Longitudinal Data Analysis

Adam J Sullivan, PhD

04/23/2018

Linear Mixed Effects Models

- We have the general framework

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i$$

- What we have here is the ability to model population characteristics (fixed effects) and allow for subject specific effects (random effects).

- This allows us to understand population information as well as allow for individuals to vary differently.

Random Intercept Model

- One approach that is often used to handle the covariance among repeated measures is to assume that is comes from a random subject effect.

- This would mean that each subject has an underlying difference in change that is constant over all measurements.

- We model this as:

$$Y_{ij} = \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots \beta_p X_{ijp} + b_i + \varepsilon_{ij}$$

Random Intercept Model

- Typically we have $X_{ij1} = 1$ for all subjects so that our model is now:

$$Y_{ij} = (\beta_1 + b_i) + \beta_2 X_{ij2} + \cdots \beta_p X_{ijp} + \varepsilon_{ij}$$

Random Intercept Models

- This means that we have population attributes for all $\beta_2 - \beta_p$ but we allow for individual intercepts $|beta_1 + b_i$.

- This is why we call it the random intercept model.

- This means that the population mean is:

$$E(Y_{ij}) = \beta_1 + \beta_2 X_{ij2} + \cdots \beta_p X_{ijp}$$

- because

$$b_i \sim N(0, \sigma_b^2) \qquad\qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Random Intercept Models

- It is not necessary for the errors to be normal but this is the case for a linear regression.

- For example if we considered a simple case of a trend over time

$$Y_{ij} = (\beta_1 + b_i) + \beta_2 t_{ij} + \varepsilon_{ij}$$

What the random effects tell us

- Consider two subjects:
    1. This subject responds at a higher level than the population so they would have a $b_i > 0$.

    2. This subject responds at a lower level than the population so they would have a $b_i < 0$.

- What we would see on a plot is that Subject 1 would have a line parallel to that of the population however it would be higher.

- Subject 2 would be the opposite.

Covariance and Correlation Structure

- We said before that we have two "random" terms here:

$$b_i \sim N(0, \sigma_b^2) \qquad\qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

- This would imply that

$$Var(Y_{ij}) = Var(b_i) + Var(\varepsilon_{ij}) = \sigma_b^2 + \sigma^2$$

Covariance and Correlation Structure

- Then by introducing a subject specific effect of b_i we have induced correlation among repeated measures.

- For example consider subject i at times points j and k:

$$Cov(Y_{ij}, Y_{ik}) = \sigma_b^2 \qquad \Rightarrow \qquad Corr(Y_{ij}, Y_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2}$$

Variance Matrix

- We then have the following Covariance Matrix:

$$
\begin{bmatrix}
\sigma_b^2 + \sigma_e^2 & \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\
\sigma_b^2 & \sigma_b^2 + \sigma_e^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\
\sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma_e^2 & \cdots & \sigma_b^2 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\sigma^2 & \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 + \sigma_e^2
\end{bmatrix}
$$

Compound Symmetry Matrix

- We refer to this as a *Compound Symmetry Covariance* structure.

  - *Drawbacks*: Variances and correlations are constant across time occasions.

  - *solution*: Allow for heterogeneity in trends across times.

An Example: Sleep Study

- The average reaction time per day for subjects in a sleep deprivation study.

- On day 0 the subjects had their normal amount of sleep.

- Starting that night they were restricted to 3 hours of sleep per night.

- The observations represent the average reaction time on a series of tests given each day to each subject.

An Example: Sleep Study

- These data are from the study described in Belenky et al. (2003), for the sleep-deprived group and for the first 10 days of the study, up to the recovery period.

- Gregory Belenky, Nancy J. Wesensten, David R. Thorne, Maria L. Thomas, Helen C. Sing, Daniel P. Redmond, Michael B. Russo and Thomas J. Balkin (2003) Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. Journal of Sleep Research 12, 1–12.

