Journal of Statistical Planning and Inference 143 (2013) 202–208

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Minimum breakdown designs in blocks of size two

Shin-Fu Tsai, Chen-Tuo Liao *

Division of Biometry, Institute of Agronomy, National Taiwan University, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 20 December 2011
Received in revised form
21 April 2012
Accepted 9 July 2012
Available online 17 July 2012

Keywords:
Block design
Connectedness
Missing value
Robust design
Two-color microarray

ABSTRACT

In scientific investigations, there are many situations where each two experimental units have to be grouped into a block of size two. For planning such experiments, the variance-based optimality criteria like A-, D- and E-criterion are typically employed to choose efficient designs, if the estimation efficiency of treatment contrasts is primarily concerned. Alternatively, if there are observations which tend to become lost during the experimental period, the robustness criteria against the unavailability of data should be strongly recommended for selecting the planning scheme. In this study, a new criterion, called minimum breakdown criterion, is proposed to quantify the robustness of designs in blocks of size two. Based on the proposed criterion, a new class of robust designs, called minimum breakdown designs, is defined. When various numbers of blocks are missing, the minimum breakdown designs provide the highest probabilities that all the treatment contrasts are estimable. An exhaustive search procedure is proposed to generate such designs. In addition, two classes of uniformly minimum breakdown designs are theoretically verified.

1. Introduction

In many experimental situations, it is often necessary to split treatments into blocks of size two. Particularly, when the experimental units are drastically variable, the blocking scheme with the smallest block size can effectively control the variation over units, leading the within-block variation to be much smaller than the between-block variation. For example, an investigator may work with twins in a learning program, two plots in an agricultural field block, two parts in a machine, etc. However, when implementing a design with the smallest block size, there could exist some uncontrollable factors during experimentation such that the observations tend to be missing. A participant may quit from a learning program due to some personal reasons, the yield of a specific plot in an agricultural field experiment is unavailable because of crop diseases or insect pests, and the observation of an experimental unit is missing due to some mechanic problems. When some observations are missing, the resulting incomplete data can lead to a loss in estimation efficiency, or the most extreme situation that the treatment contrasts of interest are no longer estimable. Recently, the two-color microarray technology becomes a popular and powerful tool for simultaneously exploring thousands of gene expression measurements. Typically, the statistical design issues for two-color microarray experiments are treated as block designs of size two, and several authors construct efficient or highly efficient designs for this innovative biotechnology. The reader is referred to Kerr and Churchill (2001), Kerr (2006), Bailey (2007) and Chai et al. (2007), among others. In two-color microarray experiments, there exist various factors, such as incomplete hybridization, image resolution, image corruption, see Troyanskaya et al. (2001), which can often cause missing gene expression measurements. Therefore, the robust designs

* Corresponding author. Tel.: +886 2 3366 4760; fax: +886 2 2362 0879.
E-mail address: ctliao@ntu.edu.tw (C.-T. Liao).

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2012.07.005

against missing data appear to be a genuine need in two-color microarray experiments. Robust designs for two-color microarray experiments are explored by Mahbub Latif et al. (2009), they discussed the connectedness and average efficiency of residual designs, when some gene expression measurements are unavailable. Note that an experimental design with a single factor is said to be connected, if it ensures that all the pairwise treatment comparisons are estimable.

Robust designs against missing data dates back to the pioneering work of Ghosh (1979). The notion of robustness proposed by Ghosh (1979) is named by Dey (1993) as Criterion 1 robustness. A design is said to be Criterion 1 robust against a loss of any i observations or blocks, if all of its residual designs remain connected. Based on the Criterion 1 robustness, Ghosh (1982a) showed that the balanced incomplete block (BIB) designs are robust against the loss of any $r-1$ blocks, where r represents the replicate of each treatment in the original design. In addition, the maximal robustness of block designs against the loss of whole blocks has been extensively explored, see Baksalary and Tabis (1987), Sathe and Satam (1992), Notz et al. (1994), and Godolphin and Warren (2011), among others. Alternatively, another criterion called Criterion 2 is also commonly used for assessing design robustness against the unavailability of data. A design is said to be Criterion 2 robust, if the relative efficiency between the residual design and the original design is not too small. The reader is referred to Bhaumik and Whittinghill (1991), Lal et al. (2001), and Morgan and Parvu (2008), among others, for more details regarding the characterizations and constructions of good designs with respect to the Criterion 2 robustness. In this study, the design robustness in terms of connectedness of residual designs is explored, the estimation efficiency of residual designs will be addressed through other communication.

