

Each part is worth 10 points, for a total of 70 points.

**Problem 1**

Refer to the *airbnb.csv* dataset in the Homework 6 page.

The dataset was originally collected by Trinh and Ameri (2016), and includes information on 1561 AirBnB listings in 43 Chicago neighborhoods. For each listing, the dataset includes price per day, rating, number of reviews, type of accommodation, neighborhood, and several other variables.

Let  $Y_{ij}$  be the number of reviews for the  $i$ th listing in the  $j$ th neighborhood, and consider two potential predictors:

- $X_{1ij}$ : the visitor rating, on a scale from 1 to 5.
  - $X_{2ij}$ : the accommodation type, with three levels: private room, shared room, or entire house (or apartment).
- a. Does there appear to be an association between  $Y$  and the two predictors?
  - b. Listings within a neighborhood cannot be assumed to be independent. Therefore, fit a Poisson hierarchical model to the response variable, where each neighborhood is allowed to have its own intercept. Use `jags` or `stan` and justify your priors.
  - c. Summarize the posterior distributions for the overall intercept, the regression coefficients, and the between-neighborhood standard deviation. Interpret the results.
  - d. Check the overall fit of your model. If you generate 50 samples of size 1561 each from the model, does the distribution of the replicated  $y$  values look like the observed distribution in the sample?
  - e. Fit a different model to the same data, this time assuming that the number of reviews is distributed as a Negative Binomial variable. Carry out a posterior predictive check to see whether this model is an improvement over the Poisson model you fitted in Part b. Explain your findings.
  - f. Summarize the posterior distributions of overall intercept, the regression coefficients, and the between-neighborhood standard deviation. Interpret the results.
  - g. Obtain the posterior predictive distribution of a new listing of a private room in Avondale (one of the neighborhoods). Summarize the posterior predictive distribution and interpret what you see in English.