

DM16 SLIDES (Version 1.0)

1. Module Overview

1.1 Module Cover (START)

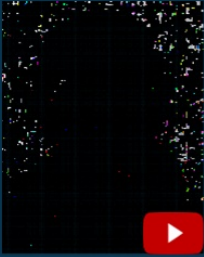


1.2 Instructors




1.3 Designers


Meet the Designers:



Xi Lu
Florida State
University




Jonathan Lehrfeld
ETS



André A. Rupp
Mindful
Measurement

1.4 Welcome



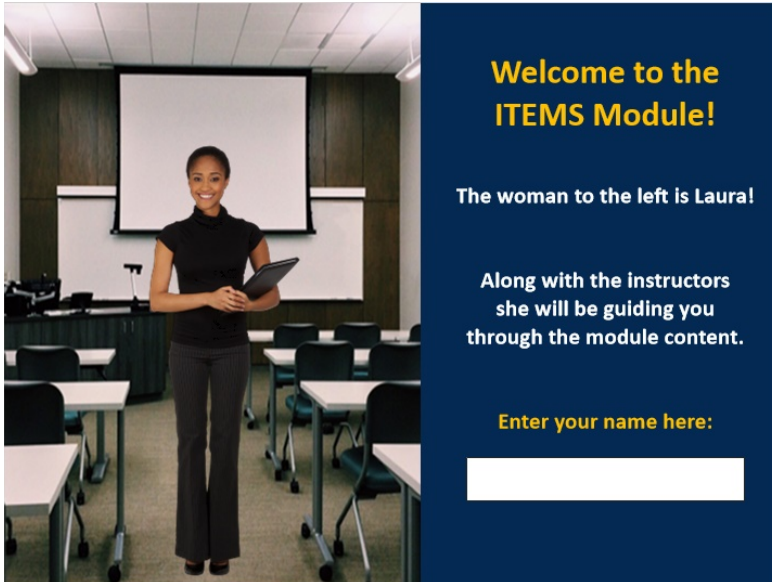
**Welcome to the
ITEMS Module!**

The woman to the left is Laura!

Along with the instructors
she will be guiding you
through the module content.

Enter your name here:

Untitled Layer 1 (Slide Layer)



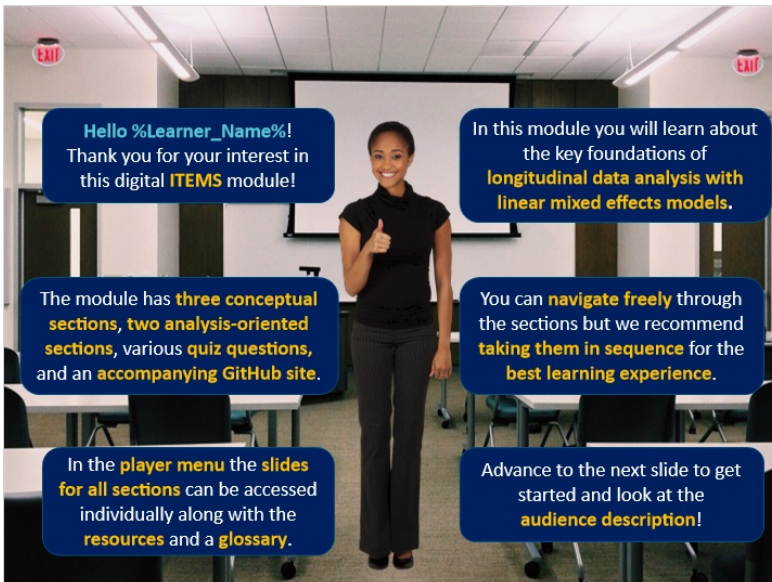
**Welcome to the
ITEMS Module!**

The woman to the left is Laura!

Along with the instructors
she will be guiding you
through the module content.

Enter your name here:

1.5 Overview



Hello %Learner_Name%!
Thank you for your interest in
this digital **ITEMS** module!

In this module you will learn about
the key foundations of
**longitudinal data analysis with
linear mixed effects models.**

The module has **three conceptual
sections, two analysis-oriented
sections**, various **quiz questions**,
and an **accompanying GitHub site.**

You can **navigate freely** through
the sections but we recommend
taking them in sequence for the
best learning experience.

In the **player menu** the **slides
for all sections** can be accessed
individually along with the
resources and a **glossary.**

Advance to the next slide to get
started and look at the
audience description!

1.6 Target Audience

Target Audience

Anyone who would like a gentle statistical introduction to this topic:

- graduate students and faculty in Master's, Ph.D., or certificate programs
- psychometricians and other measurement professionals
- data scientists / analysts
- research assistants or research scientists
- technical project directors
- assessment developers



However, we hope that you find the information in this module useful no matter what your official title or role in an organization is!

1.7 Expectations (I)



Let's discuss expectations....


1.8 Expectations (II)

ITEMS Modules in Context



1.9 Learning Objectives

Learning Objectives



1. Design research questions that can be answered with linear mixed effects models (LMEs)
2. Understand data collection protocols, missing data, and timing of measurements
3. Explore and summarize longitudinal data both numerically and graphically
4. Write out the key equations, assumptions, and components of LMEs
5. Run LME analyses using R to make appropriate output interpretations

1.10 Prerequisites

Prerequisites

- **Working knowledge of the following topics:**
 - ✓ Hypothesis testing and p -value interpretation
 - ✓ Principles of exploratory data analysis
 - ✓ Multiple linear regression models
 - ✓ General linear models / ANOVA models
- **Basic practical experience with:**
 - ✓ Summarizing data numerically and graphically
 - ✓ Working with *R* and associated *R* packages
 - ✓ Interpreting model output for decision-making

1.11 Resources


Resources

Module Citation

Harring, J. R., & Johnson, T. L. (2020). Longitudinal data analysis (Digital ITEMS Module 16). *Educational Measurement: Issues and Practice*, 39(4), XX-XX.

GitHub Tutorial

<https://tessaleejohnson.github.io/ITEMS-Longitudinal/>



Additional Resources



References (Slide Layer)

Resources



Back

1.12 Main Menu

**Main Menu**

01 Conceptual Foundations
[15 Minutes]

02 Design and Data Considerations
[15 Minutes]

03 Linear Mixed-effects Model (LME)
[15 Minutes]

04 LME Analyses
[25 Minutes]

05 Data Activity
[25 Minutes]

06 Quizzes
[10 Minutes]

Navigation

Navigation (Slide Layer)





2. Section 1: Conceptual Foundations


2.1 Cover: Section 1



2.2 Learning Objectives: Section 1





Learning Objectives




1. Identify the benefits of using mixed effects models to analyze longitudinal data
2. Propose research questions that can be answered using longitudinal data and the LME
3. Differentiate LMEs from other common modeling approaches and types of analyses
4. Understand the basic structure of the example NLSY data set

2.3 Importance of Longitudinal Research

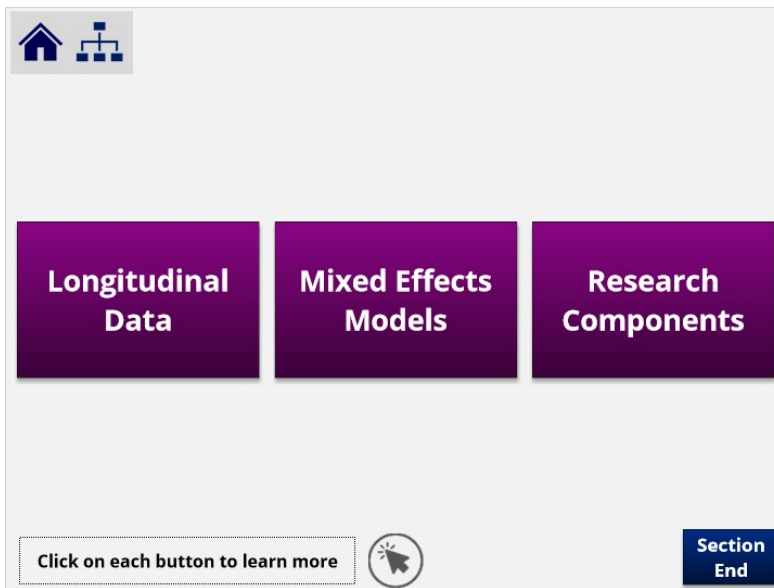


Importance of Longitudinal Research



- Recent article in a special issue in *Journal of Research on Educational Effectiveness* called for “High Quality Longitudinal Research in Education” (Watts, Bailey, & Li, 2019)
- Research with long-term follow-up helps educators **understand the impacts of interventions over time**
- Analyzing studies with long-term follow-up requires **knowledge of longitudinal methods**



2.4 Topic Selection



2.5 Bookmark: Longitudinal Data



2.6 What Are Longitudinal Data





What Are Longitudinal Data?

LONGITUDINAL DATA

- Individuals sampled from a larger population
- Same individuals measured on same variables over time
- At least 3 time points for most models

2.7 Why Collect Longitudinal Data



Why Collect Longitudinal Data?

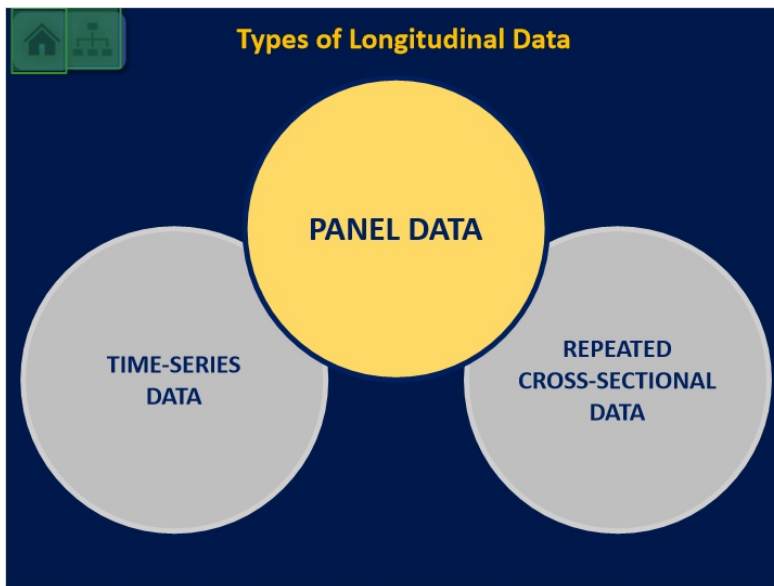
PROS OF LONGITUDINAL DATA

- Understand how phenomena change over time
- Assess individual change as well as population change

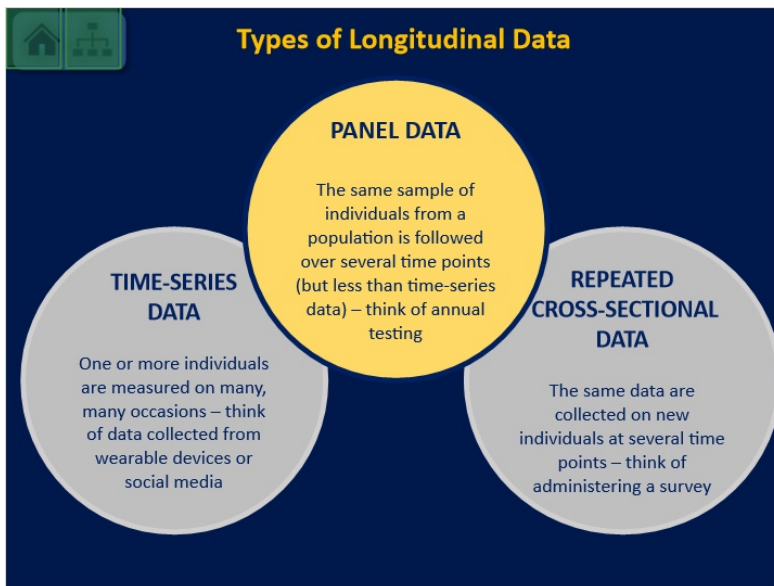
- Longitudinal data can help answer questions about **outcome trajectories** – key for research in **education, psychology, epidemiology, and more**

- Individual outcomes are assessed over time – allows for the **separation of effects** that **vary over time** and those that are **static or invariant**

2.8 Types of Longitudinal Data I



2.9 Types of Longitudinal Data II





2.10 Bookend: Longitudinal Data



2.11 Bookmark: Mixed Effects Models



2.12 What Are Mixed Effects Models





What Are Mixed Effects Models?

MIXED-EFFECTS MODELS

- Flexible statistical models separate individual effects from population effects
- Accommodate many of the messier aspects that accompany longitudinal data (e.g., “missingness”)

2.13 Why Use Mixed Effects Models



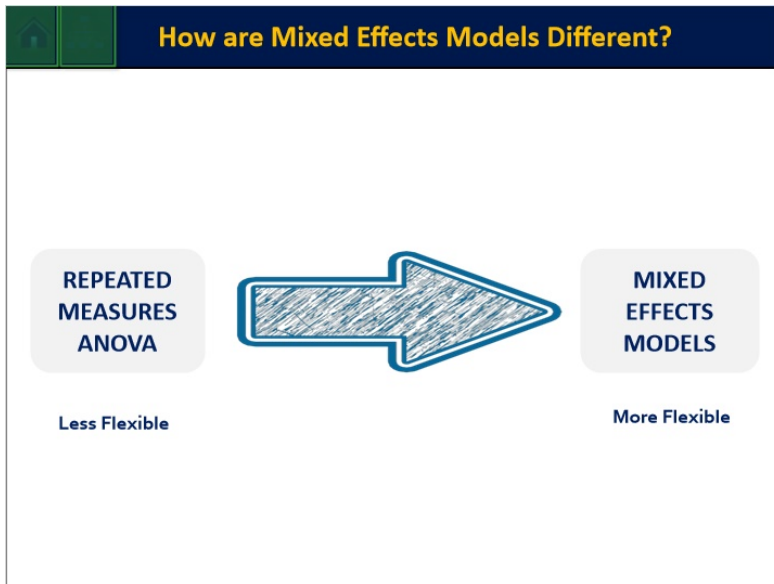
Why Use Mixed Effects Models?

