STAT 530 - Fall 2005

Homework 7 - Due: Thursday, November 13th
Homework 8 - Due: Thursday, November 20th

Both of these problems uses the data set:
`http://www.stat.sc.edu/~habing/courses/data/crabs2.txt`

The data set contains four groups of fifteen crabs each: `O` = orange male, `o` = orange female, `B`= blue male, `b` = blue female.

Each crab has eight measurements. The first four are: `FL` = frontal lobe size (mm), `RW` = rear width (mm), `CL` = carapace length (mm), and `CW` = carapace width (mm). The next four variables are indicated with the same names as the previous four, but with an `s` added at the beginning. These were created by first dividing each of the original measurements by the total body depth (a measure of overall size). They were then standardized by subtracting the mean and dividing by the standard deviation.

Homework 7

1) If you want to use the measurements to tell apart the four different groups of crab, briefly describe why you might want to divide the four individual measurements by the overall size of each crab? (e.g. what fact about the crabs might this help account for?)

2) If you want to use the measurements to tell apart the four different groups of crab apart, briefly describe why you might want to standardize the measurements?

3) If you decided to use two-dimensional multi-dimensional scaling on the divided and standardized data set, which of Classical or Isometric do you think would be best at showing the distinct clusters in this data set? Why? (Don't perform the scaling to answer this.)

4) Perform the scaling method you chose in part 3 and construct a plot of the scaling where each crab is denoted solely by a dot. By looking at the separation of points in the scaling, divide the scaling into separate clusters. (Do not do any additional analyses to help you do this! You do not need to make 4 distinct clusters if you can't see them in the plot.)

5) Re-plot the scaling in 4, this time labeling each crab by its type. Summarize how well you think these measurements separate the four groups.

6) For the method you chose in 3, compare the estimated stress for the one, two, and three dimensional multi-dimensional scalings. How many dimensions should you have used?

Homework 8

1) Conduct a cluster analyses using the four divided and standardized variables, indicting each crab on the dendogram by its group, using each of nearest neighbor, average, furthest neighbor, and Ward's linkages. Which seems to do the best job of separating the four groups? (Briefly say why it seems best to you, labeling the picture is probably helpful.)

2) Compare this to using the linkage you found best and the non-adjusted variables. Did the adjusted variables indeed seem to work better? (Briefly say why, maybe by labeling the picture.)

3) Of course, we actually know which four groups the crabs belong to. Check the assumption that the four standardized variables are multivariate normal with the same covariance for each of the four groups.

4) Select one of the four MANOVA test statistics (justify your choice) and perform the MANOVA using the four standardized variables. Be sure to state the null and alternate hypothesis, and your conclusion.