# BIOS 755: Linear Mixed Models I

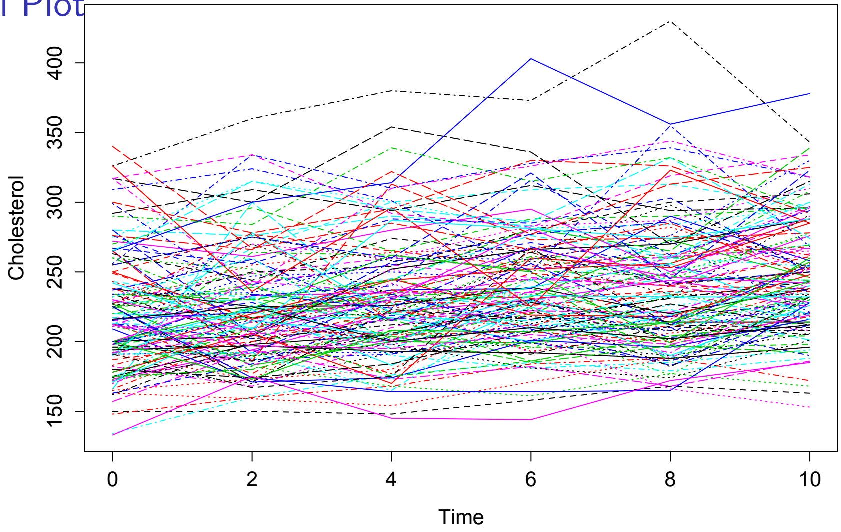
Alexander McLain

February 2, 2023

# Framingham study Cholesterol Data

- In the Framingham study, each of 2634 participants was examined every 2 years for a 10 year period for his/her cholesterol level.
- Study objectives:
  - How does cholesterol level change over time on average as people get older?
  - How is the change of cholesterol level associated with sex and baseline age?
- A subset of 200 subjects' data is used for illustrative purpose.

# Spaghetti Plot



# Introduction to Linear Mixed Models

- In the General Linear Model we focused our conceptual model on the covariance and correlation of the error terms.
- In linear mixed models, the conceptual model is based on thinking about individual behavior first.
- The possibilities for how this is represented and how the variation in the population is represented, are very flexible.
- As we'll see, linear mixed models can incorporate heterogeneity and different correlation structures (even though we don't think about them that way).

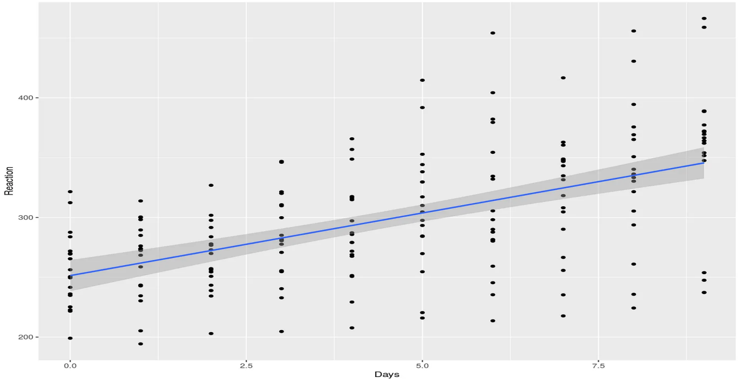
# Linear Mixed (Effects) Models

- The linear mixed model can be expressed as

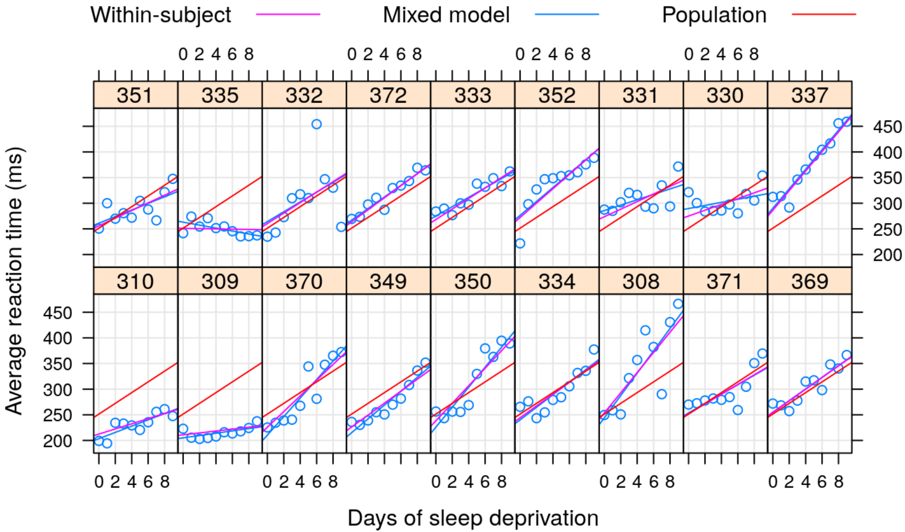
$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$$