- **Mixed effect models model both fixed (population) and random (individual) effects** using the familiar regression framework allowing researchers to answer questions about population change over time as well as individual change
- **Longitudinal mixed-effects models (LMEs) are flexible** – they handle missing data, unbalanced designs, and a plethora of mean and variance/covariance structures

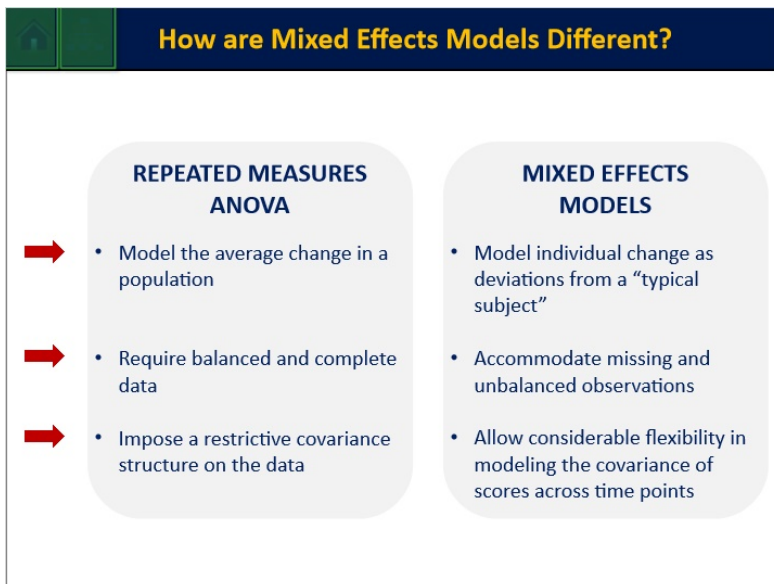
MODEL COMPONENTS

- ✓ **Means:** expected change, predictors
- ✓ **(Co)variances:** distribution of residuals across persons, times
- ✓ **Random effects:** individual change, deviations from mean

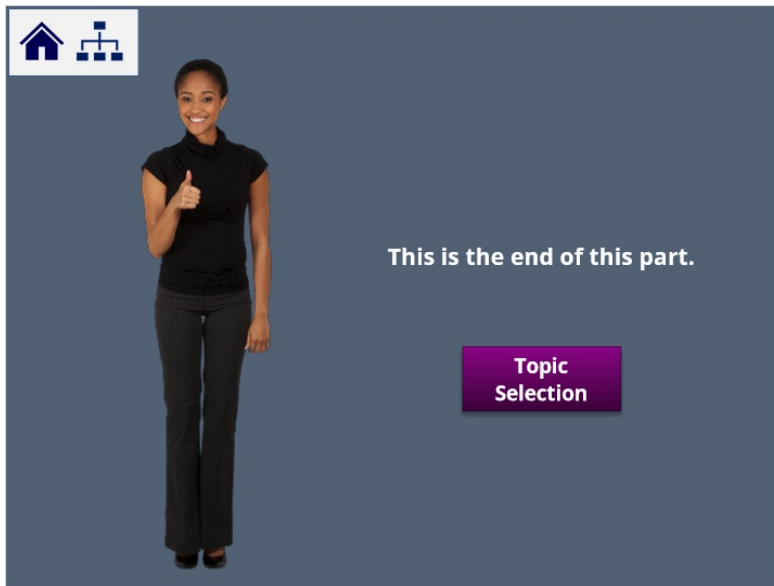
2.14 How Are Mixed Effects Models Different I



2.15 How Are Mixed Effects Models Different II



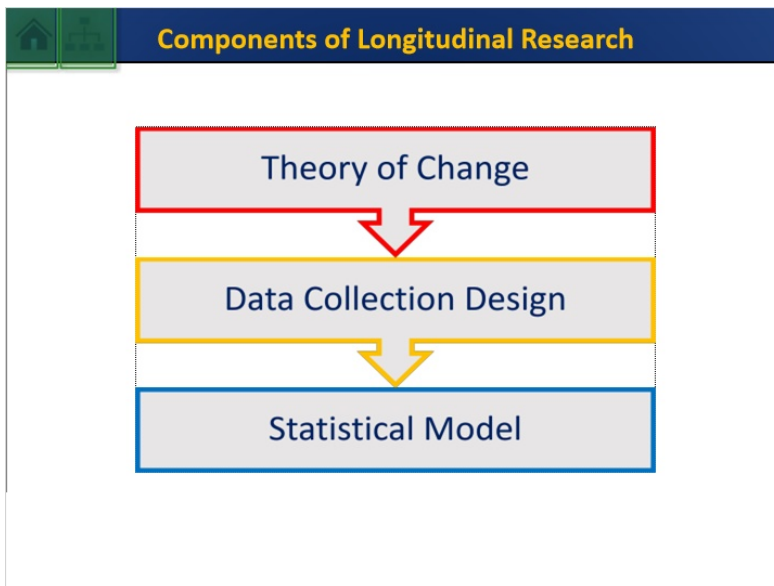
2.16 Bookend: Mixed Effects Models



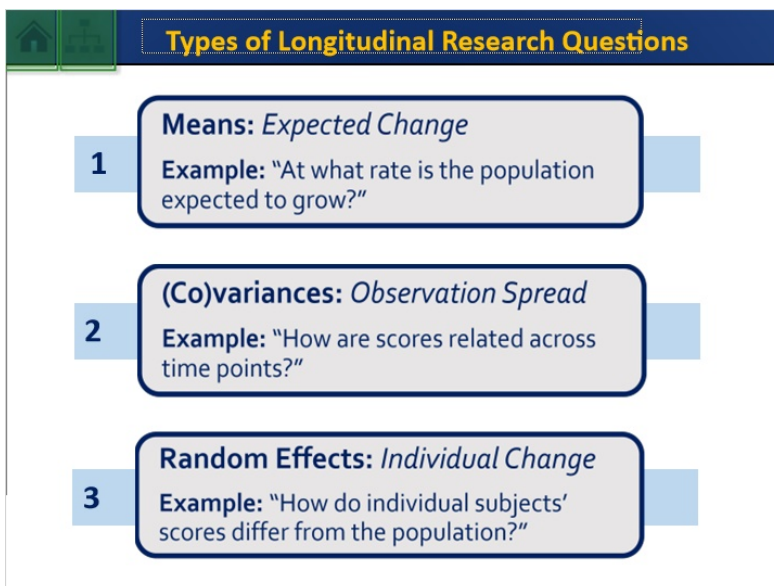
2.17 Bookmark: Research Components



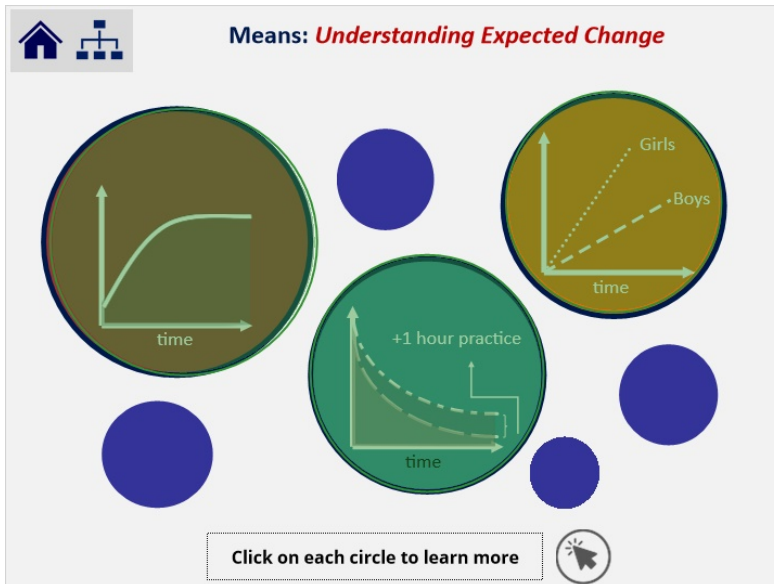
2.18 Components of Longitudinal Research



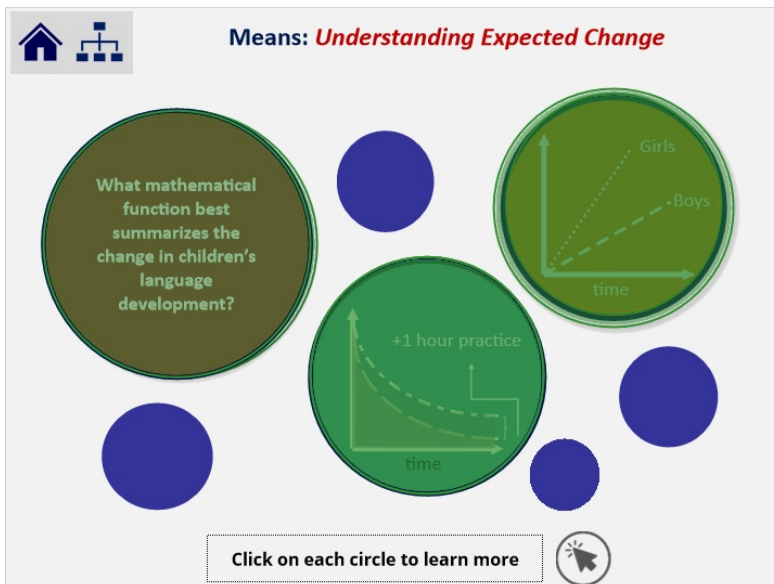
2.19 Types of Longitudinal Research Questions





2.20 Understanding Expected Change

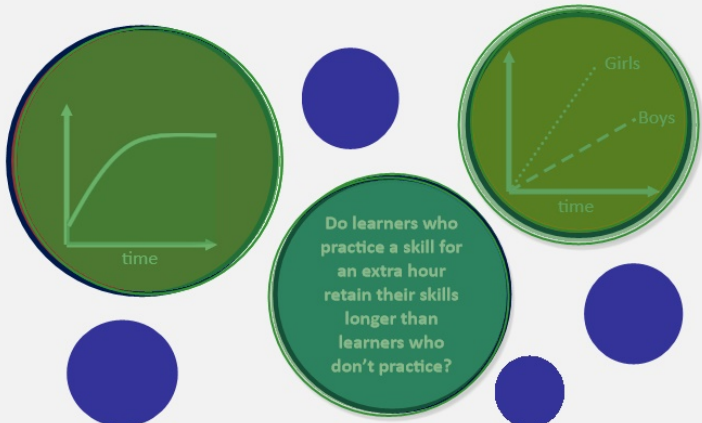


Q1 (Slide Layer)




Q2 (Slide Layer)



  Means: *Understanding Expected Change*




The slide features three large green circles, each containing a graph with 'time' on the x-axis. The left circle shows a solid curve that rises and then plateaus. The right circle shows two dashed lines, one labeled 'Girls' and one labeled 'Boys', both starting at the origin and increasing linearly, with the 'Girls' line having a steeper slope. The central circle contains the text: 'Do learners who practice a skill for an extra hour retain their skills longer than learners who don't practice?'. There are also five smaller blue circles scattered around the main content.


Click on each circle to learn more 

Q3 (Slide Layer)

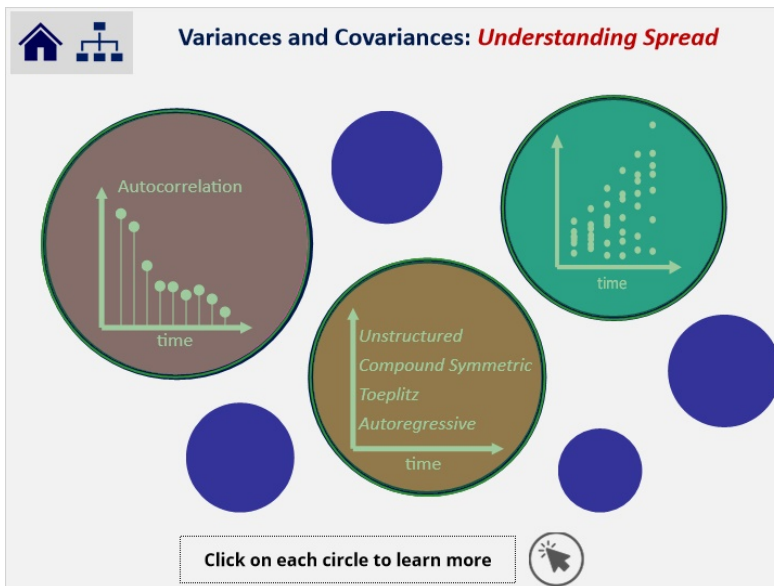
  Means: *Understanding Expected Change*



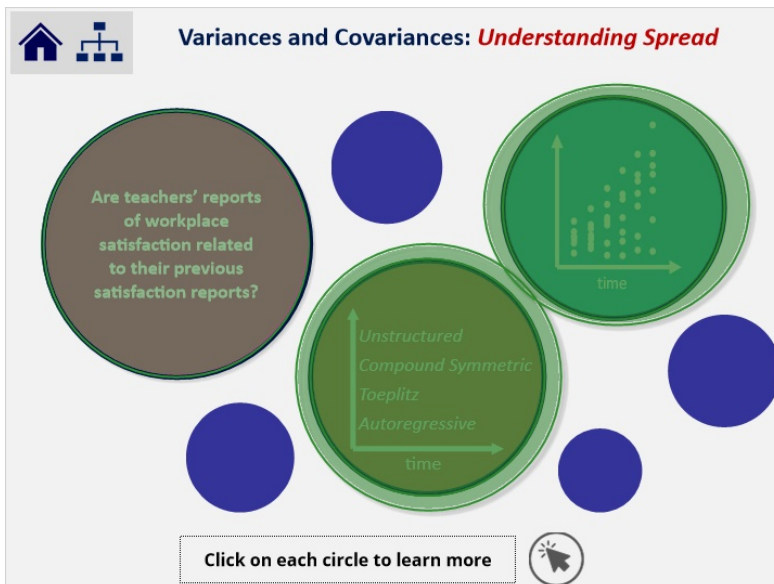
The slide features three large green circles, each containing a graph with 'time' on the x-axis. The left circle shows a solid curve that rises and then plateaus. The right circle shows a dashed curve that rises and then plateaus, with the text '+1 hour practice' above it. The central circle contains the text: 'Is the average growth rate the same for boys as it is for girls?'. There are also five smaller blue circles scattered around the main content.

