8203 D Sp 13

[*From* Biometrika, *Vol.* 46. *Parts* 3 *and* 4, *December* 1959.]

MISCELLANEA

The asymptotic efficiency of the χ_r^2-test for a balanced incomplete block design*

By PH. VAN ELTEREN AND G. E. NOETHER†
Mathematical Centre, Amsterdam, and Boston University

Friedman (1937) has shown how n treatments can be compared on the basis of the rankings of m 'observers'. This method has been extended by Durbin (1951) to cover a balanced incomplete block design, i.e. the case when each observer ranks $k \leqslant n$ treatments exactly once and each treatment is compared with any other treatment exactly λ times. We want to find the asymptotic ($m \to \infty$) relative efficiency (in the sense of Pitman, see e.g. Hannan, 1956) of Durbin's test with respect to the usual analysis of variance test for a balanced incomplete block design.

We note for future reference that

$$\lambda = \frac{l(k-1)}{n-1}, \tag{1}$$

where l is the total number of times a given treatment is used (replications).

The asymptotic relative efficiency is most easily obtained with the help of a formula given explicitly for the first time by Hannan (1956). It may be stated roughly in the following way. If both test statistics have, under the alternative hypothesis, non-central χ-square distributions with the same number of degrees of freedom, the asymptotic relative efficiency of one test with respect to the other test is equal to the ratio of the two non-centrality factors after the alternatives have been set equal.

Essentially, then, all we have to do is to compute the two non-centrality factors which we shall denote by d_r^2 and d_F^2 for the rank and F-tests, respectively. The conditions for the applicability of Hannan's formula can be shown to hold provided the underlying distributions satisfy some very general regularity conditions. These details will not be presented in this paper, since similar considerations have been given in several other papers, e.g. Andrews (1954), Benard & van Elteren (1953) and Bradley (1955).

Before computing the non-centrality factors, we have to specify the mathematical model which we are going to consider. The usual analysis of variance model suggests the following approach.

Let $F(x)$ be a continuous cumulative distribution function with density function $f(x) = F'(x)$. Let $\mathbf{x}_{\mu\nu}$ ($\mu = 1, 2, \ldots, m$; $\nu = 1, 2, \ldots, n$)‡ be the chance variable associated with the observation of the μth observer (block, in the analysis of variance) on the νth treatment. It is then assumed that the distribution $F_{\mu\nu}(x)$ of $\mathbf{x}_{\mu\nu}$ is given by

$$F_{\mu\nu}(x) = F(x+\theta_\nu+\eta_\mu),$$

where, without loss of generality, we may assume that $\sum_\nu \theta_\nu = 0$. The null hypothesis to be tested is that $\theta_1 = \theta_2 = \ldots = \theta_n = 0$. The alternatives to be considered specify that for a given number of replications l,

$$\theta_\nu = \theta_{\nu l} = \frac{\delta_\nu}{\sqrt{l}},$$

where the δ_ν are given constants satisfying

$$\sum_\nu \delta_\nu = 0. \tag{2}$$

For this model, we find (e.g. Anderson & Bancroft, 1952, §19·3) that

$$d_F^2 = lC\sum_\nu \theta_{\nu l}^2/\sigma^2 = C\sum_\nu \delta_\nu^2/\sigma^2, \tag{3}$$

where σ^2 is the variance associated with $F(x)$ and

$$C = \frac{n(k-1)}{k(n-1)} \tag{4}$$

is the *efficiency factor* of the given incomplete block design. The fact that for large m, F has approximately a non-central χ-square distribution follows from an argument similar to that given in Andrews (1954).

For computing d_r^2, let us introduce the following notation. Let $r_{\mu\nu}$ stand for the rank which the μth observer assigns to the νth treatment (assuming that he considers the νth treatment), and let $R_\nu = \sum_\mu^{(\nu)} r_{\mu\nu}$

* Report SP 63 of the Mathematical Centre, Amsterdam.

† Work supported by the Office of Naval Research.

