36-617: Applied Linear Models
Fall 2019
HW08 – Due Fri Nov 8, 11:59pm
Grace until Sun Nov 10, 11:59pm

- Please turn the homework as a pdf online in our course webspace at canvas.cmu.edu, using the gradescope link for hw08 under Assignments.
- Reading for next week: Sheather Ch 10, *Section 1 only*. There will be a quiz on Monday.
- I will be out of town at a meeting Thursday through Sunday this week.

Exercises

1. Please list the people you worked with on this assignment.
2. Sheather, pp. 328–329, #2. The data set bookstore.txt can be found in the "0 - Sheather stuff" directory in the files area for our class on Canvas.
3. Consider the data for Sheather, p. 329, #3. The data set CarlsenQ.txt can be found in the "0 - Sheather stuff" directory in the files area for our class on Canvas. Use summary(), View(), plot(), pairs() etc. to familiarize yourself with the data (you don't have to turn in any of this initial exploration).

   Instead of what is asked for by Sheather, please do the following:

   (a) Fit the ordinary linear model predicting Sales from all variables *except* Case and Time (Case is just an observation number and Time is redundant with the Q variables). Use summary(), plot(), acf() and any other tools to assess the strengths and weaknesses of this model.

   (b) Use the function gls() from library(nlme) to fit the same model but now allowing the residuals to have an AR1 correlation. Note: your function call is going to look something like this:

   ```
   gls(Sales ~ (whatever), data=carlsen, correlation=corAR1())
   ```

   i. What is the estimated lag-1 autocorrelation?

   ii. Compare the estimated coefficients in the models from part (a) and part (b). Are there changes in what is or is not a significant predictor?

   (c) Use the `logLik()` function to construct a likelihood ratio test for whether the autocorrelation should be 0 (model from part (a)) or nonzero (model from part (b)).

   (d) Transform the model from part (a) by premultuplying by the inverse square root of the autocorrelation matrix for the residuals, assuming an AR(1) model and using the value you found in part (b)(i) above, and fit the new model. *(Hint: use the recipe from lecture 21 in class.)* Use summary(), plot(), acf() to assess the strengths and weaknesses of this model.

[Continued on next page]

4. **[Not quite multilevel models]** Consider again the `cdi.dat` data from Project 02. Construct the variable `pct.hs.grad` $<-$ (`hs.grad` / `pop`) $\times$ 100%. Then, using `state` as the cluster (or group) variable, create four plots like those on slides 12–15 of lecture 22 (intro to mlm I), using `pct.hs.grad` as the x variable, and `per.cap.income` as the y variable:

   (a) Ignore `pct.hs.grad` and only look at `mean(per.capita.income)` or `per.cap.income ~ 1` in each state

   (b) Ignore states and fit a single linear regression `per.capita.income ~ pct.hs.grad`

   (c) Use same slope on `pct.hs.grad` for all states, different intercepts

   (d) Fit a different regression `per.capita.income ~ pct.hs.grad` in each state, ignoring all the other states

   Print out the plots, and write two sentences for each graph: The first sentence should describe good and bad features of the plot; the second sentence should provide a comparison of this plot with the other three.

   Nb., None of these is a multilevel model yet. The multilevel model would provide fits to the individual states which are a compromise between plots (b) [a single linear fit for all data, ignoring states] and (d) [a different linear fit for each state, ignoring the other states].