

Homework 3

Biost 540

General Instructions:

- When working in groups, elect one member as the leading member who will be responsible for uploading the homework assignment solutions. Please note that all members in the group are expected to equally contribute to the assignment!
- Be sure to show work for all problems! R code (and output) should not appear in the main body of the homework; however, the code should appear at the end of the assignment as an Appendix. It should be possible for someone to use the code to reproduce any figures or numeric results.

Part A: Sixcity Data - Complete Data

Download the dataset `sixcity.csv` from the class website that contains longitudinal measurements on respiratory infections occurring among children in one of six cities.

1. Conduct an exploratory analysis of the longitudinal dataset that focuses on the relationship between the prevalence of respiratory disease and age within each of the two smoking (mother's) groups.
2. Analyze the effect of mother's smoking on prevalence of respiratory disease as a function of age using a *marginal* regression model. Interpret the estimated coefficient for mother's smoking status.
 - a. Experiment with a number of different "working" correlation matrices including, at a minimum, independence, exchangeable and autoregressive. Which if any of these seems most suitable for use with these data?
 - b. Is there any evidence for an interaction between smoking and age? How do you interpret this interaction?
3. Estimate a *conditional* model for the probability of infection. Explicitly state your model and modeling assumptions. Interpret the estimated coefficient for mothers' smoking status.
4. Estimate a *transition* model for the probability of infection. Write out the model and state your modeling assumptions. What is the interpretation of the estimated coefficient for mother's smoking status?
5. Compare the estimated coefficients for mothers' smoking status from the fitted models in (2), (3) and (4). Comment upon the magnitude of the estimates and potential problems in their interpretations.

Part B: Sixcity Data - Missing Data

To be posted later