QBIO 310 HW #9

We are going to analyze the attached data set "penguins.csv". This data set has size measurements on adult foraging penguins observed on islands in the Palmer Archipelago near Palmer Station, Antarctica. We're going to do some common statistical tests in R. The R notebook "commontests" that I have posted on Brightspace will be helpful for this assignment. Answer the questions below in an R notebook, and turn in the PDF document. For all figures label the axes. For all tests it is important that you understand the hypotheses and the conditions. R will compute the p-value, then you should clearly state your conclusion (use significance level $\alpha = 0.05$).

1. Were any species found on all three islands? If so, which one(s). Are the number of males and females in the data set roughly balanced for each species? Note: the "table" command will be useful for this problem.

2. Make a side-by-side boxplot comparing bill length for males and females. Do a t-test testing the null hypothesis that mean bill length is the same for males and females. What is your conclusion? Make a QQ-plot of bill length for males, make a separate QQ-plot of bill length for females. Are the t-test conditions satisfied? Also do the Wilcox-Mann-Whitney test. Do the t-test and Wilcox-Mann-Whitney test agree?

3. Make a side-by-side boxplot comparing bill length for all three species. Do a one-way ANOVA to test the null hypothesis that all three specie have the same mean bill length. Show the ANOVA table. What is your conclusion? Make a QQ-plot of the residuals. Make a plot of the residuals vs. the group means. Are the ANOVA conditions satisfied?

4. Make a side-by-side boxplot showing bill length as a function of both sex and species. Do a two-way ANOVA. Show the ANOVA table. For a two-way ANOVA there are three null hypotheses, what are your three conclusions? Make a QQ-plot of the residuals. Make a plot of the residuals vs. the group means. Are the ANOVA conditions satisfied?

5. Make a scatter plot with bill length on the x-axis and bill depth on the y-axis. Compute the correlation. Compute the least squares line, and add the least squares line to your plot. Is the slope (and the correlation) significantly different than zero? If so, does the sign of the slope (and correlation) surprise you? Make a QQ-plot of the residuals. Make a plot with bill length on the x-axis and the residuals on the y-axis. Are the linear model assumptions satisfied?

   Next repeat everything in the previous paragraph, but just for the species Adelie.

   Finally make another scatter plot with bill length on the x-axis and bill depth on the y-axis. This plot should include all three species, but the species should be color-coded (i.e., use one color for the Adelie penguins, a second color for the Chinstrap penguins,

and a third color for the Gentoo penguins). Does this figure explain the surprising observation about the sign of the slopes?

6. What is the data set you have selected for your term paper? If you want, you are welcome to discuss with me your progress on the term paper before you turn it in. It is due the last day of classes on Friday, May 2.