Title:
Multivariate Analysis: Overview

Author(s):
Ingram Olkin and Allan R. Sampson

Technical Report number (Dept. of Statistics, Stanford Univ.):
2001-2

Date:
February 2001

Abstract:

Multivariate analysis is conceptualized by tradition as the statistical study of data in which multiple measurements are made on each experimental unit and for which the relationships among multivariate measurements and their structure are important to the study's understanding. A fairly natural and modern, somewhat overlapping, categorization of multivariate analysis is the following: normal and more general multivariate models and distribution theory; the study and measurement of relationships; probability computations of multidimensional regions; and the study and exploration of data structures and patterns.

The multivariate normal distribution plays a central role in the statistics of multidimensional data in the same way that the univariate normal distribution plays a central role. The Student's-t generalizes to Hotelling T^2 and the chi-square distribution to the Wishart distribution. Newer multivariate distributions have been developed using various approaches to model multivariate data when the multivariate normal distribution does not provide an adequate model. For multidimensional data, relationships among the variables are fundamental to explore. Among the useful techniques to understand and quantify these are multivariate regression analysis and various correlational notions such as partial correlations and canonical correlations. Implementation of multivariate methods often requires computations of complex multidimensional probabilities. Approaches to these computations include both obtaining lower bounds for the probabilities and using numerical approximation techniques. The exploration of structure and patterns for complex large multivariate data sets is a crucial concern of modern data analysis and data mining.

Multivariate tools that have proven to be useful in this context, include principal components analysis, canonical analysis, factor analysis, path analysis, structural equation methods, clustering, and discriminant analysis.