‡ Here and below we use bold type to indicate chance variables.

where $\sum_{\mu}^{(\nu)}$ denotes summation over those l observers who have rated treatment ν. If we then set $\mathbf{u}_\nu = \mathbf{R}_\nu - \frac{1}{2}l(k+1)$, Durbin's test statistic is, except for a constant factor given by

$$\chi_r^2 = \frac{12}{n\lambda(k+1)} \sum_{\nu=1}^{n} \mathbf{u}_\nu^2,$$

which, for large l, has approximately a non-central χ-square distribution with $(n-1)$ degrees of freedom. We then have

$$d_r^2 = \frac{12}{n\lambda(k+1)} \sum_{\nu} \{\mathscr{E}(\mathbf{u}_\nu)\}^2.$$

Now
$$\mathscr{E}(\mathbf{R}_i) = \sum_{\mu}^{(i)} \sum_{j(\mu)\neq i} P\{\mathbf{x}_{\mu i} > \mathbf{x}_{\mu j(\mu)}\} + l, \tag{5}$$

where $j(\mu)$ refers to the treatments rated by the μth observer. Since

$$\begin{aligned} P\{\mathbf{x}_{\mu i} > \mathbf{x}_{\mu j(\mu)}\} &= \int F(x+\theta_{j(\mu)}+\eta_\mu) f(x+\theta_i+\eta_\mu)\,dx \\ &= \int F(x+\theta_{j(\mu)}-\theta_i) f(x)\,dx = \int F\left(x+\frac{\delta_{j(\mu)}-\delta_i}{\sqrt{l}}\right) f(x)\,dx \sim \frac{1}{2} + \frac{\delta_{j(\mu)}-\delta_i}{\sqrt{l}} \int f^2(x)\,dx, \end{aligned}$$

and since every treatment occurs exactly λ times together with every other treatment, (5), in view of (1) and (2), becomes

$$\mathscr{E}(\mathbf{R}_i) \sim \lambda \sum_{\substack{j=1\\ j\neq i}}^{n} \left[\frac{1}{2} + \frac{\delta_j - \delta_i}{\sqrt{l}} \int f^2(x)\,dx\right] + l = \frac{l(k+1)}{2} - \frac{n\lambda}{\sqrt{l}} \delta_i \int f^2(x)\,dx.$$

Thus
$$\mathscr{E}(\mathbf{u}_\nu) \sim -\frac{n\lambda}{\sqrt{l}} \delta_i \int f^2(x)\,dx,$$

and, finally,
$$d_r^2 \sim \frac{12n(k-1)}{(k+1)(n-1)} \left[\int f^2(x)\,dx\right]^2 \sum_{\nu} \delta_\nu^2. \tag{6}$$

From (3), (4) and (6), the asymptotic relative efficiency E_k of the χ_r^2-test with respect to the F-test is found to be

$$E_k = \frac{12k}{k+1} \left[\sigma \int f^2(x)\,dx\right]^2. \tag{7}$$

It is interesting to note that (7) depends only on the block size k, and not on the number of treatments.

We obtain the relative efficiency of the Friedman test by replacing k by n in (7).

In the special case when $f(x)$ is normal, (7) becomes

$$E_k(N) = \frac{3k}{\pi(k+1)}, \tag{7N}$$

some values of which are tabulated below:

k	2	3	4	5	6	7	8	9	10	17	∞
$E_k(N)$	0·64	0·72	0·76	0·80	0·82	0·84	0·85	0·86	0·87	$>0·90$	0·95

When $k = 2$, we have the case of paired comparisons where each observer is asked to compare only two treatments. In this case, (7) becomes

$$E_2 = 8\left[\sigma \int f^2(x)\,dx\right]^2, \tag{8}$$

and (7N),
$$E_2(N) = \frac{2}{\pi} = 0·637. \tag{8N}$$

The problem of paired comparisons has been of considerable interest during the last few years and many methods for treating the problem have been offered. It is easy to show that the χ_r^2-test for $k = 2$ is equal to the asymptotic form of the Bradley–Terry test discussed by Bradley (1955).