Click on each circle to learn more 



2.21 Understanding Spread

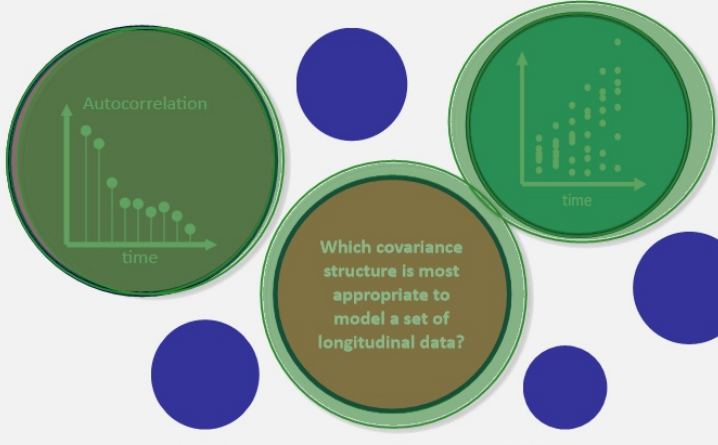


Q1 (Slide Layer)



Q2 (Slide Layer)

  Variances and Covariances: *Understanding Spread*




Autocorrelation



time

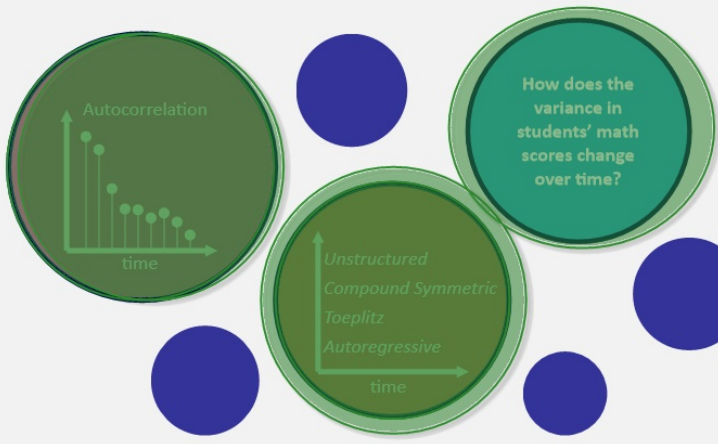
time

Which covariance structure is most appropriate to model a set of longitudinal data?

Click on each circle to learn more 

Q3 (Slide Layer)

  Variances and Covariances: *Understanding Spread*



Autocorrelation


time

time

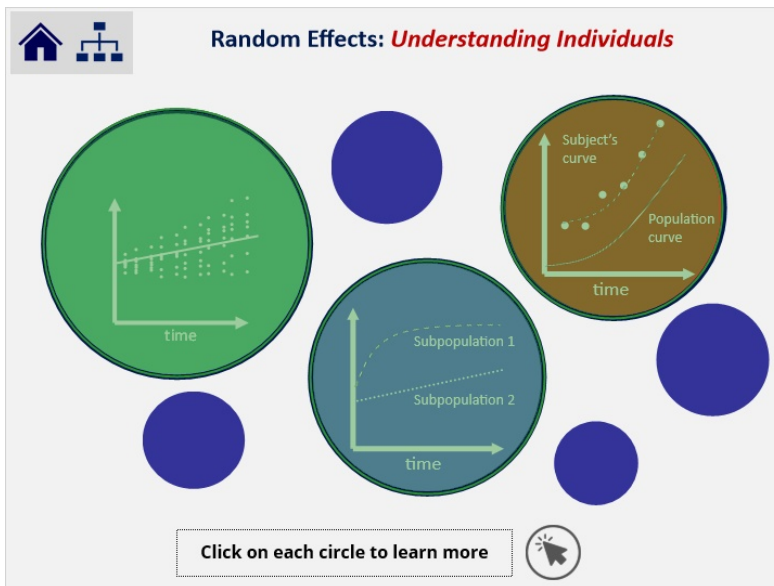
Unstructured
Compound Symmetric
Toeplitz
Autoregressive

time

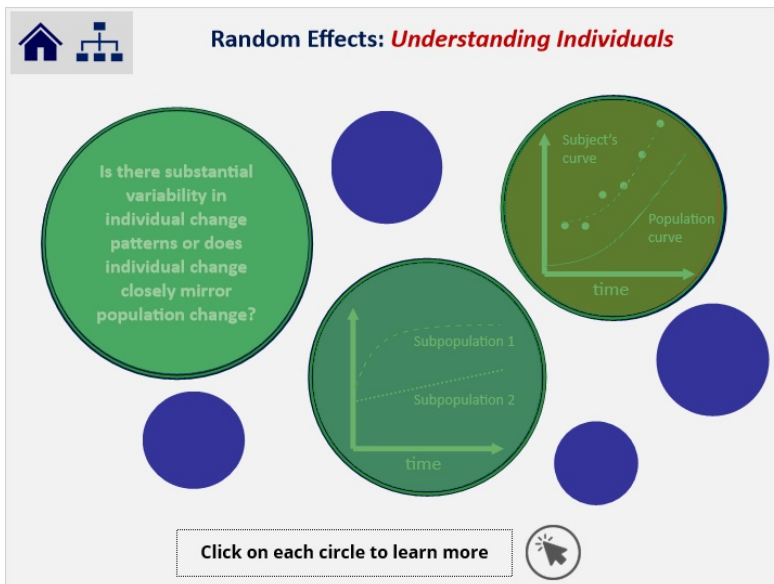
How does the variance in students' math scores change over time?

Click on each circle to learn more 



2.22 Understanding Individuals

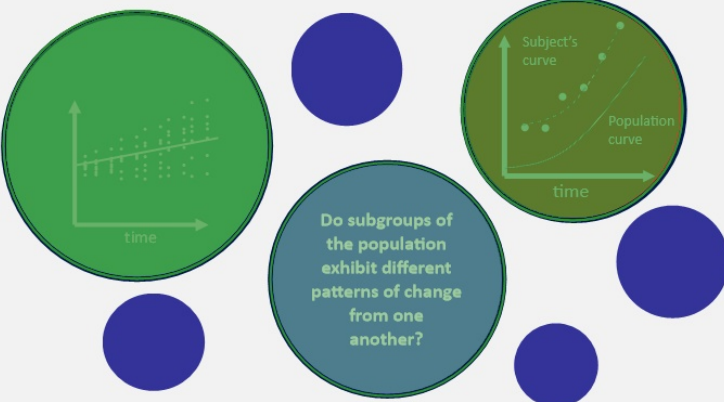


Q1 (Slide Layer)




Q2 (Slide Layer)



  Random Effects: *Understanding Individuals*

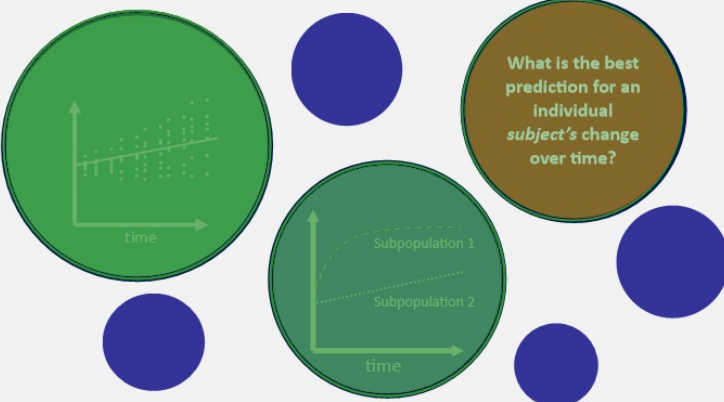


Do subgroups of the population exhibit different patterns of change from one another?


Click on each circle to learn more 

Q3 (Slide Layer)

  Random Effects: *Understanding Individuals*



What is the best prediction for an individual subject's change over time?

Click on each circle to learn more 

2.23 Bookend: Research Components



2.24 Example NLSY

National Longitudinal Survey of Youth

#A tibble: 405 x 17



ID	GEN	MOMYR	COG	EMO	AN1	AN2	AN3	AN4	R1
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1						1	0	27
2	1						0	1	31
3	0						5	3	36
4	1						-99	-99	18
5	1						3	1	23
6	0						0	0	21
7	0						0	10	21
8	0						0	4	13
9	0						2	0	29
10	0	28	9	11	2	3	6	5	45

... with 395 more rows

Outcome

The youths' antisocial behavior scores were recorded at four time points (missing = -99)



2.25 Summary




Summary

- Longitudinal data are made up of repeated observations on the *same individuals*
- Mixed-effects models are *flexible statistical tools* for analyzing longitudinal data
- Research questions may investigate the three components of the mixed effects model: *means, (co)variances, or random effects*
- Throughout this module, we will use the *NLSY dataset* to demonstrate longitudinal data analysis using the mixed-effects model

2.26 Bookend: Section 2





This is the end of the section.

Quiz

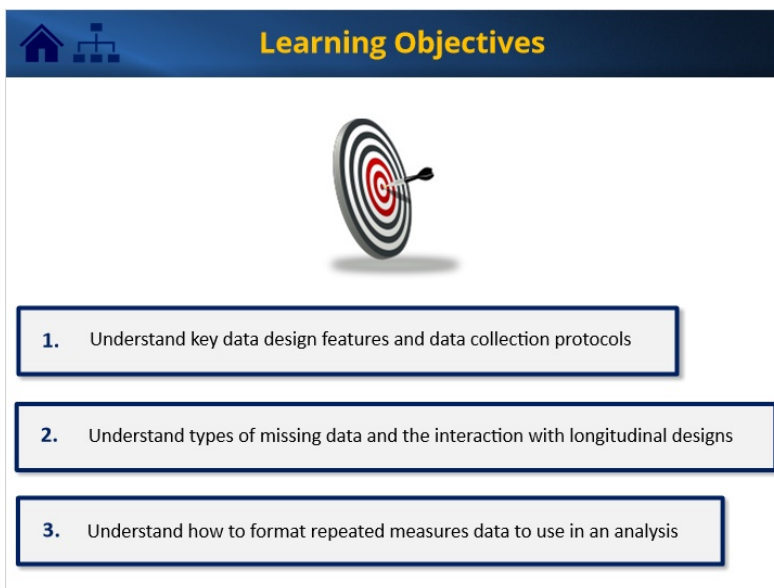
Main Menu

3. Section 2: Design and Data Considerations

3.1 Cover: Section 2





3.2 Learning Objectives: Section 2

The slide has a dark blue header with a small icon of a house with a tree on the left and the text "Learning Objectives" in yellow. Below the header is a large white area. In the center of this area is a 3D illustration of a target with a red bullseye and a black arrow hitting the center. Below the target are three white rectangular boxes with blue borders, each containing a numbered learning objective.

