HW3

Avinash Ramu

September 25, 2016

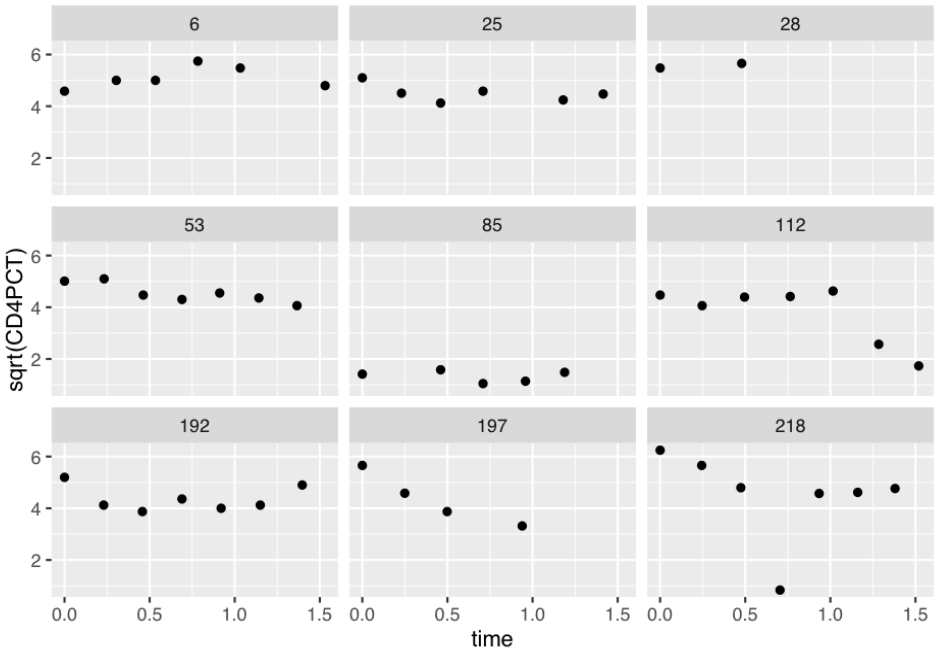
Q11.4

A

Let's use a random sample of 10 patients to plot since 250 is a large number to display

```
cd4 <- read.table("dat/allvar.csv", head = T, sep = ",")
cd4$time <- cd4$visage - cd4$baseage
cd4_subset <- cd4[cd4$newpid %in% sample(cd4$newpid, 10), ]
cd4$newpid <- as.factor(cd4$newpid)
(plot1 <- ggplot(cd4_subset,aes(x=time, y=sqrt(CD4PCT))) + geom_point() + facet_wrap(~newpid))
```

```
## Warning: Removed 5 rows containing missing values (geom_point).
```

