Vol. IV Part II | The British Journal of Psychology, Statistical Section | June, 1951

# INCOMPLETE BLOCKS IN RANKING EXPERIMENTS

By J. DURBIN
London School of Economics

I. *Introduction.* II. *The Coefficient of Concordance.* III. *Test of Independence.* IV. *A Numerical Example.*

## I. INTRODUCTION

*Problem.* Suppose we have $n$ objects, and wish to investigate the preferences shown by $m$ different observers : one method will be to ask each observer to rank the objects in order of preference. This is the ordinary case of $m$ rankings described by Kendall (1, p. 80). If, however, $n$ is large, the method is often unsatisfactory, since it is difficult to rank a large number. Further, if the objects are presented as a list of names, the order in which they are presented may become a disturbing factor as $n$ increases. In such cases a better method is to present them in small blocks. In this paper the use of ' balanced incomplete blocks ' will be examined, and expressions calculated for the mean and variance of the coefficient of concordance. These will enable approximate tests of independence among the rankings to be constructed. Finally, the optimal properties of the estimate of the true ranking will be discussed.

The need for some such procedure is obvious ; and the method to be described has a number of applications in social and psychological research. For instance, Hall and Caradoc Jones (3) have recently described a survey in which 1,399 people were asked to grade 30 occupations in order of social standing. The difficulty of grading as many as 30 items was admitted at the outset ; and it was therefore suggested that the respondents should first assign each occupation to one of five groups, and then order the members of each group separately. This, however, offers only a partial relief ; and in many similar enquiries it would seem to be preferable for the investigator to arrange the objects in smaller groups beforehand. Much the same device may also be needed in studying preferences and in experiments on discrimination.

*Balanced Incomplete Blocks.* To be workable in practice, a ranking layout which uses incomplete blocks should yield results that permit of a fairly straightforward analysis. This requirement gives rise to the following two conditions, which seem requisite in themselves on grounds of common sense :

(*a*) Each object should occur an equal number of times in the experiment as a whole.

(*b*) The number of times two particular objects occur together in the same block should be the same for all possible pairs of objects.

Designs in which these conditions are fulfilled are called ' balanced incomplete block designs.' An excellent account of them, together with a detailed list of the appropriate layouts, will be found in the recent book by Cochran and Cox (2).

As stated above, it will sometimes be desirable to eliminate from the comparisons the effect of the order of presentation within the block. This can be done by arranging the layout so that every object occurs equally often in each position in a block. Such an arrangement can always be found by taking a sufficient number of blocks to cover all the requisite combinations ; but this number will frequently be impracticably large. However, certain compact designs satisfying the condition can be arranged for multiples of $p^2 + p + 1$ objects in blocks of $p + 1$ ; these are called ' Youden square designs,' and are fully described by Cochran and Cox.

It is unnecessary for each observer to rank the objects in *all* the blocks in a single repetition. For instance, an observer may be called upon to rank only the objects in a single block. The tests of independence described below apply to this case equally with that in which he is required to rank all the $n$ objects. If, however, the purpose is to estimate the mean ranking for a population of observers, it may be desirable to ask the observers to rank the objects more than once, using different arrangements in the blocks. In this way a greater number of independent comparisons between pairs can be obtained. With such an arrangement the greatest precision is achieved when each observer ranks an equal number of blocks. There will then be no question of testing for independence among the different rankings, except when the hypothesis under consideration is that each observer allocates his preferences at random in successive rankings.

*A Simple Example.* Suppose it is desired to rank seven objects A B C D E F G in blocks of three. From the list of Youden square designs in Cochran and Cox we take the following design :

| | | | | | | |
|---|---|---|---|---|---|---|
| A | B | C | D | E | F | G |
| B | C | D | E | F | G | A |
| D | E | F | G | A | B | C |

Each column represents a block, and each row a single replication of the objects. Since only three objects out of seven appear in each column, the 'blocks' are 'incomplete'; and, since the arrangement fulfils conditions *a* and *b*, it may be described as ' balanced.' By permuting the letters A, B, . . ., or by repeating the design as it stands, we can obtain further sets of seven blocks, if we wish. The number of blocks assigned to each observer is irrelevant to the problem of testing independence ; and in any practical case would depend on the purpose of the experiment.

