Problem Set #6: Multivariate Stats

(1) (10 pts) Mark Steele and colleagues studied communities of coral reef fishes in the Bahamas. Fishes were surveyed along 5-8 transects at each site. This data set, in the file BahamasFish.cs*v*, includes densities of only the four most abundant species. Convert the data to z-scores and then use Principal Components Analysis to evaluate whether variation in the densities of the four different species can be summarized by new, derived variables (components). How much of the variation in density is explained by the first two principal components?

(2) (10 pts) Using the same Bahamas fish community data, make a biplot of PC1 vs PC2. Add vectors that show the relationship between each fish species and these two components.

How is density of each of the 4 species related to PC1 and PC2?

(3) (10 pts) The Steele lab wanted to determine whether fish communities differed among 5 study sites. Use ANOVA on PC1 scores from the Bahamas fish community to determine whether fish communities differ among sites. The independent transects at each site will serve as replicates.

(4) (10 pts) Jeanne Robertson is examining lizards (*Liolaemus darwinii*) living on different types of soils. She wants to know if lizard color is adapted to the color of the soil on which they’re found. She collects several lizards from areas with black, white, and red soils. Then she measures the color of each lizard using three metrics. The first is brightness, which distinguishes dark from light colors (e.g., black vs. white). The second is hue, which is best at distinguishing red from black. The third is saturation, which is best at distinguishing red from white. Using all three of these metrics, she can quantify the color of each lizard. Instead of quantifying hue, saturation, and brightness as continuous variables, she creates eight bins for each variable and measures the percentage of pixels that fall into each bin. So pure red lizards would have pixels all in bin 1 of hue. Pure black lizards would have all pixels in bin 8 for brightness and bin 8 for hue. This dataset has 24 dependent variables: 8 bins for each of hue, saturation, and brightness. The data are in the file lizardcolor.csv.

Use discriminant function analysis to determine how well lizards can be assigned to their original collection location, based on their color. What was the percent success for each soil type?

(5) (10 pts) One cold February day in 1898, an “an uncommonly severe” storm passed over New England. After the storm, the zoologist Hermon Bumpus collected 136 house sparrows that had been brought down by the storm in the vicinity of his laboratory at Brown University in. More than half of the birds recovered, but the rest died from exposure. Bumpus took this as an opportunity to study natural selection in action, and measured a number of skeletal features on all the birds, as well as recording whether they survived the storm, their sex, and (in the males) whether they were adults or yearlings. Bumpus’s data are in the file bumpus.csv.

Bumpus measured (in order on the file: total length, alar extension (wingspan), weight, head length, humerus length, femur length, tibiotarsus length, skull width, and keel to sternum length. We already know that house sparrows are sexually dimorphic in overall size, so split the data by sex and do separate analyses on males and females. Use perMANOVA to determine whether the birds that survived the storm had different morphologies than birds that died. What trait(s) distinguish these two groups of birds the most? Do the results depend on sex? Include a NMDS plot for each sex.