Math 484 HW6 Fall 2021, due Friday, October 14. Long HW!

Long homework, maybe 3.5 hours. Problems A)-E), 4 pages.

Quiz 6 on Oct 12 covers predictor transformations, scatterplot matrices and HW6.

Exam 2 is Wed., Oct. 26, covers Exam 1, quiz 1-7 and HW1-7 material.

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

In the online book, Multiple Linear and 1D Regression, A) 3.10 is 3.9 and C) 3.9 is 3.8. **For some computers, to get the data in Arc, use commands like HW5C):** "File>Load>OSDisc(C:)>Program Files(x86)>Arc>Data>bcherry.lsp"

A) 3.10 In *Arc* enter the menu commands "File>Load>Data" and open the file *bcherry.lsp*. The menu *Trees* will appear. Use the menu commands "Trees>Transform" and a dialog window will appear. Select terms *Vol*, *D*, and *Ht*. Then select the *log* transformation. The terms *log Vol*, *log D*, and *log Ht* should be added to the data set. If a tree is shaped like a cylinder or a cone, then $Vol \propto D^2 Ht$ and taking logs results in a linear model.

a) Fit the full model with $Y = \log Vol$, $X_1 = \log D$, and $X_2 = \log Ht$. Add the output that has the LS coefficients to *Word*.

b) Fitting the full model will result in the menu *L1*. Use the commands "L1>AVP–All 2D." This will create a plot with a slider bar at the bottom that says *log[D]*. This is the added variable plot for $\log(D)$. To make an added variable plot for $\log(Ht)$, click on the slider bar. Add the OLS line to the AV plot for $\log(Ht)$ by moving the *OLS slider bar* to 1, and add the zero line by clicking on the "Zero line box". Include the resulting plot in *Word*.

c) Fit the reduced model that drops log(Ht). Make an RR plot with the residuals from the full model on the V axis and the residuals from the submodel on the H axis. Add the LS line and the identity line as visual aids. (Click on the *Options* menu to the left of the plot and type "y=x" in the resulting dialog window to add the identity line.) Include the plot in *Word*.

d) Similarly make an FF plot using the fitted values from the two models. Add the OLS line which is the identity line. Include the plot in *Word*.

e) Next put the residuals from the submodel on the V axis and $\log(Ht)$ on the H axis. Move the *OLS slider bar* to 1, and include this residual plot in *Word*.

f) Next put the residuals from the submodel on the V axis and the fitted values from the submodel on the H axis. Include this residual plot in *Word*.

g) Next put log(Vol) on the V axis and the fitted values from the submodel on the H axis. Move the *OLS slider bar* to 1, and include this response plot in *Word*.

h) Does $\log(Ht)$ seem to be an important term? If the only goal is to predict volume, will much information be lost if $\log(Ht)$ is omitted? **Beside each of the 6 plots, remark on the information given by the plot**. (Some of the plots will suggest that log(Ht) is needed while others will suggest that log(Ht) is not needed.)

In part b) move the OLS slider bar to 1 and click on the zero line box. In parts e), f) and g) move the OLS slider bar to 1. (*bcherry.lsp* data)

B) 3.5 You may also copy and paste *R* commands for this problem from (http://parker.ad.siu.edu/Olive/lreghw.txt).

a) Download the *R* function `tplot` that makes the transformation plots for $\lambda \in \Lambda_L$. To do a), copy and paste the 2 source commands near the top of the above file into *R*.

b) Use the following *R* command to make a 100×3 matrix. The columns of this matrix are the three nontrivial predictor variables.

```
nx <- matrix(rnorm(300),nrow=100,ncol=3)
```

Use the following command to make the response variable Y.

```
y <- exp( 4 + nx%*%c(1,1,1) + 0.5*rnorm(100) )
```

This command means the MLR model $\log(Y) = 4 + X_2 + X_3 + X_4 + e$ will hold where $e \sim N(0, 0.25)$.

To find the response transformation, you need the program `tplot` given in a). Type **ls()** to see if the programs were downloaded correctly.

c) To make the transformation plots type the following command.

```
tplot(nx,y)
```

The first plot will be for $\lambda = -1$. Move the cursor to the plot and hold the **rightmost mouse key** down and highlight **Stop** to go to the next plot. Repeat these *mouse* operations to look at all of the plots. The identity line is included in each plot. When you get a plot where the plotted points cluster about the identity line with no other pattern, include this transformation plot in *Word* by pressing the **Ctrl** and **c** keys simultaneously. This will copy the graph. Then in *Word* use the menu command "Paste". You should get the log transformation.

d) Type the following commands.

```
out <- lsfit(nx,log(y))
ls.print(out)
```

Use the mouse to highlight the created output and include the output in *Word.*

e) Write down the least squares equation for $\widehat{\log(Y)}$ using the output in d).

C) 3.9 a) *wool.lsp* data

The file *wool.lsp* has data from a 3^3 experiment on the behavior of worsted yarn under cycles of repeated loadings. The response Y is the number of cycles to failure and the three predictors are the length, amplitude, and load. Make five transformation plots by using the following commands.