Suppose the following ranks were allotted :

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | 2 | 1 | 1 | 2 | 2 | 3 |
| 1 | 3 | 2 | 3 | 3 | 3 | 1 |
| 3 | 1 | 3 | 2 | 1 | 1 | 2 |

The total ranks for the seven objects will be as follows :

| A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|
| 4 | 4 | 6 | 6 | 6 | 8 | 8 |

If the ranks had been allotted at random within the blocks, these totals would tend to be equal. If, on the other hand, the ranks corresponded to a strict ordering of the objects, the totals would be 3, 4, 5, 6, 7, 8, 9, and would differ among themselves as widely as possible. Thus, the variance of the total ranks should be a good statistic for testing the independence of the rankings within the different blocks. It will be defined more precisely in the following section.

## II. THE COEFFICIENT OF CONCORDANCE

*Mean and Variance of the Coefficient.* Suppose the $n$ objects are presented in blocks of $k$, and that in the experiment as a whole each is ranked $m$ times. Then the number of blocks will be $\frac{mn}{k}$. Within each block there are $\frac{1}{2}k(k-1)$ comparisons between pairs. Consequently, the total number of comparisons is $\frac{1}{2}mn(k-1)$. Let $\lambda$ be the number of blocks in which a particular pair of objects occurs. Then $\frac{1}{2}n(n-1)\lambda = \frac{1}{2}mn(k-1)$ so that $\lambda = \frac{m(k-1)}{n-1}$.

Consider first the case in which there are no ties. Let $x_{ij}$ denote the rank assigned to the $j$th object in the $i$th replication, and $S$ the sum of squared deviations from the mean of the totals $x_j = \sum_{i=1}^{m} x_{ij}$. $S$ attains its maximum when there is perfect concordance between the rankings. The total ranks are then $m$, $m+\lambda$, $m+2\lambda$, . . . $m+(n-1)\lambda$. Thus the maximum value of $S$ is $\frac{\lambda^2 n(n^2-1)}{12}$.

The ' coefficient of concordance ' (1, p. 81) is therefore defined by

$$\begin{aligned} W &= \frac{\text{observed } S}{\text{maximum } S} \\ &= \frac{12\,S}{\lambda^2 n(n^2-1)} \\ &= \frac{12\,\Sigma x_j^2 - 3\,nm^2(k+1)^2}{\lambda^2 n(n^2-1)}. \qquad (1) \end{aligned}$$

To calculate the mean and variance, we take the $x$'s as measured from the block means. Following the derivation given by Kendall (*loc. cit.*, pp. 90f.), we have,

$$W = \frac{12\,\Sigma x_{ij}^2 + 24\,\sum_j \sum_{i<p} x_{ij}\,x_{pj}}{\lambda^2\,n\,(n^2-1)},$$

where
$$\Sigma x_{ij}^2 = \frac{mn\,(k^2-1)}{12}.$$

Thus
$$W = \frac{k+1}{\lambda\,(n+1)} + \frac{24\,U}{\lambda^2\,n\,(n^2-1)}. \qquad (2)$$

We therefore require the mean and variance of $U = \sum_j \sum_{i<p} x_{ij}\,x_{pj}$. The $j$th object occurs in different blocks in the $i$th and $p$th replications; and an $x$ in any one block is independent of an $x$ in any other block. Consequently,

$$E(U) = 0\,; \text{ and}$$
$$U^2 = \Sigma\,x_{ij}^2\,x_{pj}^2 + 2\,\Sigma\,x_{ij}\,x_{pj}\,x_{ql}\,x_{rl}. \qquad (3)$$

