Introduction to Longitudinal Data Analysis

Christopher David Desjardins
Chu-Ting Chung
Jeffrey D. Long

11 January 2010

Outline

What is Longitudinal Data?

- The defining characteristic of longitudinal studies is that individuals are measured repeatedly through time.
  - In contrast, cross-sectional studies involve collecting measurements on an individual only once.
- Concerned with mean change over time.
- Longitudinal studies allow examination of age and cohort effects.
  - Longitudinal studies can distinguish changes over time within individuals (aging effects) from differences among people in their baseline levels (cohort effects).

What is Longitudinal Data?

- The defining characteristic of longitudinal studies is that individuals are measured repeatedly through time.
  - In contrast, cross-sectional studies involve collecting measurements on an individual only once.
- Concerned with mean change over time.
- Longitudinal studies allow examination of age and cohort effects.
  - Longitudinal studies can distinguish changes over time within individuals (aging effects) from differences among people in their baseline levels (cohort effects).

What is Longitudinal Data?

- The defining characteristic of longitudinal studies is that individuals are measured repeatedly through time.
  - In contrast, cross-sectional studies involve collecting measurements on an individual only once.
- Concerned with mean change over time.
- Longitudinal studies allow examination of age and cohort effects.
  - Longitudinal studies can distinguish changes over time within individuals (aging effects) from differences among people in their baseline levels (cohort effects).

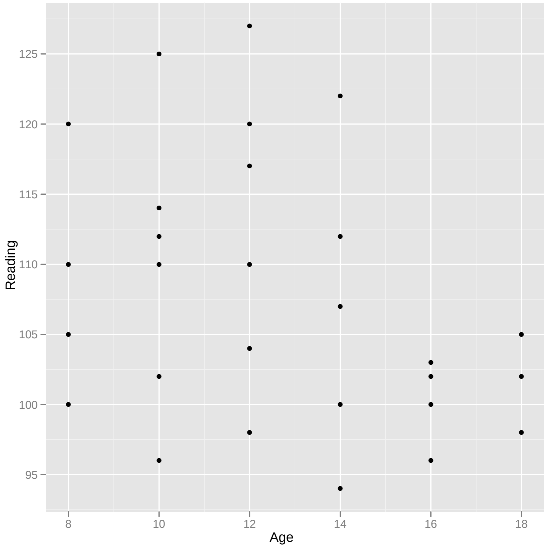
What is Longitudinal Data?

- The defining characteristic of longitudinal studies is that individuals are measured repeatedly through time.
  - In contrast, cross-sectional studies involve collecting measurements on an individual only once.
- Concerned with mean change over time.
- Longitudinal studies allow examination of age and cohort effects.
  - Longitudinal studies can distinguish changes over time within individuals (aging effects) from differences among people in their baseline levels (cohort effects).

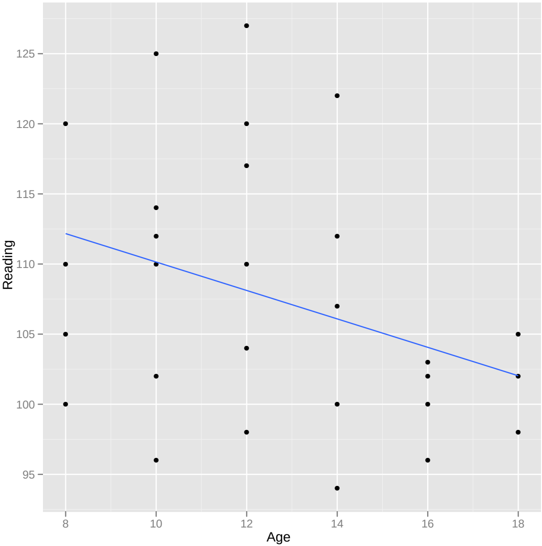
What is Longitudinal Data?

- The defining characteristic of longitudinal studies is that individuals are measured repeatedly through time.
  - In contrast, cross-sectional studies involve collecting measurements on an individual only once.
- Concerned with mean change over time.
- Longitudinal studies allow examination of age and cohort effects.
  - Longitudinal studies can distinguish changes over time within individuals (aging effects) from differences among people in their baseline levels (cohort effects).

