AML 610 Homework #6

**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****.**

**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 hwk6\_<first name>\_<initial of last name>.R and hwk6\_<first name>\_<initial of last name>.doc**

1) Read in the file <http://www.sherrytowers.com/chicago_crime_summary_c.txt> and use the code below to make a data frame bdat with NA’s for missing time stamps:

cdat=read.table("chicago\_crime\_summary\_c.txt",header=T)

#########################################

# now create a dataframe that has no missing days

# fill the missing info with NA's

#########################################

vjul = seq(min(cdat$jul),max(cdat$jul))

bmat = matrix(NA,ncol=ncol(cdat),nrow=length(vjul))

bdat=as.data.frame(bmat)

names(bdat)=names(cdat)

bdat$jul = vjul

for (icol in 1:ncol(bdat)){

 if (names(bdat)[icol]!="jul"){

 bdat[vjul%in%cdat$jul,icol] = cdat[,icol]

 }

}

2)Correct the battery data, x8, for long term trends using a cubic polynomial fit. In all the parts below, when I say to use the battery data, use the residuals of this fit that corrects the battery data for long term trends.

3)Do a partial autocorrelation plot, pacf, of the trend-corrected battery data. Based on the plot, how many lagged terms do you think are appropriate to include in an autoregression lagged model?

4)Do a lagged regression with the appropriate number of terms. Save the results in an object called myfit. How much to the variance in the data is described by the model? Plot the data and overlay the model. Does it appear to be a good description of the data?

5)Now add the temperature on the day of the batteries into the autoregressive fit (ie; with no lag). Save the result in myfitb. Looking at the AIC; is the relationship between batteries and temperature consistent with nearly instantaneous Granger causality?

6)Now also add the ozone on the day of the batteries into the fit and save the result in myfitc. Looking at the AIC; is the relationship between batteries and temperature consistent with nearly instantaneous Granger causality?

7) From the original cdat data frame above, select the data from 2008 onwards (that data is continuous). Detrend the battery data using a cubic polynomial, and use the detrended data for parts (7) to (9).

Use spec.pgram() to do a Fourier analysis of the spectrum and plot the power spectrum vs the period. Which periods appear to correspond to the most distinct peaks in the power spectrum?

8) Using the function harmonic\_regression() in [www.sherrytowers.com/fall\_course\_libs.R](http://www.sherrytowers.com/fall_course_libs.R) fit a harmonic with period 7 to the data. Is the amplitude of the harmonic significantly different than zero? Now test the amplitude of a harmonic with period 113 days. Is it significantly different than zero? How about the amplitude of a harmonic with period 320 days? What do you think the explanation for the last finding is?

9) Using the function lomb() in [www.sherrytowers.com/fall\_course\_libs.R](http://www.sherrytowers.com/fall_course_libs.R) create a plot of the harmonic amplitude significance (1-pvalue) vs periods 2 to 400 in steps of 1. Are the amplitudes at period=7 and period=365 significantly different than zero?