STAT/BIOST 571: Homework 6

Philip Pham

February 25, 2019

Problem 1: Fitting and interpreting the results of a linear mixed effects model; robust standard error estimation (20 points)

Download the `creatinine.csv` *dataset from the course website. This file contains repeated observational data for 619 subjects, some of whom have hypertension and some of whom have a hereditary kidney disease, as indicated by the* `group` *variable, according to the coding in the Table 1. The outcome variable*

Group	Kidney disease	Hypertension	Sample size
1	Yes	Yes	294
2	Yes	No	103
3	No	Yes	73
4	No	No	149

Table 1: Measurements of serum creatinne reciprocals from 619 subjects in four groups

is `scr`*, the reciprocal of serum creatinine. Serum creatinine is a measure of kidney function, with lower values indicating better kidney function. Higher values of the reciprocal reported in* `scr` *indicate better kidney function. The observations were taken at arbitrary times from each subject, with the number of observations ranging from 1 to 22. Ignoring hypertension status, we are interested in estimating the rate of change of* `scr` *for subjects with and without hereditary kidney disease. Thus, the only fixed effect covariates in your model should be age, kidney disease status, and possibly an interaction between these. In order to account for correlation within subjects, you will be fitting a linear mixed effects model with uncorrelated random slopes and intercepts and serial correlation of residuals that follows a spherical correlation model, including a nugget (correlation should be based on the timing of observations). Please use the* `lme()` *function in the* `nlme` *package in R to fit your models (i.e., do not code your own nonlinear optimization).*

(a) *Fit the model by ML and report the estimated values of all variance parameters.*

The parameter estimates of fitting the model without and with an interaction term can be seen in Tables 3 and 4. The variance parameters can be found in Table 2. Their meaning is detailed in the subsequent paragraphs.

	Model	
Parameter	Without interaction (Equation 1)	With interaction (Equation 2)
$\hat{\sigma}^2$	0.0693487	0.051000981
$\hat{\sigma}^2_{\gamma_0}/\hat{\sigma}^2$	0.03108841	0.3403312
$\hat{\sigma}^2_{\gamma_1}/\hat{\sigma}^2$	0.0003932786	0.000456134
$\hat{\alpha}_r$	7.8894707	4.6700641
$\hat{\alpha}_n$	0.1759323	0.2299764

Table 2: Variance parameters for ML-fitted models.

Let t_{ij} be the age of the subject i at observation j. Let x_i indicate whether the subject has has kidney disease. Without an interaction term, the mean model is

$$Y_{ij} = (\beta_0 + \gamma_0) + \beta_2 x_i + (\beta_1 + \gamma_1)\, t_{ij} + \epsilon_{ij}. \tag{1}$$

With the interaction term, the mean model is

$$Y_{ij} = (\beta_0 + \gamma_0) + \beta_2 x_i + (\beta_1 + \beta_3 x_i + \gamma_1)\, t_{ij} + \epsilon_{ij}. \tag{2}$$

γ_j are the random effects, where γ_0 is subject-specific adjustment to the intercept, and γ_1 is the subject-specific adjustment to the slope.

The covariance structure of a cluster i can be described by the matrix

$$\Sigma_i = \sigma^2 \left(Z_i G Z_i^\intercal + R_i\right) \tag{3}$$

Z_i is a $m_i \times 2$ matrix, where the first column entries are all 1s, and the second column entries are ages for each subject t_{ij}.

G is a 2×2 diagonal matrix that describes the variance of the random effects γ_0 and γ_1:

$$G = \frac{1}{\sigma^2} \begin{pmatrix} \sigma^2_{\gamma_0} & 0 \\ 0 & \sigma^2_{\gamma_1} \end{pmatrix} \tag{4}$$

R_i is an $m_i \times m_i$ matrix that describes the correlations between the ϵ_{ij}s for different js with a nugget parameter $0 \leq \alpha_n < 1$ and range parameter $\alpha_r > 0$. $R_{ijj} = 1$ and $R_{ijj'} = (1 - \alpha_n) \exp\left(-\frac{|t_{ij} - t_{ij'}|}{\alpha_r}\right)$. Estimates for these parameters can be found in Table 2.

(b) *Report point estimates and standard errors for all fixed effect coefficients in your model. Include three versions of standard error estimates: (i) robust/empirical sandwich SEs that correctly account for clustering of the data, (ii) bootstrap SEs that correctly account for clustering of the data, and (iii) model based SE estimates based on the assumed random effect model being correct.*

Solution: Table 3

	Estimate	ML Standard Error	Sandwich Standard Error	Bootstrap Standard Error
(Intercept)	1.532222	0.039557	0.039348	0.042201
age	-0.012915	0.000964	0.000941	0.001028
kidney.disease	-0.281916	0.026260	0.025027	0.035864

Table 3: Standard error estimates for fixed effect parameters.

(c) *Now, give point estimates and three versions of standard error estimates for the marginal rates of change in* `scr` *in subjects with and without kidney disease. As in part (b), your three versions of SE estimates should be: (i) robust/empirical sandwich SEs that correctly account for clustering of the data, (ii) bootstrap SEs that correctly account for clustering of the data, and (iii) model based SE estimates based on the assumed random effect model being correct. Depending on how you parameterized your model, these answers may or may not coincide with estimates you reported in part (b). You may parameterize the model any way you wish, but you must answer this question based on the output of a single call to* `lme()`.

	Estimate	ML Standard Error	Sandwich Standard Error	Bootstrap Standard Error
(Intercept)	1.190676	0.054232	0.050269	0.050637
age	-0.003108	0.001436	0.001194	0.001216
kidney.disease	0.313575	0.071373	0.068325	0.066354
age:kidney.disease	-0.016649	0.001857	0.001705	0.001651

Table 4: Standard error estimates for fixed effect parameters with interaction term.

Solution: Table 4