There are $\frac{1}{2}nm(m-1)$ terms in the first of these sums; and the expectation of each is $\{E(x_{ij}^2)\}^2 = \frac{(k^2-1)^2}{144}$. The only terms contributing anything to the expectation of the second sum are those for which the $j$th and $l$th objects occur together in each of two different blocks. Now $E(x_{ij}\,x_{ql}) = E(x_{pj}\,x_{rl}) = -\frac{k+1}{12}$ when the $j$th and $l$th objects are in the same block, and zero when they are not in the same block. Thus there is a contribution $2\left(\frac{k+1}{12}\right)^2$ whenever the $j$th and $l$th objects occur together in each of two different blocks. The number of such cases is $\frac{1}{2}n\,(n-1)\,\lambda\,(\lambda-1)$. Substituting in the expectation of (3) we have

$$E(U^2) = \frac{1}{288}\,\{nm\,(m-1)\,(k^2-1)^2 + n\,(n-1)\,\lambda\,(\lambda-1)\,(k+1)^2\}$$
$$= \frac{mn\,(k+1)\,(k^2-1)}{288}\,\{(m-1)\,(k-1) + \lambda - 1\},$$

since $\quad (n-1)\lambda = m\,(k-1)$.

Consequently $E(W) = \dfrac{k+1}{\lambda\,(n+1)}$

and
$$var\,(W) = \frac{2\,mn\,(k+1)\,(k^2-1)}{\lambda^4\,n^2\,(n^2-1)^2}\,\{(m-1)\,(k-1) + \lambda - 1\}$$
$$= \frac{2\,(k+1)^2}{mn\,\lambda^2\,(n+1)^2}\left(m - 1 + \frac{\lambda-1}{k-1}\right). \qquad (4)$$

When $k = n$, and therefore $\lambda = m$ (Kendall's case of '$m$ rankings'), we have

$$E(W) = \frac{1}{m},$$
$$var\,(W) = \frac{2\,(m-1)}{m^3\,(n-1)},$$

which agree with the formulæ given by Kendall (*loc. cit.*, p. 93).

The relative simplicity and the generality of these formulæ are due to the symmetry of the design, and in particular to the fact that each pair of objects occurs together an equal number of times in the same block. General formulæ for the higher moments cannot be obtained, since they would necessitate sets of three, four, etc., objects occurring equally frequently in the same block; and these conditions will only be satisfied by certain of the less useful designs. However, the higher moments are not needed, since the approximate distributions used for testing independence in the way described below require the first two moments only.

*Paired comparisons.* The case of paired comparisons, described in general terms by Kendall (pp. 121f.), is of special interest. Here $k = 2$; and to give the necessary symmetry $n$ must be even, and $m$ must be a multiple of $n - 1$. We find

$$E(W) = \frac{3\,(n-1)}{m\,(n+1)},$$

and

$$var\,(W) = \frac{18\,(n-1)^2}{m^3\,n\,(n+1)^2}\left(\frac{mn}{n-1} - 2\right).$$

If $m = n - 1$, then $E(W) = \frac{3}{n+1}$, and $var\,(W) = \frac{18\,(n-2)}{n\,(n^2-1)\,(n+1)}$.

*Ties.* The above results hold only when there are no ties in the rankings. However, for the case $k = n$, Kendall (p. 94) has shown that the presence of a small number of ties does not affect the moments of $W$ appreciably; and the result may be expected to apply equally to the incomplete block case. The appropriate modifications for a large number of ties are given by Kendall for the case $k = n$; but for large $m$ they are already rather complicated and for other values of $k$ they would be still more difficult to apply. A simpler procedure is to allocate ranks of tied objects at random, e.g., by tossing a coin. This does not affect the validity of tests based on $W$, and enables the same formulæ to be used in all cases.