An Example: Sleep Study

- `Reaction` - Average reaction time (ms)

- `Days` Number of days of sleep deprivation

- `Subject` Subject number

Getting the Data

```
library(lme4)
sleepstudy
```

```
##    Reaction Days Subject
## 1       250    0     308
## 2       259    1     308
## 3       251    2     308
## 4       321    3     308
## 5       357    4     308
## 6       415    5     308
## 7       382    6     308
## 8       290    7     308
## 9       431    8     308
## 10      466    9     308
## 11      223    0     309
## 12      205    1     309
## 13      203    2     309
## 14      205    3     309
## 15      208    4     309
## 16      216    5     309
## 17      214    6     309
## 18      218    7     309
## 19      224    8     309
```
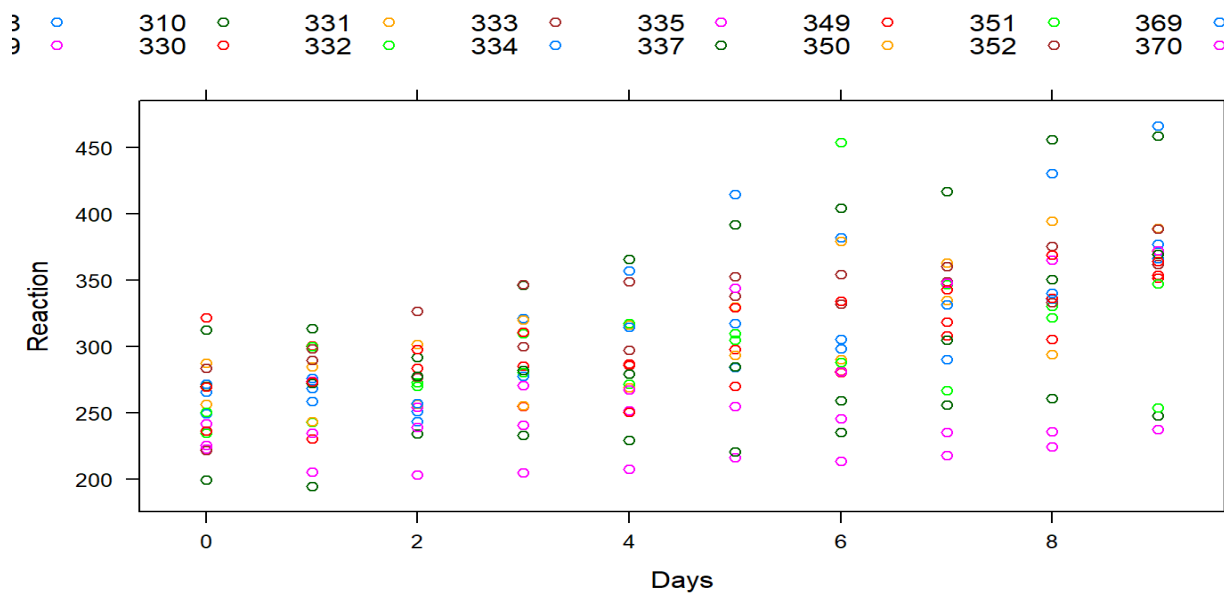
