October 2015

CSUMB 550 Homework 6a
Principal Components Analysis

California sea lions may have different foraging strategies based on the type of prey they are catching since different types of prey live at different depths. The foraging strategy is always linked to the limitation that the sea lions have to come to the surface to breathe, which creates different diving characteristics. The diving characteristics are related to each other, and your job is to describe those relationships.

The data file "CalSeaLionDivesLtd.csv" contains dive data (and some other stuff) for 14 California sea lions. The dive data are:

Maxdepth = maximum depth of a dive
Dduration = dive duration
Botttime = duration at depth
DescRate = descent rate
AscRate = ascent rate
PDI = post dive interval (time at the surface after a dive)
DWigglesBott = up and down movement at depth (higher values are an indication of diving mesopelagically since benthic divers can't wiggle)
TotVertDistBot = total amount of distance traveled while at the bottom
BottRange = maximum – minimum depth while at the bottom
Efficiency = Botttime/(Dduration + PDI)
IDZ = consistency of depth (1 = current dive depth was within ± 10 m of the previous dive depth; another potential indication of benthic diving)

Steps and questions to answer along the way. Provide answers to the questions and turn in R code.

1. Define your entities and variables. How many entities and variables do you have?
2. Create a mean value by entity for each variable. A thought question: If you look at the raw data for each variable, do you think the mean is a good representation of that data? Why or why not?
3. Test the assumptions of PVA using the mean values by entity. Have we met the assumption of linearity? Why or why not? If not, is there a way to correct for it? Are there outliers? Would you remove them? Why or why not? Does removal of outliers change the outcome?
4. Calculate the eigenvalues and eigenvectors using the mean values by entity.
5. Based on the methods we discussed (latent root criterion, scree plot, broken stick crition), how many principal components would you keep?
6. Create code for at least one significance test method to determine the number of principal components to keep. Do the results differ from the other methods?
7. Calculate the loadings. Based on the loadings, interpret what each component is describing and name the components.
8. Calculate the PC scores and plot them against each other. What are they telling you?