Statistics 203
Introduction to Regression and Analysis of Variance
Assignment #4
Due Thursday, March 10

Prof. J. Taylor

USE R FOR ALL CALCULATIONS. PROVIDE COPIES OF YOUR CODE IN THE ASSIGNMENT.

Q. 1) Consider the one-way random effects ANOVA model

$$Y_{ij} \sim \mu_\cdot + \alpha_i + \varepsilon_{ij}, \qquad 1 \leq i \leq r, 1 \leq j \leq n$$

where $\alpha_i \sim N(0, \sigma^2_\alpha)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$ are independent. Suppose one is interested in the quantity

$$\theta = \frac{\sigma^2_\mu}{\sigma^2_\mu + \sigma^2}.$$

This quantity describes the relative contribution of the random effects variance to the total variance of one observation.

Based on the ANOVA table below, construct a $(1-\alpha)\cdot 100\%$ confidence interval for θ.

Source	SS	df	$E(MS)$
Treatments	$SSTR = \sum_{i=1}^r n\left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2$	$r-1$	$\sigma^2 + n\sigma^2_\mu$
Error	$SSE = \sum_{i=1}^r \sum_{j=1}^n (Y_{ij} - \overline{Y}_{i\cdot})^2$	$(n-1)r$	σ^2

Q. 2) The webpage

http://lib.stat.cmu.edu/DASL/Stories/ResinsRidTermitesfromTrees.html

describes a dataset involved in the study of the effect of a particular resin from a tropical tree that is resistant to termites. The response variable counts the number of termites alive as a function of time in 16 different dishes: 8 dishes treated with a dose of 5mg of resin, the other 8 with a does of 10mg.

The data is also available at

http://www-stat.stanford.edu/~jtaylo/courses/stats203/data/termite.table

(a) Plot the traces as a function of time using `type='b'` to get both points and lines in your plot. Use two different colors for the two dose levels? Does there appear to be an effect of dose? Is it appropriate to treat time as a continuous variable?

(b) Is there anything unusual about the plot? Based on the plot, are there likely mistakes in the data? Fix these mistakes and use the updated data for further calculations.

(c) Fit a two way random effects ANOVA model with random intercepts within dish and main effects for dose and time, as well an interaction. This enforces some correlation between all of the observations in a given dish. What are your conclusions about these effects?

(d) Does the correlation within dish large relative to the residual standard error? Test at level $\alpha = 0.05$ whether a model without a random effect can be used instead of the model in (c).

Q. 3) The webpage

http://lib.stat.cmu.edu/DASL/Stories/hwfatal.html

describes a dataset comparing the highway fatality rate in the United States and in New Mexico over a 40 year period. The data is also available at

http://www-stat.stanford.edu/~jtaylo/courses/stats203/data/hwfatal.table

(a) Plot the data using different colors or line types for the entire United States and New Mexico. Does there appear to be a difference in the rate?

(b) For each time series separately, choose the best fitting auto-regressive model of order up to 5. Verify the goodness of fit with diagnostic plots for each time series.

(c) Fit a regression model to the data set including a linear trend for time, testing at level $\alpha = 0.05$ whether there is indeed a difference in intercept between the two rates. What about an interaction effect?