Stat 505 Assignment 11 Solutions

1. Rodents

   (a) Multilevel logistic regression with a different intercept for each community district.

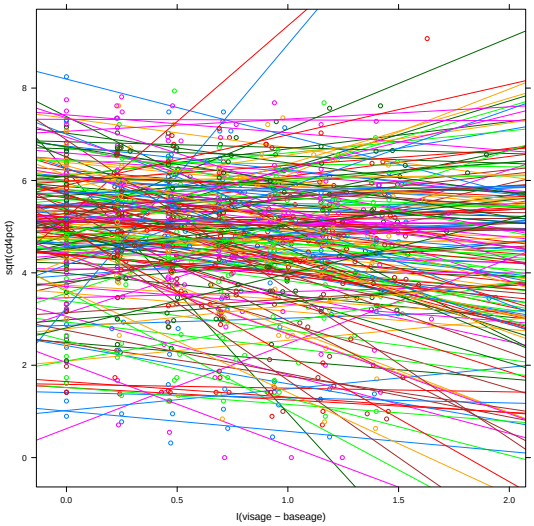
$$\begin{aligned} Pr(\text{rodent2}_i = 1) &= \text{logit}^{-1}(\alpha_{j[i]} + \beta_1 \text{defects}_i + \beta_2 \text{poor}_i + \beta_3 I_{\text{race}_i=2} + \\ &\quad \beta_4 I_{\text{race}_i=3} + \beta_5 I_{\text{race}_i=4} + \beta_6 I_{\text{race}_i=5} + \beta_7 I_{\text{race}_i=6} + \\ &\quad \beta_8 I_{\text{race}_i=7} + \beta_9 \text{floor}_i); \quad i = 1, \ldots 15894, \\ &\quad \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2), \ j = 1, \ldots, 55 \end{aligned}$$

   I am not including borough because if you know which district the apartment is in, then you know its borough. I also omit building because this is a nested random effect and would require a variable for each of many buildings.

   (b) Add in group level (district) coefficents for defects and poor.
   Same as above with $\alpha_j \sim N(\gamma_1 \text{dist.defects}_{j[i]} + \gamma_2 \text{dist.poor}_{j[i]}, \ \sigma_\alpha^2); \quad j = 1, \ldots, 55$
   With this model, I expect to get smaller estimaates of σ_α*, since the district variables should explain part of the variance of* α_j*'s.*

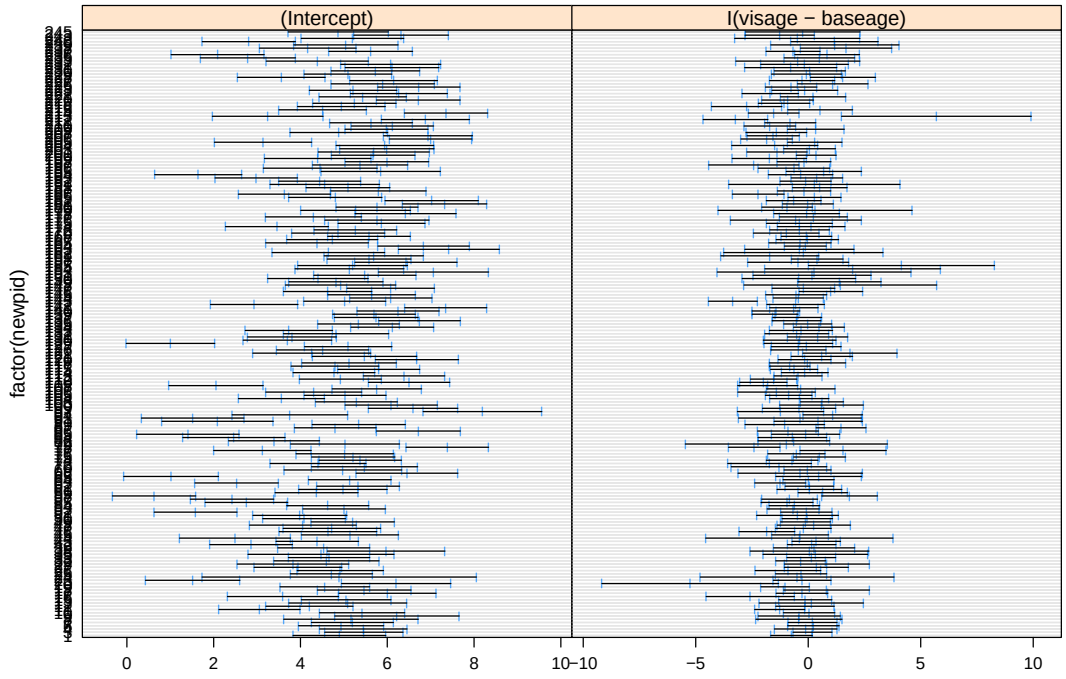
2. Exercise 11.4 on CD4 data.

   (a) Plot square root CD4 percentage over time with regression lines.



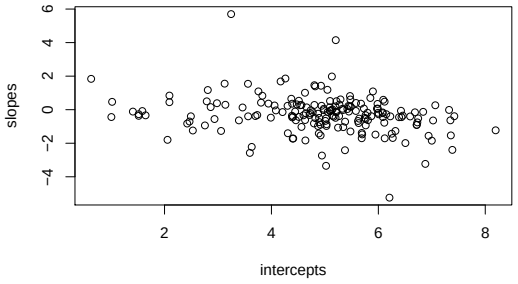
   It makes a little more sense to use age after first visit as the time variable. Even so there is little discernible pattern.

   (b) Create a fit for each kid using `lmList` in `nlme` package.