1. Understand key data design features and data collection protocols
2. Understand types of missing data and the interaction with longitudinal designs
3. Understand how to format repeated measures data to use in an analysis

3.3 NLSY Reading Recognition Skill

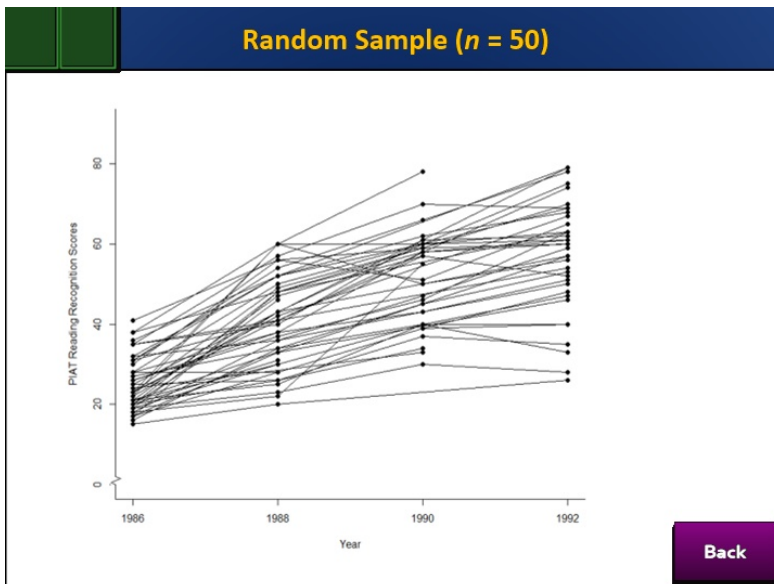


NLSY Reading Recognition Skill

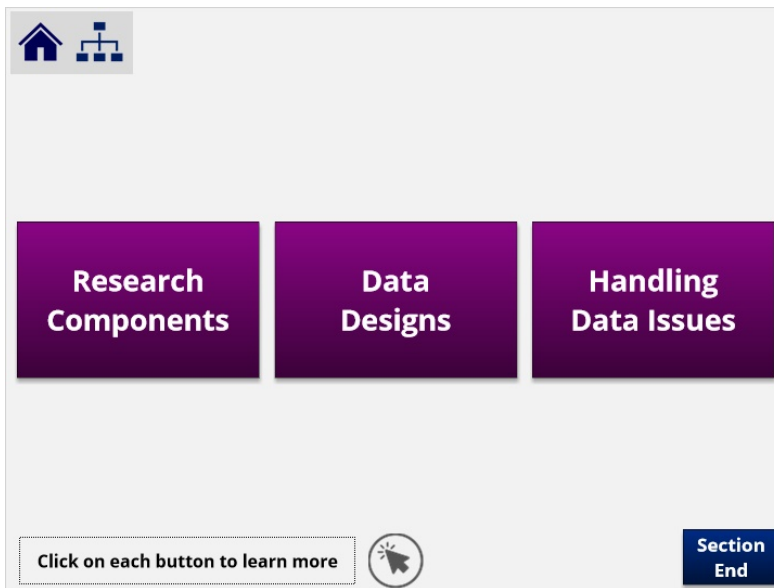
- Sample of $m = 405$ children from the *National Longitudinal Study of Youth* (1986 cohort) who were assessed at **4 time points** in **2-year increments**
- Repeated outcome measure is the **sum score** of correct responses to 84 items of *Peabody Individual Achievement Test* (PIAT) *Reading Recognition* subtest
- Click on the button to see a graph for a **random sample of $n = 50$** sum scores

Sample Data

Sample Data (Slide Layer)



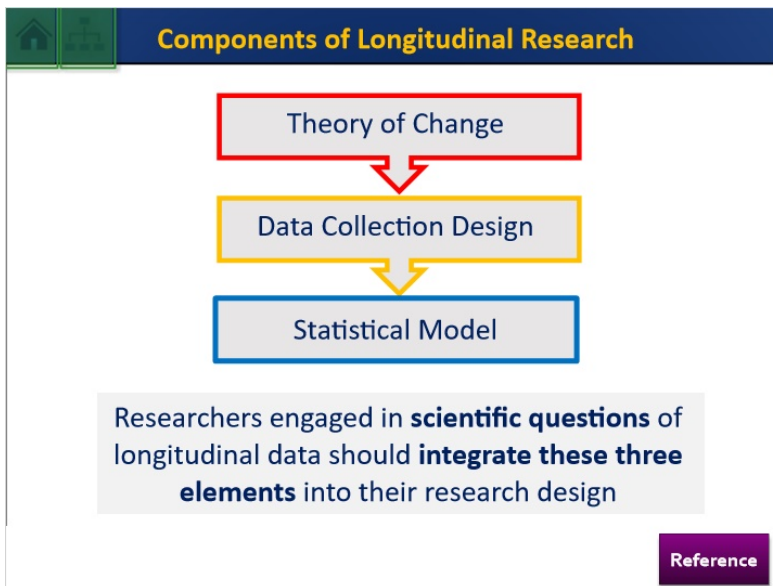
3.4 Topic Selection



3.5 Bookmark: Research Components



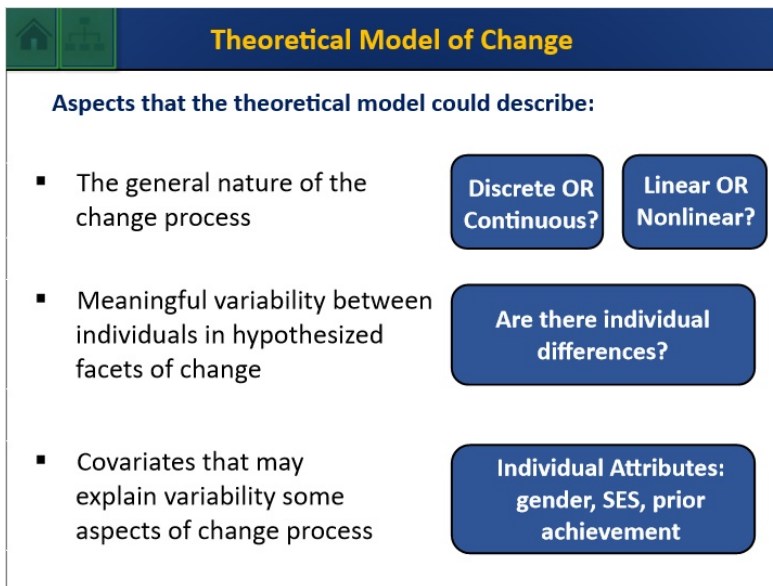
3.6 Components of Longitudinal Research



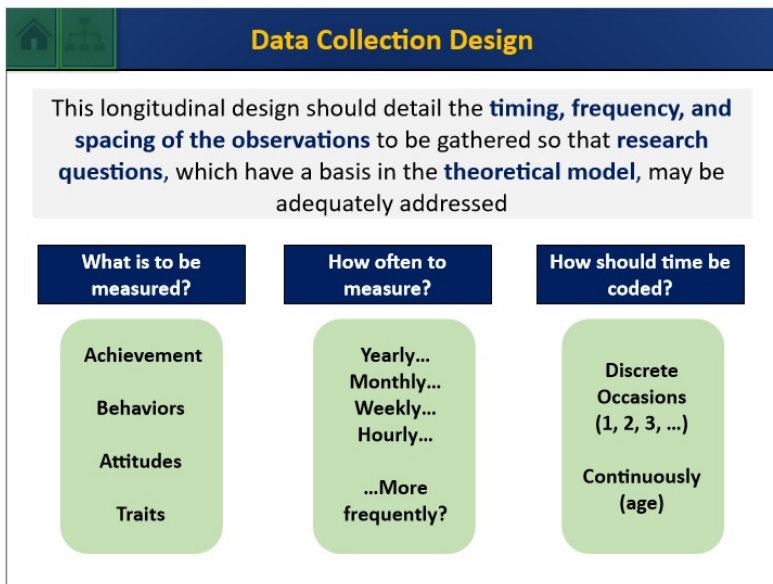
Reference (Slide Layer)

The screenshot shows the title page of a research article. The title is 'Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model'. The author is Linda M. Collins. The article is from the 'Annual Review of Psychology', Vol. 57, 509-528, published January 2006. The abstract states: 'This article argues that ideal longitudinal research is characterized by the seamless integration of three elements: (a) a well-articulated theoretical model of change observed using (b) a temporal design that affords a clear and detailed view of the process, with the resulting data analyzed by means of (c) a statistical model that is an operationalization of the theoretical model. Two general varieties of theoretical models are considered: models in which the time-related change of primary interest is continuous, and those in which it is characterized by movement between discrete states. In addition, two general types of temporal designs are considered: the longitudinal panel design and the intensive longitudinal design. For each general category of theoretical models, some of the analytic possibilities available for longitudinal panel designs and for intensive longitudinal designs are discussed. The article concludes with brief discussions of two issues particularly relevant to longitudinal research—missing data and measurement—and a few words about exploratory research.' The keywords are 'Intensive longitudinal', 'longitudinal panel', 'growth models', 'latent transition analysis'. A 'Click on the image to go to the publisher website' button and a 'Back' button are at the bottom.

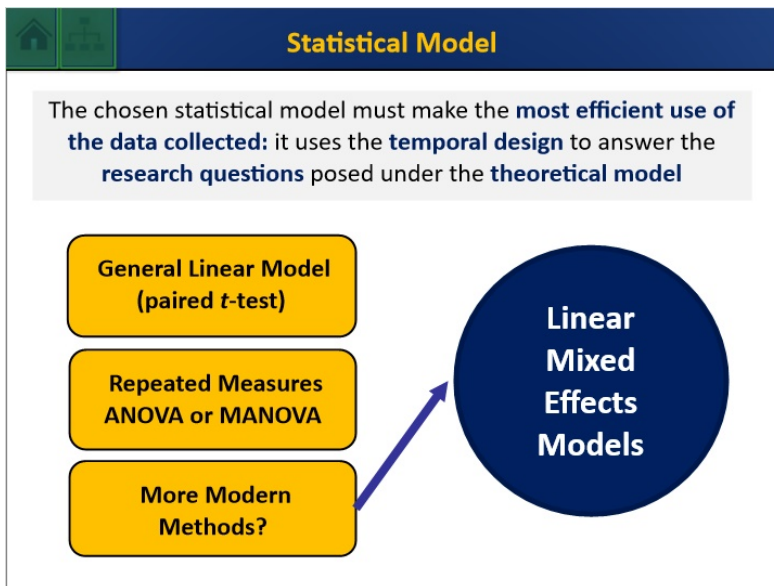
3.7 Theoretical Model of Change



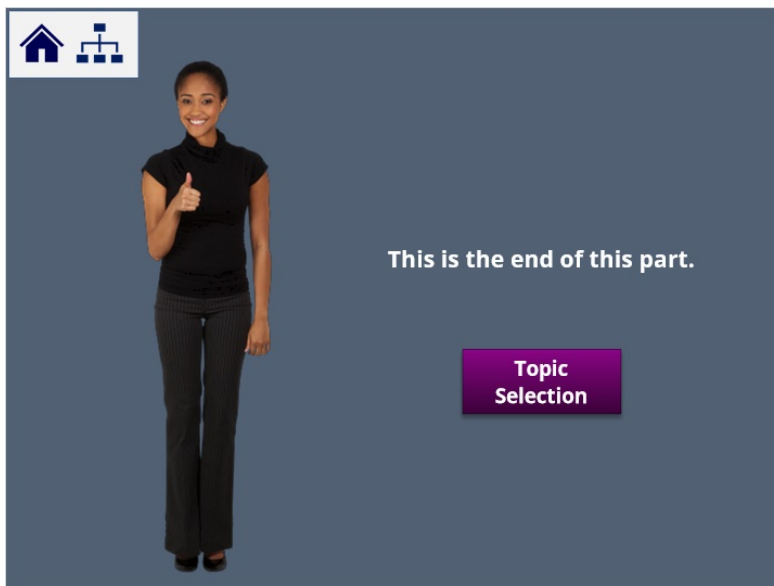
3.8 Data Collection Design



3.9 Statistical Model



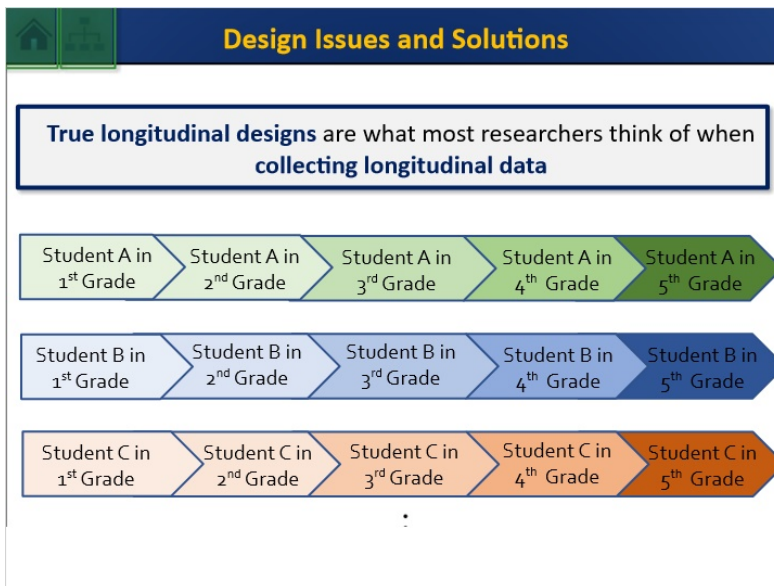
3.10 Bookend: Research Components



3.11 Bookmark: Data Designs





3.12 Design Issues and Solutions I



3.13 Design Issues and Solutions II

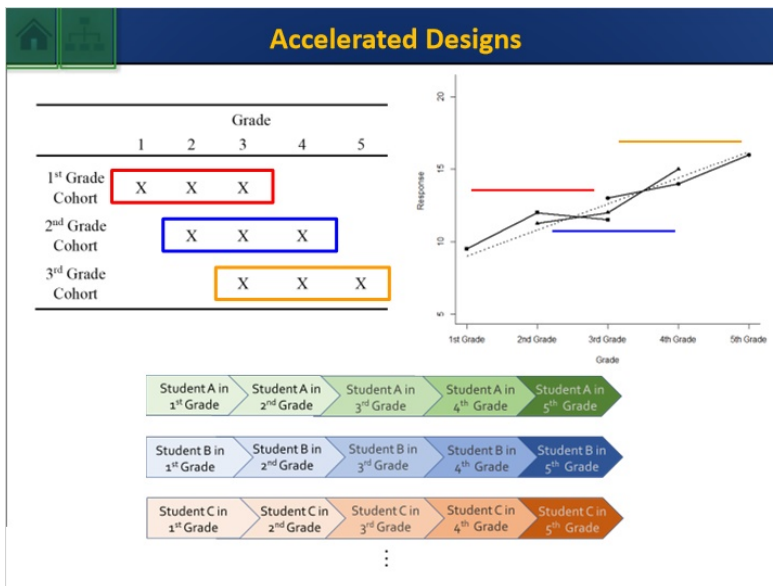
Design Issues and Solutions

Planned missingness designs such as **cohort-sequential** or **accelerated designs** provide a way to link adjacent segments of limited longitudinal data from different age cohorts to determine the existence of a **common developmental trend or growth curve**

