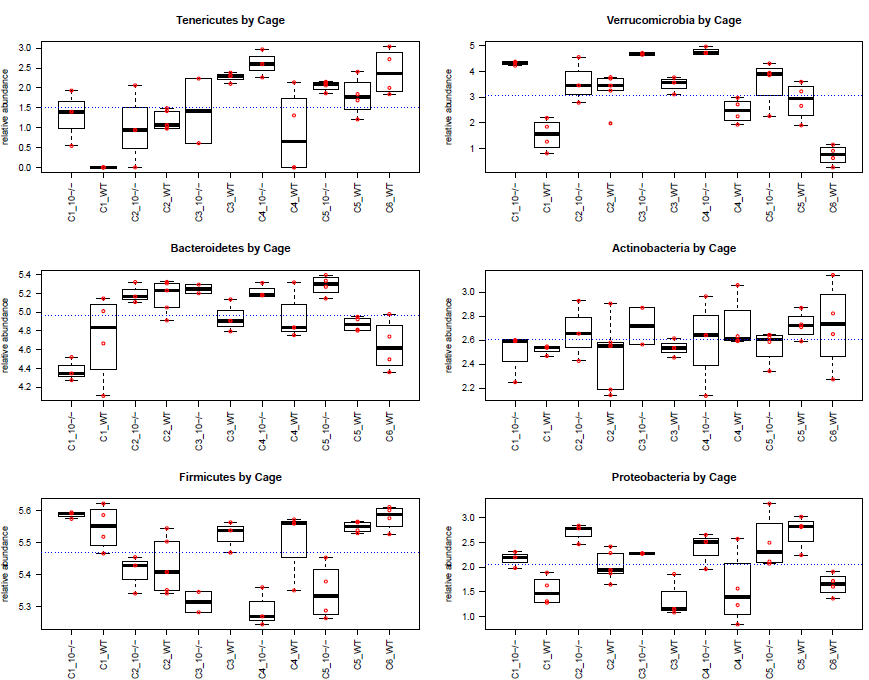
This week’s dataset is again (same as last week):

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

**For the POST timepoints only:**

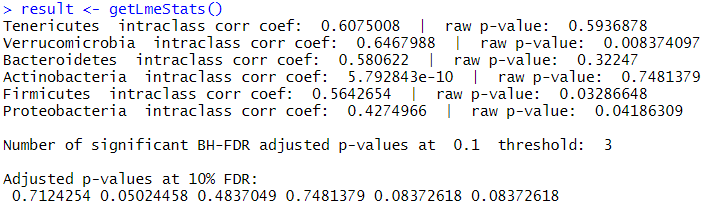
1. For each phyla, graph the relative abundance of that phyla vs. cage. Does there appear to be a cage effect across different phyla?

* The R script that generated the following set of box plots/strip charts at figure 1 is located at URL <https://github.com/jyoung67/advstatistics-labs/blob/master/labs/Lab12/generateBoxPlots.R>.
* Yes, there appears to be “a noticeable” cage effect across four different phyla. My four observations are as follows.
  1. Within the **Verrucomicrobia** graph at figure 1, the median value of all   
     10-/- cages are above the mean relative abundance value (blue dashed line), while most WT cages (4 out of 6) are below the mean.
  2. Within the **Firmicutes** graph at figure 1, the median value of nearly all WT cages (5 out of 6) are above the mean relative abundance value, while nearly all of 10-/- cages (4 out of 5) are below the mean.
  3. Within the **Proteobacteria** graph at figure 1, the median value of all   
     10-/- cages are above the mean relative abundance value, while nearly all of WT cages (5 out of 6) are below the mean.
  4. Within the **Bacteroidetes** graph at figure 1, the median value of nearly all 10-/- cages (4 out of 5) are above the mean relative abundance value, while nearly all of WT cages (5 out of 6) are below the mean.

**Figure 1: Plots of different phyla partitioned by cage**

1. 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?

* The R script for building mixed linear models, calculating intraclass correlation coefficient for each phyla, and identifying significant p-values at 10% FDR is located at URL <https://github.com/jyoung67/advstatistics-labs/blob/master/labs/Lab12/getLmeStats.R>. Output of this script is at figure 2.
* The intraclass correlation coefficient and raw & FDR-adjusted p-values for each phyla are listed in table 1.
* The following three phyla are significantly different for genotype in the mixed model at a 10% false discovery rate. Also, see highlighted entries in table 1.
  + Verrucomicrobia
  + Firmicutes
  + Proteobacteria

**Figure 2: Output of R script getLmeStats.R**

|  |  |  |  |
| --- | --- | --- | --- |
| **Phyla** | **Intraclass Correlation Coefficient** | **Raw P-Value** | **Adjusted P-value at 10% FDR** |
| Tenericutes | 0.6075008 | 0.593687 | 0.7124254 |
| Verrucomicrobia | 0.6467988 | 0.008374097 | 0.05024458 |
| Bacteroidetes | 0.580622 | 0.32247 | 0.4837049 |
| Actinobacteria | 5.792843e-10 | 0.7481379 | 0.7481379 |
| Firmicutes | 5642654 | 0.03286648 | 0.08372618 |
| Proteobacteria | 0.4274966 | 0.04186309 | 0.08372618 |

**Table 1: Calculated intraclass correlation coefficient and raw & FDR adjusted p-values for each phyla**

**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]

Getting the rho(intraclass correlation coefficient) out of a GLS model can be done with:

coef(M.gls$modelStruct[1]$corStruct,unconstrained=FALSE)[[1]]

(4) You can have both points and boxplots on a scatter graph with something like:

boxplot(myFrame$bug ~myFrame$cage)

stripchart(bug ~ cage, data = myFrame,vertical = TRUE, pch = 21, add=TRUE