From the menu "Wool" select "transform" and double click on *Cycles*. Select "modified power" and use $p = -1, -0.5, 0$, and 0.5. Use the menu commands "Graph&Fit>Fit linear LS" to obtain a dialog window. Next fit LS five times. Use *Amp*, *Len*, and *Load* as the predictors for all 5 regressions, but use Cycles^{-1}, $\text{Cycles}^{-0.5}$, log[Cycles], $\text{Cycles}^{0.5}$, and Cycles as the response.

Use the menu commands "Graph&Fit>Plot of" to create a dialog window. Double click on L5:Fit-Values and double click on Cycles, double click on L4:Fit-Values and double click on $\text{Cycles}^{0.5}$, double click on L3:Fit-Values and double click on log[Cycles], double click on L2:Fit-Values and double click on $\text{Cycles}^{-0.5}$, double click on L1:Fit-Values and double click on Cycles^{-1}.

a) You may stop when the resulting plot in linear. Let Z = Cycles. Include the plot of $\widehat{Y}$ versus $Y = Z^{(\lambda)}$ that is linear in *Word*. Move the OLS slider bar to 1. What response transformation do you end up using?

D) 3.15 a)–f) So do not do parts g), h) and i).

Activate the *cement.lsp* data, on the course webpage. Act as if 20 different samples were used to collect this data. If 5 measurements on 4 different samples were used, then experimental design with repeated measures or longitudinal data analysis may be a better way to analyze this data.

a) From *Graph&Fit* select *Plot of*, place $x1$ in H, y in V, and $x2$ in the *Mark by* box. From the OLS menu, select *Fit by marks–general* and move the slider bar to 2. Include the plot in *Word*.

b) A quadratic seems to be a pretty good MLR model. From the *cement* menu, select Transform, select x1, and place a 2 in the p box. This should add x1^2 to the data set. From *Graph&Fit* select *Fit linear LS*, select x1 and x1^2 as the terms and y as the response. Include the output in *Word*.

c) Make the response plot. Again from the OLS menu, select *Fit by marks–general* and move the slider bar to 1. Include the plot in *Word*. This plot suggests that there is an interaction: the CM cement is stronger for low curing times and weaker for higher curing times. The plot suggests that there may not be an interaction between the two new types of cement.

d) Place the residual plot in *Word*. (Again from the OLS menu, select *Fit by marks–general* and move the slider bar to 1.) The residual plot is slightly fan shaped.

e) From the *cement* menu, select *Make factors* and select x2. From the *cement* menu, select *Make interactions* and select x1 and (F)x2. Repeat, selecting x1^2 and (F)x2. From *Graph&Fit* select *Fit linear LS*, select x1, x1^2, (F)x2, x1*(F)x2, and x1^{2}*(F)x2 as the terms and y as the response. Include the output in *Word*.

f) Include the response plot and residual plot in *Word*.

E) 3.12 The following data set has 5 babies that are "good leverage points:" they look like outliers but should not be deleted because they follow the same model as the bulk of the data.

a) In *Arc* enter the menu commands "File>Load>Removable Disk (G:)" and open the file *cbrain.lsp*. Select *transform* from the *cbrain* menu, and add $size^{1/3}$ using the power transformation option ($p = 1/3$). From

Graph&Fit, select *Fit linear LS.* Let the response be *brnweight* and as terms include everything but *size* and *Obs.* Hence your model will include $size^{1/3}$. This regression will add *L1* to the menu bar. From this menu, select *Examine submodels.* Choose *forward selection.* You should get models including $k = 2$ to 12 terms including the constant. Find the model with the smallest $C_p(I) = C_I$ statistic and include all models with the same k as that model in *Word.* That is, if $k = 2$ produced the smallest C_I, then put the block with $k = 2$ into *Word.* Next go to the *L1* menu, choose *Examine submodels* and choose *Backward Elimination.* Find the model with the smallest C_I and include all of the models with the same value of k in *Word.*

b) What was the minimum C_p model was chosen by forward selection?

c) What was the minimum C_p model was chosen by backward elimination?

d) Which minimum C_p model do you prefer? Explain.

e) Give an explanation for why the two models are different.

f) Pick a submodel and include the regression output in *Word.*

g) For your submodel in f), make an RR plot with the residuals from the full model on the V axis and the residuals from the submodel on the H axis. Add the OLS line and the identity line y=x as visual aids. Include the RR plot in *Word.*

h) Similarly make an FF plot using the fitted values from the two models. Add the OLS line which is the identity line. Include the FF plot in *Word.*

i) Using the submodel, include the response plot (of $\hat{Y}$ versus Y) and residual plot (of $\hat{Y}$ versus the residuals) in *Word.*

j) Using results from f)-i), explain why your submodel is a good model.

For part d) explain your choice. For f), I liked a model with $p = 4$ so 3 nontrivial predictors and a constant. The k in *Arc* is p, that is, $k = p$.

See answers in Section 14.2.