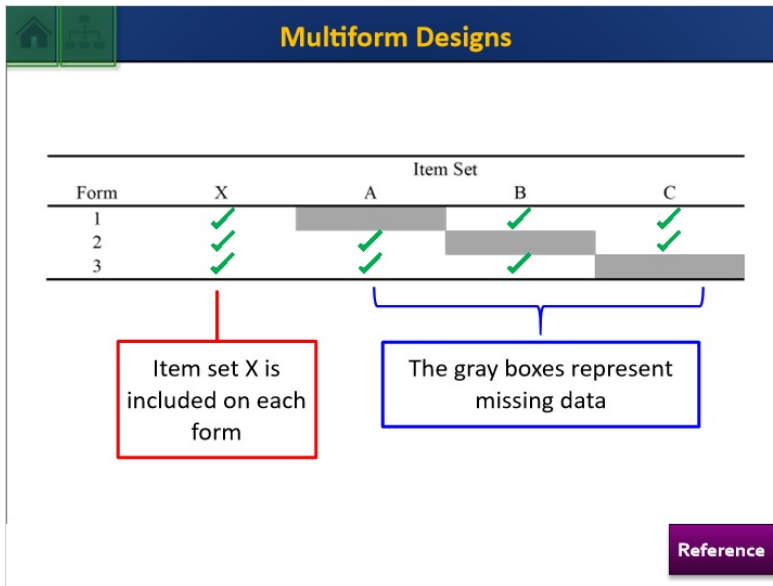



In this way, the researcher **approximates a long-term longitudinal study** by simultaneously conducting and connecting **several short-term longitudinal studies** of different age cohorts

3.14 Accelerated Designs



3.15 Multiform Designs



Reference (Slide Layer)

Reference

Original Article

Planned Missing Data Designs in Educational Psychology Research

Melissa R. Barmann & Gregory R. Hancock
Pages 505-518 | Published online 12 May 2016

Download citation | <https://doi.org/10.1080/00461520.2016.1208094> | [Check for updates](#)

Full Article | Figures & data | References | Citations | Metrics | Reprints & Permissions

Get Access

Abstract

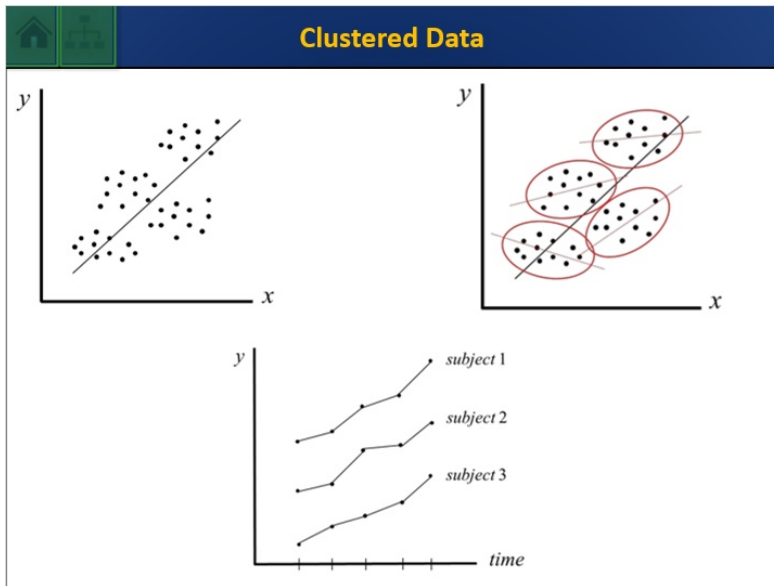
Although missing data are often viewed as a challenge for applied researchers, in fact missing data can be highly beneficial. Specifically, when the amount of missing data on specific variables is carefully controlled, a balance can be struck between statistical power and research costs. This article presents the issue of planned missing data by discussing specific designs (i.e., multiform designs, longitudinal wave-missing designs, and 2-method measurement designs), introducing the power and cost benefits of such scenarios to applied education and educational psychology researchers.

People also read

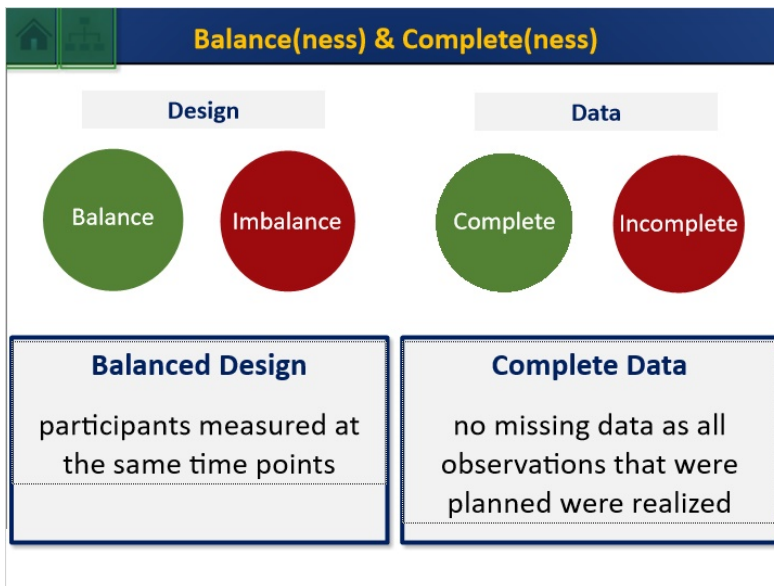
Click on the image to go to the publisher website

Back

3.16 Clustered Data



3.17 Balanceness & Completeness



3.18 Hypothetical Example I

Hypothetical Example 1

Subject 1	
Age	Score
10	9.2
12	10.5
13	9.8
16	12.6

Subject 2	
Age	Score
9	10.1
11	11.6
14	10.8
16	13.9

Subject 3	
Age	Score
9	11.1
10	12.8
15	15.3
16	12.2

Imbalanced Design

Complete Data

3.19 Hypothetical Example II

Hypothetical Example 2

Subject 1	
Age	Score
10	9.2
12	10.5
13	9.8
16	12.6

Subject 2	
Age	Score
9	10.1
11	11.6
14	10.8
16	13.9



Subject 3	
Age	Score
9	11.1
10	12.8
15	15.3
16	12.2

Age								
Subject	9	10	11	12	13	14	15	16
1	○	9.2	○	10.5	9.8	○	○	12.6
2	10.1	○	11.6	○	○	10.8	○	13.9
3	11.1	12.8	○	○	○	○	15.3	12.2

Balanced Design

Incomplete Data

3.20 Hypothetical Example III





Hypothetical Example 3


Subject	Age						
	9	10	11	12	13	14	15
1	9.2	10.5	9.8	12.6	.	.	.
2	10.1	.	11.6	.	.	.	13.9
3	11.1	.	12.8	.	15.3	.	12.2

Imbalanced Design

Incomplete Data

3.21 Bookend: Data Designs







This is the end of this part.

Topic Selection

3.22 Bookmark: Handling Data Issues





3.23 Handling Design and Data Nuances





Handling Design and Data Nuances

- It is critical to determine how you want to **treat data** when either a **design is imbalanced** and/or when there are **missing data**



- This has implications for the **structure of the data set**, how the data will be **investigated and analyzed** as well as the **validity of the inferential analyses**

3.24 Not Recommended




Not Recommended



	Age							
Subject	9	10	11	12	13	14	15	16
1	.	9.2	.	10.5	9.8	.	.	12.6
2	10.1	.	11.6	.	.	10.8	.	13.9
3	11.1	12.8	15.3	12.2

	Wave			
Subject	1	2	3	4
1	9.2	10.5	9.8	12.6
2	10.1	11.6	10.8	13.9
3	11.1	12.8	15.3	12.2

In this example, **ignoring age may be indefensible** especially if the scores reflect some type of **developmental phenomenon** that is naturally tied to age



3.25 Recommended



Recommended

- It is common, and often preferred, in longitudinal studies not to **force data to be complete and balanced**
- This allows the observations to be **anchored to the chronological metric** rather than the order in which the observations were obtained
- In the previous example, if the **age distinction** is an important substantive feature, then the data should be treated as **incomplete within a balanced design**
- Plan ahead for a **sensible data collection procedure** and **anticipate issues** that will arise

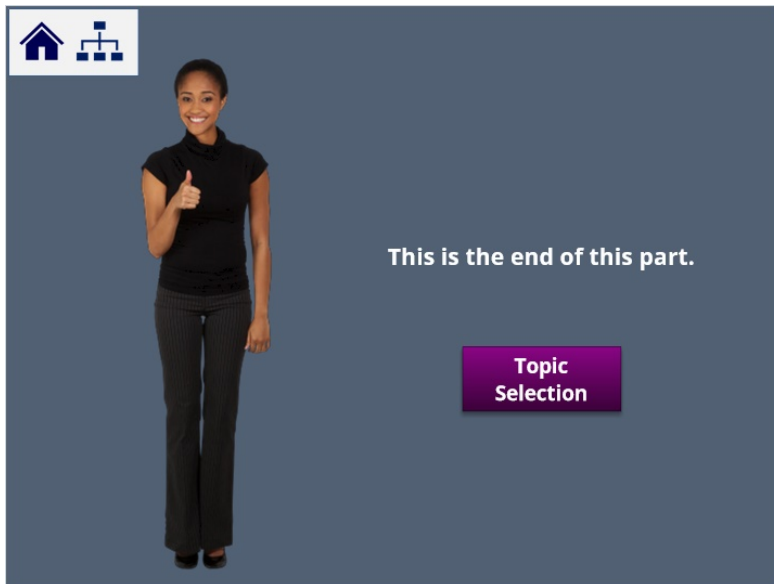
3.26 Data Structures I

Data Structures													
VARIABLES													
SUBJECTS	id	gen	momage	homecog	homeemo	an1	an2	an3	an4	r1	r2	r3	r4
	1	1	27	7	11	1	0	1	0	27	49	50	-99
	2	1	27	10	7	1	1	0	1	31	47	56	64
	3	0	27	7	7	5	0	5	3	36	52	60	75
	4	1	24	8	8	1	1	-99	-99	18	30	-99	-99
	5	1	26	8	8	2	3	3	1	23	49	-99	77
	6	0	25	6	11	1	0	0	0	21	-99	45	53
	7	0	22	5	5	3	-99	-99	10	21	-99	48	-99
	8	0	23	1	4	0	-99	-99	4	13	-99	37	-99
	9	0	24	3	7	5	1	0	0	29	-99	35	38
	10	0	28	9	11	3	1	6	5	45	58	76	80
	11	0	25	8	8	1	1	-99	-99	18	38	47	-99
	12	0	27	12	12	1	0	0	0	22	47	38	57
	13	1	28	9	12	1	0	1	3	26	42	53	69
	14	1	26	9	10	2	1	4	1	18	22	-99	40
	15	1	27	12	0	0	-99	0	0	34	46	51	-99
	16	1	29	11	10	0	0	2	1	18	38	40	-99
	17	0	24	10	11	2	2	2	2	17	37	50	65
	18	0	25	8	9	0	1	1	2	20	40	-99	57
	19	1	25	9	6	5	4	-99	5	27	60	-99	-99
	20	0	26	11	5	2	4	4	0	27	34	40	47
	21	0	27	8	4	0	0	0	0	25	37	41	72
	22	0	27	14	12	0	1	2	0	31	49	61	67
	23	0	25	9	11	0	1	1	3	32	40	44	59
	24	1	27	8	10	2	0	0	2	18	35	48	-99



3.27 Data Structures II

Data Structures													
Wide Format													
id	gen	momage	homecog	homeemo	an1	an2	an3	an4	r1	r2	r3	r4	
1	1	27	7	11	1	0	1	0	27	49	50	-99	
2	1	27	10	7	1	1	0	1	31	47	56	64	
3	0	27	7	7	5	0	5	3	36	52	60	75	
4	1	24	8	8	1	1	-99	-99	18	30	-99	-99	
Long / Stacked Format													
id	gen	read											
1	1	27											
1	1	49											
1	1	50											
1	1	-99											
2	1	31											
2	1	47											
2	1	56											
2	1	64											
3	0	36											
3	0	52											
3	0	60											
3	0	75											
4	1	18											
4	1	30											
4	1	-99											
4	1	-99											

