Goals

RMANOVA, RIP

Linear Mixed Effects Models

Random Intercepts

Random Intercepts and Slopes

Let's do it!

Repeated Measures ANOVA, RIP

Mariel Finucane, PhD

Gladstone Bioinformatics Core
September 24, 2013

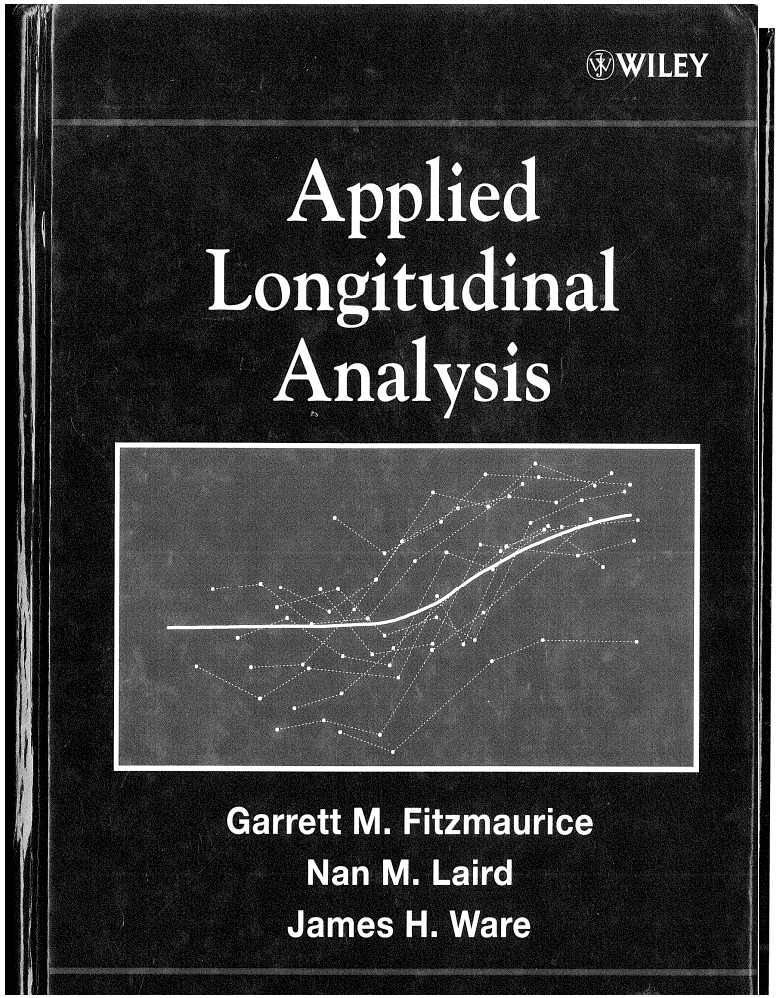
Motivation: Say we have a continuous outcome that's changing over time...

- Should we use regression or ANOVA?
- Is simple linear regression appropriate?

Goals

1. Realize that repeated measures ANOVA should RIP.
2. Learn about a flexible, powerful alternative: linear mixed effects models.
   (a) Write down the regression equation
   (b) Fit and plot the regression in R.
   (c) Interpret the regression output.

Repeated measures ANOVA: an "Historical Approach"



among the repeated measures that are somewhat less restrictive. In addition, these models permit the variance to change over time in a smooth fashion. Indeed, random effects models provide both very flexible and parsimonious models for the covariance and are particularly well suited to handling longitudinal data that are irregularly timed. These models are discussed at length in Chapter 8.

3.6 HISTORICAL APPROACHES

We conclude this chapter with a brief survey of some of the earliest developments in methods for analyzing longitudinal and clustered data. Historically, a variety of relatively simple methods have been developed for the analysis of repeated measures data. Some, but not all, of these happen to be special cases of the regression models for longitudinal data that are the focus of later chapters of this book. As a result, in this section we provide only a brief historical survey of some of these approaches, highlighting their relation to more general models, and noting some of their potential limitations. Many of the shortcomings of these methods alluded to here will be more readily apparent when the methods are viewed as special cases of the regression models considered in later chapters.