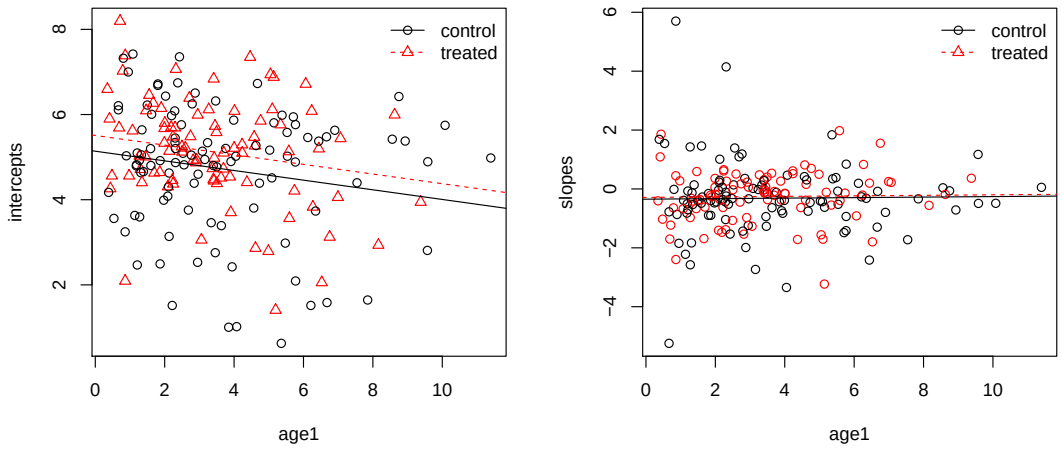


I see more spread in Intercepts than in slopes. Slopes line up well right over the zero point. There is more evidence of intercept variation than of variation of slope from one kid to another.

(c) Extract coefficient vectors for each kid and the two predictors we need.
Plot of individuals slopes and intercepts shows little relationship.



We then look for treatment and age effects on intercepts and slopes.



Interpret your results.

Intercepts are approximate cd4 measurements taken at the time kids come in for their first visit. I would expect to see no treatment effect, however the summary results tell me that the treated group has a higher intercept than the control group (0.48 is 2.3 SEs from 0). I am not surprised to see that the intercept depends on base age, because the kids all were exposed to HIV at birth, and the older they are at first visit, the more chance the disease has had to knock down their CD4 counts. The effect estimate is -0.12 with SE of 0.05.

Slopes are generally negative, as HIV does reduce CD4 counts. Neither base age (estimate of 0.02 with SE 0.04) nor treatment (estimate of -0.02 with SE 0.17) seem to have much effect on the slope. I hope that this was an early trial. Now we do have drug cocktails which can indefinitely delay effects of HIV.

R Code

```
> options(digits = 3, width = 120)
> require(xtable)
> require(arm)
> cd4 <- read.csv("http://www.math.montana.edu/~jimrc/classes/stat505/data/mycd4.csv")
> cd4$vdate <- as.Date(cd4$vdate)
> require(lattice)
> print(xyplot(sqrt(cd4pct) ~ I(visage - baseage), cd4, group = newpid, type = c("p", "r")))
> require(nlme)
> cd4.kidfit <- lmList(sqrt(cd4pct) ~ I(visage - baseage) | factor(newpid), data = subset(
    cd4,
+     !is.na(baseage))[, -9])
> print(plot(intervals(cd4.kidfit)))
> intercepts <- summary(cd4.kidfit)$coefficients[, 1, 1]
> slopes <- summary(cd4.kidfit)$coefficients[, 1, 2]
> plot(slopes ~ intercepts)
> age1 <- with(subset(cd4, !is.na(baseage)), tapply(baseage, newpid, min))
> trt <- with(subset(cd4, !is.na(baseage)), tapply(treatmnt, newpid, min))
> summary(intFit <- lm(intercepts ~ age1 + factor(trt)))
```

```
Call:
lm(formula = intercepts ~ age1 + factor(trt))

Residuals:
```

```
   Min      1Q  Median      3Q     Max
-3.904  -0.686   0.114   1.021   2.766

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.1388     0.2100   24.47   <2e-16
age1          -0.1133     0.0444   -2.55    0.012
factor(trt)2   0.3702     0.2001    1.85    0.066

Residual standard error: 1.36 on 184 degrees of freedom
Multiple R^2: 0.0532,    Adjusted R^2: 0.0429
F-statistic: 5.17 on 2 and 184 DF,  p-value: 0.00655
```

```
> summary(slopeFit ← lm(slopes ~ age1 + factor(trt)))
```

```
Call:
lm(formula = slopes ~ age1 + factor(trt))

Residuals:
   Min      1Q  Median      3Q     Max
-4.902  -0.439   0.003   0.490   6.040

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.34755    0.17353   -2.00    0.047
age1          0.00881    0.03671    0.24    0.811
factor(trt)2  0.06036    0.16533    0.37    0.715

Residual standard error: 1.13 on 184 degrees of freedom
Multiple R^2: 0.000998,  Adjusted R^2: -0.00986
F-statistic: 0.0919 on 2 and 184 DF,  p-value: 0.912
```

```
> par(mfrow = c(1, 2))
> plot(intercepts ~ age1, col = trt, pch = trt)
> legend("topright", pch = 1:2, c("control", "treated"), bty = "n", col = 1:2, lty = 1:2)
> abline(intFit, 1:2)
> abline(sum(coef(intFit)[c(1, 3)]), coef(intFit)[2], col = 2, lty = 2)
> plot(slopes ~ age1, col = trt)
> legend("topright", pch = 1:2, c("control", "treated"), bty = "n", col = 1:2, lty = 1:2)
> abline(slopeFit, 1:2)
> abline(sum(coef(slopeFit)[c(1, 3)]), coef(slopeFit)[2], col = 2, lty = 2)
```