that the mean of u is zero.

We have directly from (17), (18) and (19)

$$\beta_1 = \frac{8}{N}\frac{(m-2)^2}{m(m-1)},$$

$$\beta_2 = \frac{4}{Nm(m-1)}\left\{3m^2-15m+17+\frac{3N}{4}m(m-1)\right\}.$$

For constant m, as $N \to \infty$,

$$\beta_1 \to 0, \quad \beta_2 \to 3$$

and for constant N, as $m \to \infty$,

$$\beta_1 \to \frac{8}{N}, \quad \beta_2 \to \frac{12}{N}+3,$$

confirming the tendency towards the Type III distribution.

19. The first four moments of the Type III distribution

$$dF = ke^{-px}\,x^{q-1}\,dx$$

are

$$\frac{q}{p}, \frac{q}{p^2}, \frac{2q}{p^3}, \frac{3q(q+2)}{p^4}.$$

Equating the second and third moments to those given by (17) and (18) we find

$$q = \frac{Nm(m-1)}{2(m-2)^2}, \qquad \text{......(20)}$$

$$p = \frac{2}{m-2}. \qquad \text{......(21)}$$

To make the first moments correspond we move the origin of the Σ distribution a distance $\frac{1}{2}N\binom{m}{2}\frac{m-3}{m-2}$ to the right. We thus reach the approximation to the Σ distribution, coinciding in the first three moments

$$dF = ke^{-\frac{2x}{m-2}}x^{\frac{Nm(m-1)}{2(m-1)^2}-1}dx,$$

where

$$x = \Sigma - \tfrac{1}{2}N\binom{m}{2}\frac{m-3}{m-2}$$

or, transforming to the more usual χ^2 form by putting $\chi^2 = 4x/(m-2)$, we find that

$$\left\{\Sigma - \tfrac{1}{2}N\binom{m}{2}\frac{m-3}{m-2}\right\}\frac{4}{m-2} \qquad \ldots\ldots(22)$$

is distributed as χ^2 with

$$\nu = \frac{Nm(m-1)}{(m-2)^2} \qquad \ldots\ldots(23)$$

degrees of freedom.

The fourth moments of Σ and the χ^2 approximation differ by terms of order N^{-1} and m^{-1} compared with their absolute values.

20. It only remains to be seen how large m and n must be for this to provide a satisfactory approximation.

Consider first the distributions for $m = 3$. When $n = 8$, $N = 28$, we have, for the approximation, 4Σ distributed with 168 degrees of freedom. From Table III we see that for $\Sigma = 54$, $P = 0{\cdot}011$ and for $\Sigma = 58$, $P = 0{\cdot}0011$. Applying a continuity correction by deducting unity from Σ we find for the χ^2 approximation with $\chi^2 = 4 \times 53$, $\nu = 168$, $P = 0{\cdot}011$, and with $\chi^2 = 4 \times 57$, $P = 0{\cdot}00114$. The correspondence is very close, in spite of the low value of m.

For $m = 4$, $n = 5$, $N = 10$, the approximation gives $2\Sigma - 30$ distributed with 30 degrees of freedom. For $\Sigma = 40$ and 41, this gives, with continuity corrections of 0·5, half the variate-interval, $\chi^2 = 49$ and 51, $\nu = 30$. From the diagram given in Yule & Kendall's "Introduction to the Theory of Statistics" (1937) it is seen that these values lie one on either side of the 1% value; and this is in accordance with the exact values of P, which are seen from Table IV to be 0·016 and 0·0088. Similarly we find that the values of Σ, 37 and 38, lie on either side of the 5% level, which is again in accordance with the exact values, $P = 0{\cdot}060$ and $0{\cdot}038$.

For $m = 6$, $n = 4$, $N = 6$, the approximation gives $\Sigma - 33{\cdot}75$ distributed with 11·25 degrees of freedom. For $\Sigma = 59$ and 60 the corresponding χ^2 values are seen to lie on either side of the 1% point, which accords with the exact value of Table VI.

We conclude that the χ^2 approximation provides an adequate test of significance for the values of m and n outside the range for which Tables III–VI give exact values.

21. As a matter of theoretical interest we may record the results for the distribution of u when the data are ranked. It appears that in this case

$$\frac{12}{m-2}\frac{2n+5}{2n^2+6n+7}\left[\Sigma-\frac{1}{2}\binom{m}{2}\binom{n}{2}\left\{1-\frac{1}{3(m-2)}\frac{(2n+5)^2}{2n^2+6n+7}\right\}\right]$$

is distributed approximately as χ^2 with

$$\binom{m}{2}\frac{2}{(m-2)^2}\binom{n}{2}\frac{(2n+5)^3}{(2n^2+6n+7)}\text{ degrees of freedom.}$$

This result is not of much practical value. The case of m rankings can be more simply treated by other methods.

INTERPRETATION OF RESULTS OF PAIRED COMPARISONS

22. In the light of the foregoing theory we may discuss the interpretation of the results of a paired comparison experiment.