How does the Data look?

```
xyplot(Reaction ~ Days,groups=Subject,data=sleepstudy,auto.key=list(columns=9))
```
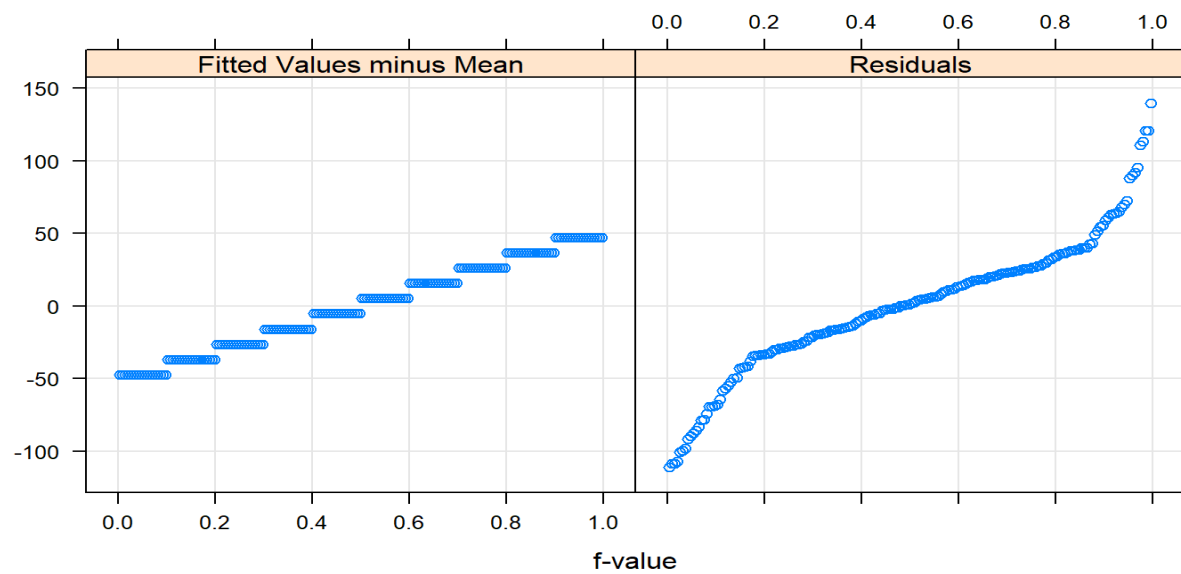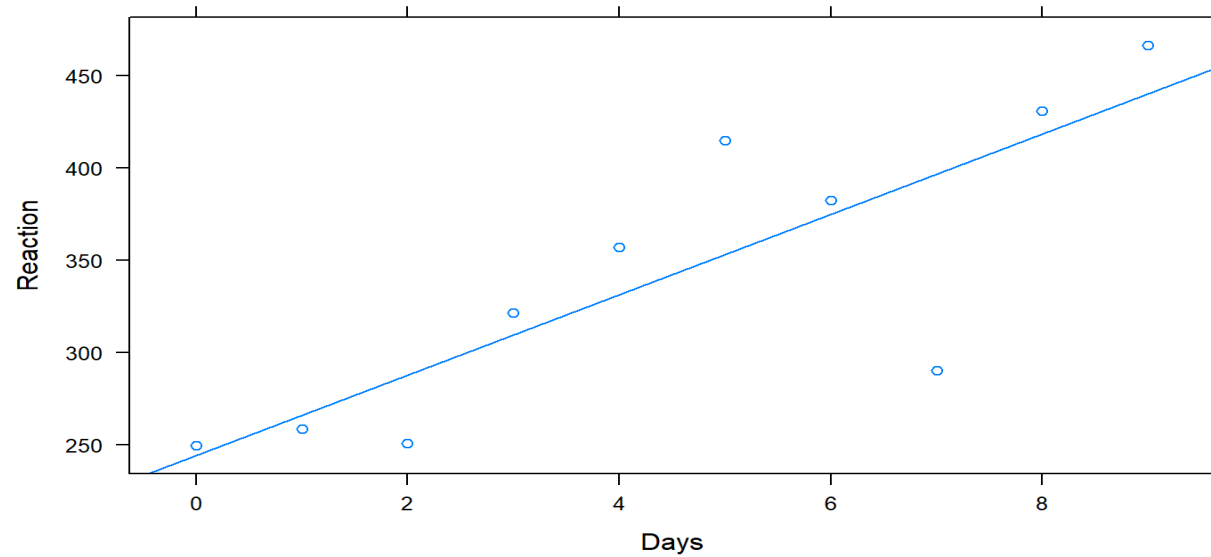
How does the Data look?
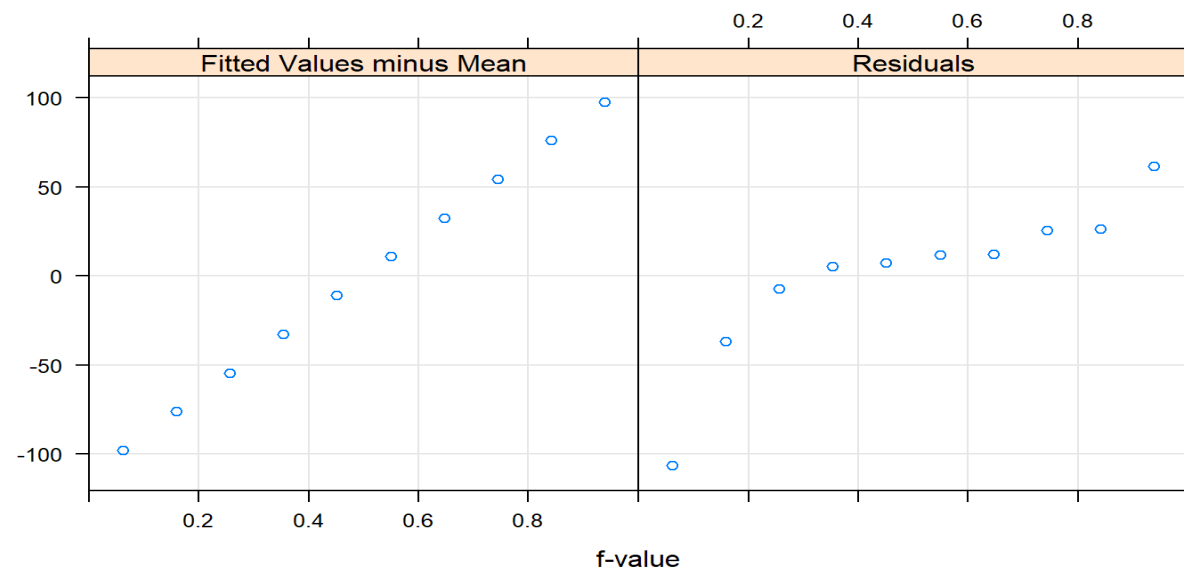
What if we just did linear?

```
library(simr)
rfs(model)
```
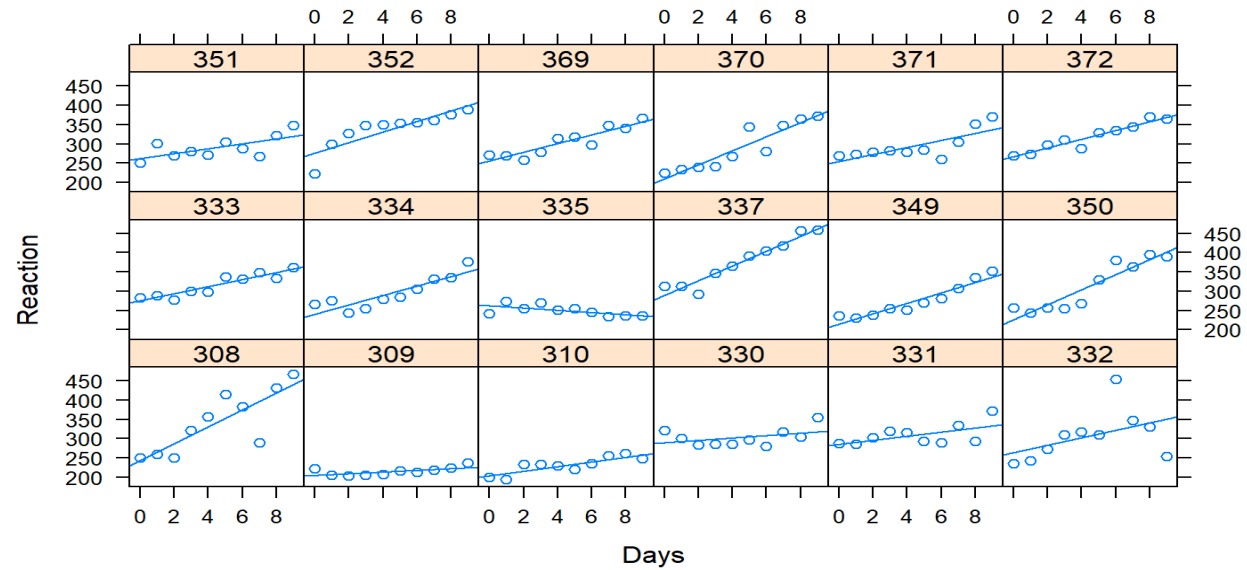
Linear Model SIngle Subject

