Multilevel models

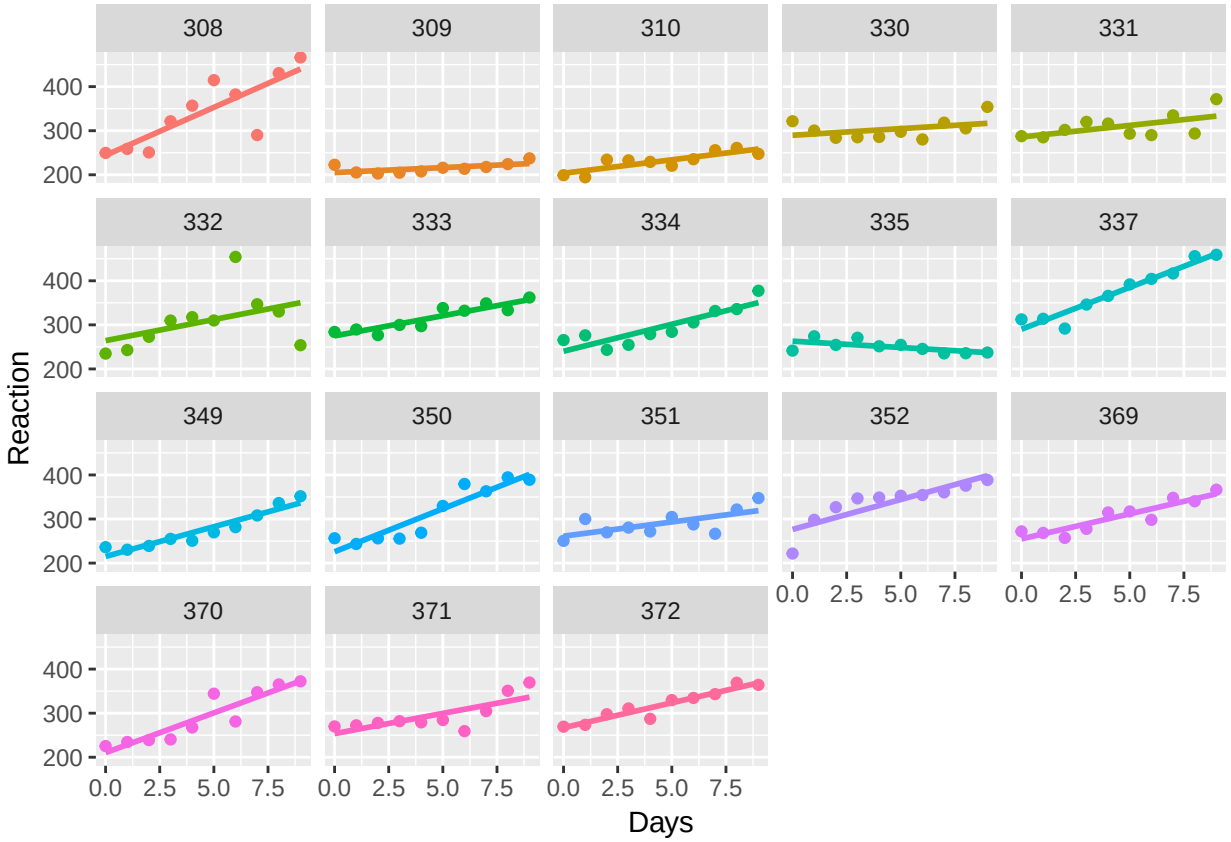
Mark Andrews

November 14, 2018

```
library(lme4)
library(lmerTest)
library(ggplot2)
library(readr)
library(dplyr)
library(tidyr)
```

Let's consider the following data:

```
ggplot(sleepstudy,
       aes(x=Days, y=Reaction, col=Subject)
) + geom_point() +
  stat_smooth(method='lm', se=F) +
  facet_wrap(~Subject) +
  guides(col=F)
```



If we label our the reaction time, subject, and day on observation i by y_i, $s_i \in \{1, 2 \dots J\}$, and x_i, respectively, a random intercepts model of this data would be

$$y_i \sim N(a_{s_i} + bx_i, \sigma^2), \quad \text{for all } i \in 1, 2 \dots n$$
$$a_j \sim N(\alpha, \tau^2), \quad \text{for all } j \in 1, 2 \dots J$$

```
M_0 <- lmer(Reaction ~ Days + (1|Subject),
            data = sleepstudy)
```