If for each observer the coefficient of consistence is unity the comparisons reduce to rankings and may be discussed by known methods. But if some or all the coefficients are not unity we have to consider the following possibilities:

(*a*) Some of the observers may be bad judges and the inconsistences reflect their shortcomings in making comparisons.

(*b*) Some of the objects may differ by amounts which fall below the threshold of distinguishability for some observers.

(*c*) The property under judgment may not be a linear variate at all and we may be getting the sort of confusion which would result if observers were asked to compare English towns according to the bivariate concept "geographical position".

(*d*) Several of the effects may be operating simultaneously.

23. If we have only one observer and have no prior knowledge of his capabilities it is not in general possible to apportion his inconsistences among these causes. Exceptions may occur when the inconsistences are of a marked and peculiar kind; for instance if they involve only four objects out of 15, we may suspect that the four are practically indistinguishable rather than that the observer is unable to make distinctions at all and avoided inconsistences among the others by sheer chance. But even here conclusions drawn *a posteriori* after inspection of the data are dubious. Table II gives a test of the hypothesis that an observer is incapable of making judgments. For example, with $n = 7$, the chances are 983 in 1000 that if the preferences are made at random there will be more than two inconsistent triads, so that if we find two or less, it is improbable that the observer is completely incapable of judgment. We might then be led to suppose that his small deviation from internal consistence is due to fluctuation of attention, very close resemblance to the objects giving rise to the inconsistences, or both.

24. With m observers the investigation can be taken a good deal further. If all the observers show inconsistences we suspect that the objects are at fault or that the observers are being asked to perform an impossible task. On the other hand, if most of the observers show a small or zero inconsistence we suspect that the others are just bad judges and may reject their data accordingly.

As between indistinguishability of objects and non-linearity of variate, a choice of explanations would depend largely on the extent to which inconsistences were concerned with the same set of objects. If there is a high value of u, indicating concordance of judgment, we expect to find most of the inconsistences confined to certain objects, and common to observers. In this case we suspect that the objects are close together in the degree to which they exhibit the quality under consideration. But if the observers scatter their inconsistences over the whole field u will be moderate or low and we suspect that the observers are being asked to do something beyond their capacity; and this brings us to question the validity of regarding the quality as a linear variable.

25. When a quality such as "bravery" or "intelligence" is insusceptible to measurement there is frequently doubt of this kind. But this has not deterred investigators from assuming that such statistical variables exist, or from requiring observers to rank objects according to them, or in some cases from replacing such rankings by quantiles of the normal curve. We are never tired of criticizing this Principle of the Hypostasis of Plausible Terminology. Hitherto it has flourished largely because of the difficulty of adducing evidence against it; and we hope that the inconsistence of paired comparisons will provide a criterion, however rough, of the legitimacy of the methods to which it leads.

But we would emphasize that our approach to the method of paired comparisons has a somewhat different object from that elaborated by Thurstone (1927 and many subsequent papers). As we understand it, his method is appropriate where one is entitled to assume *a priori* or by reason of precautions taken in the selection of material that a linear variable is involved and that there exist perceptible differences between the items presented for comparison. Our object is to make it possible to dispense with such assumptions and precautions.

26. A few words may be added about the case in which an objective order is known to exist (as, for instance, in judging individuals according to age or weight). In such circumstances the appearance of inconsistences will indicate unreliability of the part of the observer or subliminal differences between objects. A measure of the observer's reliability may be obtained by calculating u between known and observed comparisons. If ζ is high enough to enable us to accept his judgments as internally consistent on the whole, u may still be low enough to reject his judgments as accurate.

27. We conclude with an example of the application of the foregoing theory to some experimental material.

Classes of children (ages 11 to 13 inclusive) were asked to state their preferences with respect to certain school subjects. Each child was given a sheet on which were written the possible pairs of subjects and asked to underline the one preferred in each case. Two classes gave the following results:

(*a*) 21 boys, 13 school subjects. The preferences are shown in Table VII, which is in the form described in section 13; e.g. there were 18 boys who preferred Art to Religion.

TABLE VII

Preferences of 21 *boys in* 13 *subjects*

	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
1. Woodwork	—	14	20	15	15	16	16	18	18	18	20	21	20	211
2. Gymnastics	7	—	14	12	13	18	14	16	16	20	16	18	19	183
3. Art	1	7	—	10	14	10	16	18	16	16	17	16	19	160
4. Science	6	9	11	—	11	12	15	14	13	13	17	17	16	154
5. History	6	8	7	10	—	14	11	12	14	15	13	14	16	140
6. Geography	5	3	11	9	7	—	14	14	13	13	16	15	17	137
7. Arithmetic	5	7	5	6	10	7	—	9	11	13	15	13	15	116
8. Religion	3	5	3	7	9	7	12	—	12	14	14	16	14	116
9. English Literature	3	5	5	8	7	8	10	9	—	10	13	13	15	106
10. Commercial subjects	3	1	5	8	6	8	8	7	11	—	10	10	14	91
11. Algebra	1	5	4	4	8	5	6	7	8	11	—	10	13	82
12. English Grammar	0	3	5	4	7	6	8	5	8	11	11	—	13	81
13. Geometry	1	2	2	5	5	4	6	7	6	7	8	8	—	61
													Total	1638

The calculation of Σ for this table, in which the objects are arranged in order of total number of preferences, may be shortened by noting that Σ as given by equation (7) may be transformed into the form

$$\Sigma = S(\gamma^2) - mS(\gamma) + \binom{m}{2}\binom{n}{2},$$

where the summation now takes place over the half of the table below the diagonal. Since the numbers in this half are smaller than those in the other half there is a considerable saving in arithmetic.