A further method consists in allocating the ranks of tied members in such a way that the significance of $W$ is reduced. The effect of this is to make the test rather more stringent than the significance level would suggest.

## III. TEST OF INDEPENDENCE

Following Kendall we note that there are two approximations to the distribution of $W$ which are suitable for testing the hypothesis of independence among the $m$ rankings. The more accurate of these is the Beta distribution:

$$dF = \frac{1}{B(p,q)}\, W^{p-1}\,(1-W)^{q-1}\, dW. \tag{5}$$

Kendall (p. 93) has shown by calculating the third and fourth moments that for the case $k = n$ this gives a good approximation. By analogy one would expect the approximation to be satisfactory in the incomplete block case. Equating the first two moments with the mean and variance of $W$ we find we must take

$$\left.\begin{aligned} p &= \frac{mn\left(1 - \dfrac{k+1}{\lambda\,(n+1)}\right)}{2\left(\dfrac{nm}{n-1} - \dfrac{k}{k-1}\right)} - \frac{k+1}{\lambda\,(n+1)} \\ \text{and}\qquad q &= \left(\frac{\lambda\,(n+1)}{k+1} - 1\right) p \end{aligned}\right\}, \tag{6}$$

where $\lambda = \frac{m\,(k-1)}{n-1}$ as before.

When $k = n$, these values reduce to $p = \frac{1}{2}\,(n-1) - \frac{1}{m}$ and $q = (m-1)\,p$, as given by Kendall.

On putting $F = \frac{\left(\frac{\lambda\,(n+1)}{k+1} - 1\right) W}{1 - W}$, equation (5) reduces to Fisher's variance ratio distribution with $\nu_1 = 2p$, and $\nu_2 = 2q$ degrees of freedom, the values of $p$ and $q$ being given by (6).

For moderate $n$ and large $m$, $W$ tends to be distributed as a multiple of $\chi^2$ with $n - 1$ degrees of freedom, since $S$ is a sum of squares of $n$ variates connected by a linear relation, where the variate values are themselves sums of $m$ independent observations.

Putting $W = a\,\chi^2$, we have $E(W) = \frac{k+1}{\lambda\,(n+1)} = a\,(n-1)$. Thus $\chi_r^2 = \frac{\lambda\,(n^2-1)}{k+1}\,W$ tends to be distributed as $\chi^2$ with $n - 1$ degrees of freedom.

It is of interest to compare the variance of $\chi_r^2$ with $2(n-1)$, the variance of the fitted distribution.

$$\begin{aligned} var\,(\chi_r^2) &= \frac{\lambda^2(n^2-1)^2}{(k+1)^2}\,\frac{2(k+1)^2}{\lambda^2 mn(n+1)^2}\left(\frac{mn}{n-1}-\frac{k}{k-1}\right) \\ &= 2(n-1)\left\{1-\frac{k(n-1)}{mn(k-1)}\right\} \\ &= 2(n-1)\left(1-\frac{1}{m}\right) \text{ approximately.} \end{aligned}$$

(In the ordinary case of $m$ rankings $k = n$, so that $var\,(\chi_r^2)$ equals this value exactly.) This illustrates the fact that the governing factor in the accuracy of the $\chi^2$ approximation is the magnitude of $m$.

## IV. A NUMERICAL EXAMPLE

Consider by way of illustration the following example with imaginary data. Using the design already given (p. 86), a manufacturer has tested the preferences of a number of persons for seven varieties of ice-cream. The letters will now refer to the seven varieties ; and, as before, each column will represent a block. The design was repeated three times ; and the resulting 21 blocks were assigned at random to 21 tasters, each of whom was asked to rank the three relevant varieties in order of preference. The totals of the ranks assigned to the seven objects were (let us suppose) as follows :

| A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|
| 20 | 13 | 18 | 25 | 22 | 12 | 16 |

We require to test whether the preferences exhibit a significant departure from randomness.

