Math 484 HW8 Fall 2022, due Friday, Nov. 4.

Problems A)-D), 3 pages.

Exam 2 is Wed, Oct. 26.

Quiz 8 on Nov 2 covers predictor and response transformations, scatterplot matrices, variable selection and models with interactions, factors and powers and whatever is covered on Friday, Oct. 26. (HW6-8).

Final is Fri., Dec. 16, 10:15-12:15.

Read the tips for making a good MLR model below problem F) on HW7.

Shortcut for getting files from my webpage into *Arc*: For problem A) below, open the *pollution.lsp* file with *Notepad*. Then use the menu commands "File>Save As". A window appears. On the top "Save in" box change what is in the box to "My Documents" using the menu on the right of the box. Then go to *Arc* and use the menu commands "File>Load". A window appears. Change "Arc" to "My Documents" and open *pollution.lsp*.

A) 3.17 Get pollution.lsp from my website, save on a flash drive, as you have done for cbrain.lsp on HW1 and lab 1, and cyp.lsp for HW3 C.

Get the McDonald and Schwing (1973) data *pollution.lsp* from (http://parker.ad.siu.edu/Olive/lregbk.htm), and save the file on a flash drive. Activate the *pollution.lsp* dataset with the menu commands "File > Load > Removable Disk (G:) > pollution.lsp." Scroll up the screen to read the data description. Often simply using the log rule on the predictors with $\max(x)/\min(x) > 10$ works wonders.

a) Make a scatterplot matrix of the first nine predictor variables and the response *Mort*. The commands "Graph&Fit > Scatterplot-Matrix of" will bring down a Dialog menu. Select DENS, EDUC, HC, HOUS, HUMID, JANT, JULT, NONW, NOX, and MORT. Then click on *OK*.

A scatterplot matrix with slider bars will appear. Move the slider bars for NOX, NONW, and HC to 0, providing the log transformation. In *Arc*, the diagonals have the min and max of each variable, and these were the three predictor variables satisfying the log rule. Open *Word.*

In *Arc*, use the menu commands "Edit > Copy." In *Word*, use the menu command "Paste." This should copy the scatterplot matrix into the *Word* document. Print the graph.

b) Make a scatterplot matrix of the last six predictor variables and the response *Mort*. The commands "Graph&Fit > Scatterplot-Matrix of" will bring down a Dialog menu. Select OVR65, POOR, POPN, PREC, SO,
WWDRK, and MORT. Then click on *OK*. Move the slider bar of SO to 0 and copy the plot into *Word*. Print the plot as described in a).

c) Click on the *pollution* menu and select *Transform*. Click on the *log transformations* button and select HC, NONW, NOX, and SO. Click on *OK*.

Then fit the full model with the menu commands "Graph&Fit > Fit linear LS". Select MORT for the response. For the terms, select DENS, EDUC, log[HC], HOUS, HUMID, JANT, JULT, log[NONW], log[NOX], OVR65, POOR, POPN, PREC, log[SO], and WWDRK. Click on *OK*.

This model is the full model. To make the response plot use the menu commands "Graph&Fit >Plot of". Select MORT for the V-box and L1:Fit-Values for the H-box. Click on *OK*. When the graph appears, move the OLS slider bar to 1 to add the identity line. Copy the plot into *Word.*

To make the residual plot use the menu commands "Graph&Fit >Plot of". Select L1:Residuals for the V-box and L1:Fit-Values for the H-box. Click on *OK*. Copy the plot into *Word.* Print the two plots.

d) Using the "L1" menu, select "Examine submodels" and try forward selection. Using the "L1" menu, select "Examine submodels" and try backward elimination. You should get a lot of output including that shown in Example 3.7.

Fit the submodel with the menu commands "Graph&Fit > Fit linear LS". Select MORT for the response. For the terms, select EDUC, JANT, log[NONW], log[NOX], and PREC. Click on *OK*.

