Application of BIBDR in health sciences using R

Cite as: AIP Conference Proceedings **2293**, 420109 (2020); https://doi.org/10.1063/5.0031657
Published Online: 25 November 2020

Teresa A. Oliveira, and Amílcar Oliveira

Application of BIBDR in Health Sciences using R

Teresa A. Oliveira[1,2, a)] Amílcar Oliveira[1,2]

[1]Universidade Aberta, Palácio Ceia, Rua da Escola Politécnica, Lisboa, Portugal
[2]Center of Statistics and Applications, University of Lisbon, Portugal

Corresponding author: [a)] Teresa.Oliveira@uab.pt

Abstract. The role of Experimental Design is very well known, considering applications to a broad range of areas, such as Agriculture, Biology, Medicine, Industry, Education, Economy, Engineering and Food Consumption Sciences. Motivated by the variety of problems faced in the several areas and simultaneously taking advantage of the emerging technological developments, new theoretical results, as well as new designs and structures, have been developed by researchers and practitioners accordingly to the needs. Experimental Design got a place among the most important statistical methodologies and, mainly because of allowing to separate variation sources, since the last century it has been strongly recommended for Health Sciences studies. In this area, particular attention has been devoted to Randomized Complete Block Designs and to Balanced Incomplete Block Designs (BIBD) - which allow testing simultaneously a number of treatments bigger than the block size. Thus, after a brief review of some particular BIBD properties and of BIBDR - Balanced Incomplete Blocks with Block Repetition, an applications to Health Sciences simulated data is illustrated, by exploring R software.

INTRODUCTION

RA Fisher during the first part of the 20th century established the foundations of Experimental Design. Considering the implementation of a particular Design, the experimenter must focus and look for the best strategies in the following steps: planning and selecting the design, conducting the experiments, collecting the observed data and providing the analysis and a clear interpretation of results. The aim is to provide a deep understanding of the problem and a powerful experimental process at a reduced cost. Experimental design has countless application possibilities in different areas, such as Agriculture, Biology, Medicine, Health Sciences, Industry, Education, Economy, Engineering and Food Consumption Sciences.

Block designs group similar units into blocks, so that variation among it within the blocks is reduced. Blocking variables are a property of the experimental units and not something to be manipulated, as the varieties or treatments concerning their respective levels. Incomplete block designs also group units into blocks, however the block units are not enough to accommodate all the treatments. As a particular branch of Experimental Design, Balanced Incomplete Block Designs - BIBD, proposed by Yates, F. (1936), remain in the forefront of research areas and there are still many open questions on these designs at the theoretical level. The five parameters (ν, b, r, k, λ) characterize a BIBD as an arrangement of ν varieties or treatments in b blocks of size k< ν, where each variety occur r times and every pair of varieties concur in exactly λ blocks, see for example Caliński and Kageyama (2000).

The necessary but not sufficient conditions for the existence of a BIBD are given by:

$$vr = bk\ , \quad r(k-1) = \lambda(\nu-1)\ , \quad b \geq \nu$$

The efficiency factor of a design is given by E=(ν(k-1))/(k(ν-1)) and, taking into account that k<ν, then, for a BIBD, the efficiency factor is smaller than unity and has reduced efficiency by comparison with a Complete Randomized Design (CRD) with the same b blocks of size k. For a given BIBD, the maximum percentage reduction in efficiency compared to a CRD with the same numbers of treatments and observations is given by 100(1-E).

The number of distinct blocks in a design is the design cardinality and it is denoted by b*. When b*<b then the design has one or more repeated blocks. The general and optimal properties of a BIBD are not affected by the presence of block repetition and the efficiency of the design neither. Foody and Hedayat (1977) and Hedayat and Hwang (1984)

International Conference of Numerical Analysis and Applied Mathematics ICNAAM 2019
AIP Conf. Proc. 2293, 420109-1–420109-4; https://doi.org/10.1063/5.0031657
Published by AIP Publishing. 978-0-7354-4025-8/$30.00

presented some examples of practical situations for which it is desirable to have block repetitions in a design. Many authors have been developing research on exploring the properties of BIBD with repeated blocks (BIBDR), see for example Raghavarao et al. (1986) and Ghosh and Shrivastava (2001).

