THE COORDINATE-FREE APPROACH TO GAUSS-MARKOV ESTIMATION, AND ITS APPLICATION TO MISSING AND EXTRA OBSERVATIONS

WILLIAM KRUSKAL
UNIVERSITY OF CHICAGO

1. Introduction and summary

The purposes of this paper are (1) to describe the coordinate-free approach to Gauss-Markov (linear least squares) estimation in the context of Model I analysis of variance and (2) to discuss, in coordinate-free language, the topics of missing observations and extra observations.

It is curious that the coordinate-free approach to Gauss-Markov estimation, although known to many statisticians, has infrequently been discussed in the literature on least squares and analysis of variance. The major textbooks in these areas do not use the coordinate-free approach, and I know of only a few journal articles that deal with it ([2], plus some of the references in Dutch that it lists, and, to some extent, [1], [3] and [7]). The coordinate-free viewpoint is implicit in R. A. Fisher's geometrical approach to sampling problems.

The subject of missing observations in Model I analysis of variance is well understood and often discussed. This paper presents no new results here, but it does present a viewpoint different from that usually given. In contrast, the topic of extra observations, although it was briefly considered by Gauss [5], section 35 of *Theoria Combinationis* . . . , has elicited hardly any papers since. (I know only of papers by R. L. Plackett [9] and K. D. Tocher [10].) The problem of extra observations is important in its own right and also in connection with the treatment of so-called outliers. I shall discuss a method of treating extra observations that bears some resemblance to that for missing observations. In particular, it leads to possible methods for treating apparent outliers that I described briefly in [8].

There are two major motivations for emphasizing the coordinate-free approach to Gauss-Markov estimation. First, it permits a simpler, more general, more elegant, and more direct treatment of the general theory of linear estimation

The work leading to this paper was supported in part by the Logistics and Mathematical Statistics Branch of the Office of Naval Research and in part by the National Science Foundation.