Residuals: Single Subject

Linear Regression: Per Subject

```
xyplot(Reaction ~ Days|Subject,data=sleepstudy,auto.key=list(columns=9),type=c("p","r"))
```

Linear Regression: Per Subject

Estimating Models Independently

```
xyplot(Reaction ~ Days,groups=Subject,data=sleepstudy,auto.key=list(columns=9),type=c("p","r"))
```

Estimating Models Independently

```r
many <- as.data.frame(coef(models))
many$method <- "Many Models"
many$Intercept <- many[["(Intercept)"]]
many[["(Intercept)"]] <- NULL
many.wide <- many
many <- melt(many,id.vars = "method",variable.name="coef",value.name="est")

panel.density.mean <- function(x,...){
    panel.densityplot(x,...)
    panel.abline(v=mean(x),...)
}

densityplot(~est|coef,groups=method,data=many
            ,scales=list(x=(list(relation="free")))
            ,xlab="Estimate"
            ,panel=function(...) panel.superpose(panel.groups=panel.density.mean,...)
)
```

Estimating Models Independently
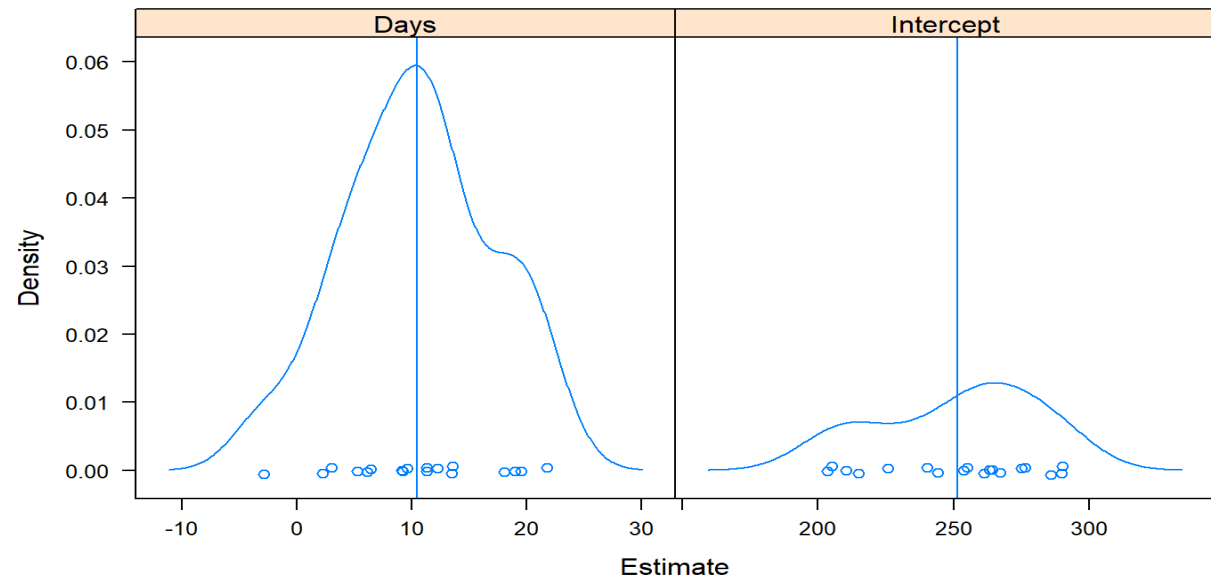
Potential Models

- per-subject correction for intercepts:

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + b_{1i} + \varepsilon_{ij} \qquad j = 1, \ldots, n_i$$

- per-subject correction for slopes:

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + b_{2i} t_{ij} + \varepsilon_{ij} \qquad j = 1, \ldots, n_i$$

- per-subject correction for both:

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + b_{1i} + b_{2i} t_{ij} + \varepsilon_{ij} \qquad j = 1, \ldots, n_i$$

Random Intercept and Slope Model Example

- Consider if we have a treatment vs control study

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + \beta_3 trt_i + \beta_4 t_{ij} \times trt_i + b_{1i} + b_{2i} t_{ij} + \varepsilon_{ij} \qquad j = 1, \ldots, n_i$$

- Then for the control group:

$$Y_{ij} = (\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) t_{ij} + \varepsilon_{ij}$$

The Treated Group

- For the treated group we would have:

$$Y_{ij} = (\beta_1 + \beta_3 + b_{1i}) + (\beta_2 + \beta_4 + b_{2i})t_{ij} + \varepsilon_{ij}$$

Covariance and Correlation Structure

- Then we would have:

