

Biost/Stat 571: Homework # 4

Due 5pm, Wed Mar 1 via Canvas

Note: Homework should be submitted in as a PDF document with problems clearly labeled and in order. Use of Latex is preferred, but hand-written math is acceptable. In the case of the latter, handwriting must be in clearly legible print. Illegible (unreadable by the TAs and instructor) and unclear work will not receive credit. Clarity in derivations and exposition count as these are essential in any professional setting.

Problem 1. Consider the six city data on the class website. The data file is sixcity.dat and the variable names are given in sixcity.docx. This data set contains 537 children from Steubenville, Ohio, each of whom was examined annually from age 7 to age 10 for the presence of wheezing (which suggests a diagnosis of asthma). Mother's smoking status was reported. Consider fitting marginal logistic models with covariates age, smoking and interaction between age and smoking.

(1) Analyze this data set using a random intercept logistic mixed model by assuming a normally distributed random intercept.

(2) Compare the results with those obtained under the GEE logistic model and discuss the differences in interpretation of the regression coefficients given by the two models.

(3) BONUS: write your own code implementing a random intercept logistic mixed model assuming a normally distributed random intercept. You may use any numerical integration approach including numerical integration functions inside of R.

Problem 2. (Open Ended Problem) An investigator is interested in conducting a study in which one of two treatments will be administered to an even number of m individuals ($m/2$ in each treatment group). Subsequently, the investigator plans to measure a dichotomous characteristic on each individual n times longitudinally. Note that n does not necessarily equal 5 in this case and these are not necessarily replicates anymore.

The investigator considers two different strategies for analyzing the data: (1) by using a logistic mixed model with random intercept and (2) using a GEE.

What are the relative merits and/or limitations of these approaches? Conduct a simulation study to assess these approaches. If you were given more time, what additional simulation scenarios would you consider?