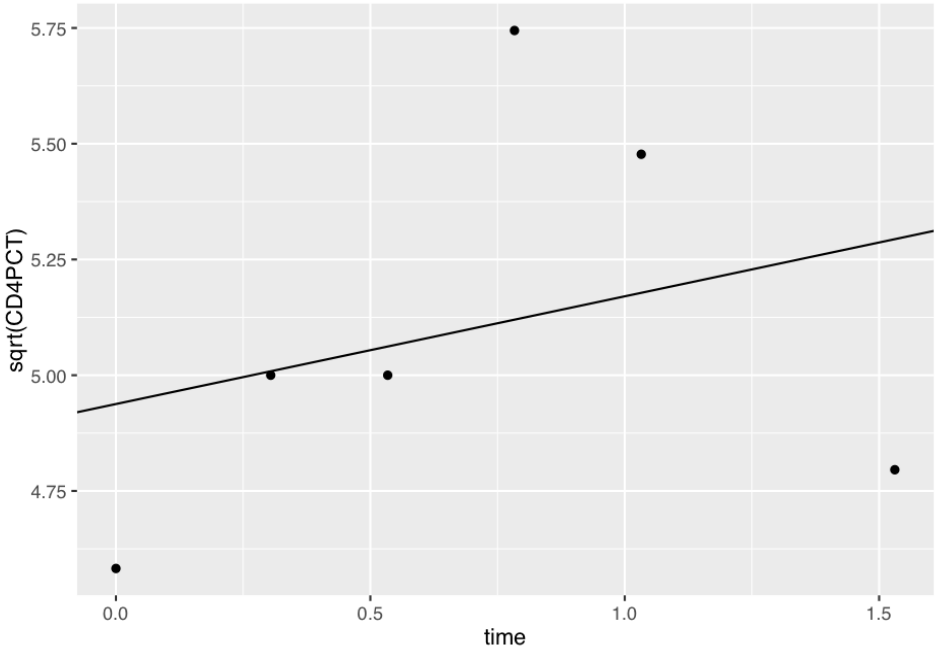


B

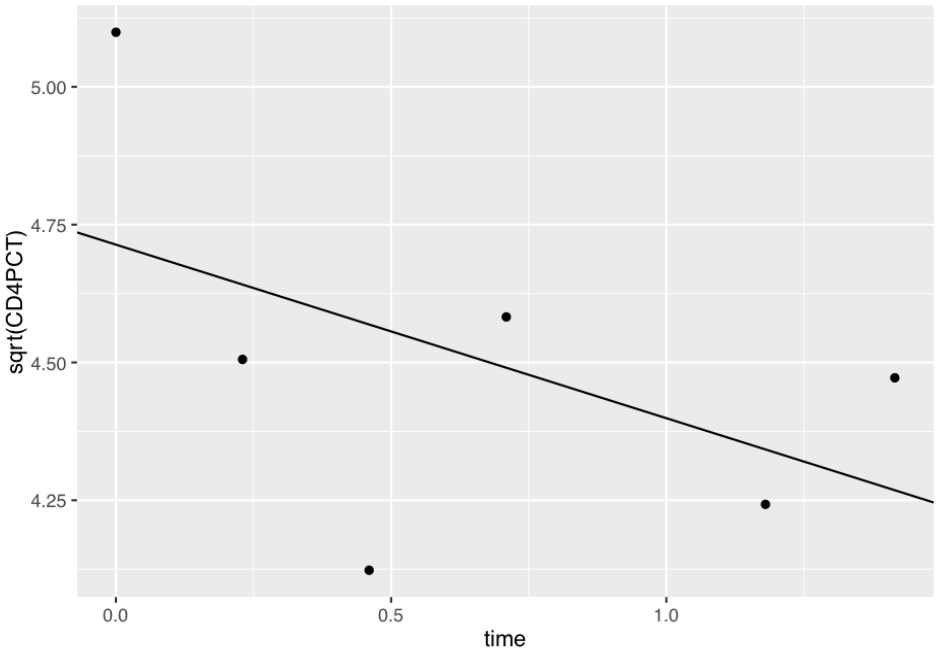
Print the models fit to the ten patients

```r
cd4_splits <- split(cd4_subset, cd4_subset$newpid)
for (i in 1:length(cd4_splits)) {
  lm1 = lm(sqrt(CD4PCT) ~ time, data = cd4_splits[[i]])
  coeffs <- lm1$coeff
  plot1 <- ggplot(data = cd4_splits[[i]]) + geom_abline(slope = coeffs[2], intercept = coeffs[1]) + geo
  print(plot1)
}
```