We find $$\Sigma = 9718$$

and hence $$u = \frac{2 \times 9718}{\binom{21}{2}\binom{13}{2}} - 1 = 0{\cdot}186.$$

There is thus a certain amount of agreement among the children, indicated by the positive value of u. Is this significant?

We note first of all that this distribution of preferences could not have arisen by chance to any acceptable degree of probability. In fact, $\chi^2 = 412{\cdot}4$ (equation (22)) and $\nu = 90{\cdot}7$. The large value of ν justifies the use of the normal approximation to the χ^2 distribution and we find $\sqrt{(2\chi^2)} - \sqrt{(2\nu - 1)} = 15\cdot3$ a very improbable result on the hypothesis of a random allocation of preferences.

The distribution of circular triads was as follows:

No. of triads	Frequency	No. of triads	Frequency
	1	12	1
1	1	17	3
4	5	21	1
6	2	25	1
7	2	29	1
8	1	39	1
10	1	Total	21

The total number of circular triads was 242 with a mean of 11·5. Only one boy was entirely consistent. On the other hand, for $n = 13$ the maximum number of circular triads is 91, with a mathematical expectation of 71·5. It is thus clear that, except perhaps for one boy, we cannot suppose that any boy allotted preferences at random. We are again led to conclude that the boys are genuinely capable of making distinctions, and that consistently on the whole. Half the boys have coefficients of consistence ζ greater than 0·92.

We conclude that the boys can make preferences and that in their view the subjects are sufficiently different to enable a reasonably consistent set of preferences to be made. So far as these data are concerned we would see no objection to the assumption that a *scale* of preferences can be set up. With this in mind we can say that the value of u indicates a certain amount of agreement, though not a strong one, between the boys as to which subjects they prefer.

(b) 25 girls, 11 school subjects. Table VIII shows the data.

TABLE VIII

Preferences of 25 *girls in* 11 *subjects*

	1	2	3	4	5	6	7	8	9	10	11	Total
1. Gymnastics	—	10	19	17	20	17	21	21	21	18	22	186
2. Science	15	—	12	15	17	15	21	19	18	16	17	165
3. Art	6	13	—	16	16	18	10	17	16	19	16	147
4. Domestic Science	8	10	9	—	16	11	13	15	14	11	14	121
5. History	5	8	9	9	—	14	18	12	13	15	18	121
6. Arithmetic	8	10	7	14	11	—	12	13	12	16	18	121
7. Geography	4	4	15	12	7	13	—	14	15	14	14	112
8. English Literature	4	6	8	10	13	12	11	—	14	13	14	105
9. Religion	4	7	9	11	12	13	10	11	—	11	17	105
10. Algebra	7	9	6	14	10	9	11	12	14	—	12	104
11. English Grammar	3	8	9	11	7	7	11	11	8	13	—	88
											Total	1375

We find $\Sigma = 8928$, $u = 0{\cdot}082$.

For the χ^2 significance test, $\chi^2 = 180{\cdot}3$, $\nu = 62{\cdot}4$ and $\sqrt{(2\chi^2)} - \sqrt{(2\nu - 1)} = 7{\cdot}9$, as before a very significant result.

The distribution of circular triads was

No. of triads	Frequency	No. of triads	Frequency
1	2	17	1
2	2	19	1
3	1	22	1
4	1	23	2
6	1	27	1
8	1	32	1
9	1	35	1
11	2	37	1
12	2	38	1
13	1		—
14	1	Total	25

The total number of circular triads is 382 with a mean 15·28. For $n = 11$ the maximum number of circular triads is 55 with an expectation of 41·25. Several of the girls come very close to this, the worst having a coefficient of consistence equal to 0·31.

We are, however, again led to conclude that the preferences were not allotted at random and that most of the girls are capable of exercising a judgment which is on the whole consistent. There is only a very slight agreement in preferences.

Thus the girls are less consistent and less alike in preferences than the boys.

REFERENCES

KENDALL, M. G. (1938). "A New Measure of Rank Correlation." *Biometrika*, **30**, 81.

KENDALL, M. G. & BABINGTON SMITH B. (1939). "On the Problem of *m* Rankings." *Ann. Math. Statist.* **10**, 273.

THURSTONE, L. L. (1927). "A Law of Comparative Judgment." *Psychol. Rev.* **34**, 275.