

Statistics 579

Applied Multivariate Methods

Exam 3

1. The file FoodNutrient.txt contains nutritional data on 27 different food items. Nutritionists desire to find clusters that contain food items that are similar enough to warrant a recommendation of interchangeability when planning menus.
 - a) Perform a Principal Components Analysis of these data, and determine the number of principal components that would be needed to explain most of the variation in these data. Plot the first three principal components.
 - b) Using the Average Linkage Method and the Centroid Method, do a cluster analysis of these data, and produced the corresponding dendrogram. Using the plot of the first three principal components, the cubic clustering criterion, the pseudo F statistic, and the pseudo T^2 statistic, determine the number of clusters that exist in this data set.
 - c) Perform a k-Means clustering of these data, using the number of clusters determined in part (b).
 - d) Identify which food items belong in each of the clusters. Do the 3 different clustering methods produce the same results?
 - e) For each of the three clustering methods used, produce a 3-D scatter plot of the first three principal components, and annotate each point with the cluster number.

2. The file Milk.txt contains data from the first phase of a study of the cost of transporting milk from farms to dairy plants. Cost information was obtained on Fuel, Repair, and Capital, all measured on a per-mile basis, for 2 Types of trucks: Gasoline and Diesel.
 - a) Create a 3-D plot of the data, using different symbols for the 2 different types of trucks.
 - b) I have identified 2 data points that are clear outliers. Find them and eliminate them before you do any of the analyses below. Make sure you clearly identify which 2 points you are eliminating, and state your reasons why.
 - c) For each of the two types of trucks separately, determine if it is reasonable to conclude that the data arose from a multivariate normal distribution.
 - d) Test for the equality of the population variance-covariance matrices for the 2 types of trucks. Use $\alpha = .01$.
 - e) Assuming that the variance-covariance matrices are equal, test whether the mean vectors are the same for both types of trucks.
 - f) If you found, in part (b), that the variance-covariance matrices are in fact different, test for the equality of the mean vectors assuming unequal variance-covariance matrices. Compare these results with the results from part (c).
 - g) Using Bonferroni t-tests, construct simultaneous 95% confidence intervals for the difference between the 2 Types of truck, for each of the 3 dependent variables separately.