BIBD with repeated blocks have been successfully used in areas like Psychiatry, Medicine, Education and Food Sciences, as a tool to assess agreement and to compare inter-rater reliability. In Health Sciences and particularly in the case of Pediatrics patients, it is extremely important that there is an agreement on the diagnosis provided by different physicians. Designs with block repetition, as a way to simplify the doctor's allocation on team daily routines for patient's examinations, are thus the preferred ones.

Consider a practical study on the inter-rater reliability of the Pneumonia diagnosis in children, based on a rate resultant from the observations of, among other variables, the percent of Peripheral Oxygen Saturation (SpO2) in environmental air. The aim is to test if there are significant differences between the specialists who evaluate the performance of patients by this rating scale, denoted by RS, in order to help on the decision whether to admit the patient to stay in the hospital or not. In this study, a practical situation is simulated, and it is considered that the patients were observed by different specialists and the RS on assessing the impact of Pneumonia disease in each patient was registered. By assigning treatments to the subjects following a particular BIBD (9, 24, 8, 3, 2) structure with repeated blocks, it is illustrated the usefulness of BIB Designs as a powerful tool to solve emerging similar problems in Health Sciences environment. Software R was used in all the computations for the example illustration.

SELLECTING A BIBD WITH REPEATED BLOCKS

Consider a BIBD with incidence matrix $\mathbf{N} = (n_{ij})$, where n_{ij} is the number of times that the i^th^ variety occurs in the j^th^ block, $i = 1,2,\dots,v$; $j = 1,2,\dots,b$, and the coefficient matrix $\boldsymbol{D}$, for estimating a vector of block effects $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_b)'$, such that:

$$\mathbf{D} = \mathrm{k}\mathbf{I}_{\mathrm{b}} - \frac{1}{\mathrm{r}}\mathbf{N}'\mathbf{N},$$

where $\mathbf{I}_b$ denotes a $b \times b$ identity matrix, see for example Raghavarao *et al*. (1986).

Let B_j and $B_{j'}$ represent any two blocks of a BIBD with h treatments in common. The $(j, j')^{\text{th}}$ element of the matrix $\mathbf{N}'\mathbf{N}$ is equal to h and the variance of the difference of block effects is a known result:

$$\mathrm{Var}\left(\hat{\beta}_j - \hat{\beta}_{j'}\right) = 2\sigma^2\,(v\lambda + k - h)/vk\lambda \tag{1.1}$$

Once h can take at most $k+1$ values, namely $0,1,\dots,k$, so there will be at most 4 possible variance values for the estimated elementary block effect contrasts, in the example of BIBD(9, 24, 8, 3, 2).

It is easy to show the equivalence between (1.1) and the following expression:

$$\mathrm{Var}\left(\hat{\beta}_j - \hat{\beta}_{j'}\right) = 2\sigma^2\left(\frac{1}{k} + \frac{k-h}{vk\lambda}\right),\ h = 0,1,\dots,k \tag{1.2}$$

A total of 36 non-isomorphic BIBD (9, 24, 8, 3, 2) are presented in CRC Handbook of Combinatorial Designs (2006), pag.28. Consider the structures 1 and 36 (from the referred page), corresponding to b*=12 and b*=24. For the comparison of variance block effects in the design, between the case with twelve repeated blocks and the case with no repeated blocks, the treatments 1, 2, 3, 4, 5, 6, 7, 8, 9 were considered.