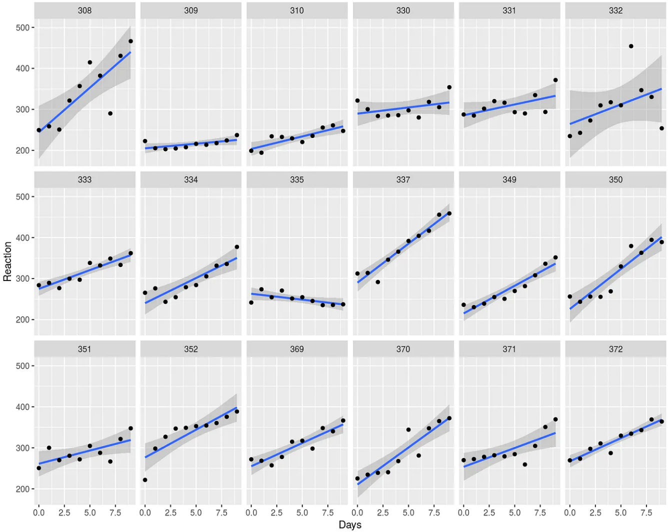
where

- $\boldsymbol{X}_i$ – $n_i \times p$ matrix of fixed effect covariates
- $\boldsymbol{\beta}$ – $k \times 1$ vector of regression coefficients (fixed effects).
- $\boldsymbol{Z}_i$ – $n_i \times q$ matrix of random effect covariates.
- $\boldsymbol{b}_i$ – $q \times 1$ vector of random effects, $\boldsymbol{b}_i \sim N(0, \boldsymbol{G})$,
- $\boldsymbol{e}_i$ – $n_i \times 1$ vector of errors and $\boldsymbol{e}_i \sim N(0, \boldsymbol{R}_i)$.




- Consider a sleep deprivation study where sleeping time of 18 individuals was restricted, and a Reaction of their organism on a series of tests was measured during 10 days. The data includes three variables: 1) Reaction, 2) Days, 3) Subject, i.e. the same individual was followed during 10 days.


Within-subject
Mixed model
Population
351
335
332
372
333
352
331
330
337
310
309
370
349
350
334
308
371
369
Average reaction time (ms)
Days of sleep deprivation
200
250
300
350
400
450
0 2 4 6 8


308
309
310
330
331
332
333
334
335
337
349
350
351
352
369
370
371
372
Reaction
Days
500
400
300
200
0.0
2.5
5.0
7.5

# Random intercept and slope model

The random intercept and slope model:

$$\boldsymbol{Y}_i = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\boldsymbol{t}_i + \boldsymbol{e}_i$$

where $\boldsymbol{t}_i' = \{t_{i1}, t_{i2}, \ldots, t_{in_i}\}$

- $\beta_0$ is the average intercept and $b_{0i}$ are the deviations from the average intercept.
- $\beta_1$ is the average slope and $b_{1i}$ are the deviations from the average slope.
- We could add other fixed effects to this model (sex, smoking, etc.).

# Random intercept and slope model

The random intercept and slope model:

$$\boldsymbol{Y}_i = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\boldsymbol{t}_i + \boldsymbol{e}_i$$

- $\boldsymbol{R}_i = var(\boldsymbol{e}_i)$ describes the covariance of the residuals
- In the models we've been running in the previous weeks, this is the covariance of the $i$th subject's deviations from $\beta_0 + \beta_1\boldsymbol{t}_i$ (i.e., the overall trend)
- Now it's the covariance of the $i$th subject's deviations from $\beta_0 + b_{0i} + (\beta_1 + b_{1i})\boldsymbol{t}_i$ (i.e., their individual trend)
  - Usually, it is assumed that $\boldsymbol{R}_i = \sigma^2 I$, which is the "conditional independence assumption."

# Linear Mixed (Effects) Models

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$$

The vector of regression parameters $\boldsymbol{\beta}$ are the fixed effects, which are assumed to be the same for all individuals.

- Fixed effects are constant across individuals, and random effects vary.
- For example, in a growth study, a model with random intercepts $\beta_0 + b_{0i}$ and fixed slope $\beta_1$ corresponds to parallel lines for different individuals $i$, or the model $Y_{ij} = \beta_0 + b_{0i} + \beta_1 t_{ij} + e_{ij}$

# Conditional vs marginal mean

- The **conditional** mean of $\boldsymbol{Y}_i$, given $\boldsymbol{b}_i$, is

$$E(\boldsymbol{Y}_i|\boldsymbol{b}_i) = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i$$

- The **marginal** or **population-averaged** mean of $\boldsymbol{Y}_i$ is

$$E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\beta$$

- In contrast to $\beta$, the vector $\boldsymbol{b}_i$ is comprised of subject-specific regression coefficients.
- All covariates in $\boldsymbol{Z}$ will be in $\boldsymbol{X}$, and it's rare to consider more than 2 variables in $\boldsymbol{Z}$.

# Conditional vs marginal variance

- In the mixed model
$$\boldsymbol{Y}_i = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$$
- We have the following conditional and marginal expectations
$$E(\boldsymbol{Y}_i|\boldsymbol{b}_i) = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i, \quad E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\beta$$
along with the following conditional and marginal variances
$$\begin{aligned} var(\boldsymbol{Y}_i|\boldsymbol{b}_i) &= var(\boldsymbol{e}_i) = \boldsymbol{R}_i, \quad \text{and} \\ var(\boldsymbol{Y}_i) &= var(\boldsymbol{Z}_i\boldsymbol{b}_i) + var(\boldsymbol{e}_i) \\ &= \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i' + \boldsymbol{R}_i \end{aligned}$$

# Linear Mixed (Effects) Models

- The introduction of random effects induces correlation among the $\boldsymbol{Y}_i$.
- $Var(\boldsymbol{Y}_i)$ is described in terms of a set of covariance parameters, some defining $\boldsymbol{G}$ and some defining $\boldsymbol{R}_i$.
- Linear mixed models are really just another type of covariance matrix, which can lead to some strange results.
- They are the model of choice when the data are unbalanced.