Homework6

BIOS6643 Fall 2021

8/20/2021

Question 1 GzLMM

We have learned about GzLMM's that are basically a combination of LMM's and GzLM's. One of the complexities of GzLMM's is that maximizing the (true) likelihood involves using a numerical technique like Gaussian quadrature. In not more than 2 sentences (1 might even work), explain what makes the GzLMM likelihood more complicated than the LMM likelihood, in terms of maximization. We will distinguish 'good' versus 'best' answers in grading.

Question 2 Model Specification

A study is planned where data will be collected on asthmatic subjects on every weekday for one month. There are two outcome measures of interest, (i) medication use counts and (ii) FEV1. You are the statistician and the PI is looking for your suggestions about models to use.

a. If it is anticipated that responses within subjects over time are serially correlated (but with some decay the further measurements are apart) for both outcomes, which SAS procedure (or R package/function) would you suggest using to fit the data? Answer separately for each outcome.

b. Related to a, talk about how you would set up the data and specify the REPEATED statement in SAS (or comparable code for R) for each outcome, so that the correlation between responses is accounted for properly, including gaps caused by no measurements on weekends. (NOTE: to deal with unequal spacing in GzLM/GEE, you need to include records for equally spaced time points and fill in with missing values as necessary, e.g., for weekends as described above; we will discuss this more soon.)

c. Say that we now consider an indicator of whether subjects used medication or not on a given day (no $use = 0$, at least 1 $use = 1$). In this case, the researcher is more concerned about accounting for general differences between subjects in the model (e.g., on one extreme there may be big users and on the other, very little users) than accounting for serial correlation (although the latter may still exist). What procedure would you suggest using if you wanted to account for between-subject variability of use, and also approximate the true likelihood in estimation? What are the drawbacks of this approach?

d. For **part c**, suggest a procedure you might use if you wanted to include both a random intercept for subjects in the model, as well as account for potential serial correlation of repeated measures. What are the drawbacks of this approach?

Question 3 GzLMM and GEE for Albuterol data

In class we have discussed the albuterol data, which involves children `emo::ji("child")` who take rescue medication for their asthma. They take this 'as needed' (i.e., based on how they feel) but are also prescribed to take it (i.e., 'pre-treats'). We would like to see how the daily albuterol use counts relate to air pollution measures, controlling for other covariates. The air pollution variable used here is ln(morning hourly maximum $PM_{2.5}$). To help account for the pre-treats, the indicator variable Friday was included in the model (the one day they did not receive pre-treats since there is no gym class). Often meteorological variables are also controlled for in air pollution models; here we will include temperature, pressure and humidity. Finally, also add 'date' as a predictor; treat as continuous. Complete the following.

a. Run GEE and use the MODELSE option in the REPEATED statement to incorporate the scale parameter into the GEE process. Highlight the results. Does adding the scale parameter into the process modify the SE's up or down? Use the AR(1) working covariance structure.

b. Now fit the GzLMM using RSPL approach (the default method in PROC GLIMMIX). Include a spatial power structure to account for serial correlation (use 'date' as the indexing variable). Highlight the results. How does the scale parameter in GEE compare with the residual variance in the GzLMM PL approach? (Recall that the residual variance in GzLMM PL acts as the scale parameter; make sure to compare apples-to-apples, though.)

c. Do the scale / residual variance estimates suggest over or under-dispersion in the data (considering Poisson distribution)?

d. How do slope estimates and SE's for ln(mm$PM_{2.5}$) differ between the GEE and GzLMM PL approach? What about these SE's compared to the empirical SE of GEE (which is also given in default output)?

e. In a sentence, interpret the relationship between morning particulate matter (1 hour maximum) and children's albuterol use. In order to make the slope more meaningful, interpret the effect per SD increase in the pollutant variable. Use the GzLMM PL estimate to do this.

Question 4 Longitudinal logistic regression models

Consider fitting the exacerbation data using various longitudinal logistic regression models. Note: I am providing SAS code for you to finish this homework exercise. However, if you would prefer to try R, that is fine; you will just need to create your own programs; just translate the given models into R (e.g., use the same predictors and what not). Write a 1-page summary of results that includes responses to questions below. You can compare results between models other than your 'best' ones. You can also include tables or figures, if you'd like. You will be graded on this both quantitatively and qualitatively (i.e., distinguishing 'good' versus 'best' responses).

a. For each major approach (GEE, quadrature, linearization), determine the 'best model' among those in the SAS program.

b. Now between approaches, pick the model that you think is the 'best' or more appropriate. This could be based on which model clearly meets assumptions best and/or offers the best way to model the data, in your opinion. Note: it isn't really possible to compare goodness-of-fit statistics between modeling approaches, so you will need to use other ways to make your decision. However, talk about how you made your decision.

c. Consider 2 particular effects, day and weekend (former is 'continuous' and the latter is binary). Interpret effects for these predictors in words, and mention which type of interpretation it has (population averaged or subject specific).