Lecture 4 - Correlated and uncorrelated random effects

Richard Wilkinson

sleepstudy dataset

Lets now consider the sleepstudy dataset from the lme4 package. This is from a report on a study of the effects of sleep deprivation on reaction time for a number of subjects chosen from a population of long-distance truck drivers. These subjects were divided into groups that were allowed only a limited amount of sleep each night. We consider here the group of 18 subjects who were restricted to three hours of sleep per night for the first ten days of the trial. Each subject's reaction time was measured several times on each day of the trial.

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
str(sleepstudy)
```

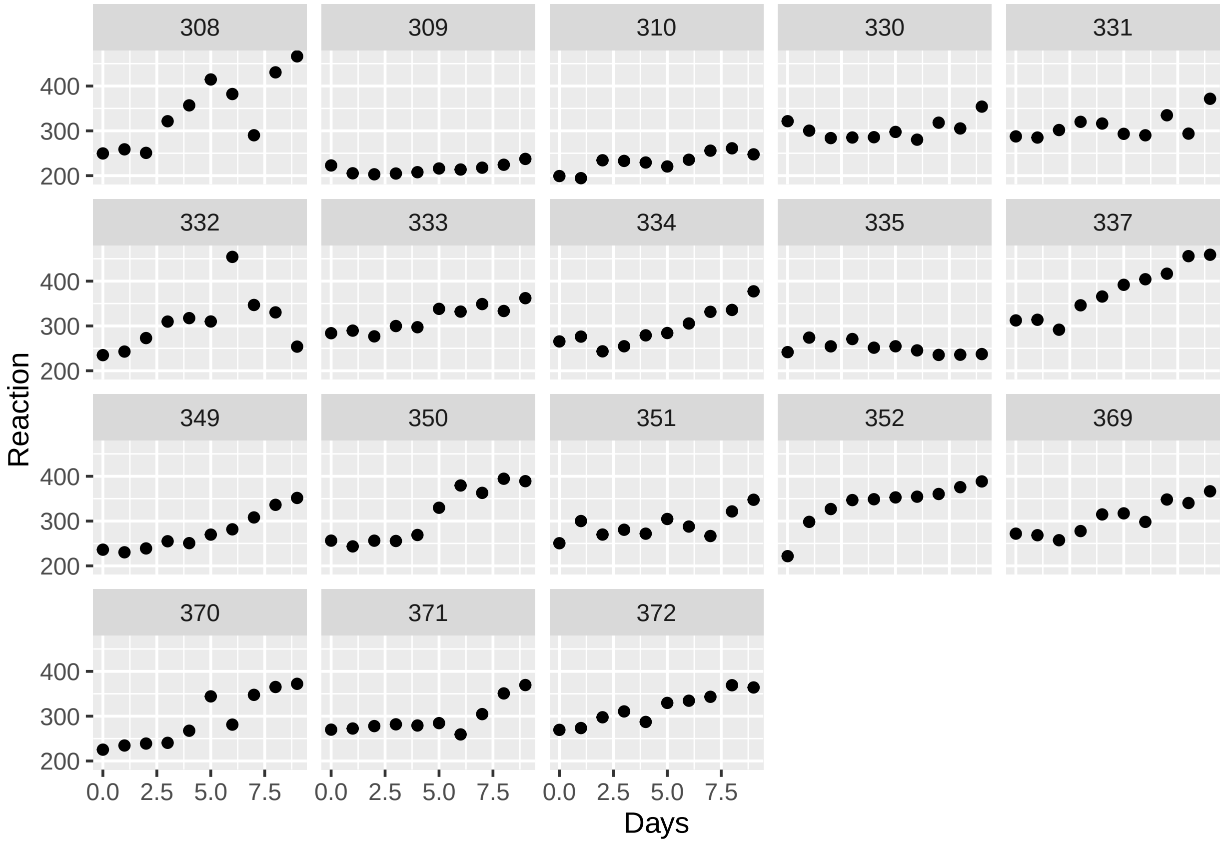
```
## 'data.frame':    180 obs. of  3 variables:
##  $ Reaction: num  250 259 251 321 357 ...
##  $ Days    : num  0 1 2 3 4 5 6 7 8 9 ...
##  $ Subject : Factor w/ 18 levels "308","309","310",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
head(sleepstudy)
```

