Beyond Balanced Incomplete Block Designs: Addressing challenges, connections, applications and R

Oliveira, T.A.[1,3], Oliveira, A.[1],Correia, H.[2]
[1]Universidade Aberta and CEAUL, Lisbon,
[2]MEMeC – Universidade Aberta, Lisbon, Portugal
[3]Corresponding author: Teresa Oliveira, email: toliveir@uab.pt

Abstract

Experimental Designs have been widely used, among others, in agricultural, industrial and computer experiments, in order to reduce experimental cost and to provide efficient parameter estimation. Balanced Incomplete Block Designs (BIBD) play a key role in this context: in addition to their optimal properties and to the possible applications in various areas of science, these designs present extremely relevant links, highlighted in branches of Statistics like Sampling Theory. Some literature review is presented exploiting the combinatorial properties of various BIBD to construct controlled sampling designs with minimum number of distinct blocks, as well the most relevant methods on the construction of BIBD with repeated blocks (BIBDR). An application of BIBDR in the area of Education was performed using software R, with the aim of comparing five fields of algebraic thinking in a sample consisting of students in the 1st year of higher education in Cape Verde.

Key Words: BIBD, Education Sciences, Experimental Design, Sampling

1. Introduction

The use of blocks is very important in experimental design since it allows to reduce or eliminate the variability introduced by factors that can influence the experience but in which the researcher is not interested. Balanced Incomplete Block Designs with block Repetition (BIBDR) arise in order to reduce the costs of experiment implementation, allowing also to get viable results even when some experimental units are lost. Such designs were presented by Foody and Hedayat (1977), and the trade-off method for the construction of BIBDR was developed by Hedayat and Li (1979).
Currently, many researchers in various areas demonstrate the importance of BIBD applications as a contribution to the advancement of Science. The use of BIBD in Sampling Theory first appeared in the work presented by Chakrabarti (1963), in which the author relates the structure of such designs with the notion of Support of a Sampling Plan. The Support of a Sampling Plan is the number of distinct samples with positive probability of being selected. Also in the controlled sampling, BIBD are used to obtain samples with a minimum support size - minimum number of distinct blocks - and thus to identify the maximum number of distinct blocks with non-preferred samples, see Avadhani and Sukhatme (1973). In the area of Education Sciences, there are yet few published papers using BIBD applications. We refer to Yang (1985) who presents an application of BIBD in which the aim is to estimate the reliability of an oral examination of the faculty of medicine at the University of Iowa. Also Van der Linden et al. (2004) showed how the problem of calculating an optimal BIBD can be regarded as a combinatorial optimization problem, considering a sample of 1996 eighth graders in Mathematics from the NAEP (National Assessment of Educational Progress). In our work we present an original application of BIBDR in the area of Education, using R on computations, aiming to compare five fields of algebraic thinking, considering as sample the results of a questionnaire performed to students in the 1st year of higher education in Cape Verde.

2. Balanced Incomplete Block Designs (BIBD) and block repetition

2.1 The concept

When in the experimental design the number of repetitions for all pairs of ν varieties (or treatments) is the same, we represent it by λ, and we have the so-called Balanced Incomplete Block Designs (BIBD). A BIBD is a proper equi-replicated binary design. In a BIBD all the b blocks have the same number of treatments, k, concurring r times along the blocks and usually this is represented as BIBD(ν,b,r,k,λ). A BIBD is then characterized by five non-independent integer parameters, such that:

(i) $n = r\nu = bk$, (ii) $r(k-1) = \lambda(\nu-1)$, (iii) $b \geq \nu$ (Fisher inequality)

These conditions are necessary but not sufficient for the existence of a BIBD, and still this is an open problem. For example considering $\nu = b = 6$ and $r = k = 3$ we obtain $\lambda = 6/5$ which is not integer. Also if we consider $\nu = 15, b = 21, r = 7, k = 5$ we have $\lambda = 2$, which is integer and however it was proved the inexistence of such BIBD, see Hedayat and Hwang (1984).

If in a BIBD there are less than b distinct blocks, then the design has repeated blocks. The set of all distinct blocks in a BIBD is called the support of the design, and the design cardinality is represented by b*. The notation BIBD$(\nu,b,r,k,\lambda|b^*)$ is used to denote a BIBD(ν,b,r,k,λ) with precisely b* distinct blocks. Oliveira (2010) present some developments on this issue.

2.2 The statistical model

Consider a BIBD(ν,b,r,k,λ) satisfying the conditions (i), (ii) and (iii).

