Michigan Math. J. 58 (2009)

Variance and Concurrence in Block Designs, and Distance in the Corresponding Graphs

R. A. Bailey

In memory of D. G. Higman

1. Introduction

When Fisher initially advocated partitioning the units in an experiment into blocks of similar units in [14, Sec. 48], he proposed that each treatment should occur on one unit in each block. (In statistical contexts, the points of a design are usually called *treatments.*) Such designs were eventually called *randomized block designs* or *complete-block designs.* However, natural blocks may not be large enough to contain every treatment and so, in [41], Yates introduced designs with incomplete blocks. He had the intuition to propose designs in which each pair of treatments occurs together in the same number of blocks. He called these *symmetrical incomplete randomized block arrangements*; nowadays, statisticians call them *balanced incomplete-block designs* while pure mathematicians call them *2-designs.* The last phrase was first used in print by D. R. Hughes [18], who told me in 2001 that it had been suggested by his then colleague D. G. Higman.

When a design is used for an experiment, there is one observation for each treatment in each block. These data are analyzed to estimate the relative effects of the different treatments. In this context, 2-designs have three clear advantages: (i) analysis of data from an experiment using such a design does not require matrix inversion (this consideration was important in pre-computer days); (ii) the variance of the estimator of the difference between the effect of treatment i and the effect of another treatment j is independent of the pair $\{i, j\}$; and (iii) the design minimizes the average value of these pairwise variances.

For a given practical experiment, there may not exist a 2-design with the required parameters. What design should one use then? In [42], Yates introduced *square lattice* designs for n^2 treatments in rn blocks of size n, where $2 \le r \le n+1$. The treatments are the cells of an $n \times n$ square array; the blocks correspond to the rows and columns of the array and to the letters of $r-2$ mutually orthogonal $n \times n$ Latin squares. In such a design, each pair of treatments occurs together in either one or zero blocks, and the variance of the estimator of their difference depends only on this number, being slightly smaller in the former case. Use of these designs led statisticians to believe that, in any incomplete-block design, the variance would always depend on the number of blocks containing a given pair

Received August 17, 2007. Revision received March 14, 2008.