

Name _____

Psychology 516
Applied Multivariate Analysis
Homework 8
Due November 1, 2018

I've just obtained a zoo! Well, at least all of the animals from a zoo; the land was lost because of failure to pay property taxes and the owner decided to get out of the zoo business. I've always wanted to own a zoo, but I don't know beans about it. My first job is to organize the creatures in some sensible way so that the staff can care for them easily and visitors can find what they want to see without too much random walking around.

The file, Set_6.csv, contains one animal per row. The columns are 20 different features that define the animals. All are self-explanatory except the last one (catsize). This is an indicator of whether the animal is smaller than a typical housecat or larger than a typical housecat. All variables are binary with 0 indicating the absence of the feature (e.g., does not produce milk) and 1 indicating the presence of the feature (e.g., has fins).

1. Analyze these data using each of the following hierarchical clustering procedures: single linkage, complete linkage, average linkage, centroid method, Ward's method.
2. Choose the method that you believe provides the easiest interpretation of the data. Based on that method, how many sections or major areas do I need to build in my zoo so that I can house similar animals together? Provide a justification for this conclusion.
3. Based on the dominant features of the clusters you have identified, try to name the major clusters.
4. Just for fun. I am really going to get in trouble for one of the animals I will have in my zoo. The previous owner gave me a big discount on the animal collection provided I take this one, which was causing him all kinds of grief. Can you identify it?