The statistical model for this design is given by:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad \begin{cases} i = 1,..,\nu \\ j = 1,...,b \end{cases}$$

where y_{ij} represents the i-th observation in j-th block, μ is the general mean, τ_i denotes the effect of i-th variety (or treatment), β_j denotes the effect of j-th block, and ε_{ij} represents the random error. The model errors are assumed to be to be normally and independently distributed random variables with mean zero and variance σ^2, $\varepsilon_{ij} \approx N(0,\sigma^2)$. In our application the analysis intra-block was performed since the differences between blocks are eliminated and all of the contrast of treatment effects can be expressed as comparison of observations in the same block.

3. BIBD and Sampling Theory

3.1 Some connections between BIBD and Sampling Theory

BIBD with reduced support size are of particular importance in sampling theory. Indeed, the smaller the number of different blocks in a BIBD, the better is the possibility of adapting it to a given situation in controlled sampling, see Ouyang and Srivastava (1992). Controlled sampling is a sample collection method which reduces the probability of selecting certain undesirable combinations while maintaining the properties of a non controlled associated plan. Investigators such as Avadhani and Sukthame (1973) used BIBD in controlled sampling and the results are a reflect of the work already presented by Chakrabarti (1963). BIBD are used, considering the parameters $\nu = N, b < \binom{N}{n}, r, k = n, \lambda$, on the construction of Simple Random Controlled Samples. Selecting a block randomly, which means with probability $1/b$, the inclusion probabilities of order 1, π_i, and of order 2, π_{ij} remain the same. In fact, we can write: $\pi_i = \frac{r}{b} = \frac{k}{v} = \frac{n}{N}$ and $\pi_{ij} = \frac{\lambda}{b} = \frac{k(k-1)}{v(v-1)} = \frac{n(n-1)}{N(N-1)}$.

These probabilities are the same obtained to Simple Random Non-Controlled Sampling without replacement.

3.2 Example

Consider a situation so that it is desired to conduct a sample survey to obtain reliable estimate of the yield rate and the total production of olive oil, in a hilly district with wide spread villages. Our example is based on Avadhani and Sukhatme (1973). There are N=7 villages in the district, i=1,…,7, located as it is represented by the diagram in figure 1 and we want to select a sample of 3 from these 7 villages.

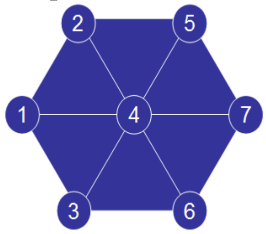

Figure 1: diagram for 7 units possible spatial distribution

Any two units (villages) which are connected by a single line are considered as neighbours. A sample $w=(i_1,i_2,i_3)$ is considered preferred if and only if there is a line between i_1 and i_2 , as well as a line between i_2 and i_3.

The total number of possible samples is $C_3^7 = 35$, as represented in the Table 1.

Table 1. Representation of all possible C_3^7 (each column also represents a block size 3)

3	4	5	6	7	4	5	6	7	5	6	7	6	7	7	4	5	6	7	5	6	7	6	7	7	5	6	7	6	7	7	6	7	7	7
2	2	2	2	2	3	3	3	3	4	4	4	5	5	6	3	3	3	3	4	4	4	5	5	6	4	4	4	5	5	6	5	5	6	6
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	4	4	4	5

The total number of preferred samples is 21 and the number of non preferred samples is 14. If each of the 35 combinations has equal chance of being selected as in Simple Random Sampling, with probability 1/35, then the probability of obtaining a non preferred sample is 14/35=0.4. Consider a BIBD(7,7,3,3,1) with varieties 1,…7, with Structure 1 represented by (246),(367),(147),(123),(156),(345). In this case only the block(156) is non preferred, since there is no line connecting 1 to 5 and 5 to 6, see figure 2. The probability of obtain a non preferred sample is 1/7<14/35. Consider the BIBD(7,7,3,3,1) with Structure 2 (134),(245),(356),(467),(126),(237),(157). Then, Blocks (356),(157),(126) and (237) are non preferred, see figure 3. Thus, the probability to obtain a non preferred sample is 4/7>14/35.