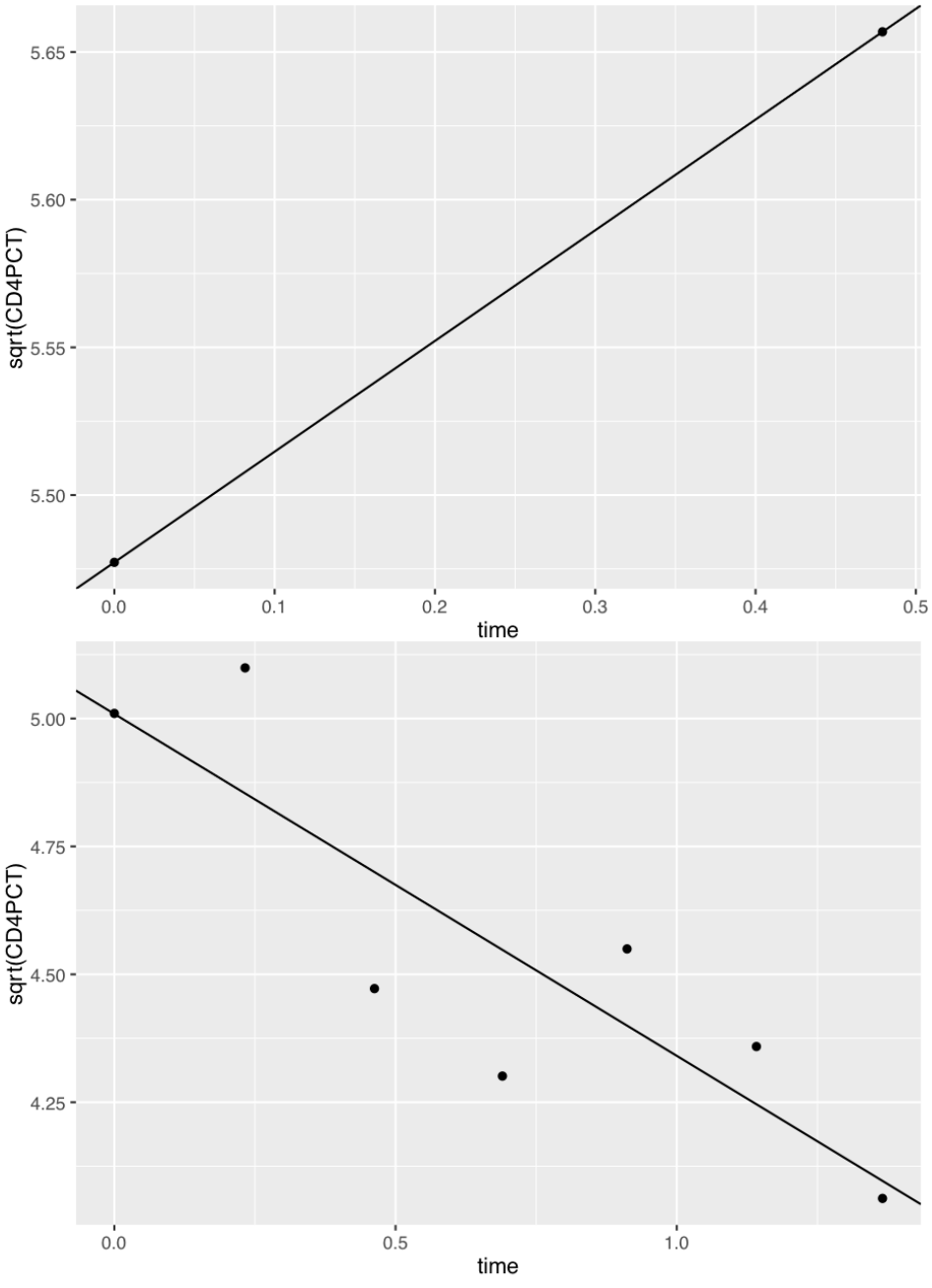
```
## Warning: Removed 1 rows containing missing values (geom_point).
```



