useful brushing technique. There is not much I can say about the example except that the data seems to have an exceedingly simple and well-defined structure. The authors were indeed fortunate in finding such strong linear structure which did not require the transformation of even one variable. Given the importance of measurement scales, it would have been nice if the authors had published the complete data set. That TOTORG and SUMDYE dominate the analysis is not very surprising as these seem to be totals over variables 1–14. (I think TOTORG is the sum of variables 1–14, but what SUMDYE is, is not clear to me.) What is clear is that the data have strong linear features, and that some of this linearity is inbuilt. How would these linear methods have fared if the samples had occupied a nonlinear manifold in 29-dimensional space? The detection of manifolds is one of the fundamental problems of multivariate data analysis. Projection pursuit is one attempt to help here, but I believe that transformations are likely to have more to offer, especially in nonlinear cases. Years ago, when Prim 9 was new, I asked the following question. Suppose I have a sample of, say, 1000 3×3 orthogonal matrices. These each give nine observations, so their space may be explored by Prim 9. Because sums-of-squares of all rows and all columns are unity, the points will lie on six three-dimensional spheres embedded in the nine-dimensional space. Further, sums-of-products of rows and columns vanish, so the points also lie on three-dimensional hyperboloids. Two-dimensional cross-sections will show circles and hyperbolae and as the cutting-planes move dynamically, the circles will grow larger, then smaller and finally vanish; similarly for the hyperbolae. How would a user observing these strange phenomena interpret what he saw? I have yet to receive a satisfactory answer to the question.

I believe that graphical methods for multivariate data analysis have much to offer. In the linear case, quite good progress has been made and I thank Drs. Weihs and Schmidli for their interesting contribution. Nonlinear multivariate analysis still has a long way to go. Progress will go hand-in-hand with good software, and I see that as a development of general-purpose statistical software.

Comment

Werner Stuetzle

This paper starts with a valid premise: many techniques for exploratory data analysis have been developed in an artificial context and illustrated using contrived and unconvincing examples. There is little experience as to which methods are useful in practice. Serious assessment of this issue would undoubtedly be valuable. However, the authors do not provide such an assessment. Their choice of building blocks for what they call the OMEGA pipeline appears to be largely driven by the computing environment at their disposal, and not by actual experience with a wide range of techniques. In addition to a case study, the paper presents a survey of methods and software. While such a survey could be helpful, the authors' attempt appears somewhat haphazard and incomplete. An encouraging aspect of the paper is the suggestion that techniques such as point cloud rotation, plot interpolation and Grand Tour, and brushing of scatterplots might eventually make their way from the esoteric realms of academia and research laboratories to actual consumers. I will first comment on the methodological part of the article and then on the data analysis.

Werner Stuetzle is Associate Professor, Department of Statistics, and Adjunct Professor, Department of Computer Science. His mailing address is Department of Statistics, GN 22, University of Washington, Seattle, Washington 98195.

COMMENTS ON METHODOLOGY

Simplification might be a useful idea. It comes up in other contexts, for example in Projection Pursuit (Friedman and Stuetzle, 1981), where one wants the chosen directions to involve as few of the variables as possible. The authors explain how the first principal component is simplified, although the properties of their procedure are not entirely clear. I do not see how they propose to simplify the second and higher principal components.

The motivation behind "p% resampling" is unclear. What is the distribution to be estimated? Why not simply do bootstrap resampling? Bootstrapping estimates the variability arising from repetitions of the experiment, assuming that the data can be interpreted as an iid sample from some distribution. One would then check how many principal component projections of bootstrap samples show some interesting