BU-663-M.

A balanced incomplete, repeated blocks design is a balanced incomplete block design for which one or more of the blocks have been repeated. To illustrate, there are 35 distinct blocks of 7 items taken 3 at a time. From this set, one may select 21 distinct blocks to form a variance balanced incomplete block design for 7 treatments in blocks of 3. Or, one may select 7 distinct blocks with each block repeated 3 times to obtain the 21 blocks. Likewise, it is possible to select 11, 13, 14, 15, 17, 18, 19, or 20 distinct blocks where some of the blocks are repeated. Thus, a class of variance balanced incomplete block designs is formed. The research in the literature to date has been on constructing this class of designs with emphasis on obtaining the minimum number of distinct blocks. For experimentation, it would be desirable to know which members of the class are better, in some sense, than others. This paper addresses the problem of devising criteria which distinguish among members of a class of designs. It was found that none of the presently used statistical criteria distinguish between members of a class. That is, the NN' matrix, where N is the treatment by block design matrix, is the same for all members of a class. However, N and $N'N$ are different for each design. Using $N'N$, or some function of $N'N$, members of the class of repeated block designs are distinguishable.

DISTINGUISHING CRITERIA FOR REPEATED BLOCK DESIGNS

by

W. T. Federer, F.-C. H. Lee, D. Raghavarao, and E. Seiden

BU-663-M* December, 1978

1. Introduction

Given that there are $\binom{v}{k} = b^*$ distinct blocks for a given block design, a repeated block design is one for which there are $d \leq b$ distinct blocks, for which the j^{th} block, $j=1,2,\cdots,b^*$, is repeated $w_j=0,1,2,\cdots$ times, and for which $\Sigma_1^{b^*} w_j = b$, the number of blocks in the design. Herein we consider only balanced incomplete block designs with repeated blocks, equal sized blocks, and equally replicated treatments. The balanced incomplete block designs have parameters v, b, d, k, r, and λ, where v is the number of treatments, b is the number of blocks, d is the number of distinct blocks, k is the block size, $k < v$, r is the number of times the i^{th} treatment, $i=1,2,\cdots,v$, is repeated, and λ is the number of times any pair of treatments occurs together in the b blocks. Also, we shall restrict the occurrence of the i^{th} treatment in the j^{th} block to be $n_{ij} = 0$ or 1, or the binary design.

One may construct a class of repeated block designs for a given set of parameters v, b, k, r, and λ. This then brings up the problem of obtaining criteria which distinguish between members of the class. Our purpose here is to consider a number of criteria and to demonstrate which will distinguish members of the class from each other and which will not. We first consider the class of repeated block designs for $v = 7$, $b = 21$, $k = 3$, $r = 9$, and $\lambda = 3$. Then, in the third section we present two theorems, the first showing which characteristics do not distinguish between members of the class and the second giving characteristics which distinguish among some or all members of the class.

* In the Mimeo Series of the Biometrics Unit, Cornell University.

2. Repeated Block Designs for $v = 7$, $b = 21$, $r = 9$, $k = 3$, $\lambda = 3$ and d Distinct Blocks

The total class of repeated block designs for $v = 7$, $b = 21$, $k = 3$, $r = 9$, $\lambda = 3$ which are balanced incomplete block designs for which $n_{ij} = 0$ or 1, is given in Table II.1. Any other designs may be shown to be isomorphic to the designs given. For example, there are several designs for $d = 7$ distinct blocks, but these are all isomorphic to each other. For the class of designs given in Table II.1, we note the following:

(i) $\lambda = 3$ for every design,

(ii) $N_{v\times b}N'_{b\times v}$ is identical for every design, where $N_{v\times b}$ is the incidence or design matrix of n_{ij}'s,

(iii) $rI - \frac{1}{k} NN' + \frac{\lambda}{k} J$, where J is a $v \times v$ matrix of ones, is identical for all designs,

(iv) intrablock and interblock means and variances of differences between these means are identical for all designs,

(v) the expected value of blocks (eliminating treatment effects) sum of squares is identical for all designs,

(vi) the two eigenvalues for $(kI_{b\times b} - \frac{1}{r} N'N + \frac{1}{r} J_{b\times b})^{-1}$ are 0.4286 and 0.3333 and are the same for all designs,

(vii) the multiplicities of the two eigenvalues in (vi) are identical for all designs, and are 7 and 14 respectively, and

(viii) the matrices N, $N'N$, $(kI_{b\times b} - \frac{1}{r} N'N + \frac{1}{r} J_{b\times b})^{-1}$ are different for all designs.