3.28 Bookend: Handling Data Issues



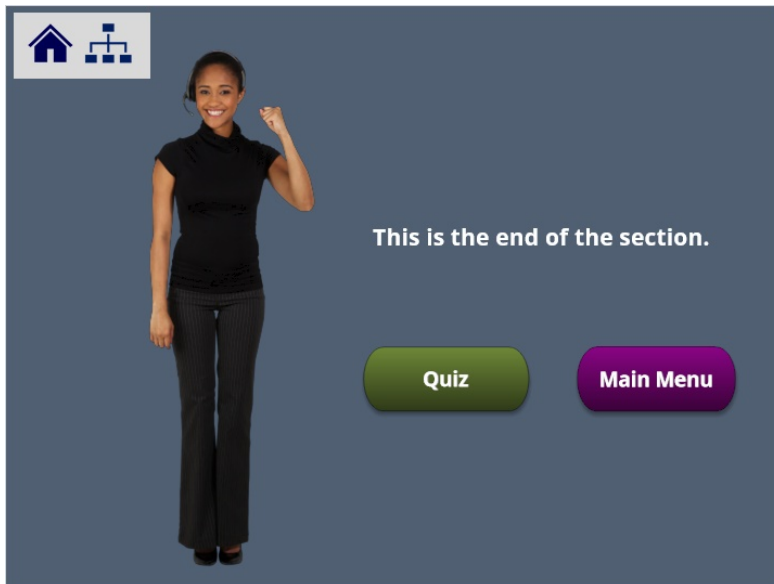
3.29 Summary



Summary

- In designing longitudinal studies, **three interconnected components must be in concert**: the theoretical model of change, the data collection design, and the statistical model
- A chosen **data-collection protocol** must keep in mind **spacing, timing, and frequency of the measurements** that align with the **theoretical model of change**
- **Balance(ness)** and **complete(ness)** are **design and data considerations** that must align with the **statistical model**
- The data, once collected, are usually stored in a **wide format** while most **analytic activities** require the data to be **stacked**

3.30 Bookend: Section 2





4. Section 3: Linear Mixed Effects Models


4.1 Cover: Section 3



4.2 Learning Objectives: Section 3





Learning Objectives

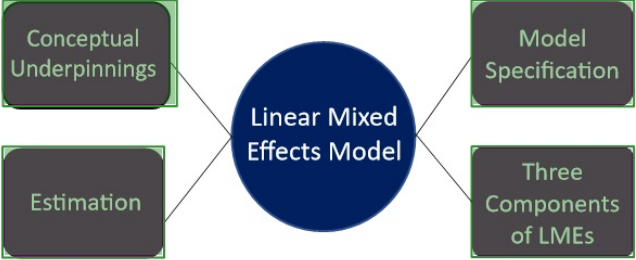



1. Understand the basic terminology and notation of the linear mixed effects model
2. Specify each of the three inter-connected components of the model
3. Understand the distributional assumptions of the model
4. Understand — conceptually — the method of maximum likelihood

4.3 Topic Selection



Section Overview



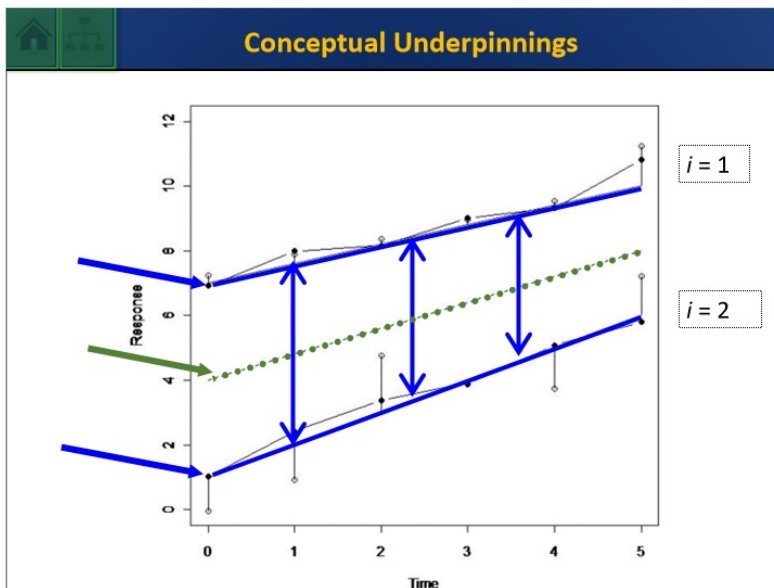
Click on each button to learn more 

Section Summary

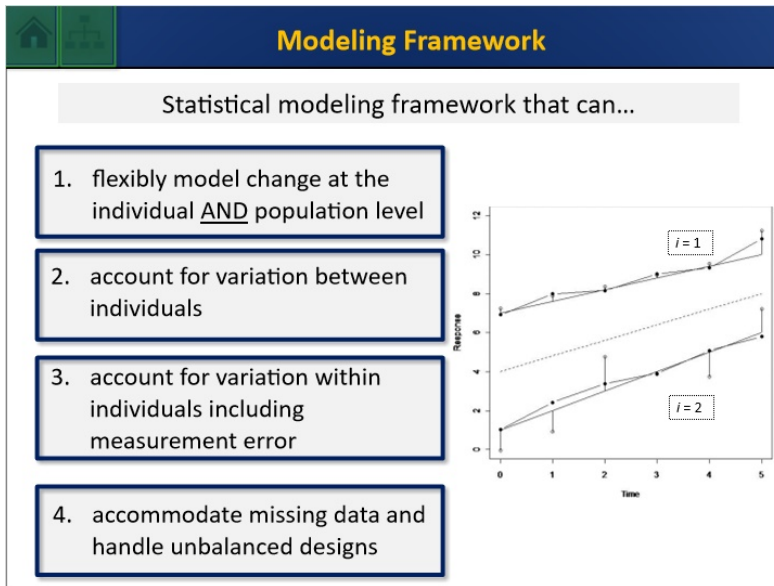
4.4 Bookmark: Conceptual Underpinnings



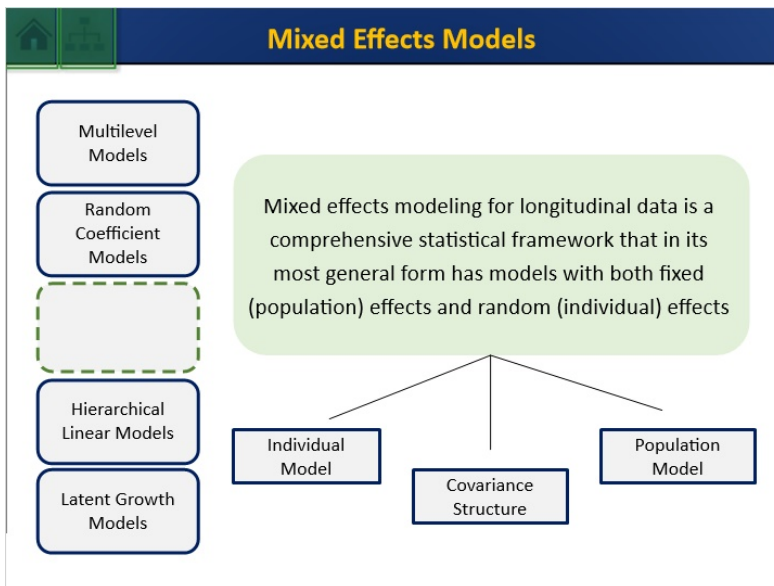
4.5 Conceptual Underpinnings



4.6 Modeling Framework



4.7 Mixed Effects Models I



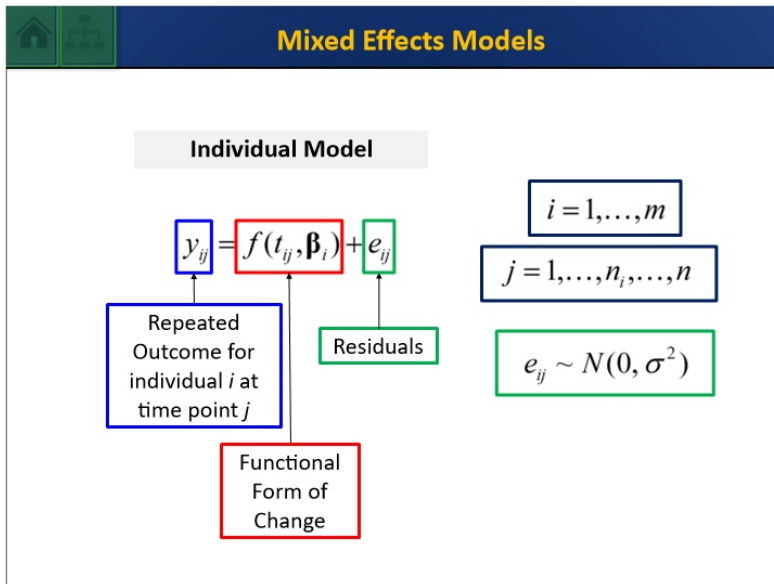
4.8 Bookend: Conceptual Underpinnings



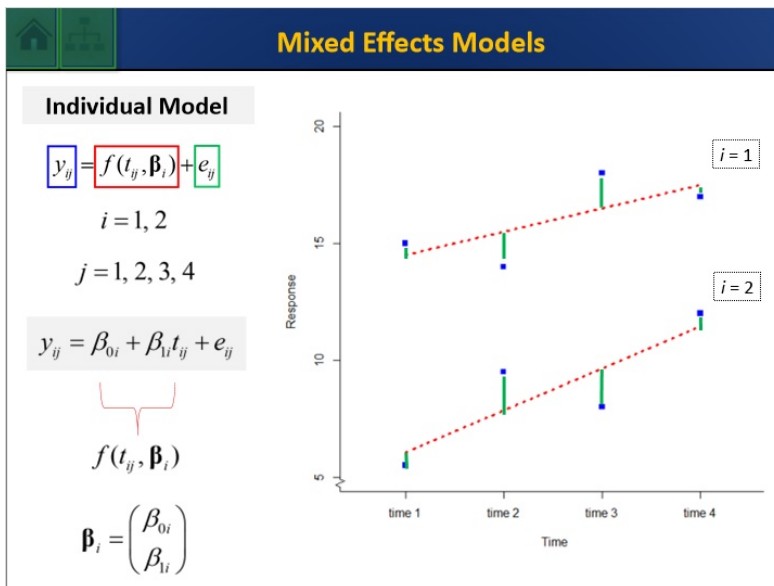
4.9 Bookmark: Model Specification



4.10 Mixed Effects Models II



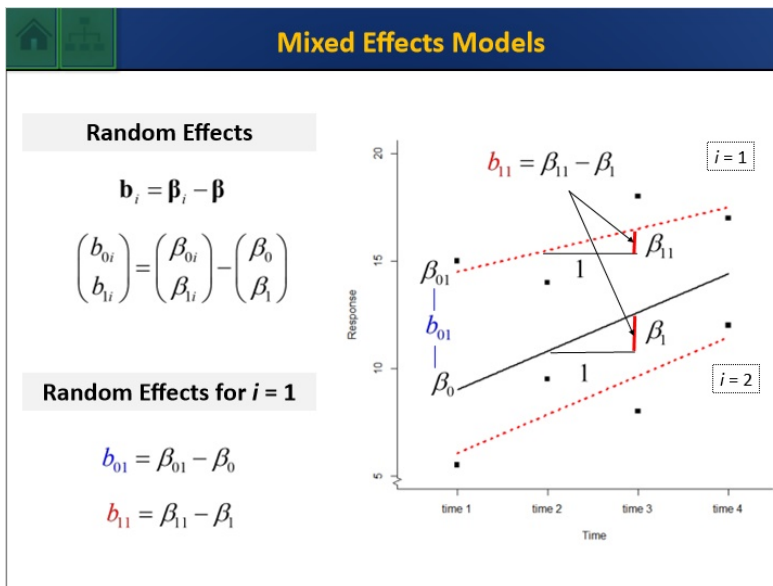
4.11 Mixed Effects Models III



4.12 Mixed Effects Models IV

Mixed Effects Models	
Population Model $\beta_i = g(z_i, \beta, b_i)$ <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid red; padding: 2px; text-align: center;">Fixed Effects</div> <div style="border: 1px solid green; padding: 2px; text-align: center;">Random Effects</div> </div>	Unconditional (No Covariates) $\beta_{0i} = \beta_0 + b_{0i}$ $\beta_{1i} = \beta_1 + b_{1i}$
$b_i \sim MVN(0, \Phi)$ $\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim MVN \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \varphi_{00} & \\ \varphi_{10} & \varphi_{11} \end{pmatrix} \right]$	Conditional (with Covariates) $\beta_{0i} = \beta_0 + \beta_2 z_{1i} + \dots + b_{0i}$ $\beta_{1i} = \beta_1 + \beta_3 z_{1i} + \dots + b_{1i}$

4.13 Mixed Effects Models V





4.14 Mixed Effects Models VI

Mixed Effects Models	
Individual Model	Unconditional Population Model
$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij}$	$\beta_{0i} = \beta_0 + b_{0i}$ $\beta_{1i} = \beta_1 + b_{1i}$
$y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + e_{ij}$	
Mixed Effects Model	residual
	random effect for slope
	population slope
	random effect for intercept
	population intercept