  - $b_{1i} \sim N(0, \sigma_{b_1}^2)$

  - $b_{2i} \sim N(0, \sigma_{b_2}^2)$

  - $Cov(b_{1i}, b_{2i}) = \sigma_{b_1, b_2}$

  - $\varepsilon_{ij} \sim N(0, \sigma^2)$

Covariance and Correlation Structure

- Thus we have that

$$
\begin{aligned}
Var(Y_{ij}) &= Var(b_{1i} + b_{2i}t_{ij}\varepsilon_{ij}) \\
&= Var(b_1 i) + 2t_{ij}Cov(b_{1i}, b_{2i}) + t_{ij}^2 Var(b_{2i}) + Var(\varepsilon_{ij}) \\
&= \sigma_{b_1}^2 + 2t_{ij}\sigma_{b_1,b_2} + t_{ij}^2\sigma_{b_2}^2 + \sigma^2
\end{aligned}
$$

Covariance and Correlation Structure

We can also show that

$$Cov(Y_{ij}, Y_{ik}) = \sigma_{b_1}^2 + (t_{ij} + t_{ik})\sigma_{b_1,b_2} + t_{ij}t_{ik}\sigma_{b_2}^2 + \sigma^2$$

Estimation Techniques

- We already saw that linear regression used least squares estimation unfortunately there is no simple expression for the maximum likelihood estimator of the covariance components that we have.

- We rely on an iterative algorithm instead.

Maximizing the Likelihood

- Then what happens with this is that if we maximize the likelihood we end up with the same least squares estimator that we have before however our estimate of σ^2 is

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(y_i - X_i\beta)^2}{N}$$

- where N is the sample size, we could then show that

$$E(\hat{\sigma}^2) = \left(\frac{N-p}{N}\right)\sigma^2$$

Residual Maximum Likelihood

- This means that if we have small samples of size N than we underestimate the σ^2.

- Due to this we use what is called *Residual Maximum Likelihood (REML)*

Residual Maximum Likelihood

- Basically the bias we saw came from the fact that we do not know the true β so we replace this term with a $\hat{\beta}$ however this is an estimated value.

- We saw a similar problem when we tried to originally estimate variance and had to adjust for sample size.

- We will find that R, SAS and Stata use this as the default method.

Random Intercept Model

```
library(lme4)
model.intercepts <- lmer(Reaction ~ 1 + Days + (1|Subject),data=sleepstudy)
```

Random Intercept Model

```
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ 1 + Days + (1 | Subject)
   Data: sleepstudy

REML criterion at convergence: 1786

Scaled residuals:
    Min     1Q Median     3Q    Max
-3.226 -0.553  0.011  0.519  4.251

Random effects:
 Groups    Name        Variance Std.Dev.
 Subject  (Intercept) 1378      37.1
 Residual              960      31.0
```

Random Intercept Model

Number of obs: 180, groups: Subject, 18

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.405	9.747	25.8
Days	10.467	0.804	13.0

Correlation of Fixed Effects:

	(Intr)
Days	-0.371

Random Slope Model

```
model.slopes <-    lmer(Reaction ~ 1 + Days + (0 + Days|Subject),data=sleepstudy)
```

Random Slope Model

```
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ 1 + Days + (0 + Days | Subject)
   Data: sleepstudy

REML criterion at convergence: 1766

Scaled residuals:
    Min      1Q Median      3Q     Max
-3.510 -0.559  0.054  0.624  4.602

Random effects:
 Groups    Name Variance Std.Dev.
 Subject  Days  52.7      7.26
 Residual       842.0     29.02
```

Random Slope Model

Number of obs: 180, groups: Subject, 18

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.41	4.02	62.5
Days	10.47	1.87	5.6

Correlation of Fixed Effects:

	(Intr)
Days	-0.340

Random Intercept and Slope

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

Random Intercept and Slope

```
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ 1 + Days + (1 + Days | Subject)
   Data: sleepstudy

REML criterion at convergence: 1744

Scaled residuals:
    Min      1Q Median      3Q     Max
-3.954 -0.463  0.023  0.463  5.179

Random effects:
 Groups    Name        Variance Std.Dev. Corr
 Subject  (Intercept) 612.1     24.74
          Days          35.1      5.92    0.07
 Residual              654.9     25.59
```

