Optimum designs with leverage control

Marcelo Andrade da Silva*
Interinstitutional Graduate Program in Statistics - USP/UFSCar, Sao Carlos, Brazil - marceloddr@gmail.com

Luzia Aparecida Trinca
Institute of Bioscience - Unesp, Botucatu, Brazil - ltrinca@ibb.unesp.br

Abstract

In many experiments in agriculture and biology, several factors are investigated at the same time. Limited material, work-people, physical space, equipment use, time or financial resources may restrict the number of units to be tested such that the use of full factorials are unviable. Careful planning is important in order to help the researcher obtaining the desired information. Usually the desired information is translated to a function of the information matrix in order to produce optimum designs. Composite design criteria incorporating several desired properties are very promising such that the design chosen presents good performances under several aspects. For the classic linear model, an optimum design specifies the design matrix, such that some function of interest of the information matrix or of its inverse is optimized. These functions, named "criterion functions", have the purpose of making sure the researcher meets his experimental objectives. Single property functions may produce designs that are too tight and lack robustness to missing data. Thus, we propose to include the H property, as defined below, in the expression of a compound criterion to prevent the inclusion of points in the design that are too influential in the model fitting. A fairly simple measure is based on the diagonal elements of the H matrix, sometimes also called influence matrix, projection matrix or hat matrix. These diagonal elements of the H are simple measures of the influence of each observation in the fitting of the model. Thus, we propose to minimize the variability of the diagonal elements of the H matrix. In the literature, the most familiar method for optimizing the design is the exchange algorithm. This method is a heuristic that perform swaps of treatments (point exchange) or factor levels (coordinate exchange) in an initial design until the criterion value stops improving. For experiments with large number of factors, computation is expensive and thus computational efficiency is an issue. We implemented the method in C++ language and produced designs for some examples varying the composition of the design criterion. In general, the new compound criterion that incorporates the H property produced quite attractive designs since their efficiency under usual properties are very high and their leverages are more homogeneous indicating that the designs are more robust to missing data and prevents the data from having influential observations.

Keywords: optimum factorial designs; robust designs; missing observation; compound criteria.