Lab #7:

Please e-mail code, graphs and answers to questions to [afodor@uncc.edu](mailto:afodor@uncc.edu) with “Lab #7” in the subject line.

The due date is before lab on Wed March 29th.

Due to adverse weather on the Eastern seaboard, we have class as normal on Wed. March 15rh so please come to lecture at 9:30 and lab at 12:30 as usual.

(1) Consider the list price vs. best price of a new GMC pickup (data are from http://college.cengage.com/mathematics/brase/understandable\_statistics/7e/students/datasets/slr/frames/frame.html)

|  |  |
| --- | --- |
| ListPrice | BestPrice |
| 12.4 | 11.2 |
| 14.3 | 12.5 |
| 14.5 | 12.7 |
| 14.9 | 13.1 |
| 16.1 | 14.1 |
| 16.9 | 14.8 |
| 16.5 | 14.4 |
| 15.4 | 13.4 |
| 17 | 14.9 |
| 17.9 | 15.6 |
| 18.8 | 16.4 |
| 20.3 | 17.7 |
| 22.4 | 19.6 |
| 19.4 | 16.9 |
| 15.5 | 14 |
| 16.7 | 14.6 |
| 17.3 | 15.1 |
| 18.4 | 16.1 |
| 19.2 | 16.8 |
| 17.4 | 15.2 |
| 19.5 | 17 |
| 19.7 | 17.2 |
| 21.2 | 18.6 |

(You can copy this data in R with read.delim("clipboard") if you have the data in your clipboard).

Plot out the list price vs. the best price. Fit a linear model. What is the p-value for the null hypothesis that the two prices are not associated? What is the r-squared for the association?

(2) Consider the case-control file for the colorectal adenomas data set that is here:

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

A separate file gives obesity (BMI) data for these same subjects:

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

For each OTU in the spreadsheet, generate a p-value from linear regression comparing BMI to the relative abundance of each OTU. Graph out all the p-values. Do they appear uniformly distributed? Does the microbial community appear to be influencing body weight in this cohort? Are any of these associations significant at a 10% false discovery rate?

Hints: To lookup the ids in the BMI table, you will need to some processing on the “sample” column in the caseControl file. The following code will convert the a sampleID so that it will match the BMI file.

# remove case and control

key <- sub("case", "", sampleID)

key <- sub("control", "", key)

# remove extraneous information from the suffix

key <- strsplit( key, "\_")[[1]][1]

Also, to get the p-value out of the linear model try:

anova( myLm)$"Pr(>F)"

We’ll see why that work shortly in future lectures.