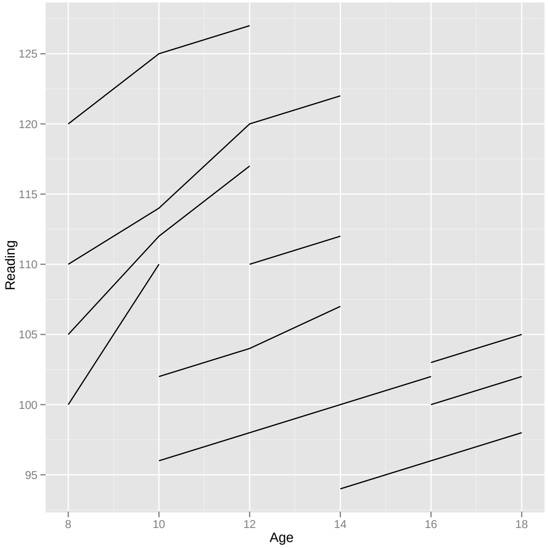
Age or cohort effects?



Age or cohort effects?



Age or cohort effects?



Benefits of Linear Mixed Models

- Fit intercepts and slopes for individuals
- Accounts for correlations between time points
- Accommodates missing data
- Accommodates measurements at unequal intervals
- Relatively few parameters required to account for dependency due to repeated measures
- Can handle both continuous and categorical covariates

Benefits of Linear Mixed Models

- Fit intercepts and slopes for individuals
- Accounts for correlations between time points
- Accommodates missing data
- Accommodates measurements at unequal intervals
- Relatively few parameters required to account for dependency due to repeated measures
- Can handle both continuous and categorical covariates

Benefits of Linear Mixed Models

- Fit intercepts and slopes for individuals
- Accounts for correlations between time points
- Accommodates missing data
- Accommodates measurements at unequal intervals
- Relatively few parameters required to account for dependency due to repeated measures
- Can handle both continuous and categorical covariates

Benefits of Linear Mixed Models

- Fit intercepts and slopes for individuals
- Accounts for correlations between time points
- Accommodates missing data
- Accommodates measurements at unequal intervals
- Relatively few parameters required to account for dependency due to repeated measures
- Can handle both continuous and categorical covariates

Benefits of Linear Mixed Models

- Fit intercepts and slopes for individuals
- Accounts for correlations between time points
- Accommodates missing data
- Accommodates measurements at unequal intervals
- Relatively few parameters required to account for dependency due to repeated measures
- Can handle both continuous and categorical covariates

Benefits of Linear Mixed Models

- Fit intercepts and slopes for individuals
- Accounts for correlations between time points
- Accommodates missing data
- Accommodates measurements at unequal intervals
- Relatively few parameters required to account for dependency due to repeated measures
- Can handle both continuous and categorical covariates

Correlations between time points

- Traditional regression does not account for this correlation
- This results in bogus standard error estimates and p-values.
- This is done in the covariance-variance matrix among the repeated measures known as $\mathbf{\Sigma_i}$
- Recall a correlation matrix is a standardized covariance-variance matrix.
- This matrix is involved directly in estimating the parameters unlike in traditional regression.
- This matrix is broken down further and includes the random effects matrix.

Correlations between time points

- Traditional regression does not account for this correlation
- This results in bogus standard error estimates and p-values.
- This is done in the covariance-variance matrix among the repeated measures known as $\mathbf{\Sigma_i}$
- Recall a correlation matrix is a standardized covariance-variance matrix.
- This matrix is involved directly in estimating the parameters unlike in traditional regression.
- This matrix is broken down further and includes the random effects matrix.

Correlations between time points

- Traditional regression does not account for this correlation
- This results in bogus standard error estimates and *p*-values.
- This is done in the covariance-variance matrix among the repeated measures known as $\mathbf{\Sigma_i}$
- Recall a correlation matrix is a standardized covariance-variance matrix.
- This matrix is involved directly in estimating the parameters unlike in traditional regression.
- This matrix is broken down further and includes the random effects matrix.

Correlations between time points

- Traditional regression does not account for this correlation
- This results in bogus standard error estimates and *p*-values.
- This is done in the covariance-variance matrix among the repeated measures known as $\mathbf{\Sigma_i}$
- Recall a correlation matrix is a standardized covariance-variance matrix.
- This matrix is involved directly in estimating the parameters unlike in traditional regression.
- This matrix is broken down further and includes the random effects matrix.