From a historical perspective, three methods for the analysis of repeated measures data can be distinguished: (1) univariate repeated measures analysis of variance (ANOVA), (2) multivariate repeated measures analysis of variance (MANOVA), and (3) methods based on summary measures. All three of these approaches have enjoyed varying degrees of popularity, and some are still in widespread use, in different areas of application. Many of these approaches are unnecessarily restrictive in their assumptions and their analytic goals. For example, ANOVA and MANOVA focus on comparing groups in terms of their mean response trend over time but provide little information about how individuals change over time. Also, as we will see later, ANOVA and MANOVA have numerous features that limit their usefulness for the analysis of longitudinal data. In contrast, the regression models that are discussed throughout the remainder of this book make more realistic assumptions and can address the major scientific questions of interest in a longitudinal study. For all of the reasons that were outlined in Section 1.4, we view the regression paradigm as being the most useful, general, and versatile approach for analyzing longitudinal data arising from the health sciences.

Repeated Measures Analysis by ANOVA

One of the earliest proposals for analyzing correlated responses was the repeated measures analysis of variance (ANOVA), sometimes referred to as the "univariate" or "mixed-model" analysis of variance. The analysis of variance paradigm was developed in the early part of the twentieth century by R. A. Fisher. Although many of the early applications of ANOVA were to designed experiments in agriculture, since then it has found widespread application in many other disciplines. In the repeated

Repeated measures ANOVA: "RIP"

First page preview

Download full text

Should we quit using repeated measures analysis of variance?

Repeated Measures ANOVA, R.I.P.?

Charles E. McCulloch

It is a difficult experiment to run and to analyze: What are the effects of alcohol on sleepiness and does a hormone, pregnenolone, which has been shown to enhance memory in rat experiments, help alleviate the sleepiness? Each person is tested under each of four conditions on four different visits in random order: a placebo for the drug and for the hormone, alcohol alone, hormone alone, and the combination. Each subject is also queried multiple times within a visit in the minutes after alcohol (or placebo) ingestion. Some subjects drop out of the protocol without completing all the conditions and some of the sleepiness scores are not recorded within a visit because of difficulties executing the protocol. How should the data be analyzed?

This is an opportunity for the professional statistician to pull out any of a number of impressive and more recent tools of the trade: generalized estimating equations, mixed model analyses, imputation, and inverse probability weighting. Gone are the Huynh-Feldt or Geisser-Greenhouse corrections, expected mean squares, figuring out the "correct" error term, and filling in missing data. Or are they? The investigator, after being led through the results of a SAS Proc MIXED analysis bemoans, "Can't you run a repeated measures analysis of variance? That is all I can really understand." Are the new methods overkill, or do they offer significant advantages?

Analysis of Variance

Analysis of variance (ANOVA) has a long history dating back to R.A. Fisher (1925). The key idea is to divide up the variability in a data set into interpretable components. Consider a simplification of the above scenario: We record Y_{ijk} = the average sleepiness score between 60 and 120 minutes after administration of alcohol or placebo for subject $i(= 1, 2, \ldots, n)$ under alcohol condition $j(= 1, 2)$ and pregnenolone condition $k(= 1, 2)$. One of the simplest models we might entertain for such a situation would be

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + p_i + \varepsilon_{ijk}, \qquad (1)$$

where μ is the overall mean, α_j represents the alcohol effect, β_k the pregnenolone effect, $\alpha\beta_{jk}$ the interaction effect, p_i is the person effect, and ε_{ijk} is an error term. This model hypothesizes simple person effects that raise or lower (if the effect is negative) the average sleepiness in all four conditions. Interest focuses on the interaction, because the scientific question is whether pregnenolone helps to reduce the sleep-inducing effect of alcohol.

With a mean, error term and four explanatory factors in the model, the analysis of variance would partition the variability in Y_{ijk} into four sources: person, alcohol, pregnenolone, and the interaction. An ANOVA would generate a table as outlined in Table 1. A typical assumption is that the ε_{ijk} are all independent and follow a $N(0, \sigma^2)$ distribution and that the rest of the terms in (1) are fixed. If these assumptions are correct, the F-statistic of interest, namely F(Int)= MS(Int)/MS(Err), has an exact F-distribution with 1 and $3(n-1)$ degrees of freedom. So this forms a straightforward significance test of the null hypothesis of no interaction, with p-value given by $P\{F^1_{3(n-1)} \geq F(\text{Int})\}$, with F representing an F-distributed random variable with the given degrees of freedom.