Since the matrix $(kI_{b\times b} - \frac{1}{r} N'N + \frac{1}{r} J)^{-1}$ differs for each design, one needs to study the matrices to determine how they differ and how they can be used to differentiate between the various block designs. Except for the two designs for

Table II.1 BIB(7,21,9,3,3)

Block composition	Number of distinct blocks (frequency of occurrence) 7	11	13	14	15	17	18	19	20	21
1 2 3	3	3	3	2	3	2	2	3	2	1
1 2 4	-	-	-	1	-	1	-	-	-	1
1 2 5	-	-	-	-	-	-	-	-	-	1
1 2 6	-	-	-	-	-	-	-	-	1	-
1 2 7	-	-	-	-	-	-	1	-	-	-
1 3 4	-	-	-	-	-	-	-	-	1	-
1 3 5	-	-	-	-	-	-	-	-	-	-
1 3 6	-	-	-	1	-	1	1	-	-	1
1 3 7	-	-	-	-	-	-	-	-	-	1
1 4 5	3	3	-	2	1	2	2	1	1	1
1 4 6	-	-	1	-	1	-	-	1	-	-
1 4 7	-	-	2	-	1	-	1	1	1	1
1 5 6	-	-	2	-	1	-	1	1	1	1
1 5 7	-	-	1	1	1	1	-	1	1	-
1 6 7	3	3	-	2	1	2	1	1	1	1
2 3 4	-	-	-	-	-	-	-	-	-	1
2 3 5	-	-	-	-	-	1	1	-	-	-
2 3 6	-	-	-	-	-	-	-	-	-	-
2 3 7	-	-	-	1	-	-	-	-	1	1
2 4 5	-	-	1	-	1	-	-	1	1	-
2 4 6	3	2	2	2	2	1	2	1	1	1
2 4 7	-	1	-	-	-	1	1	1	1	-
2 5 6	-	1	-	1	-	1	1	1	1	1
2 5 7	3	2	2	2	2	1	1	1	1	1
2 6 7	-	-	1	-	1	1	-	1	-	1
3 4 5	-	-	2	1	1	-	1	1	1	1
3 4 6	-	1	-	-	-	1	1	1	1	1
3 4 7	3	2	1	2	2	2	1	1	-	-
3 5 6	3	2	1	2	2	1	-	1	1	1
3 5 7	-	1	-	-	-	1	1	1	1	1
3 6 7	-	-	2	-	1	-	1	1	1	-
4 5 6	-	-	-	-	-	1	-	-	-	-
4 5 7	-	-	-	-	-	-	-	-	-	1
4 6 7	-	-	-	1	-	-	-	-	1	1
5 6 7	-	-	-	-	-	-	1	-	-	-

d = 7 and d = 21, there are four different variances among block means adjusted for treatment effects. These are of the form 2(a-b), 2(a-c) = 2(a-b/2), 2a, and 2(a+c). The frequency of occurrence of the various variances are given in Table II.2 where it may be noted that they differ for various values of d in most cases. It should be noted for d = 14 and d = 15, that the frequencies are identical. The same is true for the pair of designs d = 18 and d = 19. Thus, the frequency of occurrence of the various types of variances may be used to distinguish among most members of the class. The frequency of variances 2(a-b) decreases from 21 to zero as the number of distinct blocks increases from 7 to 21. The reverse situation holds for the variance 2(a+c). The frequency of the variance 2(a-c) is three times the frequency of the variance 2(a+c) for any given number of distinct blocks. The frequency of the variance 2a is simply $\binom{21}{2}$ = 210 minus the sum of frequencies of the other three types of variance.

Table II.2. Frequency of occurrence of variance of differences between block means adjusted for treatment effects.