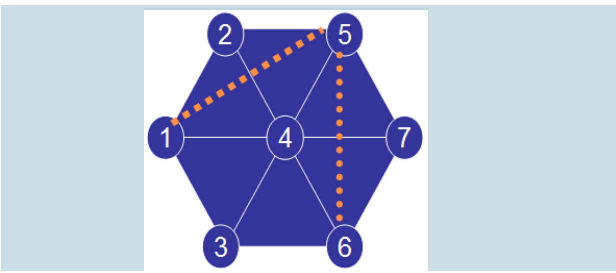

Figure 2. Structure 1, non preferred block

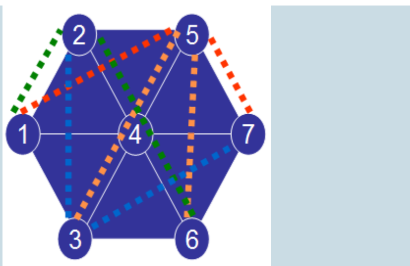

Figure3. Structure 2, non preferred blocks

As we can observe, the wrong choice of the BIBD may even make the controlled case worse than the uncontrolled case. It is then crucial to look for a selection criteria allowing to choose the best structures in each situation.

4. BIBDR: revisiting some construction methods

4.1 Method Trade-off

The main idea behind the trade-off method, developed by Hedayat and Li (1979), is to compensate for, or replace, some blocks with other blocks, so that the general characteristics that define a BIBD are maintained. Let $v > k > t \geq 1$ be three integer, and let $P_l(X)$ denote the set of all l-subsets based in a set X with dimension v.

A trade T, represented by $T(t,k,v)$, based in the elements (blocks) of $P_k(X)$ is a pair of blocks collections, non empty and disjoint, T_1 and T_2, each with *m* blocks of $P_k(X)$, so that the number of times each element of $P_t(X)$ is replaced by T_1 is equal to the number of times that is replaced by T_2. The integer numbers m,k,v,t are the basic parameters of a trade and represent respectively the volume - sum of all positive entrances of a block in the trade, the block dimension, the number of varieties and the strength of the trade. Consider a BIBD with parameters (v,k,λ) which contain the blocks collection in T_1 of a trade $T(t,k,v)$. By replacing in this design the blocks of T_2 for the blocks of T_1 we obtain a new BIBD structure that besides obeying the same parameters may have a support size different from the original design.

4.2 Juxtaposition method and method based on the complementary design

According to Calinski e Kageyama (2003), one of the easiest BIBDR construction methods is the Juxtaposition Method, which allows obtaining a BIBDR by repeating the blocks of an existent BIBD.

Another very simple method for BIBDR construction is based on the definition of complementary design and it is particularly useful in situations when blocks have big number of plots. By definition, the complementary design of a BIBD (v,b,r,k,λ) is a BIBD with parameters $(v^* = v, b^* = b, r^* = b - r, k^* = v - k, \lambda^* = b - 2r + \lambda)$. So, the existence of a BIBD with block size *k* guarantees the existence of another BIBD, the complementary one, with block size $v - k$, in which blocks are composed by varieties missing in the respective block of the original design.

5. Application of BIBDR

5.1 Description

In this work an application of BIBDR in Education Sciences is considered, based on a sample resulting from an inquiry to students in Cape Verde.

Our study aims to compare the algebraic reasoning of students starting grade school career at Higher Education in 5 different domains: algebraic technicality, generalization of arithmetics, formulating laws, geometrization of algebra and solving problems (Bridge J., 2005). For this study, a questionnaire was developed, with five groups of questions corresponding to the five considered fields of algebraic thinking: Group 1 - Algebraic Technicality, Group 2 - Generalization of Arithmetic, Group 3 - Formulation of Laws, Group 4 - Geometrization of algebra, 5 - Solving problems. Each group consists of four issues points, thus making a total of 20 questions. 100 questionnaires were randomly selected for quantitative data analysis. For data analysis we used the software R (version 2.12.1).

5.2 Model and results using R

For data analysis a BIBDR$(5,100,60,3,30|b^*=10)$ was performed. The structure of this design was obtained using Juxtaposition Method, considering multiplicity ten and reproducing copies of the original BIBD$(5,10,6,3,3)$. In table 2. we present the resulting ANOVA, considering the model with the adjusted treatment effects and the block effects ignoring treatments, given by the software R command *BIB.test():*
analysis<-BIB.test(Blocks,Treatments,test-grade,method="tukey",alpha=0.05,group=TRUE)

Table 2. ANOVA for intra-block Analysis of BIBDR$(5,100,60,3,30|b^*=10)$

Response: test-grade

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Blocks	99	112893	1140.3	1.6893	0.0009922 ***
Treatments	4	66595	16648.8	24.6639	< 2.2e-16 ***
Residuals	196	132305	675.0		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