When every subject is tested under each of the conditions and there are no missing data, the calculations leading to Table 1, though somewhat tedious, can be performed with a hand calculator. There is a certain tidiness in showing how all the variability in the data is divided up into pieces, no matter what order factors are entered into the model, the additional amount attributable to the model upon entering a new explanatory factor is given by the sum of squares (SS) in Table 1 for that factor. Also, the sums of squares for the various factors add up to SS(Tot), the "total" sum of squares.

When there is unbalancedness to the data in the sense that not every subject has data for each condition, then the calculations are much more difficult and essentially require

Click to decrease image size

Repeated measures ANOVA makes the following assumptions (which likely don't hold in your dataset):

- Discrete covariates.
- Common measurement occasions for all subjects.
- Complete data, i.e. no missing observations.
- 'Sphericity', i.e. the correlation between repeated measurements must be the same no matter how close or far apart in time the measurements are taken.

Mixed effects models are a flexible, powerful alternative

- Mixed effects models, a.k.a.:
  - (Linear) mixed (effects) models
  - Hierarchical models
  - Multilevel models
- Accommodate continuous *and* discrete covariates.
- Accommodate any degree of 'imbalance', i.e. do not require the same number of observations on each subject nor that the measurements be taken at the same set of measurement occasions.
- Accommodate data at all levels of a 'hierarchy'.
- Don't make the untenable 'sphericity' assumption.

Mixed = fixed + random

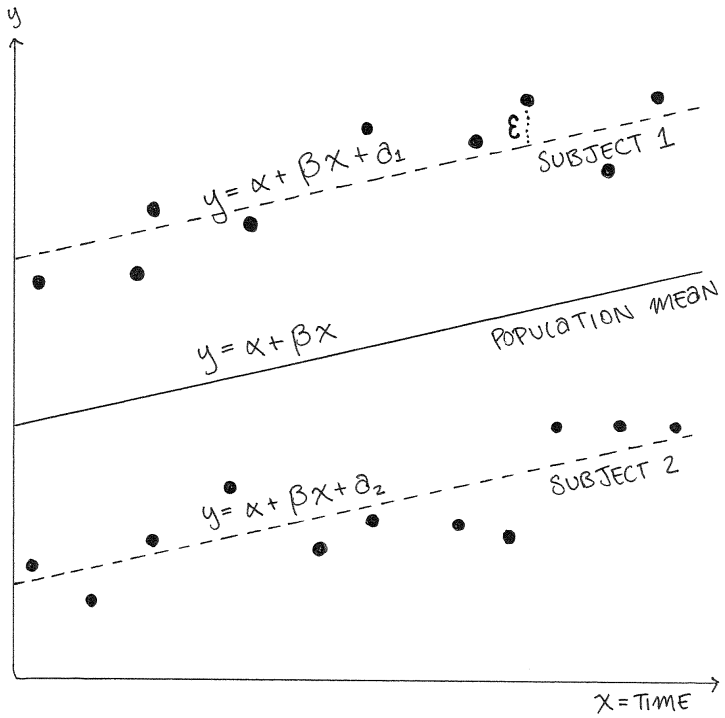
- Fixed effects (α, β)
  - Population characteristics
  - Shared by all individuals
  - Describe the mean response trajectory in the population
  - Useful to epidemiologists
- Random effects (a, b)
  - Subject-specific effects
  - Vary from one individual to another
  - Describe individuals' response trajectories
  - Useful to clinicians

The simplest case: a random intercept model

$$y = \alpha + \beta x + a + \varepsilon$$
$$= (\alpha + a) + \beta x + \varepsilon$$

- α is the fixed, population-level intercept
- β is the fixed, population-level time slope
- a is the random, subject-specific intercept:
  By how much does each individual deviate from the population average?
- ε is the error term

