AML 610 Homework #3

**For all questions, please submit your R code, and a doc file with a copy of the screen output and plots (where applicable) to** [**smtowers@asu.edu**](mailto:smtowers@asu.edu)**.**

**Due Thus Sep 26h at noon.**

**The code in your R files should exhibit all of the good coding practices mentioned in** [**http://sherrytowers.com/2012/12/14/good-programming-practices-in-any-language/**](http://sherrytowers.com/2012/12/14/good-programming-practices-in-any-language/)

**Please do not just copy and paste code from the examples given in class into your code. Plots should exhibit all of the good plotting practices mentioned in** [**http://sherrytowers.com/2013/01/04/good-practices-in-producing-plots/**](http://sherrytowers.com/2013/01/04/good-practices-in-producing-plots/)

**Please submit your files in a format hwk3\_<first name>\_<initial of last name>.R and hwk3\_<first name>\_<initial of last name>.doc**

1) Prepare a one page prospectus of your proposed term project. Describe the research question you wish to study. Describe the data sources you intend to use (including links and/or references). Provide at least three references from the literature that describe previous research related to your topic; briefly summarize the studies. Give a brief description of how your proposed study is novel and/or complements previous research. Note that you aren’t required to do any studies at this point.

2) Read in the Chicago pollution information from [www.sherrytowers.com/chicago\_pollution.txt](http://www.sherrytowers.com/chicago_pollution.txt)

a)Select two data frames that are a subset of the original; selecting data in June, and data in July.

b)Using the norm\_overlay\_and\_qq() function in [www.sherrytowers.com/fall\_course\_libs.R](http://www.sherrytowers.com/fall_course_libs.R), assess whether or not the ozone data in January and July (in your opinion) appear to be Normally distributed. Your plot should look like this:



c) Are the two sample sizes large enough that the Central Limit applies?

d) Calculate the mean and standard error on the mean of the two samples, and print to screen.

the mean and standard error of the Jun sample are 0.04525556 0.0008196768

the mean and standard error of the Jul sample are 0.04771505 0.0007368425

e) Calculate the Z statistic that is used to test the hypothesis that the sample means are the same. Calculate the p-value and print to screen. Do you reject or accept the null hypothesis that the two means are consistent with being equal?

The Z-statistic testing equality of means is -2.231479

The p-value testing the hypothesis that these two samples have the same mean is 0.01282469

f) Now calculate the Students-t statistic that is used to determine if the means of the samples are consistent with being equal and print to screen. Are the data consistent with the underlying assumptions of the Students t-test? Calculate the degrees of freedom of the statistic, and print to screen. Test to determine if the means of the distributions are consistent with being equal. Compare the p-value from this test to the one you got from the Z test… why do you think they are almost equal?

The t-statistic testing equality of means is -2.231479

with dof = 719.1918

The p-value testing the hypothesis of equality of means is 0.0129784

3) Repeat (2), this time examining the SO2 pollutant. Comment on whether or not the Z and t tests are appropriate for use with this data to test the equality of means of the two samples.