Series Editor's Introduction

In our introduction to Volume 1 of Van de Geer's book, we saw that there was a heavy emphasis on the geometry of Exploratory Multivariate Analysis. In Volume 2 there is an equally heavy emphasis on computing, and the computing is done with ready-made programs incorporated into *Categories,* a sub-package of *SPSS* (*SPSS Categories,* 1990). The emphasis on the geometry is still there, and there any many interesting pictures of data analysis results, but following GIFI (1990) the book is now structured in chapters corresponding with specific computer programs.

Any of the programs discussed is a statistical technique in the sense that one inputs data from a set of possible input data structures, and the program then generates results, which are supposedly a unique function of the input. A complicated function, to be sure, because there is numerical output, there are graphs, and there even may be error messages. It is still useful to think of a program (or a package, which consists of a number of programs) as something that transforms input to output, where the precise nature of the function implemented in the program is often determined by some additional parameters defined separately from the input data. Van de Geer discusses some of the key GIFI programs, and show what they do to input data of various kinds.

Packages, such as *SPSS,* are not very popular with statisticians. They are associated with a rigid menu of canned statistical techniques, which are applied in a routine and mechanical way by people who really don't know what they are doing. In this respect they are similar to fast food or Hollywood movies. Convenient, yes, popular, yes, but superficial and without quality. I think the disdain of statisticians for package-statistics is rather elitist and short sighted. Statisticians are under the spell of mathematical notions of optimality, of philosophical notions about inference, and of scientific notions about the proper treatment of empirical data. The packages, however, provide a language to talk about data, which is understood by many scientists (and by many journal editors). Maintaining that packages are not a good language is quite silly. The alternative languages that statisticians have come up with so far, are either ridiculously impractical (decision theory is a good example) or they can only be applied to tiny examples or in well-controlled situations. Most applied statisticians, whether they are practicing Bayesians or not, simply use the packages on their clients' data sets. And many applied statisticians have found out, to their dismay, that there are a lot of things in data manipulation that can best be done in those packages.

Badmouthing the packages is like saying that if you want to talk about food, you can only do so properly in French. Saying that packages are used by people who don't know what they are doing is dangerously close to the sort of arrogance that mathematicians seem to be especially prone to. The people who are using the packages are not necessarily out there to discover rock-solid truths, they are there to publish papers in scientific journals in which the packages are the norm. Survey type data are often messy, they do not satisfy the standard models, and they burst out of the proper statistical frameworks. If the food we are talking about is just your daily run-of-the-mill food, we might as well talk about it in plain English.

Exploratory data analysis is another statistical bogeyman. Somewhere, somehow, statisticians got the idea that science (proper science, that is) proceeds in two steps. The first step is exploratory. The scientist does all kinds of dirty things to his or her data, things that are certainly not allowed by the canons of statistics, and at the end of this thoroughly unrespectable phase he or she comes up (miraculously) with a theory, model, or hypothesis. This hypothesis is then tested with the proper confirmatory statistical methods. Of course, Popper or no Popper, this is a complete travesty of what *actually* goes on in all sciences some of the time and in some sciences all of the time. There are no two phases that can be easily distinguished. There is no dirty and clean work, and for that matter the distinction between exploratory and confirmatory seems to allocate all the interesting and creative work to the exploratory phase anyway.

Enough of these generalities. The basic idea behind the GIFI system is that variables can be grouped into subsets in various ways, and variables can be quantified using various types of restrictions. By combining partitionings of the variables with general classes of measurement restrictions, we recover many of the classical multivariate techniques, but also many extensions. The chapters in Volume 2 are named after several of the more important techniques, with associated computer programs. *PRIMALS, HOMALS, ANACOR,* and *PRINCALS* are different generalizations of PCA; *CANALS* generalizes canonical correlation analysis; and *OVERALS* generalizes a form of multiple-set canonical correlation analysis. In a sense, all GIFI techniques are options in the *OVERALS* program, but special techniques require special algorithms and special input-output options, and thus special implementations. *HOMALS, PRINCALS, ANACOR,* and *OVERALS* have been incorporated in *SPPS Categories* and are consequently widely available. *ANACOR* is Correspondence Analysis, *HOMALS* is Multiple Correspondence Analysis, and consequently these two techniques are also available in *SAS* and *BMDP* under different names.

Another aspect that Van de Geer's book has in common with GIFI (1990) is that there are many real-life examples. Not too big, but not too small either. With real variables, often trying to answer some relevant policy-related question. The relevant part of the book by GIFI is called *The Proof of the Pudding*. Of course, you may not like the pudding. But all the ingredients are discussed quite clearly in these two volumes, and after digesting them, you will be prepared well to go deeper into the GIFI pudding, or to prepare one of your own.

JAN DE LEEUW
SERIES EDITOR