Variations of the Linear Mixed Model (with Applications in R)

Anthony-Alexander Christidis

University of British Columbia

anthony.christidis@stat.ubc.ca

March 14, 2017

Overview

Topics to be Covered

Random Effect Model

Anthony-Alexander Christidis

Outline of Presentation
Topics to be Covered

Random Intercept Models
Alternative Names
Random Intercept Model
R Code for Random Intercepts
Usage of Random Effects (Intercepts Specifically)

Random Slope Models
Visualization of Random Slope Models
Mathematical Formulation of Random Slope

- Brief Review: Random Intercept Model (with R)
- Random Slope Model (with R)
- Hierarchical Models (with R)
- Linear Mixed Models with Serial Correlation (with R)
- Questions!

Alternative Names

The Linear Mixed Model (LMM) is also often referred to as:

- Random effects models
- Latent variable models
- Multi-level models
- Hierarchical models

Random Intercept Model

The classical linear model is characterized by:

$$
\begin{aligned}
Y_i &\sim N(\mu_i, \sigma^2) \\
\mu_i &= X_i \beta
\end{aligned}
$$

The classical linear mixed model is characterized by:

$$
\begin{aligned}
Y_{ij} | U_i &\sim N(\mu_{ij}, \sigma^2) \\
\mu_{ij} &= X_i \beta + U_i \\
[U_1, ..., U_M]^T &\sim N(0, \Sigma)
\end{aligned}
$$

- **observations:** $Y_{ij}, j = 1 ... J_i$
  - repeated measures j on individual i
  - fixed effects X_{ij}
  - random effects: $U_i; i = 1, ..., M$

Situation where Random Intercept Model could prove useful

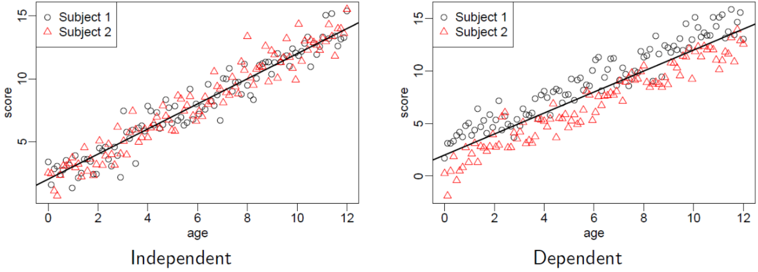


Figure: Notice how the two subjects differ throughout the measurements

Analysis of Pig Weight Data

TO R!

Usage of Random Effects (Intercepts Specifically)

There is such a thing called Shrinkage!

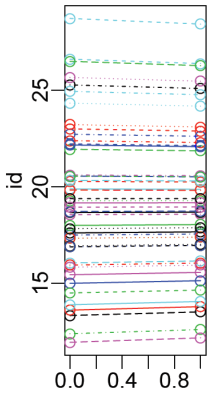


Figure: Fixed effects (left) and Random Effects (right) fitted values for the pigs weight data

Situation where Random Intercept Model could prove useful

- Notice that the random effects model shrinks values towards the mean!
- This is because the individuals random effects are drawn from a normal distribution with more mass near the mean!
- However, the random intercept model allows us to draw inference on the entire (unobserved) population of pigs regarding their weight!

Visualization of Random Slope Models

RANDOM SLOPE MODELS!

Visualization of Random Slope Models

What if instead of observing this (clear differences in the slope between subjects):

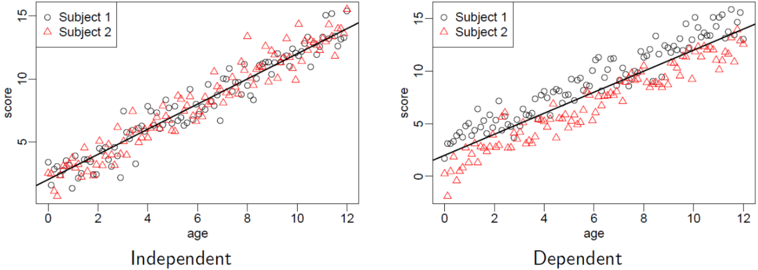


Figure: Notice how the two subjects differ throughout the measurements

Visualization of Random Slope Models

We observe this (clear differences in the slope between subjects):

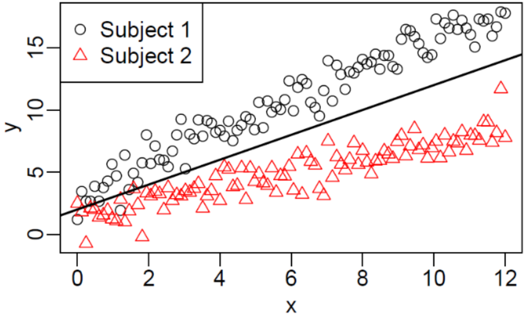


Figure: Notice how the slope of the subjects differ throughout the measurements

Mathematical Formulation of Random Slope Models

We may summarize a random slope model by the following equations:

$$
\begin{aligned}
Y_{ij}|U_i &\sim N(X_{ij} + U_{i1} + U_{i2}W_{ij}, \sigma^2) \\
\mu_{ij} &= X_i\beta + U_i \\
(U_{i1}, U_{i2}) &\sim MVN(0, \Sigma)
\end{aligned}
$$

Visualization of Random Slope Models

- X_{ijp} and W_{ij} are all covariates
- The coefficient on W_{ij} is different for each subject but come from a distribution
- U_{i1} are random intercepts
- U_{i2} are random slopes
- Note that they could be correlated
- Σ is a 2 by 2 covariance matrix

Analysis of Cow Harvest Data

TO R!