Random intercept model

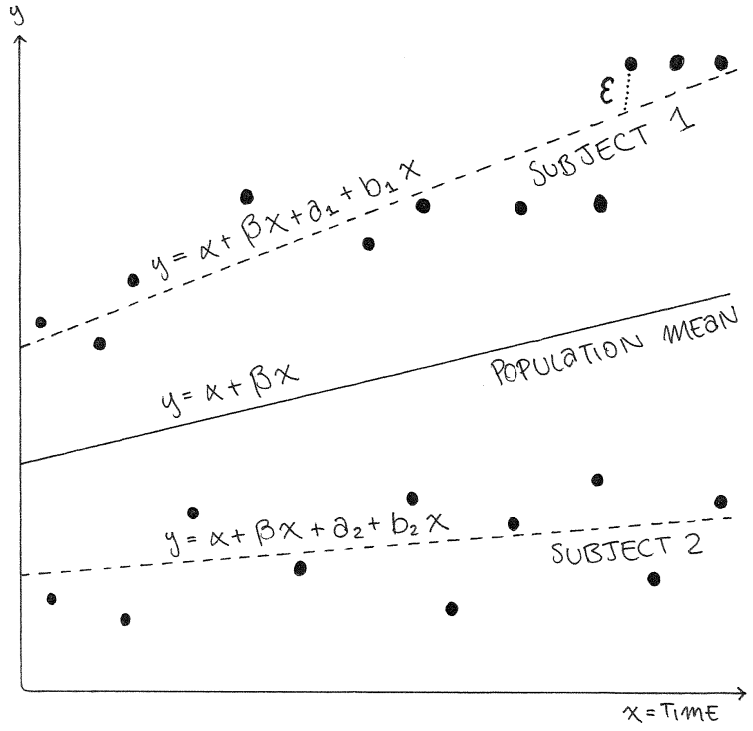


The more common scenario: a random intercept and slope model

$$y = \alpha + \beta x + a + bx + \varepsilon$$
$$= (\alpha + a) + (\beta + b)x + \varepsilon$$

- α is the fixed, population-level intercept
- β is the fixed, population-level time slope
- a is the random, subject-specific intercept:
  By how much does each individual deviate from the population average?
- b is the random, subject-specific time slope:
  By how much does the effect of time on each individual deviate from the population-average effect?
- ε is the error term

Random intercept and slope model



Let's look at some data

```
> # install.packages('arm')
> library(arm)
> head(sleepstudy)

  Reaction Days Subject
1    249.6    0     308
2    258.7    1     308
3    250.8    2     308
4    321.4    3     308
5    356.9    4     308
6    414.7    5     308

> attach(sleepstudy)
```

INSTALL AND LOAD A "PACKAGE" CALLED "arm" THAT CONTAINS THE DATA AND SOME OF THE FUNCTIONS WE'LL BE USING.

SHOW ME THE "head" OF THE DATASET CALLED "sleepstudy".

