STAT 4400/5400 Exam 2

Instructions

Due Date: **Wednesday, April 27, 2022 at 3:00 pm** Exams received after 3 pm will not be graded outside of remarkably extenuating circumstances.

This exam is worth approximately 32% of your final grade, and will be graded out of 100 points

- Please answer the following questions, and upload your responses to Canvas just as you would with a homework assignment. In your submission, please include your written (or typed) responses, along with any code and coded outputs used to answer the questions.
- You may consult the internet, textbooks, or course notes, but **do not** collaborate with other classmates.
- If you have questions about the interpretation of a question, or any other portion of the exam, please contact me as soon as possible to ensure that you receive a timely response.

Problem 1: (35 points) Multilevel Poisson regression

In the Lecture Notes 14B, from April 13, Dr. Law made a mistake when separating the 75 precincts into categories. The book suggests categorizing the precincts by percentage of ethnicity "1" in the population of the precinct.

(a) Correct the mistake and recreate the tables and plots in Section 15.1 as closely as possible. (The dataset may have changed so it's possible that you will not get the same results as in the textbook.)

(b) What are the advantages of using the level of past arrests as an offset rather than a linear predictor, if any?

Problem 2: (35 points) Optimal design

Use the dataset "allvar.csv" saved on Canvas.

(a) Suppose that the zinc study described in Section 20.5 would cost \$150 for each treated child and \$100 for each control. Under the assumptions given in that section, determine the number of control and treated children needed to attain 80% power at minimal total cost. You will need to set up a loop of simulations as illustrated for the example in the text. Assume that the number of measurements per child is fixed at $K = 7$ (that is, measuring every two months for a year).

(b) Given the cost assumptions in (a), make a generalization of Figure 20.6 with several lines corresponding to different values of the design parameter K, the number of measurements for each child.

Problem 3: (30 points) Multilevel hierarchical model

Use the dataset "schooldata.Rdata" saved on Canvas. The dataset gives measurements of extroversion, openness, agreeableness, and sociability for each subject. The subjects are nested within classes and within schools.

(a) Fit a single-level linear regression model to predict extroversion by measures of openness, agreeableness, and sociability. What inferences can be made and with what level of certainty? Does the model seem to be useful?

(b) Fit a linear model with a varying-intercept group effect using the variable school to predict extroversion by measures of openness, agreeableness, and social ability. What inferences can be made? Interpret the estimated coefficients.

(c) Fit a varying slope and varying intercept model for schools and the classrooms nested within schools to predict extroversion by measures of openness, agreeableness, and social ability. Also allow the slope of openness to vary by school. When would such a model be useful (what research question can it answer?)

(d) Provide a useful plot for either (b) or (c). Provide a full description of the plot and the concept that it depicts.