Random Intercept and Slope

```
Number of obs: 180, groups:  Subject, 18

Fixed effects:
            Estimate Std. Error t value
(Intercept)   251.41       6.82    36.8
Days           10.47       1.55     6.8

Correlation of Fixed Effects:
     (Intr)
Days -0.138
```
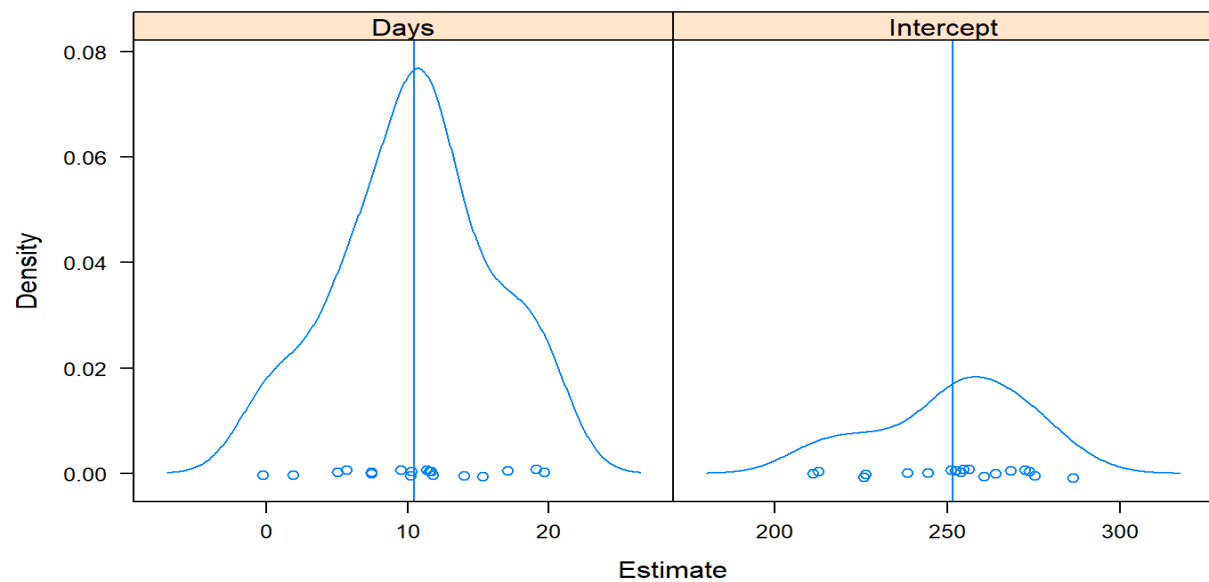
Distribution of Estimates

```
mixed <- as.data.frame(coef(model.full)$Subject)
mixed$method <- "Mixed Effects Models"
mixed$Intercept <- mixed[["(Intercept)"]]
mixed[["(Intercept)"]] <- NULL
mixed.wide <- mixed
mixed <- melt(mixed,id.vars = "method",variable.name="coef",value.name="est")

densityplot(~est|coef,groups=method,data=mixed
            ,scales=list(x=(list(relation="free")))
            ,xlab="Estimate"
            ,panel=function(...) panel.superpose(panel.groups=panel.density.mean,...)
)
```

Distribution of Estimates

How Does this Compare?

```
mixed <- as.data.frame(coef(model.full)$Subject)
mixed$method <- "Mixed Effects Model"
many$method <- "Many Models"
mixed$Intercept <- mixed[["(Intercept)"]]
mixed[["(Intercept)"]] <- NULL
mixed.wide <- mixed
mixed <- melt(mixed,id.vars = "method",variable.name="coef",value.name="est")

estimates <- rbind(many,mixed)
estimates.wide <- rbind(many.wide,mixed.wide)
estimates$method <- factor(estimates$method)
estimates.wide$method <- factor(estimates.wide$method)
densityplot(~est|coef,groups=method,data=estimates
            ,auto.key=list(columns=2),scales=list(x=(list(relation="free")))
            ,xlab="Estimate"
            ,panel=function(...) panel.superpose(panel.groups=panel.density.mean,...)
)
```