Number of distinct blocks	Variance* 2(a-b)	2(a-c)	2a	2(a+c)	Average variance
7	21	0	189	0	2a-b/5
11	13	24	165	8	2a-b/5
13	9	36	153	12	2a-b/5
14	7	42	147	14	2a-b/5
15	7	42	147	14	2a-b/5
17	4	51	138	17	2a-b/5
18	3	54	135	18	2a-b/5
19	3	54	135	18	2a-b/5
20	1	60	129	20	2a-b/5
21	0	63	126	21	2a-b/5

* $a = .36508\sigma^2$, $b = .03174\sigma^2$, $c = b/2 = .01587\sigma^2$

3. Two Theorems on Repeated Block Designs

In experiment design theory, a number of criteria have been developed to ascertain the optimality of members of a class of experiment designs. It was surprising to the authors that at least some of these did not distinguish between members of a class of repeated block designs. The following theorem encompasses these findings:

Theorem 3.1. *In a class of repeated block designs with parameters* v, b, r, k, *and* λ, *all members of the class have the following in common:*

(i) *the value of* λ,

(ii) $N_{v\times b}N'_{b\times v}$ *and* $(rI_{v\times v} - \frac{1}{k} NN' + \frac{\lambda}{k} J)^{-1}$ *matrices*,

(iii) *intrablock and interblock treatment means and variances of differences between means*,

(iv) *the two eigenvalues of* $(kI_{b\times b} - \frac{1}{r} N'N + \frac{1}{r} J_{b\times b})^{-1}$ *and their multiplicities*,

(v) *the expected value of the blocks (eliminating treatment effects) mean square, and*

(vi) A-, D-, *and* E-*optimality for both treatment and block effects*.

Proof. By definition, $\lambda = r(k-1)/(v-1)$. Since v, k and r are identical for all repeated block designs in the class, λ is a constant for all designs. Since $NN' = (r-\lambda)I + \lambda J$ for a balanced basic binary incomplete block design and since r and λ are common for all designs, the matrix NN' is common to all designs. Likewise, since NN', r, k, $I_{v\times v}$, and $J_{v\times v}$ are common for all designs, then $(rI - \frac{1}{k} NN' + \frac{\lambda}{k} J)^{-1}$ is common for all designs.

The solution for intrablock treatment effects is

$$\hat{\underline{\tau}}_{v\times 1} = (rI - \frac{1}{k} NN' + \frac{\lambda}{k} J)^{-1} (\underline{Y}_t - N\underline{Y}_b/k) \quad , \qquad (3.1)$$

where $\underline{Y}_t$ is the v × 1 vector of treatment totals, $\hat{\underline{\tau}}$ is the v × 1 vector of estimated treatment effects, and $\underline{Y}_b$ is the b × 1 vector of block totals. Note that $(rI - \frac{1}{k} NN' + \frac{\lambda}{k} J)^{-1}$ and the expected value of $(\underline{Y}_t - N\underline{Y}_b/k)$ are identical for all designs. Thus, $\hat{\underline{\tau}}_v$ the intrablock solutions are identical for all designs. Since the matrix $(rI - \frac{1}{k} NN' + \frac{\lambda}{k} J)^{-1}$ is identical for all designs, the variance of a difference between two intrablock effects , or means, are identical for all designs. To obtain interblock solutions simply replace the three k's in (3.1) by $k + \hat{\sigma}^2_\epsilon/\hat{\sigma}^2_\beta$, where $\hat{\sigma}^2_\epsilon$ is the estimated intrablock mean square and $\hat{\sigma}^2_\beta$ is the estimated component of variance for block effects. This substitution does nothing to change any conclusions obtained for intrablock solutions. Hence, interblock solutions and variances are identical for all designs in a class.

The proof that there are two eigenvalues for the matrix $(kI_{b\times b} - \frac{1}{r} NN + \frac{1}{r} J)^{-1}$, that they are the same, and that their multiplicities are identical for all designs in a class is left for D. Raghavarao to prove.

Yates (1940), Rao (1947) and Federer(1955), section XIII.2.1, give the expected value of the mean square for blocks (eliminating treatment effects) as

$$\sigma^2_\epsilon + (kb-v)\sigma^2_\beta/(b-1) \qquad (3.2)$$

or under duplication of a basic set q times as

$$\sigma^2_\epsilon + (qkb-v)\sigma^2_\beta/(qb-1) \quad . \qquad (3.3)$$

Thus, the parameters v, b, and k are the only items involved. Hence, (3.2) is the expected value of the mean square for blocks (eliminating treatment effects) for all designs in the class.

Since A-, D-, and E-optimality depend on the eigenvalues of $(rI - \frac{1}{k} NN' + \frac{\lambda}{k} J)^{-1}$, the eigenvalues will be identical for all members of the class, and hence, all designs of the class will satisfy the same A-, D-, and E-optimality standards.

