Math 484   HW 2 Fall 2022, due Friday, September 9.

3 pages: problems A-E

Quiz 2 on Friday, Sept. 9, 4 sheets of notes. Know residual plots, response plots, how to get $\hat{Y}$ given $\boldsymbol{x} = (1, x_2, ..., x_p)^T$, or for simple linear regression (SLR) given x. Know how to do the Anova F test for $\beta_2 = \cdots = \beta_p = 0$. See points 5), 6) and 10) on p. 62-3. See pages 1-2 on exam 1 review.

Get the SAS output for this problem from the Math 484 webpage.

A) 2.18. This problem shows how to use *SAS* for MLR. The data are from Kutner et al. (2005, problem 6.5). The response is "brand liking," a measurement for whether the consumer liked the brand. The variable $X1$ is "moisture content" and the variable $X2$ is "sweetness." Copy and paste the program for this problem from (http://parker.ad.siu.edu/Olive/lreghw.txt).

a) Execute the *SAS* program and copy the output file into *Notepad.* Scroll down the output that is now in *Notepad* until you find the regression coefficients and ANOVA table. Then cut and paste this output into *Word.*

b) Do the 4 step ANOVA F test.

You should scroll through your *SAS* output to see how it made the response plot and various residual plots, but cutting and pasting these plots is tedious. So we will use *Minitab* to get these plots. Find the program for this problem from (http://parker.ad.siu.edu/Olive/lreghw.txt). Then copy and paste the numbers (between "cards;" and the semicolon ";") into *Minitab.* Use the mouse commands "Edit>Paste Cells". This should enter the data in the Worksheet (bottom part of *Minitab*). Under **C1** enter **Y** and under **C2** enter **X1** under **C3** enter **X2**. Use the menu commands "Stat>Regression>Regression" to get a dialog window. Enter Y as the response variable and $X1$ and $X2$ as the predictor variable. Click on **Storage** then on **Fits**, **Residuals**, and **OK OK**.

c) To make a response plot, enter the menu commands "Graph>Plot" and place "Y" in the Y–box and "FITS1" in the X–box. Click on **OK**. Then use the commands "Edit>Copy Graph" to copy the plot. Include the plot in *Word* with the commands "Edit> Paste." If these commands fail, click on the graph and then click on the printer icon.

d) Based on the response plot, does a linear model seem reasonable?

e) To make a residual plot, enter the menu commands "Graph>Plot" and place "RESI 1" in the Y–box and "FITS1" in the X–box. Click on **OK**. Then use the commands "Edit>Copy Graph" to copy the plot. Include the plot in *Word* with the commands "Edit> Paste." If these commands fail, click on the graph and then click on the printer icon.

f) Based on the residual plot does a linear model seem reasonable?

The response is "brand liking," a measurement for whether the consumer liked the brand. The variable $X1$ is "moisture content" and the variable $X2$ is "sweetness."

B) In *ARC*, load the file *demo-bn.lsp* with the menu commands "File>Load>Data>Arcg>demo-bn.lsp." In the dialog window, enter the numbers 0 0 1 1 .7 and 50. This makes bivariate normal data where x and y are iid N(0,1) but have correlation = 0.7 and a sample size $n = 50$. Move the "OLS slider bar" to 1 to add the OLS line $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$ to the scatterplot of x versus y. (In a response plot of the fitted values versus the response, the OLS line would always be the identity line.) Click on the "Rem lin trend" box to make a "residual plot" of x versus the residuals. (A residual plot of the fitted values versus the residuals would scale the horizontal axis differently.) The point of this problem is that for each data set, the MLR model is good, but the residual plot need not be good. Also you will see several residual plots and get an idea of how much they vary from the ideal shape. Click on the "New sample" box and then click on the "Rem lin trend" box. Repeat this process several times.

a) When you obtain a residual plot that does not look very good, include it in *Word.* In *Arc*, use the menu commands "Edit > Copy." In *Word*, use the menu commands "Edit > Paste."

b) **Explain why your plot does not look good.**

C) In *ARC*, load the file *demo-prb.lsp* with the menu commands "File>Load>Data>Arcg>demo-prb.lsp." Enter the menu commands "ProbPlots>Probability Plots" and click "OK" on the resulting dialog window. For part a) the data are iid $N(0, \sigma^2)$. The normal probability plot is used for checking the normality assumption. Part a) of this problem gives you an idea of how much the normal probability plot can vary from the ideal straight line shape even when the model is good. For part b) the errors are iid χ^2_3, a highly skewed distribution. Ideally the plot will not look like a straight line, but for some data sets, the plot looks linear and so fails. In parts a) and b) we are acting as if the iid errors e_i are known. For MLR, the residuals only estimate the iid errors, and QQ or probability plots of the residuals work even worse.

a) Click on "New Sample" several times. When you get a normal probability plot that is not very straight, include it in *Word.* In *Arc*, use the menu commands "Edit > Copy." In *Word*, use the menu commands "Edit > Paste."

b) Again enter the menu commands "ProbPlots>Probability Plots" but in the dialog window click on "Chi Squared" in the Sampling distribution options. Then click on "OK." Now the data comes from a χ^2_3 distribution, and ideally, the plots should not appear straight. Click on "New sample" until you get a fairly straight plot and include it in *Word* as done in a).

D) 2.28 Get the *R* commands for this problem from (http://parker.ad.siu.edu/Olive/lreghw.txt). The data is such that $Y = 2+x_2+x_3+x_4+e$ where the zero mean errors are iid [exponential(2) - 2]. Hence the residual and response plots should show high skew. Note that $\boldsymbol{\beta} = (2, 1, 1, 1)^T$. The *R* code uses 3 nontrivial predictors and a constant, and the sample size $n = 1000$.

a) Copy and paste the commands for part a) of this problem into *R*. Include the response plot in *Word*. Is the lowess curve fairly close to the identity line?

b) Copy and paste the commands for part b) of this problem into *R*. Include the residual plot in *Word*: press the *Ctrl* and *c* keys as the same time. Then use the menu commands "Edit>Paste" in *Word*. Is the lowess curve fairly close to the $r = 0$ line?

c) The output out$coef gives $\hat{\boldsymbol{\beta}}$. Write down $\hat{\boldsymbol{\beta}}$. Is $\hat{\boldsymbol{\beta}}$ close to $\boldsymbol{\beta}$?

E) 2.4c: (just do part c) do not do parts a) or b)

```
Problem 2.4 Output, Coef. Estimates Response = height
Label            Estimate Std. Error  t-value p-value
Constant         227.351  65.1732      3.488  0.0008
sternal height   0.955973  0.0515390  18.549  0.0000
finger to ground 0.197429  0.0889004   2.221  0.0295

R Squared: 0.879324    Sigma hat:  22.0731

Summary Analysis of Variance Table
Source      df      SS       MS          F    p-value
Regression   2  259167.  129583.    265.96    0.0000
Residual    73  35567.2  487.222
```

2.4. The above output, is from the multiple linear regression of the response $Y =$ *height* on the two nontrivial predictors *sternal height* = height at shoulder, and *finger to ground* = distance from the tip of a person's middle finger to the ground.

c) From the output, are *sternal height* and *finger to ground* useful for predicting *height*? (Perform the ANOVA F test.)