Correlations between time points

- Traditional regression does not account for this correlation
- This results in bogus standard error estimates and *p*-values.
- This is done in the covariance-variance matrix among the repeated measures known as $\mathbf{\Sigma_i}$
- Recall a correlation matrix is a standardized covariance-variance matrix.
- This matrix is involved directly in estimating the parameters unlike in traditional regression.
- This matrix is broken down further and includes the random effects matrix.

Correlations between time points

- Traditional regression does not account for this correlation
- This results in bogus standard error estimates and *p*-values.
- This is done in the covariance-variance matrix among the repeated measures known as $\mathbf{\Sigma_i}$
- Recall a correlation matrix is a standardized covariance-variance matrix.
- This matrix is involved directly in estimating the parameters unlike in traditional regression.
- This matrix is broken down further and includes the random effects matrix.

Accomodates missing data

- Missing data handled through robust maximum likelihood estimation
- No need for multiple imputation when using continuous variables
- Two types of Maximum Likelihood Estimates
  - Full Maximum Likelihood
  - Restricted Maximum Likeihood - provides unbaised variance estimates

Accomodates missing data

- Missing data handled through robust maximum likelihood estimation
- No need for multiple imputation when using continuous variables
- Two types of Maximum Likelihood Estimates
  - Full Maximum Likelihood
  - Restricted Maximum Likeihood - provides unbaised variance estimates

Accomodates missing data

- Missing data handled through robust maximum likelihood estimation
- No need for multiple imputation when using continuous variables
- Two types of Maximum Likelihood Estimates
  - Full Maximum Likelihood
  - Restricted Maximum Likeihood - provides unbaised variance estimates

Accomodates missing data

- Missing data handled through robust maximum likelihood estimation
- No need for multiple imputation when using continuous variables
- Two types of Maximum Likelihood Estimates
    - Full Maximum Likelihood
    - Restricted Maximum Likeihood - provides unbaised variance estimates

Accomodates missing data

- Missing data handled through robust maximum likelihood estimation
- No need for multiple imputation when using continuous variables
- Two types of Maximum Likelihood Estimates
  - Full Maximum Likelihood
  - Restricted Maximum Likeihood - provides unbaised variance estimates

Other benefits

- Individuals can be measured at different times and do not need to measured at all times.
  - Therefore, we do not need to throw out data or impute data on individuals.
  - We throw out cases not individuals.
- Fewer parameters used in these models than ANOVA and MANOVA models. This leads to higher statistical power.
- Can handle continuous and categorical covariates simulatenously. No need for a seperate model.
  - Categorical covariates are dummy coded.

Other benefits

- Individuals can be measured at different times and do not need to measured at all times.
  - Therefore, we do not need to throw out data or impute data on individuals.
  - We throw out cases not individuals.
- Fewer parameters used in these models than ANOVA and MANOVA models. This leads to higher statistical power.
- Can handle continuous and categorical covariates simulatenously. No need for a seperate model.
  - Categorical covariates are dummy coded.

Other benefits

- Individuals can be measured at different times and do not need to measured at all times.
    - Therefore, we do not need to throw out data or impute data on individuals.
    - We throw out cases not individuals.
- Fewer parameters used in these models than ANOVA and MANOVA models. This leads to higher statistical power.
- Can handle continuous and categorical covariates simulatenously. No need for a seperate model.
    - Categorical covariates are dummy coded.

Other benefits

- Individuals can be measured at different times and do not need to measured at all times.
  - Therefore, we do not need to throw out data or impute data on individuals.
  - We throw out cases not individuals.
- Fewer parameters used in these models than ANOVA and MANOVA models. This leads to higher statistical power.
- Can handle continuous and categorical covariates simulatenously. No need for a seperate model.
  - Categorical covariates are dummy coded.

Other benefits

- Individuals can be measured at different times and do not need to measured at all times.
  - Therefore, we do not need to throw out data or impute data on individuals.
  - We throw out cases not individuals.
- Fewer parameters used in these models than ANOVA and MANOVA models. This leads to higher statistical power.
- Can handle continuous and categorical covariates simulatenously. No need for a seperate model.
  - Categorical covariates are dummy coded.

Other benefits

- Individuals can be measured at different times and do not need to measured at all times.
    - Therefore, we do not need to throw out data or impute data on individuals.
    - We throw out cases not individuals.
- Fewer parameters used in these models than ANOVA and MANOVA models. This leads to higher statistical power.
- Can handle continuous and categorical covariates simulatenously. No need for a seperate model.
    - Categorical covariates are dummy coded.