The mean of the rank totals is 18. And $S = 2^2 + 5^2 + 0^2 + 7^2 + 4^2 + 6^2 + 2^2 = 134$. Further, $n = 7$, $m = 9$, and $k = 3$ : so that $\lambda = \frac{9\times 2}{6} = 3$. Hence,

$$W = \frac{12\,S}{\lambda^2 n(n^2-1)} = \cdot 532.$$

To test whether this value is significant we apply formula (6). We obtain $p = 2{\cdot}75$, and $q = 13{\cdot}75$. To test significance we require the 5 per cent. value of $F$ for $\nu_1 = 5{\cdot}5$ and $\nu_2 = 27{\cdot}5$ degrees of freedom. This is not tabulated ; but for $\nu_1 = 5$ and $\nu_2 = 27$ we have $F = 2{\cdot}57$. The corresponding value of $W$ is given by $2{\cdot}57 = \left(\frac{3\times 8}{4} - 1\right) W/(1-W)$, so that $W = \cdot 339$. The observed value is greater than this ; and therefore must certainly be greater than the value obtained from $F$ for $\nu_1 = 5{\cdot}5$ and $\nu_2 = 27{\cdot}5$. We infer that the observed value of $W$ is significant at the 5 per cent. level, and reject the hypothesis that preferences were random.

*Estimate of a true ranking.* A basic property of Kendall's coefficient $\tau$ is that it is simply related to the number of inversions of the natural order. Thus if the order of the objects is arranged so that one of the rankings is $1, 2, \ldots, n$, then $\tau = 1 - \frac{2\,Q}{\frac{1}{4}n(n-1)}$, where $Q$ is the number of inversions in the second set of ranks. For example, in the rankings

$$\begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 1 & 3 & 5, \end{matrix}$$

$Q = 3$, one for each of the inversions 21, 41, 43.

A similar property holds for the rank correlation, namely, $\rho = 1 - \frac{12\,V}{n^3-n}$, where $V$ is the sum of the weighted inversions ; here $V$ is similar to $Q$ except that each inversion is weighted by the difference between the two ranks in one of the rankings. For instance, in the above example we score 1 for the inversion 21, 3 for the inversion 41, and 1 for the inversion 43, giving $V = 5$. This gives $\rho = 1 - \frac{6\times 10}{n^3-n}$, which agrees with the usual formula $\rho = 1 - \frac{6\,\Sigma d^2}{n^3-n}$, since $\Sigma d^2 = 10$.

The obvious way of estimating the true ranking in a balanced design is to rank the objects according to the sums of the ranks allotted to them by the different observers. Kendall (p. 97 ) has shown for the case $k = n$ that this method of estimation maximizes the average value of $\rho$ for the correlation between the estimated and the observed rankings. It follows from the preceding paragraph that it also minimizes the sum of the weighted inversions in the observed rankings relative to the estimated ranking. This conclusion evidently remains true for the more general case of $k<n$.

Similar results hold for the case of tied ranks. If two objects are ranked equally in one ranking and strictly ordered in another, we call the comparison a violated tie and assign a score equal to half the corresponding score for an inversion. If the objects are tied in both rankings, there is, of course, no score. With these rules $V$ becomes the sum of the weighted inversions plus half the sum of the weighted violated ties. Thus, if the true ranking is estimated by ranking according to the sums of ranks, where ties have been dealt with by allotting fractional ranks, the method of estimation is such as to minimize the sum of the weighted inversions in the observed rankings relative to the estimated rankings, plus half the sum of the weighted violated ties.

## REFERENCES

1. Kendall, M. G. (1948). *Rank Correlation Methods.* London : Griffin.
2. Cochran, W. G., and Cox, G. M. (1950). *Experimental Designs.* New York : Wiley.
3. Hall, T., and Jones, D. Caradoc (1950). *Brit. J. Sociol.*, I., 31.