

**Assignment #2**  
CE 764 Hydroinformatics  
Fall 2017.

Maximum marks: 10. Due Date: 11 pm, 30 Aug, 2017.

Email to: [peterck05@gmail.com](mailto:peterck05@gmail.com), [denzilroy@gmail.com](mailto:denzilroy@gmail.com) with a copy to [arpita567@gmail.com](mailto:arpita567@gmail.com)

The daily temperatures, in degree Celsius, in the months of February and March during 1980-2013 is given to you. The filename that has your roll number is the data that you will have to use for this assignment. The latitude and longitude of the location at which these temperatures are measured are also included in the filename. The format of filename is 'Temp\_YourRollNumber\_Lat\_Lon.xls'. The format of the data is 'Year Month Day Temperature'.

Q1. a) Read the data into your programming platform using an appropriate command, and compute the annual seasonal means (i.e. every year, you will have one value which is the average of all daily values of Feb and March in that year). [1]

b) Build a simple linear regression model for annual seasonal mean (ref. Q1.a)) temperature (Y) based on the year of observation (X). Present the results in a table and infer whether there is any significant linear variation of annual seasonal mean temperature with time, at 99% confidence level. What is the command to get the R-sq value directly from the fitted model? [1]

c) Using your computational package, (i) plot the scatter (between X and Y), (ii) the fitted regression line, and (iii) the confidence interval and (iv) the prediction interval of Y, at the observed annual seasonal mean values. Comment on the nature of width of these two intervals across the range of X values. [2]

d) Plot, in another figure, the variation of residuals versus the predicted Y and comment on the adequacy of the fitted regression model. [1]

e) Fit a generalized linear model (GLM) to the same dataset in Matlab with the normal distribution and its canonical link, and comment on the results. [1]

[Optional: (i) Plot a scatter of X and Y. In the same plot, (ii) plot the regression line, and the (iii) pdf of Y|X at the 2<sup>nd</sup>, 10<sup>th</sup>, and 20<sup>th</sup> values of X.]

Q2. It is said that temperatures above 33°C during Feb and March is detrimental to the growth of wheat at these locations.

a) Write a set of commands to count the number of days in each year when the daily temperature had exceeded this threshold value at your location. [1]

b) Fit a GLM for these counts (Y) with the year of observation as the independent variable (X), using the appropriate distribution and its canonical link, using your programming platform. Does the frequency of threshold-exceedance vary significantly with time, at 95% confidence level? Present the results of the model fit in a table. [2]

c) Carry out a likelihood ratio test between two models for the counts of threshold-exceedances, of which considers the variation with time while the other doesn't. Find out the AIC and BIC of these two models and comment whether the model selection based on AIC and BIC is consistent with the likelihood ratio test. [1]

[Optional: Suppose you had many predictors instead of time alone, for example, solar radiation, mean sea level pressure, and relative humidity. How would you fit and find the best GLM in one single command?]

-----Violation of honor code will result in penalty-----

\*\*\*\*\*End\*\*\*\*\*