

# CARDS SEMINAR



## Methods for Clustering Mixed Data

Seminar Speaker- Dr. JeanMarie Hendrickson

Assistant Professor of Statistics, ETSU

**ABSTRACT:** I will give a brief introduction to cluster analysis and then propose and discuss a few methods for clustering mixed data. In particular, I will discuss a model-based clustering method for mixed data based on Everitt's work and describe a simulated annealing method that was used to estimate the parameters for Everitt's model. A penalized log likelihood with the simulated annealing method is proposed as a remedy for the parameter estimates being drawn to extremes. Everitt's approach and the proposed method are compared based on their performance in clustering simulated data. I will then discuss the results of the cluster analysis using the penalized log likelihood method on a heart disease data set, in which the clustering result was compared to an expert's diagnosis using the Rand index. I will also describe an extension to Gower's coefficient, based on Kaufman and Rousseeu's definition, which allows for other types of variables to be used in the coefficient. A clustering algorithm based on the extended Gower coefficient is used to find a clustering solution for a buoy data set to see how well this method classified a variety of sites into their "true" regions. To display how the method based on the extended Gower coefficient performed in clustering data having a variety of structures, the results of a simulation study are shown.

*Friday, April 7<sup>th</sup> at 3:15 in Room 106-Gilbreath Hall*