# Outline

– Higher order random effects

– Example on longitudinal analysis

# Higher-order Random Effects

- Random intercept model implies block-specific intercepts but a common treatment effect.
- Recall a simple mixed-effects model:

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 x_i + \mu_j + \varepsilon_{ij}, \\
&= (\beta_0 + \mu_j) + \beta_1 x_i + \varepsilon_{ij},\ i = 1, \ldots, m \text{ (blocks)},\ j = 1, \ldots, r \text{ (treatments)},
\end{aligned}
$$

where $\beta_0 + \beta_1 x_i$ represents the fixed effect, $\mu_j \sim_{\text{i.i.d.}} N(0, \sigma_a^2)$ represents random intercept, and $\varepsilon_{ij} \sim_{\text{i.i.d.}} N(0, \sigma_e^2)$ represents measurement error.

- Extend the above model to **random intercept and random slope model**:

$$
y_{ij} = (\beta_0 + \mu_j) + (\beta_1 + \gamma_j) x_i + \varepsilon_{ij},
$$

with additional parameters representing random slope $\gamma_j \sim_{\text{i.i.d}} N(0, \sigma_b^2)$, where $\{\gamma_j\}$, $\{\mu_j\}$ and $\{\varepsilon_{ij}\}$ are all independent with each other.

# Longitudinal data

- Repeated measures data consist of measurements of a response (and possibly some covariates) on several experimental (or observational) units.
- Longitudinal data are repeated measures data in which the observations are taken over time.
- We wish to characterize the response over time within subjects and the variation in the time trends between subjects.
- We are interested in modeling the variability in the population from which the subjects were chosen.
- Less interested in comparing the **particular subjects** in the study.

# Sleep deprivation study (Belenky et al. 2003)

- This laboratory experiment measured the effct of sleep deprivation on cognitive performance.
- There were 18 subjects, chosen from a population of long-distance truck drivers, in the 10 day trial.
- These subjects were restricted to 3 hours sleep per night during the trial.
- On each day of the trial, each subject's reaction time was measured. The reaction time reported in average of several measurements for that subject within a day.
- These data are balanced in that each subject is measured the same number of times and on the same occasions.

# Sleep deprivation study

- A data frame with 180 observations on the following 3 variables.
  - Reaction: Average reaction time (ms)
  - Days: Number of days of sleep deprivation
  - Subject: Subject number on which the observation was made.

```
library(ggplot2)
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.4.4
```

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

```
data(sleepstudy)
str(sleepstudy)
```

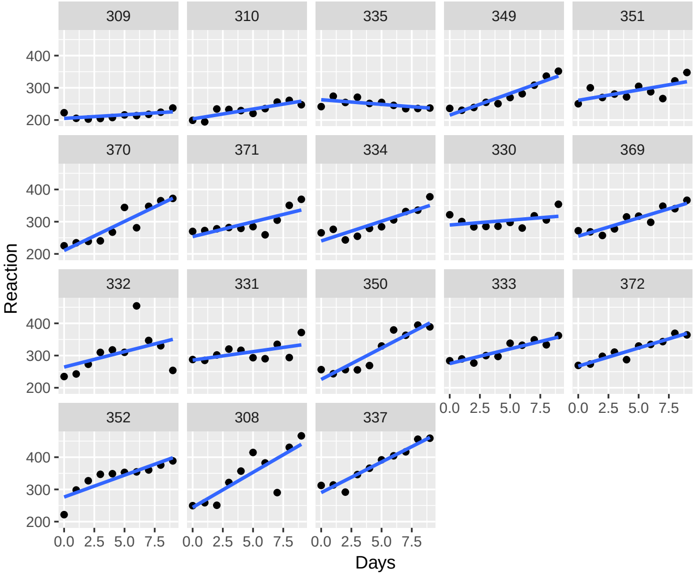
```
## 'data.frame':    180 obs. of  3 variables:
##  $ Reaction: num  250 259 251 321 357 ...
##  $ Days    : num  0 1 2 3 4 5 6 7 8 9 ...
##  $ Subject : Factor w/ 18 levels "308","309","310",..: 1 1 1 1 1 1 1 1 1 1 ...
```