4.15 Linear Mixed Effects Models I

Linear Mixed Effects Models	
$y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + e_{ij} \quad t_{i1} = 0, t_{i2} = 1, t_{i3} = 2, t_{i4} = 3 \quad n = 4$	
Written Out	$\begin{aligned} y_{i1} &= \beta_0 + \beta_1 t_{i1} + b_{0i} + b_{1i} t_{i1} + e_{i1} \\ y_{i2} &= \beta_0 + \beta_1 t_{i2} + b_{0i} + b_{1i} t_{i2} + e_{i2} \\ y_{i3} &= \beta_0 + \beta_1 t_{i3} + b_{0i} + b_{1i} t_{i3} + e_{i3} \\ y_{i4} &= \beta_0 + \beta_1 t_{i4} + b_{0i} + b_{1i} t_{i4} + e_{i4} \end{aligned}$
Matrix Form	$\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \end{pmatrix}$
Compact Matrix Form	$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i$



4.16 Linear Mixed Effects Models II




Linear Mixed Effects Models

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i$$
$$\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Phi}) \quad \text{cov}(\mathbf{b}_i, \mathbf{e}_i') = \mathbf{0} \quad \mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Theta}_i)$$
$$E[\mathbf{y}_i] = \boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} \quad \text{var}[\mathbf{y}_i] = \boldsymbol{\Sigma}_i = \mathbf{Z}_i\boldsymbol{\Phi}\mathbf{Z}_i' + \boldsymbol{\Theta}_i$$

4.17 Bookend: Model Specification

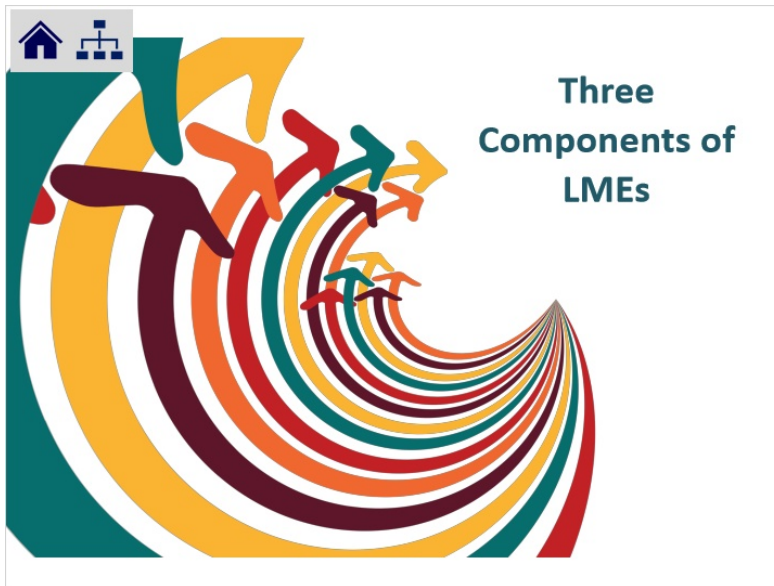




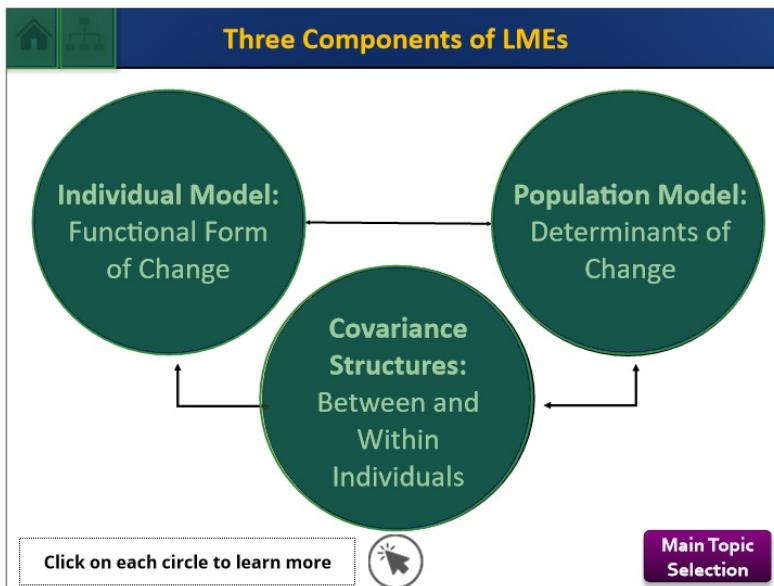
This is the end of this part.

Topic Selection

4.18 Bookmark: Three Components of LMEs



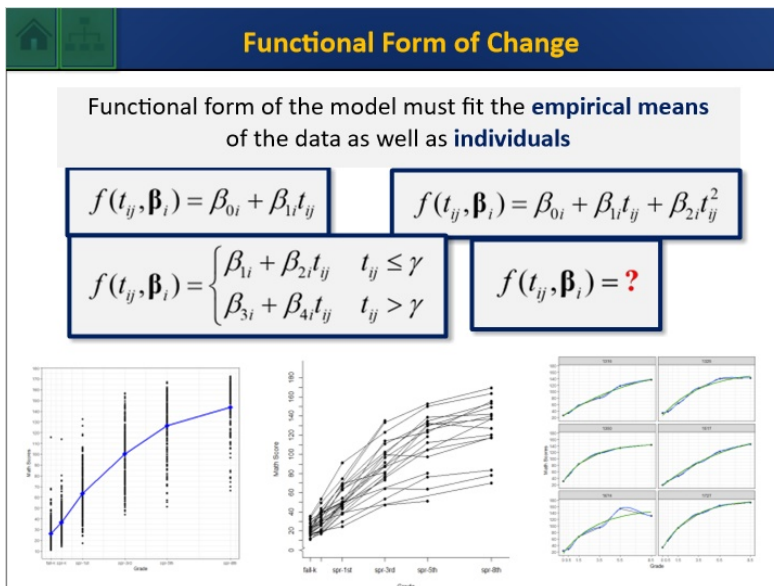
4.19 Component Selection



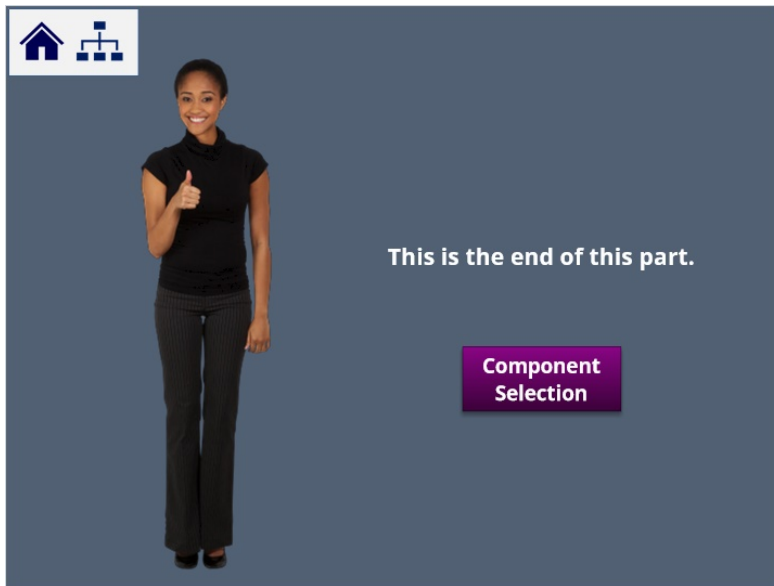
4.20 Bookmark: Functional Form



4.21 Functional Form of Change



4.22 Bookend: Functional Form



4.23 Bookmark: Variances and Covariances



4.24 Modeling Variances and Covariances I

Modeling Variances and Covariances

In a **linear mixed effects model** there are **2 sources of variation** that must be modeled

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

$\mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Theta}_i)$

Within Individual
Variation

$\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Phi})$

Between Individual
Variation

Let's assume a **linear growth process**
measured at **$n = 4$ occasions**

4.25 Modeling Variances and Covariances II

Modeling Variances and Covariances

For a linear function, there is a **random effect for intercepts** AND a **random effect for slopes**

$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij}$
 $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i$

$y_{ij} = \beta_0 + \beta_1t_{ij} + b_{0i} + b_{1i}t_{ij} + e_{ij}$
 $\mathbf{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Phi})$

$$\text{var}(\mathbf{b}_i) = \boldsymbol{\Phi} = \begin{pmatrix} \text{var}(b_{0i}) & \text{cov}(b_{1i}, b_{0i}) \\ \text{cov}(b_{1i}, b_{0i}) & \text{var}(b_{1i}) \end{pmatrix}$$

4.26 Modeling Variances and Covariances III

Modeling Variances and Covariances

For a linear function measured at 4 occasions, the **within-individual variation** is the variation associated with **individuals' residuals**

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij}$$

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

$$\mathbf{e}_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \end{pmatrix}$$

$$\text{var}(\mathbf{e}_i) = \boldsymbol{\Theta}_i = \begin{pmatrix} \text{var}(e_{i1}) & \text{cov}(e_{i2}, e_{i1}) & \dots & \text{cov}(e_{i4}, e_{i1}) \\ \text{cov}(e_{i2}, e_{i1}) & \text{var}(e_{i2}) & \dots & \text{cov}(e_{i4}, e_{i2}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(e_{i4}, e_{i1}) & \dots & \text{cov}(e_{i4}, e_{i3}) & \text{var}(e_{i4}) \end{pmatrix}$$

4.27 Modeling Variances and Covariances IV

Modeling Variances and Covariances

$\text{var}(\mathbf{e}_i) = \boldsymbol{\Theta}_i$

Homogeneous, independent

$$\boldsymbol{\Theta}_i = \begin{pmatrix} \sigma^2 & & & \\ 0 & \sigma^2 & & \\ 0 & 0 & \sigma^2 & \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

Unstructured

$$\boldsymbol{\Theta}_i = \begin{pmatrix} \sigma_1^2 & & & \\ \sigma_{21} & \sigma_2^2 & & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

Complexity

← less parameters more parameters →

Toeplitz

$$\boldsymbol{\Theta}_i = \begin{pmatrix} \sigma^2 & & & \\ \sigma_1 & \sigma^2 & & \\ \sigma_2 & \sigma_1 & \sigma^2 & \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{pmatrix}$$

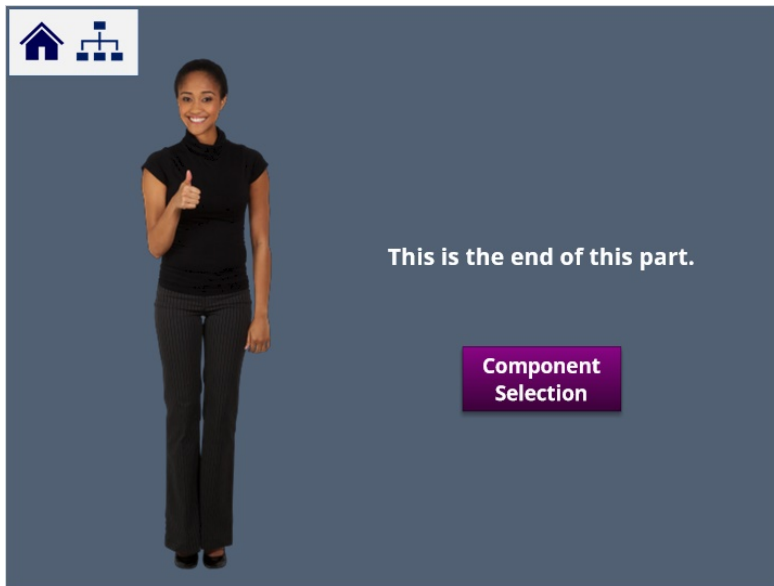
Compound symmetry

$$\boldsymbol{\Theta}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}$$

First-order autoregressive

$$\boldsymbol{\Theta}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$



4.28 Bookend: Variances and Covariances



4.29 Bookmark: Determinants of Change



4.30 Determinants of Change



Determinants of Change



Determinants of change are **time-invariant covariates** (i.e., gender, treatment condition) that help explain **why individuals differ in growth parameters**


$$\beta_i = g(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i)$$
$$\beta_{0i} = \beta_0 + \beta_2 z_{1i} + \beta_3 z_{2i} + b_{0i}$$
$$\beta_{1i} = \beta_1 + \beta_4 z_{1i} + \beta_5 z_{2i} + b_{1i}$$

gender

prior achievement

4.31 Bookend: Determinants of Change

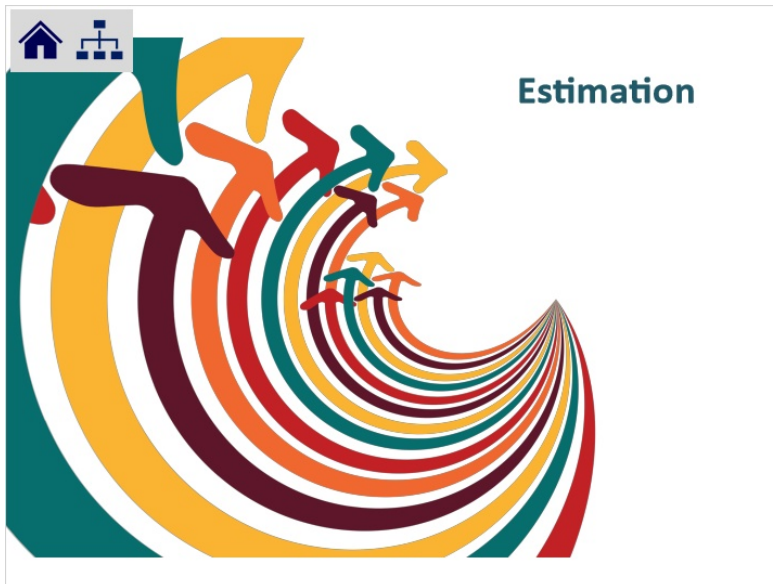






This is the end of this part.