Bradley finds $n/\{\pi(n-1)\}$ as the asymptotic efficiency of his T-test relative to the analysis of variance test in case of normality where, following our notation, n is the number of treatments involved. The difference between Bradley's result and our own result (8N) is due to the fact that Bradley assumes a one way classification for the analysis of variance with l observations in each class instead of the balanced incomplete block design which we have considered. Since the variance σ^2 in (8) clearly refers to the within block variability, the latter model seems to be more appropriate.

If we multiply Bradley's result by $1/C$, where C is given by (4) with $k = 2$, lC being the effective number of replications of the balanced incomplete block design, we get (8N).

As far as the χ_r^2-test is concerned, instead of the model

$$F_{\mu\nu}(x) = F(x+\theta_\nu+\eta_\mu),$$

we could have considered the more general model

$$F_{\mu\nu}(x) = F_\mu(x+\theta_\nu),$$

where, for different μ, the $F_\mu(x)$ may be different cumulative distribution functions. Nothing much is gained as long as we are interested in an incomplete block design, since the non-centrality parameter d_r^2, and therefore the power of the χ_r^2-test, will depend on which particular k treatments are put in which particular block. However, in the case of Friedman's test when $k = n$, easy computations show that

$$d_r^2 = \frac{12n}{m^2(n+1)}\left[\sum_{\mu=1}^{m}\int f_\mu^2(x)\,dx\right]^2 \sum_{\nu=1}^{n}\delta_\nu^2, \tag{9}$$

where $f_\mu(x) = F'_\mu(x)$. (9) can be used to compute the asymptotic power of Friedman's test for this more general alternative.

For $k = n = 2$, the χ_r^2-test is equivalent to the two-sided sign test, a test whose properties have been more thoroughly investigated than those of any other distribution-free test. (8) then becomes the asymptotic efficiency of the sign test relative to the t-test, a quantity which is usually given (e.g. Hodges & Lehmann, 1956) as

$$E_{s,t} = 4\sigma^2 f^2(0). \tag{10}$$

The explanation is that (10) refers to the one-sample sign test while (8) refers to the two-sample sign test, a distinction which is rarely made. (10) can be used for the two-sample case if f is interpreted as the density of $\mathbf{x}-\mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y}$ are the two chance variables under consideration. From a practical point of view, it is more important to know the efficiency in terms of the individual distribution of $\mathbf{x}$ and $\mathbf{y}$. Then the answer is given by (8). As far as the two-sample sign test is concerned, it is, of course, easy to obtain (8) directly without reference to the χ_r^2-test.

The results obtained in this paper are asymptotic, i.e. valid for large numbers of replications. Nothing seems to be known about the relative efficiency of the χ_r^2-test for small numbers of replications except in the case $k = n = 2$. Since in this particular case the relative efficiency for a small number of replications is known to be much higher (in the case of normality) than indicated by the asymptotic formula, it seems reasonable to assume that the asymptotic values derived in this paper can be considered minimum values for the corresponding efficiencies.

REFERENCES

Anderson, R. L. & Bancroft, T. A. (1952). *Statistical Theory in Research*. New York: McGraw-Hill.

Andrews, F. C. (1954). Asymptotic behaviour of some rank tests for analysis of variance. *Ann. Math. Statist.* **25**, 724–35.

Benard, A. & van Elteren, Ph. (1953). A generalization of the method of m rankings. *Proc. Kon. Ned. Ak. van Wet.* A, **56**. *Indag. math.* **15**, 358–69.

Bradley, R. A. (1955). Rank analysis of incomplete block designs, III. *Biometrika*, **42**, 450–70.

Durbin, J. (1951). Incomplete blocks in ranking experiments. *Brit. J. Psychol.* **4**, 85–90.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Ass.* **32**, 675–99.

Hannan, E. J. (1956). The asymptotic powers of certain tests based on multiple correlations. *J. R. Statist. Soc.* B, **18**, 227–33.

Hodges, J. L., Jr. & Lehmann, E. L. (1956). The efficiency of some non-parametric competitors of the t-test. *Ann. Math. Statist.* **27**, 324–35.

8203