This model is the submodel suggested by backward elimination. To make the response plot use the menu commands "Graph&Fit >Plot of". Select MORT for the V-box and L2:Fit-Values for the H-box. Click on *OK*. When the graph appears, move the OLS slider bar to 1 to add the identity line. Copy the plot into *Word.*

To make the residual plot use the menu commands "Graph&Fit >Plot of". Select L2:Residuals for the V-box and L2:Fit-Values for the H-box. Click on *OK*. Copy the plot into *Word.* Print the two plots.

e) To make an RR plot use the menu commands "Graph&Fit >Plot of". Select L1:Residuals for the V-box and L2:Residuals for the H-box. Click on *OK*. Move the OLS slider bar to one. On the window for the plot, click on *Options.* A window will appear. Type $y = x$ and click on *OK* to add the identity line. Copy the plot into *Word.* Print the plot.

f) To make an FF plot use the menu commands "Graph&Fit >Plot of". Select L1:Fit-Values for the V-box and L2:Fit-Values for the H-box. Click on *OK*. Move the OLS slider bar to one and click on *OK* to add the identity line. Copy the plot into *Word.*

g) Using the response and residual plots from the full model and submodel along with the RR and FF plots, does the submodel seem ok?

B) Show that the hat matrix $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ is idempotent, that is, show that $\boldsymbol{H}\boldsymbol{H} = \boldsymbol{H}^2 = \boldsymbol{H}$. Warning: $\boldsymbol{X}$ is not square, so $\boldsymbol{X}^{-1}$ does not exist.

C) 3.27 (Basically trying to find a better model than that of HW7 F: 3.11e. Try adding one or two more log terms. Try using service instead of log(service).)

Activate *big-mac.lsp* in *Arc.* Assume that a multiple linear regression model holds for $t(y)$ and some terms (functions of the predictors) where y is BigMac = hours of labor to buy Big Mac and fries. Using techniques you have learned in class find such a model. (Hint: Recall from Problem 3.11 that transforming all variables to logs and then using the model constant, log(service), log(TeachSal) and log(TeachTax) was ok but the residuals did not look good. Try adding a few terms from the minimal C_p model.)

a) Write down the full model that you use (e.g. a very poor full model is

$\exp(BigMac) = \beta_1 + \beta_2 \exp(EngSal) + \beta_3 (TeachSal)^3 + e)$ and include a response plot for the full model. (This plot should be linear.) Give R^2 for the full model.

b) Write down your final model (e.g. a very poor final model is $\exp(BigMac) = \beta_1 + \beta_2 \exp(EngSal) + \beta_3 (TeachSal)^3 + e)$.

c) Include the least squares output for your model and between 3 and 5 plots that justify that your multiple linear regression model is reasonable. Below or beside each plot, give a brief explanation for how the plot gives support for your model.

The output in c) should include k terms with their Wald t tests (so I can tell what the final submodel is) and the Anova F table. Since you have the response plot for the full model in a), in c) include the residual plot for the full model, the response and residual plots for the final submodel, and the RR and FF plots using the final submodel and full model.

I do not want to see anything for models other than the full model or your final submodel.

D) Computers are used to randomly assign treatments to units. Suppose there are 18 M484 students who do HW: 1 Aziz, 2 Bethany, 3 Itria, 4 Kuan, 5 Howard, 6 Maryam, 7 Kasun, 8 Lakni, 9 Sanjuka, 10 Jie Shi, 11 Yangge, 12 Wyatt, 13 Mina, 14 Felicia, 15 Zander, 16 Dhanushka, 17 Dianqiu, 18 Yunjia. Activate *R*. The command *sample(1:18,7)* can be used to assign 7 student of the class to treatment A and the remaining students get treatment B. For example

```
> sample(1:18,7)
[1] 6  13  9 12  1  5 18
```

means that Maryam, Mina, Sanjuka, Wyatt, Aziz, Howard, and Yunjia get treatment A.

Use the command *sample(1:18,7)* until you get treatment A. Include the output in *Word* and give the names of all 7 students that get treatment A.