Homework 5: Linear Mixed Models

your name

Due: April 7th at 11:59 PM

Problem 1: Prove that intraclass correlation between observations at the same level is formally expressed as

$$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2},$$

and discuss the implications of a value of ρ close to zero or one.

Problem 2: Do the following with the pulp data set:

(a) Analyze the pulp data in the notes using unrestricted maximum likelihood estimation and comment on the differences that such analysis produces when compared to the REML and ANOVA estimates.

(b) Interpret the diagnostic plots in the notes corresponding to the analysis with respect to the pulp data set; conclude that these plots indicate no particular problems.

Problem 3: Do the following:

(a) Compare and contrast the nonparametric bootstrap, residual bootstrap, and parametric bootstraps. Discuss the assumptions that make each bootstrap procedure appropriate. There is a useful guide embedded in the course notes.

(b) In the notes we developed a parametric bootstrap procedure to approximate the distribution of the LRT corresponding to a test between mixed-effects models. Write your own parametric bootstrap procedure with $B = 1e4$ samples and make it as fast as possible using parallel programming with either `mclapply` and/or `foreach` and whatever accompanying software packages are needed.

(c) Explain how the testing procedure using the `exactRLRT` function works. See the analysis of the irrigation data in the LMM course notes for context.

Problem 4: The `denim` dataset in the `faraway` concerns the amount of waste in material cutting for a jeans manufacturer due to five suppliers.

(a) Plot the data and comment.

(b) Fit the linear fixed effects model. Is the supplier significant?

(c) Make a useful diagnostic plot for this model and comment.

(d) Analyze the data with supplier as a random effect. What are the estimated standard deviations of the effects?

(e) Test the significance of the supplier term.

(f) Compute confidence intervals for the random effect SDs.

(g) Locate two outliers and remove them from the data. Repeat the fitting, testing and computation of the confidence intervals, commenting on the differences you see from the complete data.

(h) Estimate the effect of each supplier. If only one supplier will be used, choose the best.

Problem 5: Load in the `soybean_full` dataset and analyze the `tqM` response variable. Refer to the analysis of the `AqE` response in the course notes, and recall that the researchers are interested in determining which ID variables differ from the RC reference level. Consider response transformations if modeling assumptions are violated. Do attempts to rectify departures from modeling assumptions affect the conclusions?

Problem 6: Prove that the penalized weighted residual sum-of-squares problem can be cast as

$$r^2(\theta, \beta, u) = r^2(\theta) + \|L_\theta^T(u - \mu_{U|Y=y_{\text{obs}}}) + R_{ZX}(\beta - \hat{\beta}_\theta)\|^2 + \|R_X(\beta - \hat{\beta}_\theta)\|^2.$$

See the LMM notes for context.