# Plotting the data at the subject level

```
temp <- split(sleepstudy, sleepstudy$Subject)
sMeans <- lapply(temp, function(x){mean(x[, 1])})
oIndex <- order(unlist(sMeans))
d <- sleepstudy
d$Subject <- factor(d$Subject,
levels = names(sMeans)[oIndex])
```

# Plotting the data at the subject level

```
ggplot(d, aes(x = Days, y = Reaction)) + geom_point()  +
facet_wrap( ~ Subject) + stat_smooth(method="lm", se=FALSE)
```



# Initial observations

- In most cases, a simple linear regression provides an adequate fit to the within subject data.
- Patterns of some subjects deviate from linearity but the deviations are neither widespread nor consistent in form.
- There is considerable variation in the intercept (estimated reaction time without sleep deprivation) across subjects and in the slope (increase in reaction time per day of sleep deprivation).

# Fit simple linear regression and get CI of the slope and intercepts

```
library(reshape)
```

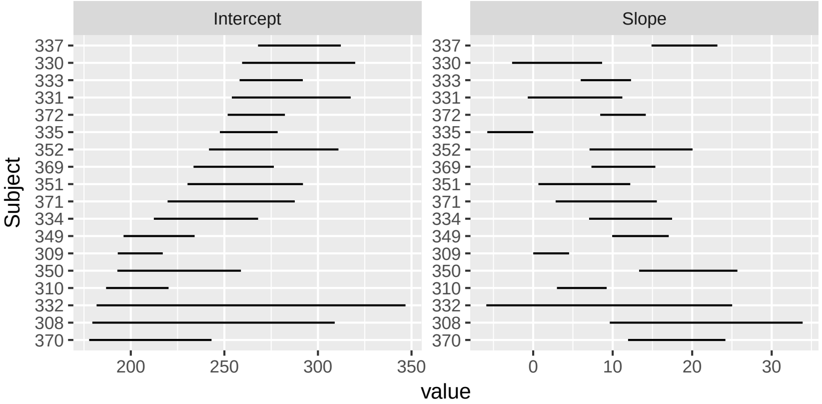
```
##
## Attaching package: 'reshape'
```

```
## The following object is masked from 'package:Matrix':
##
##     expand
```

```
out <- lapply(temp, function(x){lm1 <- lm(x[, 1] ~ x[, 2]); confint(lm1)})
sMeans <- unlist(lapply(out, function(x){x[1, 1]}))
oIndex <- order(sMeans)
#Order in increasing order of intercept
out <- out[oIndex]

dd <- melt(out)

dd2 <- dd

levels(dd2$X1)[levels(dd2$X1)=="(Intercept)"] <- "Intercept"
levels(dd2$X1)[levels(dd2$X1)=="x[, 2]"]   <- "Slope"


dd2$L1 <- factor(dd2$L1)

for (k in 1:length(oIndex)){
    levels(dd2$L1)[levels(dd2$L1)==names(sMeans)[oIndex][k]]<- k
}

dd2$L1 <- factor(dd2$L1, levels = 1:18)
```

# Subject-level estimates of slopes and intercepts

```
ggplot(dd2, aes(x = value, y = L1)) + geom_line() +
facet_wrap( ~ X1, scales = "free") + scale_y_discrete(labels=names(sMeans)[oIndex]) + labs(y = "Subject")
```



# A model with random effects for intercept and slope

- Let $y_{ij}$ denote subject $i$, day $j$.
- We will treat day as a numerical variable.
- Fixed effects model: $y_{ij} = \beta_0 + \beta_1 X_{ij} + \epsilon_{ij}$.
- Mixed effects model with both random intercept and random slope:

$$y_{ij} = \beta_0 + \beta_1 X_{ij} + b_{i1} + b_{i2} X_{ij} + \epsilon_{ij}$$

with

$$\begin{array}{rcl} (b_{i1}, b_{i2}) & \sim & \mathcal{MVN}\left( \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left( \begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{array} \right) \right) \\ (b_{i1}, b_{i2}) & \perp & \epsilon_{ij} \\ (b_{i1}, b_{i2}) & \perp & (b_{i'1}, b_{i'2}), \quad i \neq i' \end{array}$$

- In general $\sigma_{12}$ need not be zero.

