STAT*3240 F19
Assignment #4

Submit your completed assignment to Crowdmark by 11:59 pm on Friday November 15.
Each part to each question is worth 5 marks.
You can carry out the computational parts of this assignment in either SAS or R. See the document 3240_F19_A4_SASandRcode for some helpful commands.

1. Byers and Williams (1987) investigated the effect of temperature (C) on the viscosity (mPA $\cdot$ s) of toluene-tetralin blinds. The data is found in s3240_F19_viscosity is a modified version of the original data. Import the data and carry out the following.

   (a) Carry out a simple linear regression of viscosity on temperature. Draw the appropriate scatterplot, and superimpose the sample regression line.

   (b) Suppose we want to investigate possible transformations of total viscosity on the fit of the model. As part of the process, we decide to carry out the Box-Cox transformation. Carry out the Box-Cox transformation in R or SAS. What is the (approximate) confidence interval for $\lambda$? What transformation, if any, does this value of $\lambda$ suggest?

   (c) Plot a scatterplot after the appropriate transformation, and superimpose the sample regression line of the fit (the fit using the transformed values).

   Table B.2 in MPV contains data on solar thermal energy. We'll look at a subset of this data, contained in a file called 3240_F19_solar on courselink. Today we'll look at relationship between the total heat flux (kwatts) and several explanatory variables:

   - $x_1$: insolation (watts/m$^2$)
   - $x_2$: position of focal point in east direction (cm)
   - $x_3$: position of focal point in south direction (cm)
   - $x_4$: position of focal point in north direction (cm)

   There was also a fifth explanatory variable (time of day) that I removed for the purposes of this assignment (there was not a significant effect, once these 4 variables were included).

2. Carry out a regression of total heat flux on insolation and the 3 focal point variables. (As a check, the coefficient of insolation should be 0.01911.)

   a) Based on this model, predict the heat flux if $x_1 = 900$, $x_2 = 35$, $x_3 = 35$, $x_4 = 16$.

   b) Using `predict.lm`, find a 95% prediction interval for a single value of heat flux if $x_1 = 900$, $x_2 = 35$, $x_3 = 35$, $x_4 = 16$.

   c) Using `predict.lm`, find a 95% confidence interval for the true mean heat flux if $x_1 = 900$, $x_2 = 35$, $x_3 = 35$, $x_4 = 16$.

   d) Obtain the standardized coefficients. Give an interpretation of the coefficient of insolation. Be sure to relate your conclusion to the variables under study.

3. a) Get the variance inflation factors, and comment on possible multicollinearity.

   b) What observation has the greatest leverage? Comment on whether or not this is an extreme value.

   c) What observation has the externally studentized residual that is greatest in magnitude? Is this value extreme enough to be a point of concern?

   d) What observation has the largest value of Cook's distance? Is this value of Cook's distance large enough to be a point of concern?

e) Obtain partial regression plots (added variable plots) for the 4 predictors. Include them in your submission. Comment on the plots. (Is there any indication that the assumed model is inadequate?)

f) Suppose we want to investigate possible transformations of total heat flux on the fit of the model. As part of the process, we decide to carry out the Box-Cox transformation. Carry out the Box-Cox transformation in R or SAS. What is the (approximate) confidence interval for $\lambda$? What transformation, if any, does this value of $\lambda$ suggest?

4. Consider again the data from the previous question (without any transformations). Suppose we want to use the Bonferroni procedure to calculate simultaneous confidence intervals for all 4 slope parameters.

   (a) If we wish to maintain a family-wise confidence level of at least 90%, what should be the confidence level of the individual intervals?

   (b) What is the margin of error of the interval for the true slope of position of focal point in east direction in the following 3 scenarios:

   - An individual 95% interval.
   - Using the Bonferroni correction with a family-wise confidence level of 95%.
   - Using the Scheffe method with a family-wise confidence level of 95%.

# 1 Extra question that won't be grade

1. Suppose $Var(y) \propto [E(y)]^4$. What transformation of $y$ will stabilize the variance?