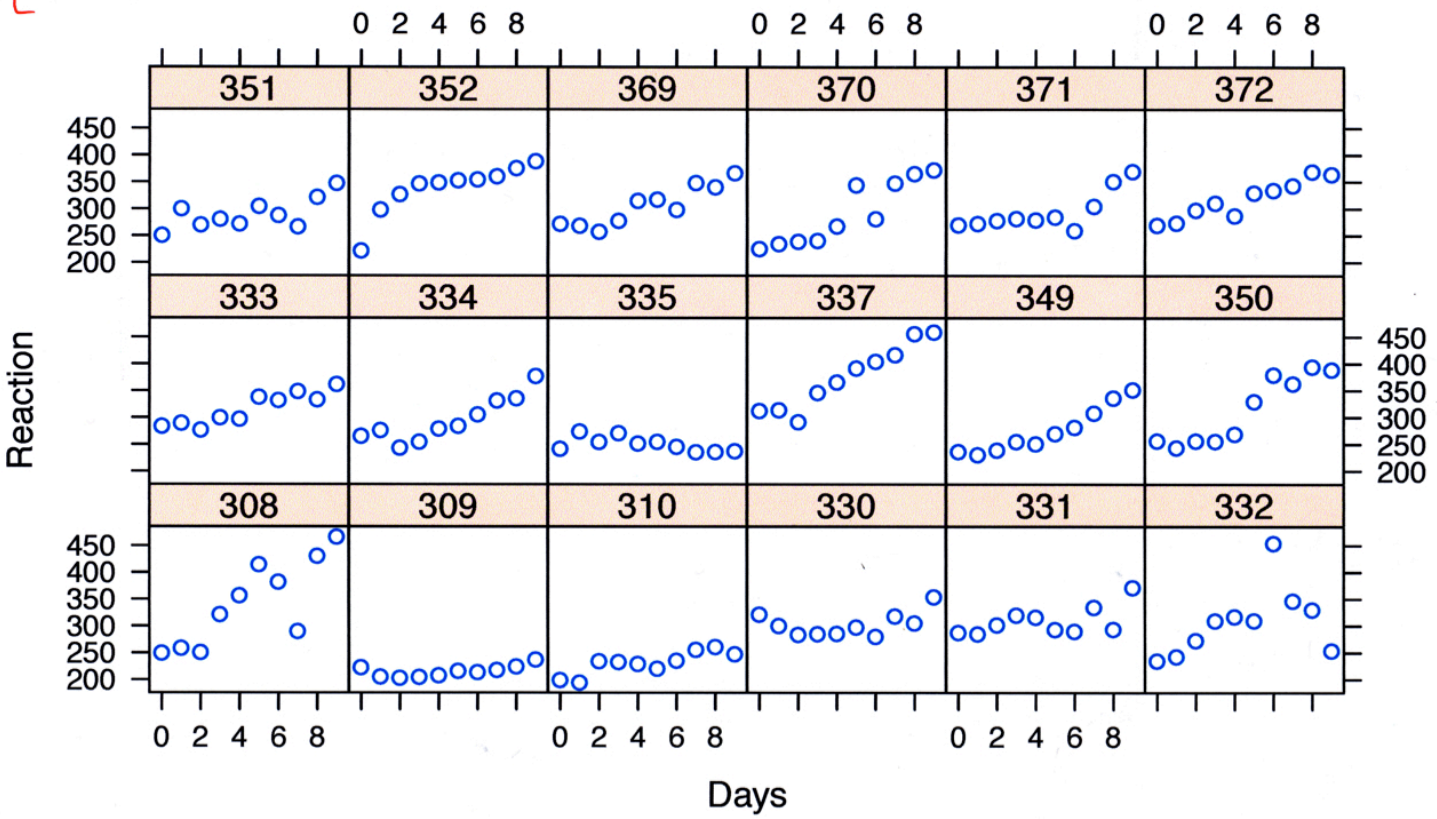
GIVE ME ACCESS TO THE VARIABLES IN THE DATASET CALLED "sleepstudy".