# A model with random effects for intercept and slope

```r
fm1 <- lmer(Reaction ~ Days + (Days | Subject), sleepstudy)
summary(fm1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ Days + (Days | Subject)
##    Data: sleepstudy
##
## REML criterion at convergence: 1743.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9536 -0.4634  0.0231  0.4634  5.1793
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 612.09   24.740
##           Days         35.07    5.922   0.07
##  Residual             654.94   25.592
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  251.405      6.825   36.84
## Days          10.467      1.546    6.77
##
## Correlation of Fixed Effects:
##      (Intr)
## Days -0.138
```

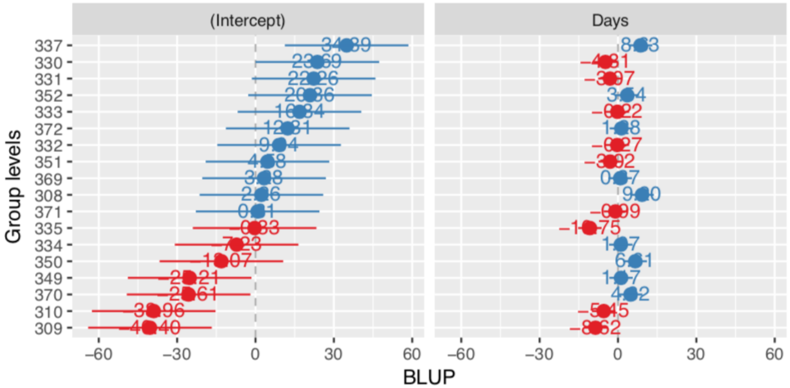
# A model with random effects for intercept and slope

- (Days | Subject) generates a vector valued random effect (intercept and slope) for each of the 18 levels of the Subject factor
- It is not clear whether there is any systematic relationship between a subject's random slope effect and the random intercept effect.
- The estimated correlation is also pretty small.

# Visualize prediction intervals for both random effects

```r
library(sjPlot)
sjp.lmer(fm1, sort = "(Intercept)")
```

```
## Plotting random effects...
```



# A model with uncorrelated random effects

- Use two random effect terms with the same grouping factor and different left hand sides.

```
fm2 <- lmer(Reaction ~ Days + (1 | Subject) + (0 + Days |Subject), sleepstudy)
summary(fm2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)
##    Data: sleepstudy
##
## REML criterion at convergence: 1743.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9626 -0.4626  0.0204  0.4653  5.1860
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Subject   (Intercept) 627.50   25.050
##  Subject.1 Days         35.86    5.989
##  Residual              653.58   25.565
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  251.405      6.885  36.514
## Days          10.467      1.560   6.711
##
## Correlation of Fixed Effects:
##      (Intr)
## Days -0.184
```

# Compare the two models

- Model fm1 contains model fm2 (fm2 is a version of fm1 where the correlation, hence the covariance, between the two random effects is set to 0.

```
anova(fm1, fm2)
```

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

```
## Data: sleepstudy
## Models:
## fm2: Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)
## fm1: Reaction ~ Days + (Days | Subject)
##     Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fm2  5 1762.0 1778.0 -876.00   1752.0
## fm1  6 1763.9 1783.1 -875.97   1751.9 0.0639      1     0.8004
```

Parsimonious fm2 is preferred (pval $> 0.05$). Consistent with the AIC and BIC results (smaller better).

See https://arxiv.org/pdf/1406.5823.pdf for more details on mixed effect models.