Some Heuristics for Analysis of Variance

By K. Ruben Gabriel*

University of Rochester

Definition of the variance of a sample as one half of the average squared difference is more intuitive than the common definition as an average squared deviation from the mean. Similarly, the one-way ANOVA F-statistic has intuitively appealing definitions as the average of squared t-statistics, either of the t's for all pairwise comparisons of means, or of the t's for comparing each mean with the average of all the others. These are distinct from Scheffé's definition of F as the square of the maximum of t's for all contrasts. Analogous definitions extend to two-way ANOVA and to MANOVA.

It is more intuitive to define overall test statistics as averages of statistics for simple components of the overall hypothesis than by the classical definitions, and there is much to be said for that in terms of heuristic appeal and didactic usefulness.

1. Introduction In statistical inference, the choice and definition of a statistic is determined by its optimality properties for estimation and testing. For example, the center of a sample is usually defined as the mean because that definition is in many ways optimal for Gaussian data. In descriptive statistics, on the other hand, the purpose of a statistic is to reveal interesting features of the data, so the choice of a statistic becomes heuristic. Thus, the median may be chosen to describe a sample's center because of its intuitive appeal as bisecting the observations into an equal number of larger and smaller ones; the mean is much less appealing as it is defined by a quite non-intuitive algorithm involving addition and division. Of course, some statistics may be intuitively motivated as well as optimally inferential in certain contexts, and in non-parametric inference heuristic criteria are often used because optimality is difficult to define.

Heuristics are also important for intuitively motivating the use of particular statistics. Thus, it is didactically preferable to introduce the median as having the simple property of bisecting by size, rather than as having the more abstract property of minimizing the sum of absolute deviations. Several other well known statistics are shown in this paper to have simple definitions with heuristic appeal, and these may be used to advantage in teaching instead of the classical less intuitive definitions.

2. The sample variance The variance of a sample of observations x_i, $i = 1, \ldots, n$, is usually defined by means of deviations from the mean $\bar{x} = \sum_{i=1}^{n} x_i/n$

*K. Ruben Gabriel died 25 May 2003. Ruben was a good friend to the editors and many of the authors whose work is represented in this volume, and we will miss him.