Lab #10:

Please e-mail code, graphs and answers to questions to [bsmit269@uncc.edu](mailto:bsmit269@uncc.edu) and [afodor@uncc.edu](mailto:afodor@uncc.edu)

Please have lab submitted (whatever you have) before class on Mon., April 13.

This week’s dataset is here:

<http://afodor.github.io/classes/stats2015/prePostPhylum.txt>

(This datset is described, albeit from a different analysis pipeline, in these papers:

<http://www.sciencemag.org/content/sci/338/6103/120.full.html>

and

<http://www.nature.com/ncomms/2014/140903/ncomms5724/full/ncomms5724.html>

For the POST timepoints only:

1. For each phyla, graph the relative abundance of that phyla vs. cage. Does there appear to be a cage effect across different phyla?
2. For each phyla build a mixed linear model with genotype as the fixed variable and cage as a random variable. Report the intraclass correlation coefficient for each phyla. Are there any phyla that are significantly different for genotype in the mixed model at a 10% false discovery rate?

Hints:

1. If you use par(mfrow=c(3,2)) you can fit all 6 plots for phyla vs. cage on one graph. You can put the p-values and intraclass correlation coefficient in the “main” text above each graph to make a nice summary figure.
2. It can be useful to make a dataframe with just the data you want before building your model. So if you are looping through columns in a “myT” that you’ve read with read.table and i is your column index..

myT <- myT[myT$time == "POST",]

bug <- myT[,i]

cage <- myT$cage

genotype <- myT$genotype

myFrame <- data.frame(bug, cage, genotype)

(and then build your models with data=myFrame…)

1. Getting a p-value out of the mixed linear model could be done with something like:

unclass(summary(M.mixed))$tTable[2,5]