# Data Depth in Multivariate Data Analysis: Usefulness and Challenges

Regina Y. Liu

Department of Statistics

Rutgers University

Piscataway, NJ 08854-8019

E-mail: rliu@stat.rutgers.edu

The advances in computer technology have facilitated greatly the collection of large scale high dimensional data, and statisticians face increasingly the task of analyzing large multivariate datasets. Simple extensions of univariate statistics to the multivariate setting do not properly capture the higher-dimensional features of multivariate data, nor do they yield geometric solutions because of the absence of a natural order for multidimensional Euclidean space. A more promising approach is the one based on "data depth", which can provide a center-outward ordering of points in Euclidean space of any dimension. Extensive developments in recent years have generated many attractive depth-based tools for multivariate data analysis, with a wide range of applications, including constructions of confidence regions, regression, and multivariate process control. Some of these applications will be demonstrated in an aviation safety analysis with some airline performance data collected by the FAA from 1993 to 1998.

The concept of data depth provides new perspectives to probabilistic as well as computational geometries. In particular, the development of implementable computing algorithms for depth-based statistics has brought about many new challenges in computational geometry. Some of these challenges will be discussed.