The above criteria are the standard ones for distinguishing between members of a class of designs. Since none distinguish among our class of repeated blocks designs, other distinguishing characteristics need to be found. Some are listed in Theorem 3.2.

Theorem 3.2. In a class of repeated block designs with parameters v, b, r, k, and λ, members of the class have the following distinguishing characteristics:

(i) the number of distinct blocks d,

(ii) the number of distinct rows of the matrices N'_{bXv} and N'N,

(iii) the frequency of the various types of variances of differences between two block effects,

(iv) the estimability of types of treatment effects from block totals, and

(v) the degrees of freedom, b - d, for an error mean square for treatment effects estimated from block totals.

Proof. By definition, one distinguishing characteristic of repeated block designs is the number of distinct blocks d. Since each row of N' appears w_j, j=1,2,$\cdots$,d, times, there will only be d distinct rows of N'. Likewise, there will be d distinct rows of N'N. Note that the rank of N'N is $v \leq b$.

The proof of (iii) does not appear to be forthcoming now. It should be pointed out that there is a relationship between the frequency of values of w_j and types of variances for the 7, 21, 9, 3, 3 designs in section two. The proof is omitted at this writing, hoping that one of the authors will be able to construct a proof for the frequency of occurrence of various types of variances for the general v, b, r, k, and λ case.

With respect to distinguishing criteria (iv) and (v), we first show that b - d is the number of degrees of freedom among blocks that are repeated. Note that the sum of squares among blocks that are repeated is $w_j - 1$ for any $w_j \geq 1$ and zero otherwise. Then $\Sigma_1^d(w_j-1) = b - d$. Thus, for the example in section 2, the numbers of degrees of freedom for various values of b - d are 21 - 7, 21 - 11, 21 - 13, 21 - 14, 21 - 15, 21 - 17, 21 - 18, 21 - 19, 21 - 20, and 21 - 21. The expected value of the mean squares for these items is $\sigma_\epsilon^2 + k\sigma_\beta^2$. The fewer the number of distinct blocks, the greater the number of degrees of freedom associated with this sum of squares.

With respect to criteria (iv), we first introduce a number of concepts. First, let us consider that block totals only are available to estimate treatment effects. This is the situation when mixtures of k items are used to form a single response, e.g., a mixture of k drugs, k varieties, k nutrients, k programs, etc. Various response model equations are possible. One such is:

$$Y_j = \mu + \sum_{i_1=1}^{v} n_{i_1 j}\alpha_{i_1}/k + \sum_{\substack{i_1=1 \\ i_1 \neq i_2}}^{v} \sum_{i_2=1}^{v} n_{i_1 j} n_{i_2 j} \gamma_{i_1 i_2}/k + \epsilon_j \quad , \tag{3.4}$$

where Y_j is the j^{th} block total, μ is an overall mean effect, α_{i_1} is a general mixing effect plus the effect of the treatment effect, $\gamma_{i_1 i_2}$ is an interaction or bispecific mixing effect of a pair of treatments $i_1 i_2$, $i_1 < i_2 = 2,\cdots,v$, $n_{i_1 j} = 0$ or 1, ϵ_j is IID$(0,\sigma_\epsilon^2 + k\sigma_\beta^2)$, and the other symbols are as defined previously. Given a set of restraints, say $\Sigma\, \alpha_{i_1} = 0 = \Sigma_{i_1 \neq i_2}\, \gamma_{i_1 i_2} = \Sigma_{i_2 \neq i_1}\, \gamma_{i_1 i_2}$, unique solutions to the effects are possible when $d \geq v(v-1)/2$. Suppose that the response model equation is of the form:

$$Y_j = \mu + \sum_{i_1=1}^{v} n_{i_1 j}\alpha_{i_1}/k + 2\sum_{i_1=1}^{v-1}\sum_{i_2=2}^{v} n_{i_1 j}n_{i_2 j}\gamma_{i_1 i_2}/k$$
$$+ 3\sum_{i_1=1}^{v-2}\sum_{i_2=2}^{v-1}\sum_{\substack{i_3=3 \\ i_1 < i_2 < i_3}}^{v} n_{i_1 j}n_{i_2 j}n_{i_3 j}\pi_{i_1 i_2 i_3}/k + \epsilon_j \quad , \qquad (3.5)$$

where $\pi_{i_1 i_2 i_3}$ is a trispecific mixing effect of the triplet of treatments $i_1 i_2 i_3$ and $i_1 < i_2 < i_3 = 3, \cdots, v$. In order to obtain solutions for all parameters under the additional restraints $\Sigma^{v}_{\substack{i_1=1 \\ i_1 \neq i_2 \; i_1 \neq i_3 \; i_2 \neq i_3}} \pi_{i_1 i_2 i_3} = 0$, it is necessary that $d \geq v(v-1)(v-2)/6$ and that the block size be $k \geq 3$. It is also required that $v \geq 6$.