When considering the special class of designs in blocks of size two, missing a single observation of a particular block and missing the whole block are equivalent, since a single observation of a block does not provide any information in estimating treatment contrasts. Consequently, it would be sufficient to only consider the robustness against missing blocks for the class of designs in blocks of size two. In practical applications, there could be several candidate designs which are all Criterion 1 robust against the loss of any i blocks. To further discriminate these candidates of equal performance, the Criterion 1 robustness is generalized to quantify their robustness that the number of missing blocks is greater than i. Even though this article focuses on only the designs of size two, the proposed criterion is actually applicable to all general designs, whether variance balanced/unbalanced or equally/unequally replicated.

The rest of this article is organized as follows. In Section 2, a new criterion, called minimum breakdown criterion, is proposed for quantifying the robustness of designs in blocks of size two. Section 3 explores the robustness of designs in blocks of size two. An exhaustive search procedure is provided to generate the minimum breakdown designs. Concluding remarks are given in the final section.

2. Minimum breakdown criterion

Suppose that a robust design in seven blocks of size two is required for estimating the pairwise treatment contrasts among five treatments. The following three designs are considered as candidates:

$$d_1=\{(1,2),(2,3),(2,4),(2,5),(3,4),(3,5),(4,5)\},$$
$$d_2=\{(1,4),(1,5),(2,4),(2,5),(3,4),(3,5),(4,5)\},$$
$$d_3=\{(1,3),(1,4),(1,5),(2,3),(2,4),(2,5),(3,5)\}.$$

Note that the pair (1,2) represents a block consisting of treatments 1 and 2, and so on. Since there is only one replicate of treatment 1 in d_1, any residual design of d_1 is disconnected, if the block (1,2) is missing. Alternatively, all the residual designs of d_2 and d_3 are all connected, when any one block is deleted. In other words, both d_2 and d_3 are Criterion 1 robust against missing one block. To further discriminate the robustness between d_2 and d_3, the connectedness of their residual designs for the removal of two blocks is explored. The residual designs of d_2 are disconnected, if any one set of {(1,4),(1,5)}, {(2,4),(2,5)} and {(3,4),(3,5)} is removed from it. However, there is only one disconnected residual design of d_3, which results from eliminating the set of {(1,4),(2,4)}. Namely, although d_2 and d_3 are both not Criterion 1 robust against missing two blocks, d_3 with a higher probability to obtain a connected residual design than d_2. Based on this concept, it is reasonable to choose d_3 as the planning scheme, if the design robustness against missing data is of main interest. Consequently, a generalization of Criterion 1 robustness for block designs of size two is now presented as follows.

Definition 1. Let d be a design with v treatments and b blocks of size two. Define M_i as the number of disconnected designs among the $\binom{b}{i}$ possible residual designs, if i blocks are removed from d. The vector $\boldsymbol{M}(d)=(M_1,M_2,\ldots,M_{b-v+1})$ is defined as the disconnected pattern of d, where $b-v+1$ represents the maximal number of permissible missing blocks. Note that all the residual designs are disconnected, if any i blocks are missing for $i>b-v+1$.

According to the definition of disconnected pattern, a design d is said to be Criterion 1 robust against a loss of i blocks, if $M_1=M_2=\cdots=M_i=0$. Apparently, the disconnected pattern provides much more information than the Criterion 1 robustness. This extension is essentially in the same spirit that the minimum aberration criterion was proposed to refine the design resolution criterion for assessing two-level fractional factorial designs (Fries and Hunter, 1980). Furthermore, let P_i denote the proportion of connected residual designs, if any i blocks are deleted from d. Namely,

$$P_i=1-\frac{M_i}{\binom{b}{i}}.$$