Homework 3

STAT 697STA Spring 2023

Due March 1, 2023, 9:40am on Gradescope

1   Reading

- Read the rest of Chapter 2 of the text.

2   Questions

1. HRW 2.6

   Hint: for part (d), you want to access `exactRLRT(fitGAMM$lme)$p.value`

2. HRW 2.7

   Hint: for part (b), I used the following process

   - Create a function, based on the code from 2.6, with (200 points of) data generated as in 2.7a to calculate the power for one value of θ. I used the following structure:

     ```
     get_power <- function(theta, reps=1000){

     [insert tweaked version of all code to get power and return the same 3 numbers as in 2.6d]

             return(c(power test 1, power test 2, power test 3))
     }
     ```

     Note that the problem asks for 1000 reps. I would start with a smaller number, like 5, to test the code first. When I'm confident it's working, I'd change it to 1000. Specifying a function argument with a fixed value like this sets the default. My first test of my function looked like this:

     ```
     get_power(.5, reps=5)
     ```
   - Keeping the reps set low (5? 10?) for now, do the next few steps to make sure the code is working before waiting a long time for it to run.
   - Create a matrix to store the power for each of the 3 tests, for each level of theta
   - Run the function for each level of theta and store the values in the matrix
   - Plot the power functions as a function of theta for each of the 3 tests.
   - Once I'm satisfied the above are all working, set the reps to 1000. Remember you have to re-run it when you knit. Be ready to wait a long time when it finally runs at 1000.

3. Find a dataset not used in the textbook that it would make sense to fit with a (penalized) spline regression. Minimally, the dataset needs to have a continuous response and one continuous regressor. Ideally, the relationship should be more complicated than can be easily fit with a linear or low-level polynomial model.

   (a) Give a citation for the dataset and/or where you found it.
   (b) Define the response and regressor variables. Use 1-2 sentences to discuss what their relationship means.
   (c) Fit a (penalized) spline regression to your dataset.
   (d) Briefly describe how you interpret the relationship you found.

   Note: This question is an exercise, not yet a project. This does not need to be an especially interesting dataset to you. The analysis also doesn't need to be perfectly interpretable. But if it is interesting, it's possible it could be a good start for a project idea. When you do finally do projects, I want the datasets you use to be very interesting to you, and to involve some non-trivial elements, possibly including other predictors, dependence, multivariate data to smooth, so you can do a more nuanced analysis. I don't want your projects to be plug-and-chug 'I found these data so I stuck them in the function.' I want you to reflect more deeply than that. For this homework, it is OK to be plug-and-chug... but keep your eyes out for interesting and compelling data that might make a good project.