

APPM2720 Quiz 3

You should work on this quiz on your own and not get help from others. You are however, encouraged to use the web and any other reference materials and resources to complete these questions. This quiz is a total 100 points with the points for each section indicated by [.] .

The R data file is included in the [APPM2720 Quiz3 directory](#) you should also use the lectures and R code from the class as well Chapter 3 from [ISLR](#) for background and examples.

Please submit your work to me by email nychka@ucar.edu by midnight on Friday March 18, 2016. You can submit your answers:

- in the body of an email and attach the plots
- use the markdown format in Rstudio along with the knitr button to create an html file that you email to me (recommended!).
- add your text answers as comments in an R script, use the notebook button and convert your R script to an html file.

The data set [NWSC.rda](#) in the Quiz3 directory was collected to determine the relationship between the mechanical power needed to cool a supercomputer and some key external variables. These data are from the NCAR Wyoming Supercomputing Center (NWSC) during October 2012 and collected about every 6 minutes over October 1 through October 20. These data were taken during a testing period when the computing and cooling equipment were first installed. The supercomputer *Yellowstone* housed at NWSC is the largest computer in the world dedicated to climate and weather related geophysical research and is funded by

the National Science Foundation and the State of Wyoming.

NWSC is a data.frame with the variables:

Day The time in days (and fractions of a day).

Yload Power load on the Yellowstone supercomputer in Kw.

Mpower Power used to cool the computer in Kw

Otemp Outside temperature in degrees F

RH Outside relative humidity (0-100 scale)

Wbulb Wet bulb outside temperature (related to humidity).

The cooling of Yellowstone is done by large chilling units outside the buildings and these are effected not only by the outside temperature but also by the humidity. The goal is to predict the amount of mechanical power needed based on these variables and also based on the power load on the supercomputer. (It is possible that the more power Yellowstone draws the more it will need to be cooled.) There seem to be some periods of spiky power usage and you should not only consider the entire time window but also omit two sections. The first question omits this period from the data set.

You can do this by

```
I1<- (NWSC$Day>= 5) & (NWSC$Day<= 6.765620)
I2<- (NWSC$Day>= 17) & (NWSC$Day<= 18.5)
strange<- I1 | I2      # | means logical OR
OK <- !strange         # ! means logical NOT
NWSC2 <- NWSC[OK,]
```

To create the omitted data frame variables:

```
NWSComit<-NWSC[ strange, ]
```

- 1) [20] Using NWSC2 create a model in **lm** to predict **Mpower** from the other variables. Include a **summary** of your fit. Be sure to include as many variables as you think are important. How useful is the power consumption of Yellowstone for predicting **Mpower**? What is the most important variable?
- 2) [10] Create a plot of the residuals as function of time.
Are there any patterns?
- 3) [15] Apply this model to the full data (**NWSC**) how does your fit change? How does the residual standard error change?
- 4) [15] Use the predict function and your **lm** fit to **NWSC2** to predict the omitted data. (See lectures 8a and 8b from class for some background on using **predict**). Find the residuals for these points and report the standard deviation (**sd** function).
- 5) [10] How are the "residuals" you found in 4) different from the ones in 3)?
- 6) [10] Add the residuals from 4) to your plot in 2) and comment on any patterns. (e.g. use the **points** function with a different color to add these new values.)
- 7) [10] Find a data set on your that you think is suitable to answer a specific question or hypothesis using regression (i.e. the **lm** function.). Clearly describe the source of your data set and your question.
- 8) [10] Analyze your data set and explain how this analysis addresses your question. It may be that in the final analysis the data is not adequate to give a final answer to your question. This is OK but be sure to explain why the data falls short.