For the significance levels of 1% and 5% there is statistical evidence to reject the null hypothesis, that is, we find that there are significant differences in test grades, obtained for the different domains of algebraic thinking. Thus, having rejected the null hypothesis of treatments equality, we perform the Tukey test for the comparison of the treatment contrasts. Using again the command *BIB.test()* in R, the following results were obtained for the significance levels of 1% and 5%, respectively:

Table 3. Results of Tukey test

Tukey	**Tukey**
Alpha : 0.01	Alpha : 0.05
Std.err : 3.674303	Std.err : 3.674303
HSD : 17.14834	HSD : 14.30704
Means with the same letter are not significantly different.	Means with the same letter are not significantly different.
Comparison of treatments	**Comparison of treatments**
Groups, Treatments and means	Groups, Treatments and means
a 2 55.83333	a 2 55.83333
a 1 48.56667	a 1 48.56667
b 3 28.03333	b 3 28.03333
b 5 25.03333	bc 5 25.03333
b 4 11.03333	c 4 11.03333

It follows that:

- Among the treatments 1 and 2, 3 and 5, 4 and 5, there is no evidence of significant differences, at the significance levels of 1% and 5%;
- Among the treatments 3 and 4, at a significance level of 1%, there is no evidence of significant differences;
- The remaining pairs of treatments are significantly different levels of significance of 1% and 5%.

6. Final considerations, remarks and future work

There are still many open problems concerning BIBD, which remains a very promising research area. Besides their optimal characteristics, the diverse fields of applications and the connections of BIBD to other areas, make this kind of designs very attractive. In this work we approach the non existence of sufficient condition for a BIBD to hold, and the need of selection criteria looking for the best structures, among the many possible structures of a BIBD, in particular situations. For designs with block repetition some of the most common construction methods were reviewed. Connections between BIBD and Sampling Theory were illustrated through an example. An application of BIBDR was performed in the area of Education Sciences, seeking practical purposes for the development of education and scientific knowledge in Cape Verde.

In the future work we intend to approach the problem of BIBD and BIBDR considering different block sizes and to look for various applications of these designs.

Acknowledgements: Research partially supported by National Funds through FCT-Fundação para a Ciência e a Tecnologia, project PEst-OE /MAT /UI0006 /2011 (CEAUL)

References

Avadhani, M.S.; Sukhatme, B.V. (1973) “Controlled sampling with equal probabilities and without replacement”, *International statistical Review*, 41, 175-182.

Calinsky, T. and Kageyama, S.(2003) *Block Design: A Randomization Approach, Volume II: Design*, Lecture Notes in Statistics, Springer.

Chakrabarti, M.C. (1963) “On the use of incidence matrices in sampling from finite populations”. *J.Ind. Statist.Assoc.*, 1, 78-85.

Foody, W. and Hedayat, A.S. (1977) “On theory and applications of BIB Designs with repeated blocks”. *The Annals of Statistics*, Vol 5, no5, 932 - 945.

Hedayat, A.S. and Hwang, H.L. (1984) “BIB(8,56,21,3,6) and BIB(10,30,9,3,2) Designs with Repeated Blocks”.*Journal of Combinatorial Theory*, Series A36,73 - 91.

Hedayat, A.S.; Shuo-Yen and Li, R. (1979) “The trade off method in the construction of BIB Design with variable support sizes”, *The Annals of Statistic*, Vol. 7, Issue 6, 1277 – 1287.

Oliveira, T. (2000) *Planeamento de Experiências: Novas perspectivas,* Tese de Doutoramento, Universidade de Lisboa.

Oliveira, T. A. (2010) “BIB designs with repeated blocks: review and perspectives”, *The Islamic Countries Society of Statistical Sciences*, Vol. I, 82-96.

Ponte, J. P. (2005) “Álgebra no currículo escolar”, *Educação & Matemática: Revista da Associação de professores de matemática.* nov-dez. 2005. p. 36-42.

Srivastava, J.and Ouyang, Z.(1992)“Sampling theory using experimental design concepts”, *Lecture notes-Monograph series*, Institute of Mathematical Statistics, vol 17, 241-264.

Yang, J. (1985) “An example of the application of a Balanced Incomplete Block Design”. *Assessment and Evaluation in Higher Education*, Vol. 10, Issue 3, pages 250-253.

Van Der Linden, W.J., Veldkamp, B.P. and Carlson, J.E. (2004) “ Optimizing Balanced Incomplete Block Designs for Educational Assessments”. *Applied Psychological Measurement*, Vol.28, nº5,317-331.