Due: Apr. 12 by midnight

(1) The file "ProfEvaltnsBeautyPublic.csv" contains data described in "ProfEvaltnsBeautyPublic.txt". Fit a multilevel model to predict course evaluations from beauty and other predictors allowing the intercept and coefficient for beauty to vary by course category:

- (a) Write the model in statistical notation.
- (b) Fit the model using lmer() and discuss the results: the coefficient estimates and the estimated standard deviation and correlation parameters. Identify each of the estimated parameters with the notation in your model from (a).
- (c) Display the estimated model graphically in plots that also include the data.

(2) Models for adjusting individual ratings: a committee of 10 persons is evaluating 100 job applications. Each person on the committee reads 30 applications (structured so that each application is read by three people) and gives each a numerical rating between 1 and 10.

- (a) It would be natural to rate the applications based on their combined scores; however, there is a worry that different raters use different standards, and we would like to correct for this. Set up a model for the ratings (with parameters for the applicants and the raters).
- (b) It is possible that some persons on the committee show more variation than others in their ratings. Expand your model to allow for this.

(3) The file "olympics1932.txt" has seven judges' ratings of seven figure skaters on technical merit and artistic impression from the 1932 Winter Olympics. Continuing from Homework 5, problem 2, we use a non-nested model.

- (a) Write the notation for a non-nested multilevel model (varying across skaters and judges) for the technical merit ratings and fit using lmer().
- (b) Fit the model in (a) using the artistic impression ratings.
- (c) Display your results for both outcomes graphically.
- (d) Use posterior predictive checks to investigate model fit in (a) and (b).

(4) The file "wells.dat" contains data on arsenic level in wells in Bangladesh. For these questions, you may use and modify the R code from the vignette found here: https://github.com/stan-dev/rstanarm/blob/master/vignettes/binomial.Rmd.

(a) Fit a Bayesian version of the logistic regression model using the "stan_glm" function. Use a Student t prior with 7 degrees of freedom and a scale of 2.5 as the prior distribution for the coefficient for distance.

(b) Use the "posterior_interval" function and compute 85% Bayesian uncertainty intervals.

(c) Plot the predicted probability of switching together with the observed outcomes and interpret your results.

(5) The file "rodents.dat" contains data on rodent infestation in a sample of New York City apartments. See "rodents.doc" for a description of the dataset. We conduct a three-level logistic regression.

(a) Build a varying intercept logistic regression model (varying over buildings) to predict the presence of rodents (the variable rodent2 in the dataset) given indicators for the ethnic groups (race) as well as other potentially relevant predictors describing the apartment and building. Fit this model using lmer() and interpret the coefficients at both levels.

(b) Now extend the model in (b) to allow variation across buildings within community district and then across community districts. Also include predictors describing the community districts. Fit this model using lmer() and interpret the coefficients at all levels.

(c) Compare the fit of the models in (a) and (b).