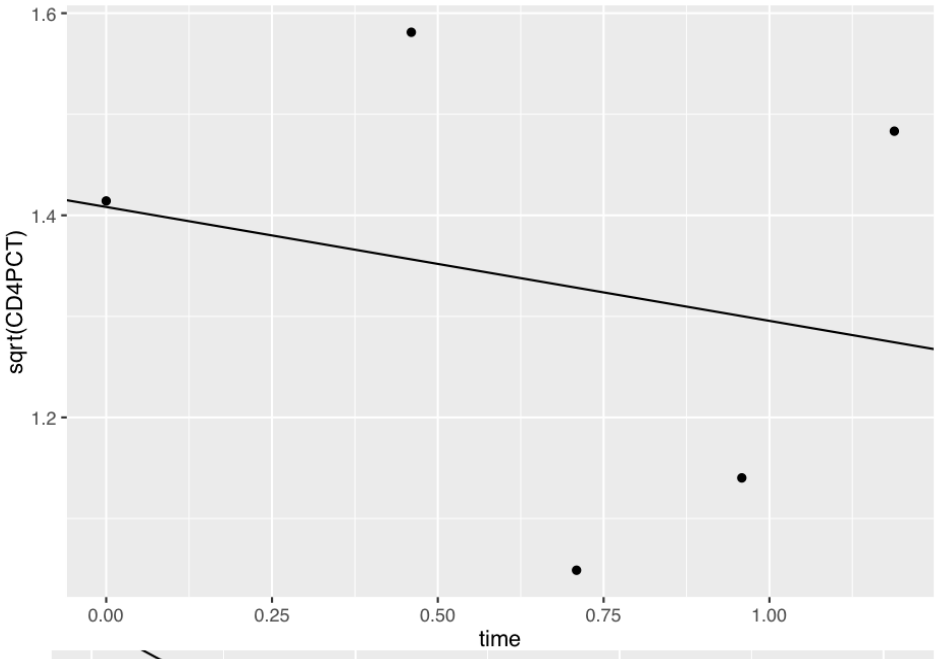

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

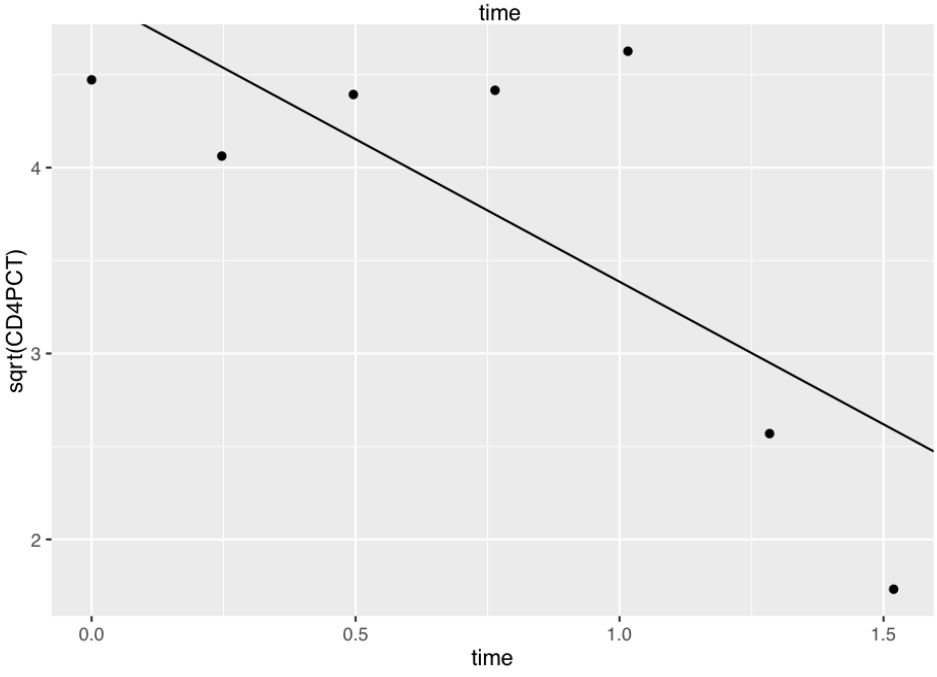


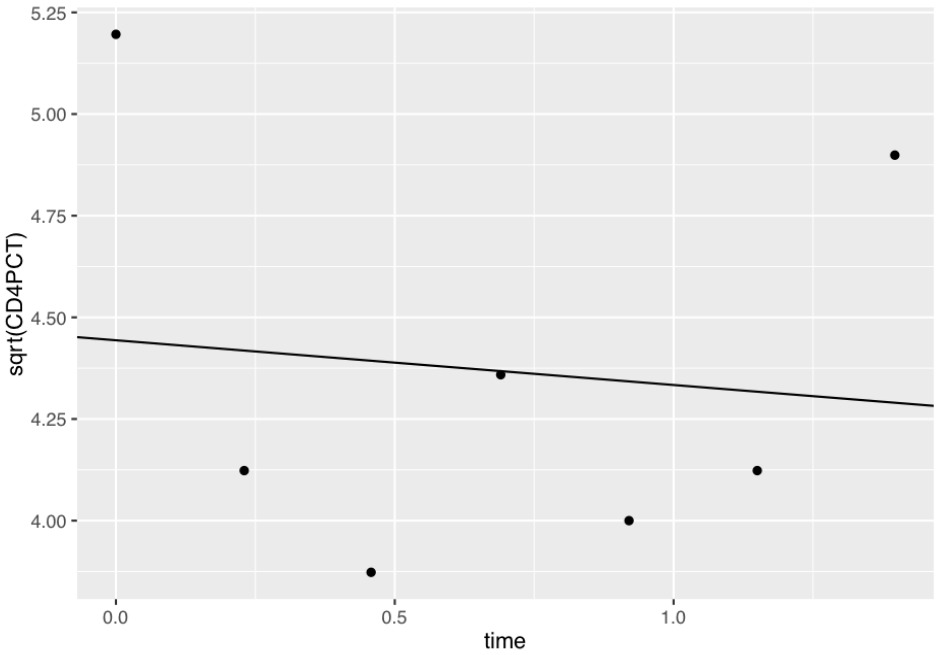

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



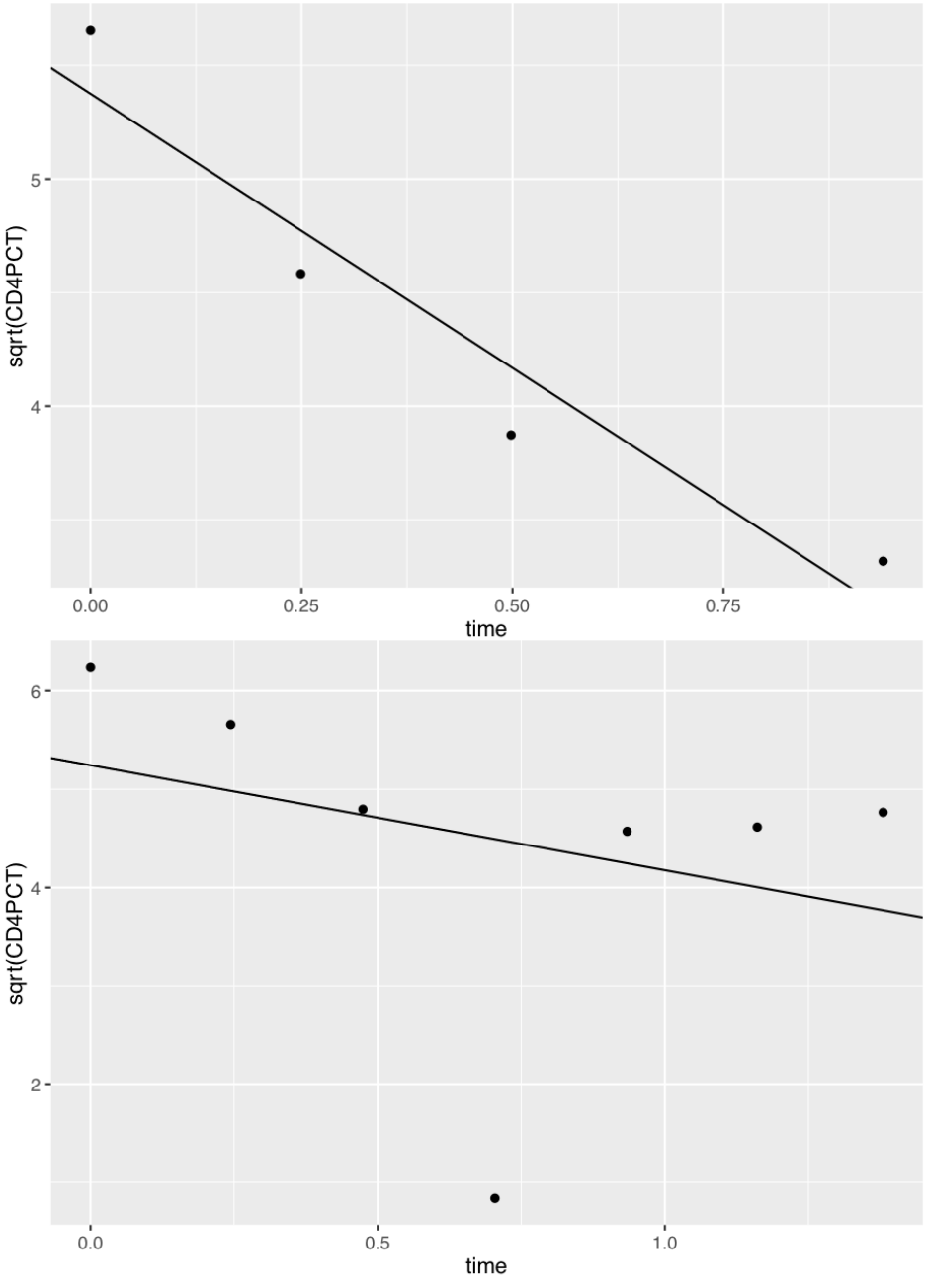

```
## Warning: Removed 1 rows containing missing values (geom_point).
```


1.6
1.4
1.2
sqrt(CD4PCT)
0.00
0.25
0.50
0.75
1.00
time


4
3
2
sqrt(CD4PCT)
0.0
0.5
1.0
1.5
time




```
## Warning: Removed 1 rows containing missing values (geom_point).
```


sqrt(CD4PCT)
5
4
0.00
0.25
0.50
0.75
time
sqrt(CD4PCT)
6
4
2
0.0
0.5
1.0
time

C

Setup a model in which the treatment and age affect outcome

```
lm1 = lm(sqrt(CD4PCT) ~ treatmnt + time + factor(newpid) - 1, data = cd4)
#print(summary(lm1))
```