One can continue extending response model equations (3.4) and (3.5) to include quater specific mixing effects for quartets of treatments, $\cdot\cdot\cdot$, n^{th} specific mixing effect for n treatments. The more effects in the response model equation, the larger the value of d required to obtain unique solutions for the effects. Thus, for statistical designs for mixtures of items, and for block totals only available, a distinguishing criterion is available for any class of repeated block designs.

For those unfamilar with statistical designs for mixtures, we have included a bibliography of such work. It should be noted that there is a large literature on statistical designs for mixtures of size $k = 2$ under the heading of diallel crossing, matched pairs, and tournaments. Randall (1976) did a literature coverage for diallel crossing designs and presented unifying and extended procedures. Davidson and Farquhar (1976) gave a bibliography on the method of paired comparisons. Both bibliographies made use of the Federer and Balaam (1972) bibliography.

Moon (1968) gave a bibliography on tournaments. The bibliography at the end of the paper contains most of the references for statistical designs for mixtures for the case $k \geq 3$ and for equal proportions in the mixture.

4. Discussion

It was surprising that at least one of the criteria listed in Theorem 3.1 did not distinguish among members of a class of repeated block designs, while at the same time it was gratifying to be able to state five characteristics in Theorem 3.2 which would distinguish among them. It is felt that more work is needed in this area. One should not feel secure that present criteria will suffice for future needs. There has been considerable emphasis on such criteria as A-, D-, and E-optimality. However, these were not usable for distinguishing among the members of a class of repeated block designs.

5. Bibliography on Statistical Designs for $k \geq 3$ mixtures

Cooper, J. (1968). A statistical analysis for data from an experiment involving all possible mixtures of pairs of p varieties applied to mixtures of 8 bean varieties. B.S. Thesis, Biometrics Unit, Cornell University.

Davidson, R. R. and P. H. Farquhar. (1976). A bibliography of paired comparisons. Biometrics 32:241-252.

Federer, W. T. (1967). Analyses for diallel crossing and matched pairs schemes when individual contributions are measurable. No. BU-246-M in the Mimeo Series of the Biometrics Unit, Cornell University.

Federer, W. T. (1975). Statistical designs for mixtures of crops and other applications. No. BU-560-M in the Mimeo Series of the Biometrics Unit, Cornell University.

Federer, W. T. and L. N. Balaam. (1972). Bibliography on Experiment and Treatment Design Pre 1968. Published for International Statistical Institute by Oliver and Boyd, Edinburgh.

Federer, W. T. and D. B. Hall. (1975). Statistical designs for mixtures. No. BU-570-M in the Mimeo Series of the Biometrics Unit, Cornell University.

Federer, W. T., A. Hedayat, C. C. Lowe, and D. Raghavarao. (1976). An application of statistical design theory to crop estimation with special reference to legumes and mixtures of cultivars. Agronomy Journal 68(6):914-919.

Free, Jr., S. M. and J. W. Wilson. (1964). A mathematical contribution to structure-activity studies. Journal Medical Chemistry 7:395-399.

Hall, D. B. (1976). Mixing designs: A general model, simplifications, and some minimal designs. M.S. Thesis, Biometrics Unit, Cornell University.

Hall, D. B. and W. T. Federer. (1977). Model equations, treatment population, and mixtures. No. BU-605-M in the Mimeo Series of the Biometrics Unit, Cornell University.

Moon, J. W. (1968). Topics on Tournaments. Holt, Rhinehart, and Winston, New York.

Raghavarao, D. and W. T. Federer. (1979). Block total response as an alternative to the randomized response method in surveys. Journal Royal Statistical Society, B, 41(1): (in print).

Smith, L. L., W. T. Federer, and D. Raghavarao. (1974). A comparison of three techniques for eliciting truthful answers to sensitive questions. Social Statistics Section Proceedings of the American Statistical Association, pp. 447-452.