STAT 700
Homework 7 Problems
due Wed. Nov. 7

3 Problems. Please follow the Lab report directions off the homework web page and work in HW Groups.

1. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables from the $N(\mu, \sigma^2)$ distribution. Then $\boldsymbol{Y} \sim N(\mu \boldsymbol{1_n}, \sigma^2 I_n)$

In class, we derived the REML estimator in matrix notation for σ^2 as

$$\frac{\boldsymbol{Y}' K(K'K)^{-1} K' \boldsymbol{Y}}{n-1}$$

For $n = 4$, write down in matrix notation and derive the above by giving (a)-(d). Be sure to show your work in (d).
(a) $\boldsymbol{Y}$
(b) K
(c) $(K'K)^{-1}$
(d) $\boldsymbol{Y}'K(K'K)^{-1}K'\boldsymbol{Y}/(n-1)$

2. This problem will fit growth curves to data for the change in an orthodontic measurement over time for several young subjects. Investigators at the University of North Carolina Dental School followed the growth of 27 children (16 males, 11 females) from age 8 until age 14. Every two years (age 8, 10, 12, 14) they measured the distance between the pituitary and the pterygomaxillary fissure, two points that are easily identified on x-ray exposures of the side of the head. It appears that there are qualitative differences between boys and girls, so we will just model the data from the male subjects. We will use the R dataset Orthodont (in the `nlme` library).

Follow the class example and make a dataset that consists of only the males. Call it `OthoM`.

(a) Fit two mixed models, `fit1` and `fit2`, one with a random effect for the intercept and one with random effects for both the intercept and slope. Provide the summary for each model.

(b) First test if the random intercept is needed in the model `fit1` and test $\mathrm{H}_0 : \sigma^2_{intercept} = 0$ against $\mathrm{H}_1 : \sigma^2_{intercept} \neq 0$.

(c) Use the two models `fit1` and `fit2` test $\mathrm{H}_0 : \sigma^2_{slope} = 0$ against $\mathrm{H}_1 : \sigma^2_{slope} \neq 0$. State your conclusion and state which is the "best" model.

(d) Using your answer to part (c). Give the estimates for all parameters in the "best" model. Give corresponding 95% CI for all the parameters using the R `intervals` function.

(e) How well does the "best" model fit the data? Include and examine the diagnostics plot of the residuals. You should also look at a Q-Q plot of the residuals.

(f) Use the R `augPred` function to plot the fitted values for the "best" model.

(g) The dataset `OthoM` is an example of `groupedData` and there is a formula associated with this type of data. Use the R function `formula(OrthoM)` to see the formula. There is also a `lmList` function that will fit a linear model to each Subject separately (see the help file). Obtain a fitted object using `lmList(OrthoM)`. Use the `plot` and `intervals` functions to plot the 95% CI of intercept and slope for each for subject. Examine the plot. Does your "best" model seem reasonable? (i.e., Does the plot indicate that there is a clear indication that one or both random effects is needed to account for the subject-to-subject variability?)

3. Orthodontic Growth Curves (cont.): We will use the R dataset Orthodont (in the `nlme` library).

(a) Following the Rat Pup R code, make a new variable sex1 which is a 1 for Female and 0 for Male. Let's center the ages, so use `I(age-11)` in your R function calls. Fit a l inear mixed model using REML, that has fixed effects that include sex1, I(age-11), sex1:I(age-11) interaction, an intercept and a random effect for just the intercept. Call this `fit1`.

Note: You can read about centering in your textbook Section 2.9.5 Centering Covariates.

(b) Use the LRT to test $H_0 : \sigma^2_{intercept} = 0$ against $H_1 : \sigma^2_{intercept} \neq 0$. State your conclusion and state which is the "best" model. Include and examine the plot of the residuals and the Q-Q plot of the residuals for the best model. Does it appear that the residuals are iid $N(0, \sigma^2)$?

(c) Examine the summary of the model `fit1` from part (a). Is there evidence to suggest that boys and girls have significantly different growth patterns? Explain.

(d) Make a plot of the standardized residuals versus fitted values by gender. (Hint: see the R code for the `ergoStool` data analysis and use `resid(., type=''p'')` option). Does it look like Males and Females have the same residual variance, σ^2?

(e) You can test part (d) as in the Rat Pup Lab, but for this homework just use your model from part (a) with the `VarIdent` option to allow for different residual variances for each gender. Call this model `fit1b`. Make a scatter plot of the standardized residuals versus fitted values for your heteroscedastic fit by gender. Now, does it look like Males and Females have the same residual variance?

(f) Which model `fit1` or `fit1b` in part (e) is the "best" model based on AIC? Plot the predicted values over time for all the subjects from the best model.