The coefficient for patient 31 is 0.70309, we can fit this to that patients data and set this coefficient as the outcome variable to get group level alphas and betas.

```
patient_31 <- cd4[cd4$newpid == 31, ]
```

Q12.2

A

```
lmer1 <- lmer(sqrt(CD4PCT) ~ time + (1|newpid), data=cd4)
print(summary(lmer1))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: sqrt(CD4PCT) ~ time + (1 | newpid)
##    Data: cd4
##
## REML criterion at convergence: 3140.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7379 -0.4379  0.0024  0.4324  5.0017
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  newpid   (Intercept) 1.9569   1.3989
##  Residual             0.5968   0.7725
## Number of obs: 1072, groups:  newpid, 250
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  4.76341    0.09648   49.37
## time        -0.36609    0.05399   -6.78
##
## Correlation of Fixed Effects:
##      (Intr)
## time -0.278
```

The coefficient for time is -0.36609 with a SE of 0.05399, the coefficient is less than zero in the +/- 2 * SE interval hence significant. This indicates a decrease in CD4PCT with time. The intercepts vary by patient.

B

```
lmer2 <- lmer(sqrt(CD4PCT) ~ time + (treatmnt + baseage | newpid), data=cd4)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
## eigenvalues
```

```
summary(lmer2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: sqrt(CD4PCT) ~ time + (treatmnt + baseage | newpid)
##    Data: cd4
##
## REML criterion at convergence: 3133.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7461 -0.4587  0.0112  0.4423  5.0311
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  newpid   (Intercept) 3.14107  1.7723
##           treatmnt    0.07792  0.2791   -1.00
##           baseage     0.06231  0.2496   -0.40  0.40
##  Residual             0.59649  0.7723
## Number of obs: 1072, groups:  newpid, 250
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  4.84612    0.09337   51.90
## time        -0.36493    0.05392   -6.77
##
## Correlation of Fixed Effects:
##      (Intr)
## time -0.287
## convergence code: 0
## unable to evaluate scaled gradient
## Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
```

The coefficient for time is still the same as part A and indicates a decrease in CD4PCT with time. Note - The model fails to converge. The variance of the treatment and baseage coefficients for the different groups is close to zero. This is indicates that the treatment and baseage have a similar effect across groups.

C

```
intercepts <- fixef(lmer2)[1] + ranef(lmer2)$newpid[,1]
slope1 <- ranef(lmer2)$newpid[,2] #slope for treatmt
slope2 <- ranef(lmer2)$newpid[,3] #slope for baseage
```

Q12.5

```
radon <- read.table("dat/mn-radon-2.txt")
head(radon)
```

```
##   radon log.radon floor          county.name county  uranium log.uranium
## 1   2.2 0.7884574     1 AITKIN                    1 0.502054  -0.6890476
## 2   2.2 0.7884574     0 AITKIN                    1 0.502054  -0.6890476
## 3   2.9 1.0647107     0 AITKIN                    1 0.502054  -0.6890476
## 4   1.0 0.0000000     0 AITKIN                    1 0.502054  -0.6890476
## 5   3.1 1.1314021     0 ANOKA                     2 0.428565  -0.8473129
## 6   2.5 0.9162907     0 ANOKA                     2 0.428565  -0.8473129
##   county_sample_size
## 1                  4
## 2                  4
## 3                  4
## 4                  4
## 5                 52
## 6                 52
```

```
lmer1 <- lmer(log.radon ~ 1 + (1 + county_sample_size | county), data = radon)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
## $checkConv, : Model failed to converge with max|grad| = 0.00467979 (tol =
## 0.002, component 1)
```

```
summary(lmer1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: log.radon ~ 1 + (1 + county_sample_size | county)
##    Data: radon
##
## REML criterion at convergence: 2258.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4634 -0.5664  0.0552  0.6388  3.3650
##
## Random effects:
##  Groups   Name               Variance  Std.Dev. Corr
##  county   (Intercept)        1.156e-01 0.339980
##           county_sample_size 2.164e-05 0.004651 -0.72
##  Residual                    6.361e-01 0.797552
## Number of obs: 919, groups:  county, 85
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.30911    0.04812   27.21
## convergence code: 0
## Model failed to converge with max|grad| = 0.00467979 (tol = 0.002, component 1)
```

```
#coef(lmer1)
```

The county sample size has a very small sigma compared to the residual sigma, this grouping doesn't seem to help the model a lot.