Let's look at some data

["FOR EACH"]
["VS."]

```
> xyplot(Reaction ~ Days | Subject)
```

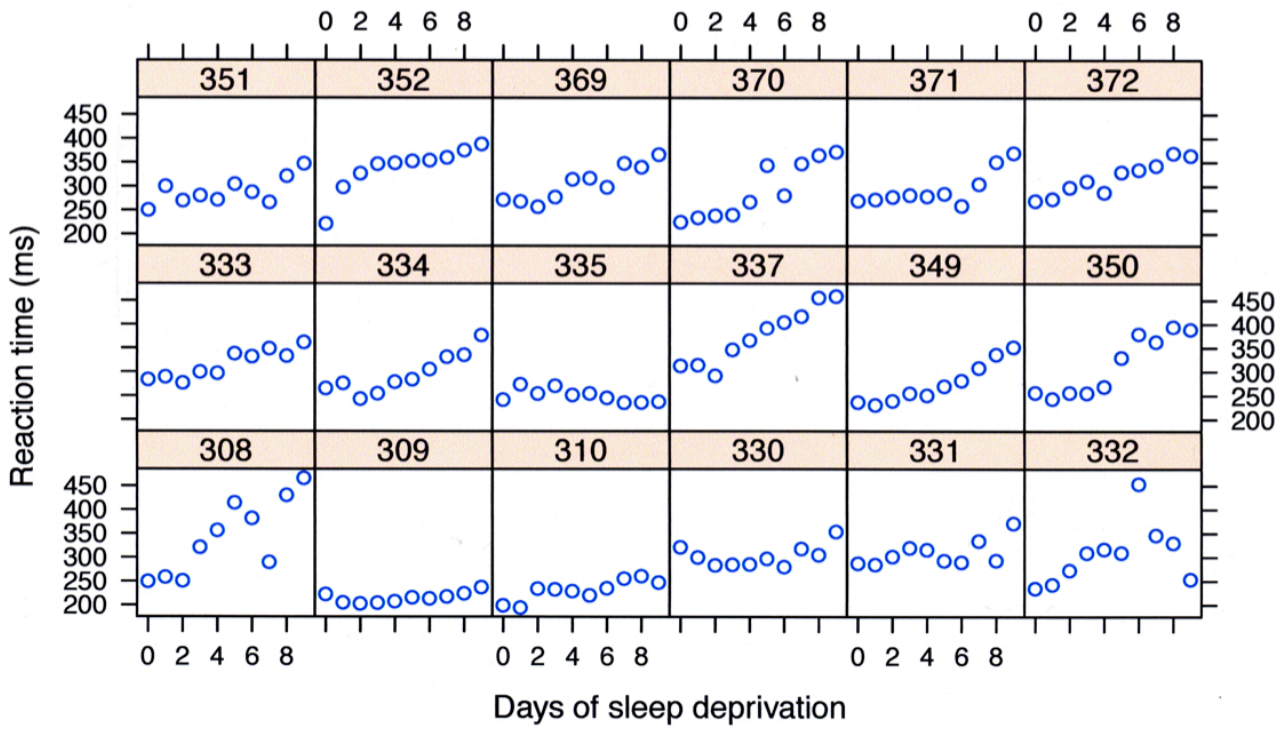
[DRAW A "LATTICE" PLOT OF y VS. x FOR EACH INDIVIDUAL.]



[ADD X- AND Y-AXIS LABELS.]