Other names for these models

- Hierarchical linear models (HLM)
  - HLM are generally cross-sectional and involve multiple levels, such as school effects
  - In longitudinal models, measurements over time are nested within individuals rather than a school
- Linear Mixed Models
- Multilevel Models
- Mixed Effects Models

Other names for these models

- Hierarchical linear models (HLM)
  - HLM are generally cross-sectional and involve multiple levels, such as school effects
  - In longitudinal models, measurements over time are nested within individuals rather than a school
- Linear Mixed Models
- Multilevel Models
- Mixed Effects Models

Other names for these models

- Hierarchical linear models (HLM)
  - HLM are generally cross-sectional and involve multiple levels, such as school effects
  - In longitudinal models, measurements over time are nested within individuals rather than a school
- Linear Mixed Models
- Multilevel Models
- Mixed Effects Models

Other names for these models

- Hierarchical linear models (HLM)
  - HLM are generally cross-sectional and involve multiple levels, such as school effects
  - In longitudinal models, measurements over time are nested within individuals rather than a school
- Linear Mixed Models
- Multilevel Models
- Mixed Effects Models

Other names for these models

- Hierarchical linear models (HLM)
  - HLM are generally cross-sectional and involve multiple levels, such as school effects
  - In longitudinal models, measurements over time are nested within individuals rather than a school
- Linear Mixed Models
- Multilevel Models
- Mixed Effects Models

Other names for these models

- Hierarchical linear models (HLM)
  - HLM are generally cross-sectional and involve multiple levels, such as school effects
  - In longitudinal models, measurements over time are nested within individuals rather than a school
- Linear Mixed Models
- Multilevel Models
- Mixed Effects Models

LMM formula

Recall a simple multiple regression model is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi} + e_i$$

In a LMM, a regression model measured j times might become:
$Y_{ij} = \beta_0 + b_{0i} + \beta_1 X_{1i} + (\beta_2 + b_{2i}) t_{ij} + e_{ij}$

LMM formula explained

Nomenclature	Definition
Y_{ij}	refers to the outcome variable, e.g. infant's weight, for individual i at time j
β_0	refers to the intercept
b_{0i}	refers to the random intercept for individual i. This allows individuals to vary in their initial values.
$\beta_1 X_{1i}$	the parameter estimate associated with a covariate, such as primary feeding type.
β_2	the fixed effect estimate associated with time
b_{2i}	refers to the random slope for individual i.This allows individuals to vary in their slopes.
t_{ij}	time variable, months since birth.
e_{ij}	is the residual error term.

Assumptions of LMMs

- Obseravations of individuals are independent
- Random effects are normally distributed
- Random residuals are normally distributed

Assumptions of LMMs

- Obseravations of individuals are independent
- Random effects are normally distributed
- Random residuals are normally distributed

Assumptions of LMMs

- Obseravations of individuals are independent
- Random effects are normally distributed
- Random residuals are normally distributed

Running LMMs

$$Y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})t_{ij} + e_{ij}$$

- Use the {lme4} package in **R** and the *lmer* function.
- If you can write your model as a LMM then you can program it in **R**.
- Basic **R** Formula: $m1 < -lmer(Outcome \sim 1 + Time + (1 + month|ID), data = data))$

Running LMMs

$$Y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})t_{ij} + e_{ij}$$

- Use the {lme4} package in **R** and the *lmer* function.
- If you can write your model as a LMM then you can program it in **R**.
- Basic **R** Formula: $m1 < -lmer(Outcome \sim 1 + Time + (1 + month|ID), data = data))$

Running LMMs

$$Y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})t_{ij} + e_{ij}$$

- Use the {lme4} package in **R** and the *lmer* function.
- If you can write your model as a LMM then you can program it in **R**.
- Basic **R** Formula:
  $m1 < -lmer(Outcome \sim 1 + Time + (1 + month|ID), data = data))$

Running LMMs

```
> library(lme4)
> data(sleepstudy)
> m1 <- lmer(Reaction ~ 1 + Days + (1 + Days|Subject), sleepstudy)
> summary(m1)

Linear mixed model fit by REML
Formula: Reaction ~ 1 + Days + (1 + Days | Subject)
   Data: sleepstudy
  AIC  BIC logLik deviance REMLdev
 1756 1775 -871.8     1752    1744
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept) 612.090  24.7405
          Days         35.072   5.9221  0.066
 Residual             654.941  25.5918
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
```

Getting Help with LMMs and R

- Help with LMMs
  - Fitzmaurice, G. M., Laird, N.M., & Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley-Interscience.
  - Diggle, P., Heagerty, P, Liang, K-Y, & Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
  - Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Help with **R** and {lme4}.
  - General **R** help. `www.r-project.org`
  - Nice intro to **R**. `http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf`
  - {lme4} help. Pinheiro, J. C. & Bates, D. (2002). *Mixed Effects Models in S and S-Plus*. Springer Press. CAN DOWNLOAD FROM LIBRARY!

