`Math 484   HW5 Fall 2022, due Friday, October 7.`

Quiz 5 on Oct 5 has a problem like HW4 F and also covers HW5 material. RR and FF plots and a scatterplot matrix could be on the quiz. **2 pages, A)-E)**

Exam 2 is Wed., Oct. 26, covers Exam 1, quiz 1-6 and HW1-6 material.
Final is Fri., Dec. 16, 10:15-12:15.

A) 2.13 The normal error model for simple linear regression through the origin is

$$Y_i = \beta X_i + e_i$$

for $i = 1, ..., n$ where $e_1, ..., e_n$ are iid $N(0, \sigma^2)$ random variables.

a) Show that the least squares estimator for β is

$$\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}.$$

b) Find $E(\hat{\beta})$.

c) Find VAR$(\hat{\beta})$.

(Hint: Note that $\hat{\beta} = \sum_{i=1}^n k_i Y_i$ where $k_i = X_i / \sum_{i=1}^n X_i^2$. Hence $E(\hat{\beta}) = \sum_{i=1}^n k_i E(Y_i) = \sum_{i=1}^n k_i \beta X_i$. The X_i which are treated as constants.
Similarly, $VAR(\hat{\beta}) = V(\sum_{i=1}^n k_i Y_i) = \sum_{i=1}^n k_i^2 V(Y_i)$ by independence, and $V(Y_i) = \sigma^2$.)

B) Activate cyp.lsp as in HW3 C).

The response variable Y is *height*, and the explanatory variable is $X_2 =$ *sternal height* (probably height at shoulder). Want to see if regression through the origin (no intercept) is an appropriate model for this data.

Enter the menu commands "Graph&Fit>Fit linear LS" and fit the model: enter *sternal height* in the "Terms/Predictors" box, and *height* in the "Response" box. Uncheck the "Fit Intercept" box, and click on *OK*.

a) Include the output in *Word*.

b) The t-test and no intercept Anova F test for this model will be equivalent since $p = 1$. Show that $t^2 = Fo$ (or $t^2 \approx Fo$ due to rounding).

c) Perform the 4 step no intercept Anova F test.

d) Predict Y if sternal height = 1390.

e) Include the response and residual plots in *Word*. (See problem 2.24 if you do not know how to make these plots. Move the OLS slider bar to 1 for both plots.)

f) The no intercept model is not appropriate, a constant is needed. How does the residual plot show that the no intercept model is not appropriate (ignoring 2 potential outliers)?

C) a) In *ARC* enter the menu commands "File>Load>Data" and open the file `big-mac.lsp` (sometimes you need to use the commands "File>Load>OSDisc(C:)>Program Files(x86)>Arc>Data>big-mac.lsp"). Scroll up the window to read the description of the data. Use the commands "Graph&Fit>Scatterplot Matrix of". In the dialog window select Bread, Busfare, TeachSal, TeachTax and BigMac (last). Click on "OK" and include the scatterplot matrix in *Word*.

b) The response BigMac is the minutes of labor needed to buy a BigMac in various countries. Are any of the marginal *predictor relationships* nonlinear (so ignore the top row and the rightmost column of the scatterplot matrix)?

c) Is $E(BigMac|TeachSal)$ linear or nonlinear?

d) Which variables satisfy the log rule?

e) Consider the 3rd plot in the top row which has BigMac on the vertical axis and TeachSal on the horizontal axis. To make this plot linear, does the ladder rule suggest that the power λ for the power transformation of BigMac should be increased or decreased? You can move the slider bar for BigMac to check your answer.

D) a) In *ARC* enter the menu commands "File>Load>Data" and open the file `brains.lsp` (sometimes you need to use the commands "File>Load>OSDisc(C:)>Program Files(x86)>Arc>Data>brains.lsp"). Scroll up the window to read the description of the data. Use the commands "Graph&Fit>Scatterplot Matrix of". In the dialog window select BodyWt and BrainWt. Click on "OK" and include the scatterplot matrix in *Word*.

b) These variables are highly right skewed and both satisfy the log rule. The small values of both variables need spreading. Move both slider bars to 0 to perform the log transformation, and include the scatterplot matrix in *Word*. Are the plots linear?

E) 2.25 In *Arc* enter the menu commands "File>Load>Removable Disk (G:)" and open the file *cyp.lsp* (obtained as in Problem 2.22).

The response variable Y is *height*, and the explanatory variables are X_2 = *sternal height*, X_3 = *finger to ground*, X_4 = *bigonal breadth*, X_5 = *cephalic index*, X_6 = *head length*, and X_7 = *nasal height*. Enter the menu commands "Graph&Fit>Fit linear LS" and fit the model: enter the 6 predictors (in order: X_2 1st and X_7 last) in the "Terms/Predictors" box, *height* in the "Response" box and click on *OK*. This gives the *full model*. For the *reduced model*, only use predictors 2 and 3.

a) Include the ANOVA tables for the full and reduced models in *Word*.

b) Use the menu commands "Graph&Fit>Plot of..." to get a dialog window. Place *L2:Fit-Values* in the H–box and *L1:Fit-Values* in the V–box. Place the resulting plot in *Word*.

c) Use the menu commands "Graph&Fit>Plot of..." to get a dialog window. Place *L2:Residuals* in the H–box and *L1:Residuals* in the V–box. Place the resulting plot in *Word*.

d) Both plots should cluster tightly about the identity line if the reduced model is about as good as the full model. Is the reduced model good?

e) Perform the 4 step partial F test (of Ho: the reduced model is good) using the 2 ANOVA tables from part a).