Due: Mar. 29 by midnight

(1) The file "apt.txt" contains data on rodent infestation in a sample of New York City apartments. See "rodents.doc" for a description of the dataset.

   (a) Write the notation for a varying-intercept multilevel logistic regression (with community districts as the groups) for the probability of rodent infestation using the individual-level predictors but no group-level predictors.

   (b) The file "dist.txt" contains data on the 55 "community districts" (neighborhoods) in the city. Expand the model in (a) by including the variables in "dist.txt" as group-level predictors.

(2) The file "olympics1932.txt" has seven judges' ratings of seven figure skaters on two criteria: *technical merit* and *artistic impression* from the 1932 Winter Olympics.

   (a) Construct a $7 \times 7 \times 2$ array of the data (ordered by skater, judge, and judging criterion).

   (b) Reformulate the data as a 98×4 array (similar to the top table in Figure 11.7), where the first two columns are the technical merit and artistic impression scores, the third column is a skater ID, and the fourth column is a judge ID.

   (c) Add another column to this matrix representing an indicator variable that equals 1 if the skater and judge are from the same country, or 0 otherwise.

(3) The file "allvar.csv" has CD4 percentages for a set of young children with HIV who were measured several times over a period of two years. CD4 cells are white blood cells called T lymphocytes or T cells that fight infection and play an important role in immune system function. The dataset also includes the ages of the children at each measurement.

   (a) Graph the outcome (the CD4 percentage, on the square root scale) for each child as a function of time.

   (b) Each child's data has a time course that can be summarized by a linear fit. Estimate these lines and plot them for all the children.

   (c) Set up a model for the children's slopes and intercepts as a function of the treatment and age at baseline. Estimate this model using the two-step procedure: (1) first estimate the intercept and slope separately for each child, and then (2) fit the between-child models using the point estimates from the first step.

(4) Continuing with the analysis of the data "allvar.csv":

   (a) Write a model predicting CD4 percentage as a function of time with varying intercepts across children. Fit using lmer() and interpret the coefficient for time.

   (b) Extend the model in (a) to include child-level predictors (that is, group-level predictors) for treatment and age at baseline. Fit using lmer() and interpret the coefficients on time, treatment, and age at baseline.

   (c) Investigate the change in partial pooling from (a) to (b) both graphically and numerically.

   (d) Compare results in (b) to those obtained in part (c).

(5) Using data of your own that are appropriate for a multilevel model, write the model in the five ways discussed in Section 12.5.