Let's look at some data

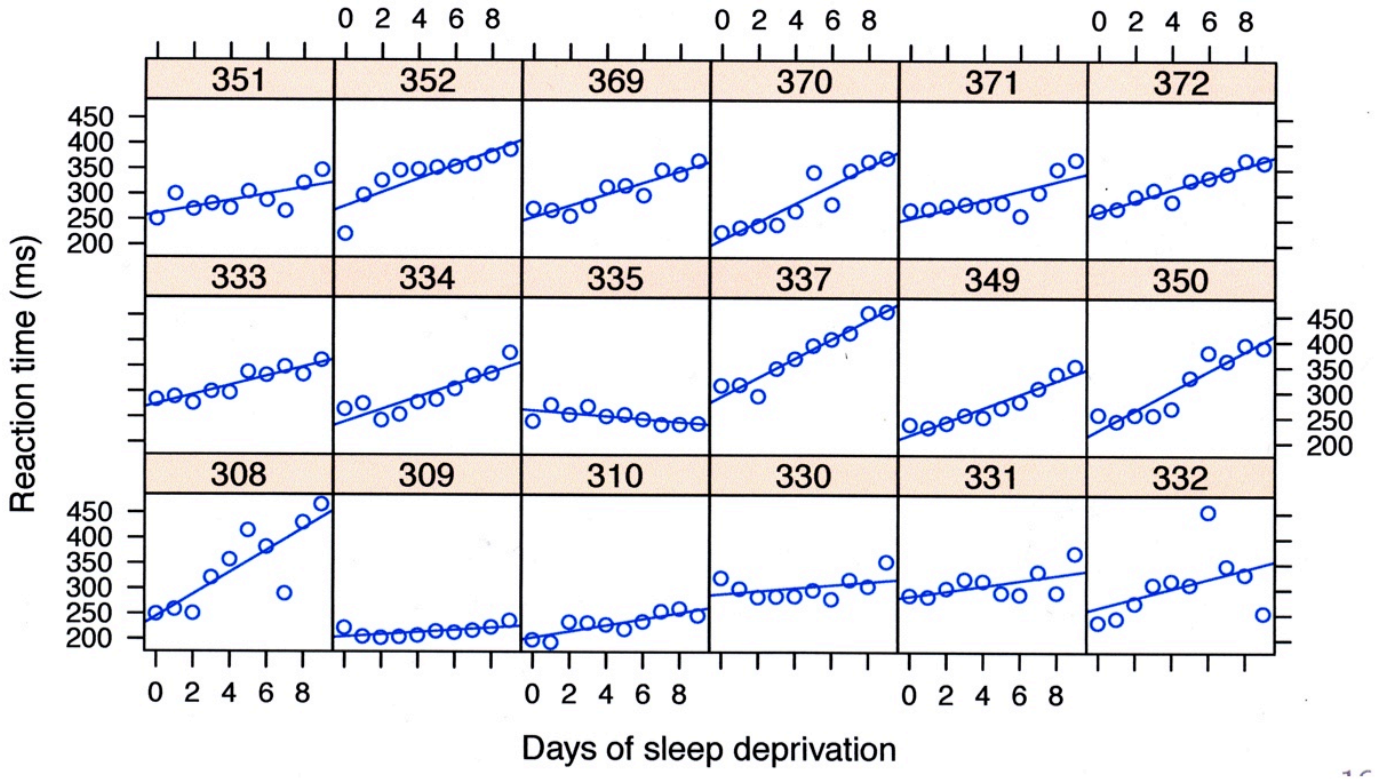
```
> xyplot(Reaction ~ Days | Subject,
+        xlab = "Days of sleep deprivation",
+        ylab = "Reaction time (ms)")
```