Getting Help with LMMs and R

- Help with LMMs
  - Fitzmaurice, G. M., Laird, N.M., & Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley-Interscience.
  - Diggle, P., Heagerty, P, Liang, K-Y, & Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
  - Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Help with **R** and {lme4}.
  - General **R** help. `www.r-project.org`
  - Nice intro to **R**. `http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf`
  - {lme4} help. Pinheiro, J. C. & Bates, D. (2002). *Mixed Effects Models in S and S-Plus*. Springer Press. CAN DOWNLOAD FROM LIBRARY!

Getting Help with LMMs and R

- Help with LMMs
  - Fitzmaurice, G. M., Laird, N.M., & Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley-Interscience.
  - Diggle, P., Heagerty, P, Liang, K-Y, & Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
  - Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Help with **R** and {lme4}.
  - General **R** help. `www.r-project.org`
  - Nice intro to **R**. `http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf`
  - {lme4} help. Pinheiro, J. C. & Bates, D. (2002). *Mixed Effects Models in S and S-Plus*. Springer Press. CAN DOWNLOAD FROM LIBRARY!

Getting Help with LMMs and R

- Help with LMMs
  - Fitzmaurice, G. M., Laird, N.M., & Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley-Interscience.
  - Diggle, P., Heagerty, P, Liang, K-Y, & Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
  - Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Help with **R** and {lme4}.
  - General **R** help. `www.r-project.org`
  - Nice intro to **R**. `http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf`
  - {lme4} help. Pinheiro, J. C. & Bates, D. (2002). *Mixed Effects Models in S and S-Plus*. Springer Press. CAN DOWNLOAD FROM LIBRARY!

Getting Help with LMMs and R

- Help with LMMs
  - Fitzmaurice, G. M., Laird, N.M., & Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley-Interscience.
  - Diggle, P., Heagerty, P, Liang, K-Y, & Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
  - Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Help with **R** and {lme4}.
  - General **R** help. `www.r-project.org`
  - Nice intro to **R**. `http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf`
  - {lme4} help. Pinheiro, J. C. & Bates, D. (2002). *Mixed Effects Models in S and S-Plus*. Springer Press. CAN DOWNLOAD FROM LIBRARY!

Getting Help with LMMs and R

- Help with LMMs
  - Fitzmaurice, G. M., Laird, N.M., & Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley-Interscience.
  - Diggle, P., Heagerty, P, Liang, K-Y, & Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
  - Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Help with **R** and {lme4}.
  - General **R** help. www.r-project.org
  - Nice intro to **R**. http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf
  - {lme4} help. Pinheiro, J. C. & Bates, D. (2002). *Mixed Effects Models in S and S-Plus*. Springer Press. CAN DOWNLOAD FROM LIBRARY!

Getting Help with LMMs and R

- Help with LMMs
  - Fitzmaurice, G. M., Laird, N.M., & Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley-Interscience.
  - Diggle, P., Heagerty, P, Liang, K-Y, & Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
  - Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Help with **R** and {lme4}.
  - General **R** help. `www.r-project.org`
  - Nice intro to **R**. `http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf`
  - {lme4} help. Pinheiro, J. C. & Bates, D. (2002). *Mixed Effects Models in S and S-Plus*. Springer Press. CAN DOWNLOAD FROM LIBRARY!

Getting Help with LMMs and R

- Help with LMMs
  - Fitzmaurice, G. M., Laird, N.M., & Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley-Interscience.
  - Diggle, P., Heagerty, P, Liang, K-Y, & Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
  - Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Help with **R** and {lme4}.
  - General **R** help. `www.r-project.org`
  - Nice intro to **R**. `http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf`
  - {lme4} help. Pinheiro, J. C. & Bates, D. (2002). *Mixed Effects Models in S and S-Plus*. Springer Press. CAN DOWNLOAD FROM LIBRARY!