```
##   Reaction Days Subject
## 1 249.5600    0     308
## 2 258.7047    1     308
## 3 250.8006    2     308
## 4 321.4398    3     308
## 5 356.8519    4     308
## 6 414.6901    5     308
```

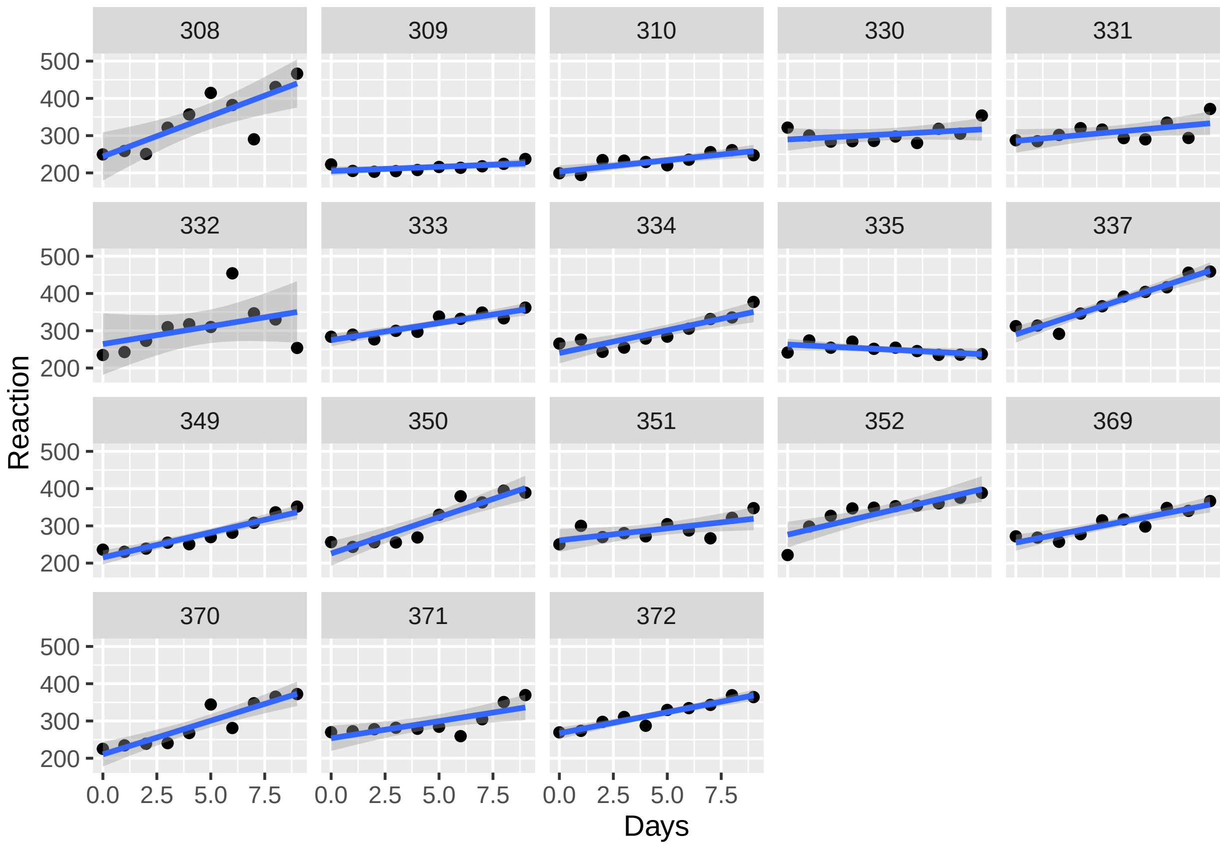
As always, we start by plotting the data

```
library(ggplot2)
qplot(Days, Reaction, facets=~Subject, data = sleepstudy)
```



qplot has an option to also draw on a fitted regression line to each person as well.

```
c <- ggplot(sleepstudy, aes(Days, Reaction))+geom_point() + facet_wrap(~ Subject, nc=5)
c+stat_smooth(method='lm')
```



Model 1: Correlated random effects

```
(fm06 <- lmer(Reaction ~ 1 + Days + (1 +Days | Subject), sleepstudy))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ 1 + Days + (1 + Days | Subject)
##    Data: sleepstudy
## REML criterion at convergence: 1743.628
## Random effects:
##  Groups   Name        Std.Dev. Corr
##  Subject  (Intercept) 24.740
##           Days         5.922   0.07
##  Residual             25.592
## Number of obs: 180, groups:  Subject, 18
## Fixed Effects:
## (Intercept)         Days
##      251.41        10.47
```

- What is the typical initial reaction time?
- How much does reaction time typically increase per day?
- What is typical subject-subject variation in the initial reaction time?
- What is the typical subject-subject variation in the slope?
- What would approximate 95% confidence intervals be for the slope and the intercept across all subjects?
- What is the typical within subject variation?
- Is there a strong relationship between a subjects initial reaction time and how strongly affected they are by sleep deprivation?

This model has used correlated random effects for the same subject, i.e., there is a correlation between the random subject intercept and the subject specific gradient. Mathematically, we have fit the model

$$y_{ij} = \alpha + a_i + (\beta + b_i)\text{Days}_j + \epsilon_{ij}$$

where

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim N(0, \Sigma) \text{ and } \epsilon_{ij} \sim N(0, \sigma^2)$$

The output above gives us estimates of the fixed effects (α, β), and the random effect variances Σ and σ^2.

Model 2: Uncorrelated random effects

```r
(fm07 <- lmer(Reaction ~ 1 + (1 | Subject)+ Days + (Days-1|Subject), sleepstudy))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ 1 + (1 | Subject) + Days + (Days - 1 | Subject)
##    Data: sleepstudy
## REML criterion at convergence: 1743.669
## Random effects:
##  Groups    Name        Std.Dev.
##  Subject   (Intercept) 25.051
##  Subject.1 Days         5.988
##  Residual              25.565
## Number of obs: 180, groups:  Subject, 18
## Fixed Effects:
## (Intercept)         Days
##      251.41        10.47
```

Note that we have explicitly ruled out the intercept term in the random effect part for Days. If we didn't do this, R would have fit an additional (unnecessary) intercept.

Note that this has fit the model

$$y_{ij} = \alpha + a_i + (\beta + b_i)\text{Days}_j + \epsilon_{ij}$$

where

$$a_i \sim N(0, \sigma_a^2), \quad b_i \sim N(0, \sigma_b^2) \text{ and } \epsilon_{ij} \sim N(0, \sigma^2)$$

with all the random effects independent of each other.

In the correlated random effects model, the correlation was estimated to be small. This suggests that the uncorrelated random effects model is a good choice. We will look later at how to formally test which model is better.