Structure 1: In this case b*=24 and the design is given by the blocks:

123,124,135,146,157,168,179,189,236,245,257,269,278,289,348,349,359,367,378,458,467,479,568,569

There are no repeated blocks and the possible values for h are then 0, 1 or 2. To compare any two blocks of the design there are 276 possible comparisons, and between them there are:

(i) Blocks which have two common treatments, for example 123 and 124 (block 1 and block 2). For such cases:

$$\mathrm{Var}\left(\hat{\beta}_1 - \hat{\beta}_2\right) = \frac{19}{27}\sigma^2.$$

(ii) Blocks which have one common treatment, for example 135 and 146. For such blocks and particularly using block 3 and block 4:

$$\mathrm{Var}\left(\hat{\beta}_3 - \hat{\beta}_4\right) = 2\sigma^2\left(\frac{1}{3} + \frac{3-1}{9\cdot 3\cdot 2}\right) = \frac{20}{27}\sigma^2 \cdot$$

(iii) Blocks which have no common treatments, for example 123 and 569:

$$\mathrm{Var}\left(\hat{\beta}_1 - \hat{\beta}_{24}\right) = 2\sigma^2\left(\frac{1}{3} + \frac{3-0}{9\cdot 3\cdot 2}\right) = \frac{7}{9}\sigma^2 \cdot$$

Structure 2: In this case b*=12 and the design is given by the blocks:

123,123,145,145,167,167,189,189,246,246,258,258,279,279,349,349,357,357,368,368,478,478,569,569

Once there are twelve blocks twice repeated, the possible values of h are 0, 1, 2 and 3. For blocks which have three common treatments, for example blocks 5 and 6, (167), then:

$$\mathrm{Var}\left(\hat{\beta}_5 - \hat{\beta}_6\right) = 2\sigma^2\left(\frac{1}{3} + \frac{3-3}{9\cdot 3\cdot 2}\right) = \frac{2}{3}\sigma^2 \quad \cdot$$

The differences of the variance of block effects for $h = 0,1,2$ are the same as for Structure 1.

The simulated example structure 2 of BIBD (9, 24, 8, 3, 2) has the maximum number of blocks with minimum variance of the difference of block effects.

APPLICATION: BIBD AND R PROJECT FOR STATISTICAL COMPUTING

In order to generate balanced incomplete block designs with the aid of the computational statistical program R, the extra '*package*', '*crossdes*' was used. This '*package*' includes several functions that assist in building balanced designs. This study aims to stress the importance of BIBD with repeated blocks applications in Health Sciences and to illustrate the role of using software R for computations. On the example, data were simulated for the RS (resulting rating scale) with the objective to investigate if there are significant differences between the patients' assessments, observed by several specialists/examiners in a study of inter-rater reliability.

Suppose that for twenty four patients there is permission of participation and that each patient enrolled in the study will be examined, separately and independently by nine specialists. Each patient sets a block and each patient cannot tolerate more than three exams, so the block size is three (three exam observations per patient). Thus, the maximum block size, k=3, is smaller than the number of treatments (varieties) to be compared, v = 9 specialists/examiners. Consider that each of the nine specialists examine eight patients and that together each pair of specialists is examining two patients, for example examiners 2 and 3 examine patients 1 and 2, but both together don't examine any other patient. This case fits the conditions for the application of a BIBD (9, 24, 8, 3, 2). Using R software, the ANOVA table was displayed for testing the effect of assessments caused by different specialists (figure 1). Simulated results for RS are presented in Table 1. The considered null and alternative hypotheses are respectively,