How Does this Compare?
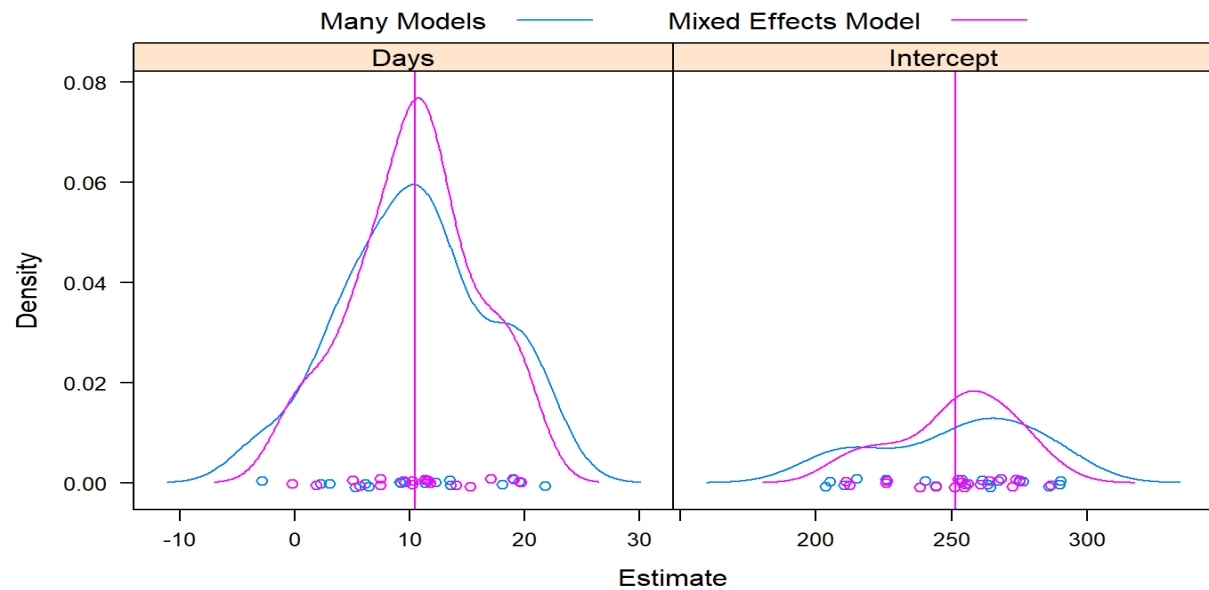
REML vs. ML

```
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ 1 + Days + (1 + Days | Subject)
   Data: sleepstudy

REML criterion at convergence: 1744

Scaled residuals:
    Min      1Q Median      3Q     Max
-3.954 -0.463  0.023  0.463  5.179

Random effects:
 Groups    Name        Variance Std.Dev. Corr
 Subject   (Intercept) 612.1    24.74
           Days         35.1     5.92    0.07
 Residual              654.9    25.59
```

REML vs. ML

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.41	6.82	36.8
Days	10.47	1.55	6.8

Correlation of Fixed Effects:

	(Intr)
Days	-0.138

REML vs ML

```
model.full.ml <- update(model.full,REML=FALSE)
```

REML vs ML

```
Formula: Reaction ~ 1 + Days + (1 + Days | Subject)
   Data: sleepstudy

     AIC       BIC    logLik deviance df.resid
    1764      1783      -876     1752       174

Scaled residuals:
    Min      1Q Median      3Q     Max
 -3.942 -0.466  0.029   0.464   5.179

Random effects:
 Groups    Name         Variance Std.Dev. Corr
 Subject  (Intercept) 565.5     23.78
          Days          32.7      5.72    0.08
 Residual              654.9     25.59
```

REML vs ML

Number of obs: 180, groups: Subject, 18

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.41	6.63	37.9
Days	10.47	1.50	7.0

Correlation of Fixed Effects:

	(Intr)
Days	-0.138

REML vs. ML

- **REML**: Restricted Maximum Likelihood

- Variance is the average squared distance to the *true mean*

- Variance measured to the *ML-estimated mean* is less than true variance in finite samples

- This is the motivation behind Bessel's correction ($n - 1$ in the denominator instead of n when calculating variance using SS)

REML vs. ML

- Similar motivation for using REML in estimation of mixed-effects models: more accurate estimation of variance (and hence random effects!)

- "log likelihood" REML-fitted models dependent on paramerization and thus not comparable across models

- similarly: AIC, BIC, other measures of fit not comparable across models

- REML-models more accurate but not comparable with each other:

  - determine model structure with comparisons between ML-estimates

  - present final model with REML-estimates

  - `lme4` has some built-in protections to prevent REML-based comparisons, but don't depend on these!