A random slopes model of this data would be

$$\begin{aligned} y_i &\sim N(a + b_{s_i} x_i, \sigma^2), \quad \text{for all } i \in 1, 2 \dots n \\ b_j &\sim N(\beta, \tau_\beta^2), \quad \text{for all } j \in 1, 2 \dots J \end{aligned}$$

```
M_1 <- lmer(Reaction ~ Days + (0 + Days|Subject),
            data = sleepstudy)
```

A random slopes and random intercepts model would be

$$\begin{aligned} y_i &\sim N(a_{s_i} + b_{s_i} x_i, \sigma^2), \quad \text{for all } i \in 1, 2 \dots n \\ a_j &\sim N(\alpha, \tau_\alpha^2), \quad \text{for all } j \in 1, 2 \dots J \\ b_j &\sim N(\beta, \tau_\beta^2), \quad \text{for all } j \in 1, 2 \dots J \end{aligned}$$

```
M_1 <- lmer(Reaction ~ Days + (1 + Days|Subject),
            data = sleepstudy)
```

These models can be re-written using different, but equivalent, notation. For example, the random slopes, random intercepts models is equivalent to the following:

$$\begin{aligned} y_i &= \alpha + \beta x_i + \nu_{s_i} + \psi_{s_i} x_i + \epsilon_i, \quad \text{for all } i \in 1, 2 \dots n \\ \nu_j &\sim N(0, \tau_\alpha^2), \quad \text{for all } j \in 1, 2 \dots J \\ \psi_j &\sim N(0, \tau_\beta^2), \quad \text{for all } j \in 1, 2 \dots J. \end{aligned}$$

In this, $\nu_j = a_j - \alpha$ and $\psi_j = b_j - \beta$.

Nested models

Sometimes we have groups nested in other groups

```
Df <- read_csv('../data/science.csv')
```

In `Df`, we have `class`, with values $\{1, 2, 3, 4\}$, nested in `school`, with values $\{1, 2 \dots 41\}$. For example,

```
group_by(Df, school, class) %>%
  summarise(n = n()) %>%
  spread(class,n, fill=0)
```

```
## # A tibble: 41 x 5
## # Groups:   school [41]
##    school   `1`   `2`   `3`   `4`
##     <int> <dbl> <dbl> <dbl> <dbl>
##  1      1    14    13    14     0
##  2      2    26     0     0     0
##  3      3    27    22     0     0
##  4      4    27     0     0     0
##  5      5    13     0     0     0
##  6      6    23    20     0     0
##  7      7    17     0     0     0
##  8      8    27     0     0     0
```

```
##  9      9    25     0     0     0
## 10     10    14     0     0     0
## # ... with 31 more rows
```

To model this nesting, we'd do the following:

```
M_1 <- lmer(like ~ sex + PrivPub + (1|school/class),
          data = Df)
```

which is identical to

```
M_2 <- lmer(like ~ sex + PrivPub + (1|school) + (1|school:class),
          data = Df)
```

However, if we use unique identifiers for `class`, i.e. `Class`, which takes values `1.1`, `1.2`, etc., then we can simply do

```
M_3 <- lmer(like ~ sex + PrivPub + (1|school) + (1|Class),
          data = Df)
```

Crossed structures

When grouping variables are not nested, they are *crossed*. For example, in a biochemistry experiment, we could have six different samples of penicillian and 24 different plates. If each sample occurs in each plate, then we have a *fully crossed* structure. But fully crossed structures are not necessary.

```
M <- lmer(diameter ~ 1 + (1|plate) + (1|sample),
          data=Penicillin)
```

Model comparison

We proceed just like in the case of generalized linear models.

```
M_null <- lmer(diameter ~ 1 + (1|sample),
               data=Penicillin)

anova(M_null, M)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: Penicillin
## Models:
## M_null: diameter ~ 1 + (1 | sample)
## M: diameter ~ 1 + (1 | plate) + (1 | sample)
##        Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## M_null  3 443.19 452.10 -218.59   437.19
## M       4 340.19 352.07 -166.09   332.19   105      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```