H_0: No differences between specialists assessment vs H_1: At least one specialist differs from the others

```
> values = c(60,65,61,27,32,25,46,52,48,43,43,42,25,25,26,60,56,57,38,40,42,44,46,40,112,98,110,83,83,82,73,70,75,
+ 111,110,105,56,60,63,88,87,90,55,65,62,72,67,70,112,106,108,78,75,76,65,60,62,78,77,78,42,37,42,65,63,66,
+ 46,50,48,76,80,77)
> examiner = c(1,2,3,1,2,3,1,4,5,1,4,5,1,6,7,1,6,7,1,8,9,1,8,9,2,4,6,2,4,6,2,5,8,2,5,8,2,7,9,2,7,9,3,4,9,3,4,9,
+ 3,5,7,3,5,7,3,6,8,3,6,8,4,7,8,4,7,8,5,6,9,5,6,9)
> |
> patient = factor(kronecker(1:24,c(1,1,1)))
> options(contrasts=c("contr.treatment", "contr.poly"))
> lm.adjust1 = lm(values~patient+examiner)
> anova(lm.adjust1)
Analysis of Variance Table

Response: values
          Df Sum Sq Mean Sq  F value Pr(>F)
patient   23  38252 1663.14 187.5727 <2e-16 ***
examiner   8     59    7.33   0.8271 0.5839
Residuals 40    355    8.87
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
> qf(0.95, df1=8, df2=40)
[1] 2.18017
```

FIGURE 1. Illustration of R commands adjusted to our example and the analysis of variance results

The p-value indicates that null hypothesis is not rejected, at the significance level 5%, thus there is no evidence of differences due to the specialists: there is concordance or agreement in their diagnosis.

	Specialists/Examiners									
Patients	1	2	3	4	5	6	7	8	9	**Mean**
1	60	65	61							62,00
2	27	32	25							28,00
3	46			52	48					48,66
4	43			43	42					42,66
5	25					25	26			25,33
6	60					56	57			57,66
7	38							40	42	40,00
8	44							46	40	43,33
9		112		98		110				106,66
14		83		83		82				82,66
11		73			70			75		72,66
12		111			110			105		108,66
13		56					60		63	59,66
14		88					87		90	88,33
15			55	65					62	60,66
16			72	67					70	69,66
17			112		106		108			108,66
18			78		75		76			76,33
19			65			60		62		62,33
20			78			77		78		77,66
21				42			37	42		40,33
22				65			63	66		64,66
23					46	50			48	48,00
24					76	80			77	77,66
Mean	42,87	77,50	68,25	64,37	71,62	67,50	64,25	64,25	61,50	

TABLE 1. Simulated observations for the example considering BIBD (9, 24, 8, 3, 2|b*=12)

RESULTS DISCUSSION

On the example, the analysis points out that there is a good consistency between the different specialists, and so the use of the RS in assessing the impact of Pneumonia disease in patient's observation could be recommended. Also the preference for using a BIBD with repeated blocks was emphasized considering the simplification of doctor's team schedule routines. These designs are useful not just because of easier implementation, reduced costs and more precision in the results, but they are preferable once the block repetition allow to have a maximum number of blocks with minimum variance of the difference of block effects. The role of BIBD with block repetition in similar applications and in other research areas will be explored in the near future.

ACKNOWLEDGMENTS

This research was partially supported by Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UID/MAT/00006/2019 (CEA/UL).

REFERENCES

1. Caliński, T. and Kageyama, S. *Block Designs: A Randomization Approach*. Vol. I: Analysis. Springer, (2000).
2. Colbourn, C.J. and Jeffrey H. Dinitz, J.H. Handbook of Combinatorial Designs, 2nd Ed., Chap.and Hall/CRC, (2006).
3. Fisher, R. A. *Statistical methods for research workers*. New York, Hafner Press, (1925).
4. Foody W., Hedayat A., *On theory and applications of BIB designs with repeated blocks*. The Annals of Statistics 5, 932–945, (1977).
5. Ghosh D.K., Shrivastava S.B. *A class of BIB designs with repeated blocks*. Journal of Applied Statistics 28, 821-833, (2001).
6. Hedayat A., Hwang H.L. *BIB (8, 56, 21, 3, 6) and BIB (10, 30, 9, 3, 2) designs with repeated blocks*. Journal of Combinatorial Theory, A36, 73 – 91, (1984).
7. Raghavarao, D., Federer W.T., Schwager S.J. (1986). *Characteristics for distinguishing among BIB designs with repeated blocks*. Journal of Statistical Planning and Inference 13, 151-163, (1986).
8. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, (2012).