Let's look at some data

```
> xyplot(Reaction ~ Days | Subject,
+        xlab = "Days of sleep deprivation",
+        ylab = "Reaction time (ms)",
+        type=c('p','r'))
```

INCLUDE BOTH "P"OINTS (THE DEFAULT) AND "R"EGRESSION LINES.



Fitting a mixed effects model

[CREATE AN OBJECT CALLED "model".]

```
> model <- lmer(Reaction ~ 1 + Days + (1 + Days| Subject))
```

$$y = \alpha + \beta x + a + bx + \varepsilon$$

[AND PUT INTO IT]

[a "L"INEAR "M"IXED "E"FFECTS "R"EGRESSION.]

SHOW ME THE "FIX"ED "EF"FECTS FROM "model".

Extracting fixed/random effects

```
> fixef(model)

(Intercept)        Days
     251.41       10.47
```

SHOW ME THE "head" OF THE "SUBJECT"-SPECIFIC "RAN"DOM "EF"FECTS FROM "model".

```
> head(ranef(model)$Subject)

    (Intercept)    Days
308       2.257  9.1993
309     -40.398 -8.6211
310     -38.961 -5.4502
330      23.692 -4.8137
331      22.261 -3.0693
332       9.040 -0.2719
```

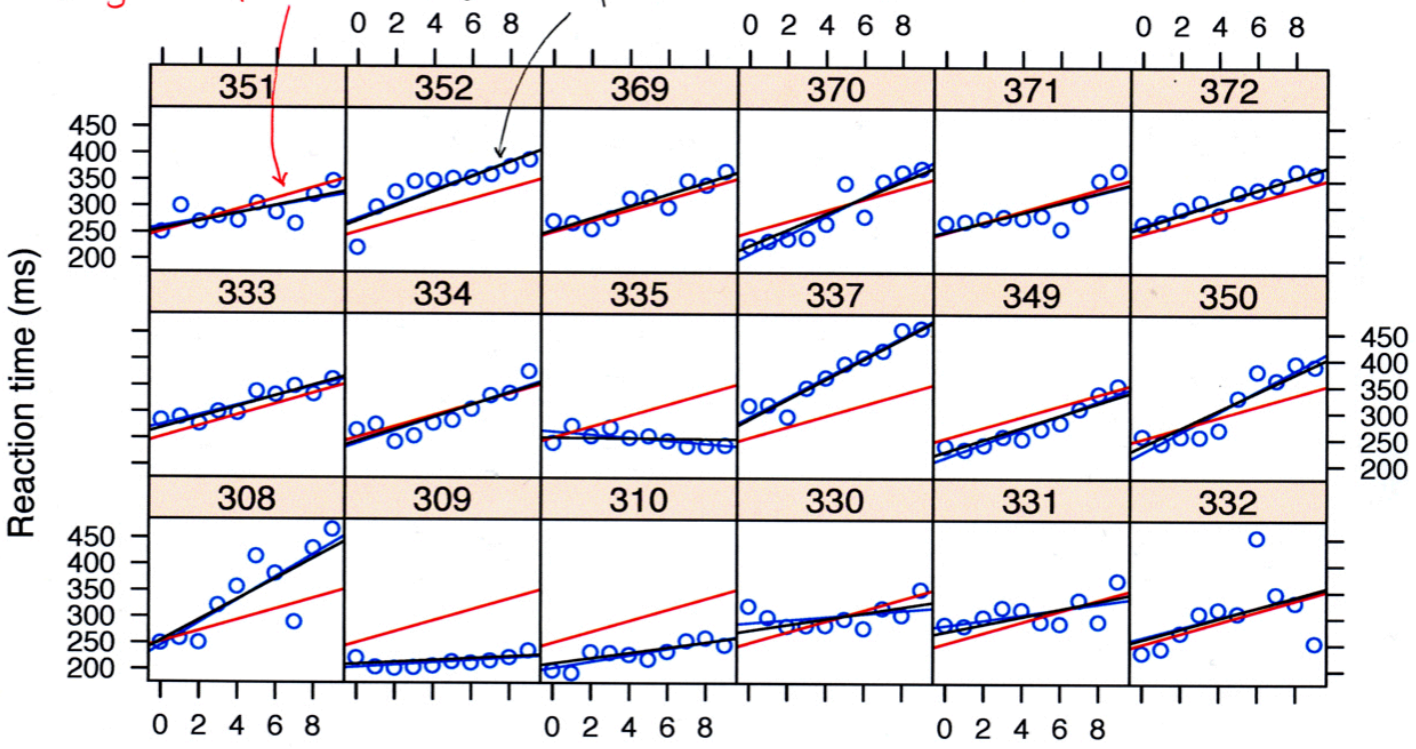
CREATE OBJECTS CALLED "alpha", "beta", "a", AND "b" CONTAINING THE RELEVANT EFFECTS.

```
> alpha <- fixef(model)[1]
> beta  <- fixef(model)[2]
> a     <- ranef(model)$Subject[,1]
> b     <- ranef(model)$Subject[,2]
```

Plotting the fitted model

```
> xyplot(Reaction ~ Days | Subject,
+        xlab = "Days of sleep deprivation",
+        ylab = "Reaction time (ms)",
+        panel = function(x, y) {
+          panel.xyplot(x, y, type=c('p', 'r'))
+          panel.curve(alpha + beta * x, col="Red")
+          i <- panel.number()
+          panel.curve(alpha+a[i] + (beta+b[i]) * x,
+                      col="Black")})
```

$[y = \alpha + \beta x]$ $[y = \alpha + \beta x + a + bx]$



Summarizing the fitted model

```
> display(model)

lmer(formula = Reaction ~ 1 + Days + (1 + Days | Subject))
            coef.est coef.se
(Intercept) 251.41     6.82
Days         10.47     1.55

Error terms:
 Groups   Name        Std.Dev. Corr
 Subject  (Intercept) 24.74
          Days         5.92    0.07
 Residual             25.59
---
number of obs: 180, groups: Subject, 18
AIC = 1755.6, DIC = 1760
deviance = 1752.0
```

α

β

[STANDARD DEVIATION OF a's]

[STANDARD DEVIATION OF b's]

Interpreting the R output

- For your Methods section:
  - *"We fit a linear mixed effects model of reaction time on the number of days of sleep deprivation, including random intercepts and slopes for each subject.*
- For your Results section:
  - *"For each additional day of sleep deprivation, there's a 10.5 ms increase in reaction time."*
  - *"There is substantial variability across subjects in their baseline reaction time (SD=24.7) and in their response to sleep deprivation (SD=5.9)."*