Analysis of Cow Harvest Data

$$E(Y_{ij}) = \begin{cases} \theta_j & \text{diet}_i = \text{barley} \\ \theta_j + \alpha_1 + \beta_1 t_j & \text{diet}_i = \text{lupins} \\ \theta_j + \alpha_2 + \beta_2 t_j & \text{diet}_i = \text{mixed} \end{cases}$$

Figure: Summary of R Random Slope Model

Visualization of Random Slope Models

HIERARCHICAL MODELS!

Mathematical Formulation of Hierarchical Models

Hierarchical (4-level) models may be summarized by the following. Note that real-life examples will follow!

$$
\begin{aligned}
Y_{ijklm} &\sim N(X_{ijklm} + B_i + C_{ij} + D_{ijk} + E_{ijkl}, \sigma^2) \\
B_i &\sim N(0, \sigma_B^2) \\
C_{ij} &\sim N(0, \sigma_C^2) \\
D_{ijk} &\sim N(0, \sigma_D^2) \\
E_{ijkl} &\sim N(0, \sigma_E^2)
\end{aligned}
$$

Mathematical Formulation of Hierarchical Models

- Suppose we have multiple observations per student, and multiple students per school
- Or multiple samples per animal and multiple animals per farm
- Or for each farm i we have several animals j, each animal was measured at more than one time k, multiple samples taken at each time l, and more than one plate m made of each sample in the laboratory.

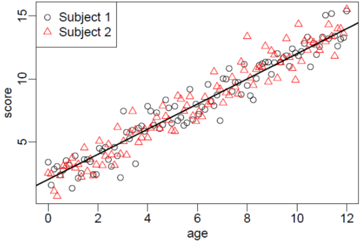
Analysis of US High School Data

TO R!

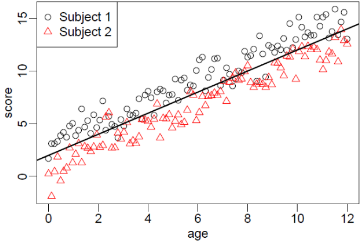
Linear Mixed Models with Serial Correlation!

Linear Mixed Models with Serial Correlation

What if instead of observing this (clear differences in the intercept between subjects):



Independent



Dependent

Figure: Notice how the two subjects differ throughout the measurements

Linear Mixed Models with Serial Correlation

Or this (clear differences in the slope between subjects):

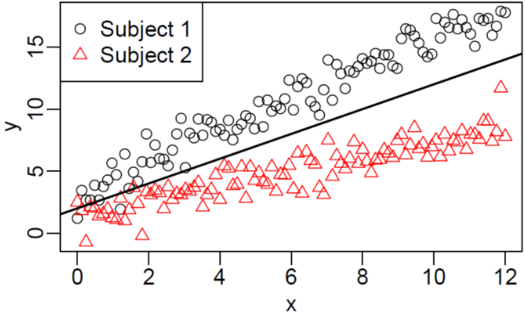


Figure: Notice how the slope of the subjects differ throughout the measurements

Linear Mixed Models with Serial Correlation

We observe this (clear differences in the slope between subjects):

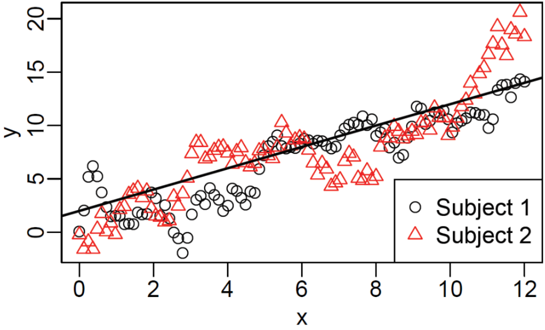


Figure: Notice serial correlation within the subjects throughout the measurements

Linear Mixed Models with Serial Correlation

- The models shown so far didnt make use of the time observations were made
- Two observations from the same individual were correlated, but the correlation was the same regardless of how far apart the observations were in time.
- We might expect observations made 1 day apart to be more similar than those made 10 days apart.
- In other words $\text{corr}(Y_{ij}, Y_{ik}) = \rho(t_{ij} t_{ik})$.

Mathematical Formulation of Linear Mixed Models with Temporal Correlation

Random effects model with temporal correlation:

$$Y_{ij}|\mathbf{U},\mathbf{V} \overset{\text{iid}}{\sim} \mathrm{N}\left[X_{ij}\beta + U_i + V_i(t_{ij}), \tau^2\right]$$

$$U_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_U^2)$$

$$\mathrm{cov}[V_i(t+h), V_i(t)] = \sigma_V^2 \exp(-|h|/\phi))$$

$$\mathrm{cov}[V_i(t+h), V_j(t)] = 0 \text{ if } i \neq j$$

- U_i is the individual level random effect
- $V_i(t)$ is the time trend
- Different for each individual
- But with the same correlation structure
- τ^2 is uncorrelated randomness (i.e. observation errors)

Analysis of pigs weight data

TO R!

R Code for Presentation

If you would like to use the code I used for this presentation, shoot me an e-mail @ anthony.christidis@stat.ubc.ca!

THANK YOU!

QUESTIONS?! ASK AWAY!