**BIOS6643 Fall 2018 HW 7 Questions to turn in and due date TBD**

Background: Children with asthma reported whether they had symptoms on a daily basis within a 2-year study period, from 2003-05 (our NJH asthma / air pollution study). The survey was collected on approximately 14 to 16 days per year, per child (see the data set to get the response patterns). Days where no collection occurred was not considered missing, since it was not the subject, but rather the design that dictated the intermittent data collection. The question is whether asthma symptoms are related to air pollution levels. [As an aside: in the past we have not found self-reported asthma symptoms to be a sensitive measure of health, particularly as it relates to more subtle environmental exposures such as air pollution. Of course, this is Denver; the exposure levels would not be so subtle in Beijing! The most sensitive health measures we have found are rescue inhaler use and LTE4 (a biomarker of inflammation). Nevertheless, let’s take a look at asthma symptoms and see what happens.] So the synopsis of your analysis is to examine the relationship between air pollution and asthma symptoms. This will be done in several steps, described below. This is real data! I have attached the data set on Canvas (see the ‘asthma data’ SAS data set).

1. Develop an ordinal logistic regression model for asthma symptoms (*Y*). The levels of *Y* are as follows: 1=I did not cough or wheeze last night; 2=I coughed or wheezed a little, but I slept well; 3=I coughed or wheezed, and it work me once or woke me early; 4=I coughed or wheezed a lot and it woke me 2 or more times; 5=I coughed or wheezed a lot, and it kept me awake most of the night. In this model, the predictors will be: l1mmaxpm25 (lag 1 of yesterday’s morning maximum PM2.5 concentration), temperature, humidity, l1cold (lag 1 indicator of presence of cold). Just for consistency, use a random intercept for subject, and run method=rmpl in SAS. Some notes: rspl did not converge, which is why I switched to rmpl; this represents a ‘marginal’ locus expansion rather than an expansion at the ‘random effects=0’ case; I tried other predictors such as date, year and date(year) but they did not seem to contribute much, so I dropped them. For the outcome of asthma, use ‘asthma(desc)’ to get more intuitive estimates (as with application in the notes); you cannot add serial correlation to the model when using the multinomial distribution. Question: how do air pollution effects compare with the common cold? How can you manipulate the slope values so that you have an apples-to-apples comparison?
2. Compare the model in (1) to one where method=laplace is used. Comment on differences in output. What are the methodological differences between these approaches, in general?
3. Determine P(*Y*=*y*) for *y*=1, 2, 3, 4 and 5. Note that your answer will depend on levels of the covariates, so use average values of the 4 covariates. To get the mean values of covariates, restrict the data so that it like data used in the OLR model. For example, you can include where asthma^=**.** and l1cold^=**.**; in the PROC MEANS statement; this should cover most of it. What if you had used, say, the first quartile of each of the covariates? How would your answer change, generally? You can answer by just thinking about it or working it out mathematically…I am not as concerned about specific numerical values here. But if you want you can just redo with 1st quartile values and then try to understand the patterns.
4. Repeat the analysis but in this case, dichotomize the asthma variable into 1 versus 2-5. For this analysis, run it the following ways:
   1. Use GzLM with GEE. What kind of interpretations do coefficients have here?
   2. Use a GzLMM with a random intercept only. What method do you choose here? What type of interpretations do coefficients have here?
   3. Use a GzLMM with a random intercept, plus see if adding serial correlation to the model is helpful. What type of interpretations do coefficients have here?
   4. Compare the results here with the OLR regression. From what you can tell, are results consistent? If not, any thoughts on what might be leading to differences?

Follow up: not to turn in unless you want to: run the model in different ways than what I proposed. Do you think you can find a better model?