Designs

Peter J. Cameron

Queen Mary, University of London, Mile End Road, London E1 4NS, U.K.

Block designs were first used in the design of experiments in statistics, as a method for coping with systematic differences in the experimental material. Suppose, for example, that we want to test seven different varieties of seed in an agricultural experiment, and have 21 plots of land available for the experiment. If the plots can be regarded as identical, then the best strategy is clearly to plant three plots with each variety. Suppose, however, that the available plots are on seven farms in different regions, with three plots on each farm. If we simply plant one variety on each farm, we lose information, because we cannot distinguish systematic differences between regions from differences in the seed varieties. It is better to follow a scheme like this: plant varieties $1, 2, 3$ on the first farm; $1, 4, 5$ on the second; and then $1, 6, 7$; $2, 4, 6$; $2, 5, 7$; $3, 4, 7$; and $3, 5, 6$.

This design can be represented by a picture:

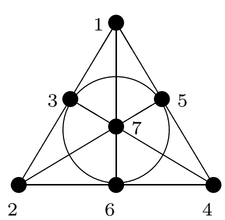

Figure 1: A block design

This arrangement is called a *balanced incomplete-block design*, or BIBD for short. The blocks are the sets of seed varieties used on the seven farms. The blocks are incomplete because not every variety can be planted on every farm; the design is balanced because each pair of varieties occur together in a block the same number of times (just once in this case). This is a 2-$(7, 3, 1)$ design: there are seven varieties; each block contains three of them; and two varieties occur together in a block once. It is also an example of a finite *projective plane*. Because of the connection with geometry, varieties are usually called "points".

Mathematicians have developed an extensive theory of BIBDs and related classes of designs. Indeed, the study of such designs predates their use in statistics. In 1847, T. P. Kirkman showed that a 2-$(v, 3, 1)$ design exists if and only if v is congruent to 1 or 3 mod 6. (Such designs are now called *Steiner triple systems*, although Steiner did not pose the problem of their existence until 1853.)

In general, given k and λ, easy counting arguments show that the values of v for which 2-(v, k, λ) designs exist are restricted to certain congruence classes. An asymptotic existence theory developed by Richard Wilson shows that this necessary condition is sufficient for the existence of a design, apart from finitely many exceptions, for each k and λ.

The concept has been further generalised: a t-(v, k, λ) design has the property that any t points are contained in exactly λ blocks. Luc Teirlinck showed that non-trivial t-designs exist for all t, but examples for $t > 3$ are comparatively rare.

However, the statisticians' concerns are a bit different. Usually, the number of varieties and the number and size of blocks are given in advance. There are various measures of the "efficiency" of a design, the amount of information about treatment differences which can be recovered from the results of the experiment.

If a BIBD exists, then it is most efficient with respect to all such measures. However, often the constraints do not admit a BIBD. For example, R. A. Fisher showed that a BIBD must have at least as many blocks as points. If no BIBD exists, there may be no design which is uniformly best, and the choice of the design can be influenced by the variability of the treatment and block effects. Combinatorial conditions guaranteeing statistical optimality are usually not known.

Block designs correspond to the simplest possible type of structure on the set of experimental units, namely a partition. For more complicated structures, more elaborate designs are required, such as row-column designs (including Latin squares) and neighbour designs. Further complication arises from the fact that the treatment set may also be structured; for example, there may be several treatment factors, each of which can be applied at one of several levels, giving rise to the theory of factorial designs.

Design theory is closely related to other combinatorial topics such as error-correcting codes; indeed, Fisher “discovered” the Hamming codes as factorial designs five years before R. W. Hamming found them in the context of error correction. Other related subjects include packing and covering problems, and especially finite geometry, where many finite versions of classical geometries can be regarded as designs.