Midterm Review STAT 506 Spring 2014

1. In linear mixed models with two levels of nested random effects, write out an expression for $\text{Var}(\boldsymbol{y}_i)$ or $\text{Var}(\boldsymbol{y}_{ij})$ in terms of Ψ_1 and Ψ_2.

2. Interpret output from R functions gls, lme, and lmer and from SAS PROC MIXED.

3. Material from Stat 505 is implicit in all we're doing in Stat 506. In particular, you should review the basics of Chapters 9 and 10 on causal inference.

   - Fundamental Problem of causal inference:
     We can't observe the same unit under more than one treatment condition. The unobserved counterfactual is not available. Work arounds:
     - Use close substitutes
     - Randomize treatment assignment
     - Adjust for lurking variables.
   - Randomized Experiments provide unbiased treatment effects contrasts.
   - Observational Studies must assume **ignorability** to make causal inference.
   - Do not control for post-treatment variables.

4. ARM Chapter 11 Multilevel Structures

   - Varying intercept and varying slope
   - Clustered data
   - Repeated Measures
   - Indicators in fixed and random models
   - Cost/benefit of multilevel modeling

5. ARM Chapter 12 Basic Multilevel linear models

   - Notation
   - Partial pooling (no predictors)
   - Partial pooling (with predictors)
   - Using lmer in R
   - Ways to write the multilevel model
   - Group level predictors
   - Model building and stat significance
   - Predictions
   - Minimal sample sizes

6. ARM Chapter 13

   - Growth curve model, $y_i \sim N(\alpha_{j[i]} + \beta_{j[i]} x_i, \sigma_y^2)$ population average intercept and slope, (μ_α, μ_β). Individual "adjustments" to each have a variance-covariance matrix. Figures like 13.2 illustrate shrinkage toward the mean.
   - Centering reduces correlations between intercept and slope.(Fig 13.7)

- Crossed effects §13.5, especially HW8
- Model Selection §13.6

7. Model Selection, Multicollinearity (from notes)

   - For each of the model comparison criteria, know what the goal is (smaller? larger? close to p?)
   - What variables are plotted in an added variable plot?
   - When is model selection not a big problem?
   - How do we tell if there is a collinearity problem.
   - What do you do about a collinearity problem which disappears in the centered data?

8. Chapter 14

   - Multilevel models work nicely with stratified data.
     - Split voter data into lots of strata using fixed effects for race, gender, interaction. Add random effects for age, education, age by education interaction, state and region.
     - Use logistic regression to estimate voter probability of voting Rep in each stratum.
     - Multiply estimates by census data on actual numbers in each stratum to scale up to state results.
     - Resulting model gives predictions which vary about as much as the actual observations. No-pooling gives too much spread, complete pooling too little spread. Partial pooling is like Goldilocks – just right.
   - Item analysis shows how well questions discriminate, and how Justices vote on more (or less) contentious issues.