Component Selection

4.32 Bookmark: Estimation



4.33 Estimation



Estimation



LMEs are typically estimated using **maximum likelihood** under the assumption that the **data are multivariate normal**

$$\mathbf{y}_i \sim N(\underbrace{\mathbf{X}_i \boldsymbol{\beta}}_{\boldsymbol{\mu}_i}, \underbrace{\mathbf{Z}_i \boldsymbol{\Phi} \mathbf{Z}_i' + \boldsymbol{\Theta}_i}_{\boldsymbol{\Sigma}_i})$$

Maximum Likelihood Estimation

- ✓ Assume the data follow a probability distribution
- ✓ Assume the data are a random sample from this distribution
- ✓ Find the parameter values that are best supported by the data

4.34 ML and REML



ML and REML

Software that estimates LMEs by maximum likelihood does so in **2 stages**

Stage 1: Estimate parameters comprising the mean component of the model— β using generalized least squares

Stage 2: Estimate parameters comprising the variance/covariance components of the model— using Φ and Θ_i

maximum likelihood (ML) or **restricted maximum likelihood** (REML)

4.35 Bookend: Estimation







This is the end of this part.

Topic Selection



4.36 Summary




Summary

- Linear mixed effects models provide a **flexible statistical framework** for longitudinal data
- **Nuanced longitudinal designs** and **missing data** can be accommodated
- **Three interconnected components** of the LME model are:
 - (1) a model for the individual
 - (2) a model for the population
 - (3) a model for the variances and covariances among the repeated measures
- The **typical assumption** is that the continuous repeated measures data are **multivariate normal**
- Model parameters are usually **estimated using ML or REML**

4.37 Bookend: Section 3





This is the end of the section.

[Quiz](#)


[Main Menu](#)

5. Section 4: LME Analyses

5.1 Cover: Section 4





5.2 Learning Objectives: Section 4



Learning Objectives

1. Understand the set of data analytic activities involved in an LME model analysis
2. Understand exploratory data analyses for LME model analyses
3. Describe the iterative framework that make up fitting and refining the LME model
4. Check model assumptions and LME model diagnostics as part of an analytic process

5.3 NLSY Data Recap





NLSY Data Recap

Repeated Measures Variables (1986, 1988, 1990, 1992)	Time-Invariant Covariates
<ul style="list-style-type: none"><u>r1-r4</u>: PIAT reading recognition skills scores<u>an1-an4</u>: BPI antisocial behavior scores<u>ag1-ag4</u>: Children's ages	<ul style="list-style-type: none"><u>gen</u>: Child's gender (0 = female, 1 = male)<u>momage</u>: Mother's age at Time 1<u>homecog</u>: Cognitive stimulation measure (0-14) at Time 1<u>homeemo</u>: Home emotional support measure (0-13) at Time 1

Reference

Reference (Slide Layer)



Reference



Collection Thesaurus
Search education resources
☐ Peer reviewed only ☐ Full text available on ERIC


The 1997 Profile of American Youth: Overview.
Curtin, Linda T.

The Department of Defense is developing norms for its Armed Services Vocational Aptitude Battery (ASVAB) and the newly created Interest Finder, a vocational interest inventory. The normative effort, called the 1997 Profile of American Youth (PAY97), will be undertaken as part of the Bureau of Labor Statistics' 1997 National Longitudinal Survey of Youth. The national norms that result from this cooperative project will be used for military personnel selection and placement purposes for enlistees and as the basis for career selection in high schools in the student testing program. The ASVAB will be normed in its recently implemented computerized form, and Interest Finder norms will also be developed from its computer version. Methodological studies are being planned to address substantive issues related to procedures and materials for the interview, test administration, and data analysis stages. A participation incentive and performance bonus study has preceded pilot testing involving more than 1,500 examinees. Other studies that are being initiated are studies of the appropriateness of the tests with adolescents, a hardware and environment study, and a pretest of the longitudinal youth survey. The ASVAB and Interest Finder will be administered in the summer of 1997 to the 19,000 subjects of the longitudinal study. Normative information will then be available for the planned uses of both measures. (Contains one figure, two tables, and eight references.) (SLD)

Descriptors: Adaptive Testing; Adolescents; Attitude Tests; Armed Forces; Career Choice; Computer Assisted Testing; High School Students; High Schools; Incentives; Interest Inventories; Military Personnel; National Norms; National Surveys; Post-Enlistment; Profiles; Selection; Test Construction; Test Use; Vocational Interests.

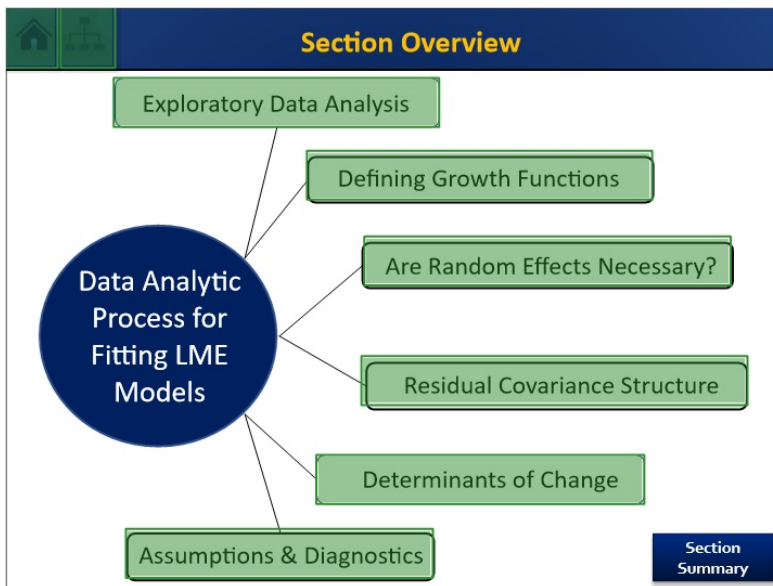
Publication Type: Reports - Evaluative; Speech/Testing Papers
Education Level: N/A
Audience: N/A
Language: English

Click on the image to go to the publisher website

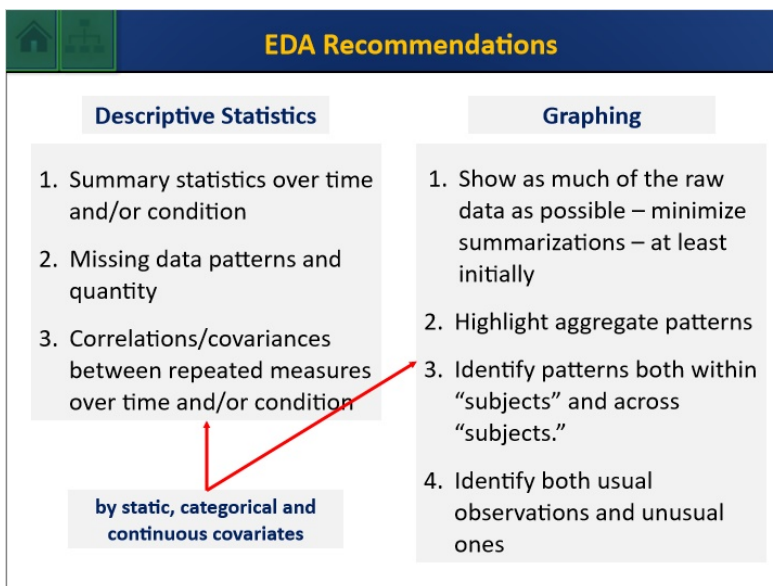


Reference

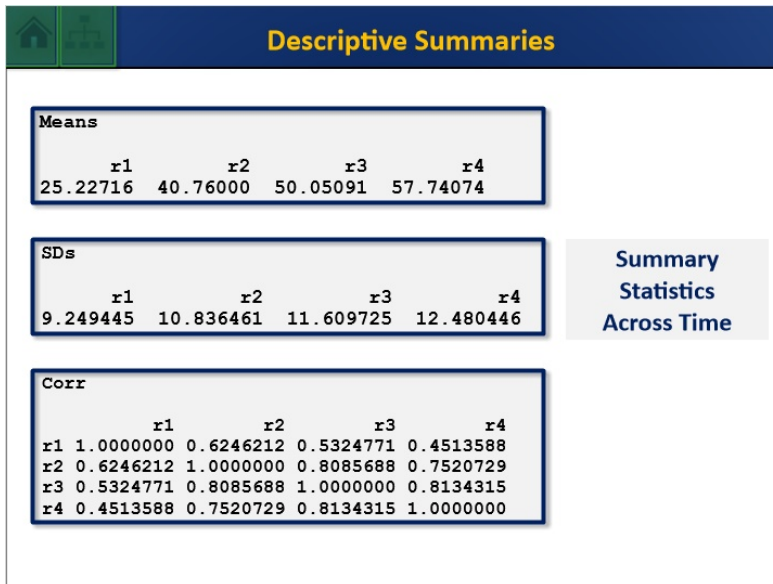
5.4 Topic Selection



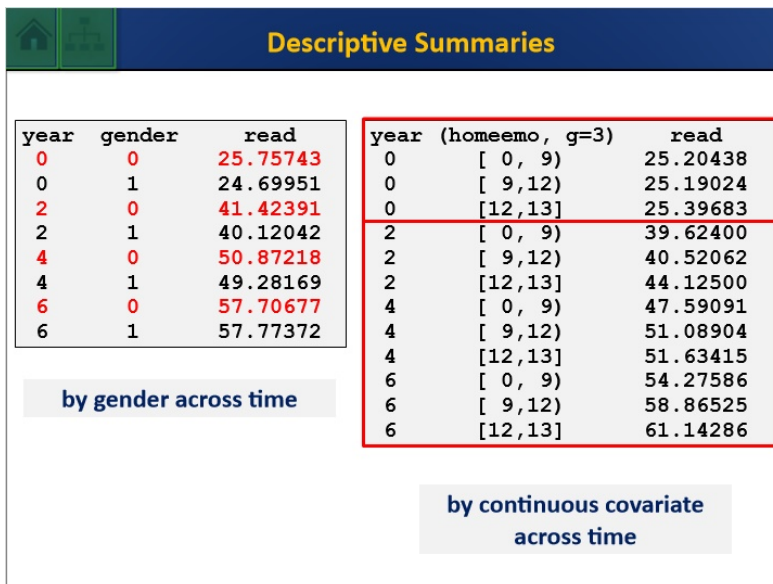
5.5 EDA Recommendations



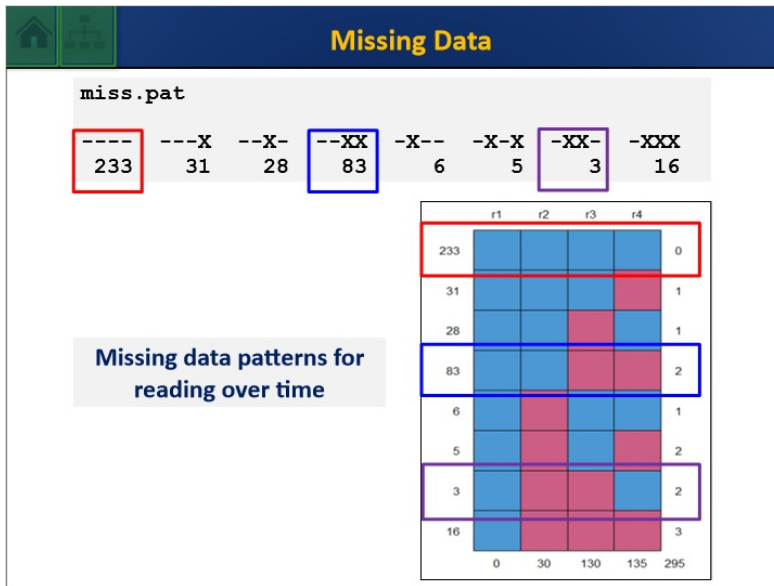
5.6 Descriptive Summaries I



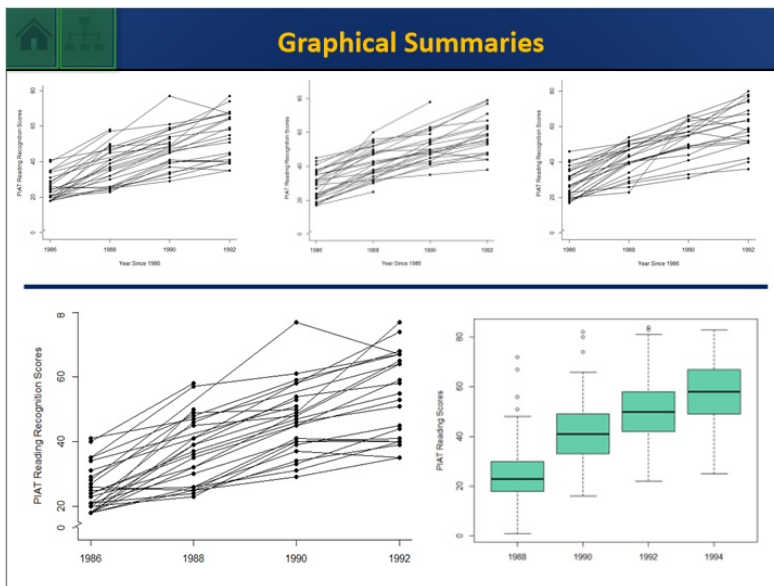
5.7 Descriptive Summaries II



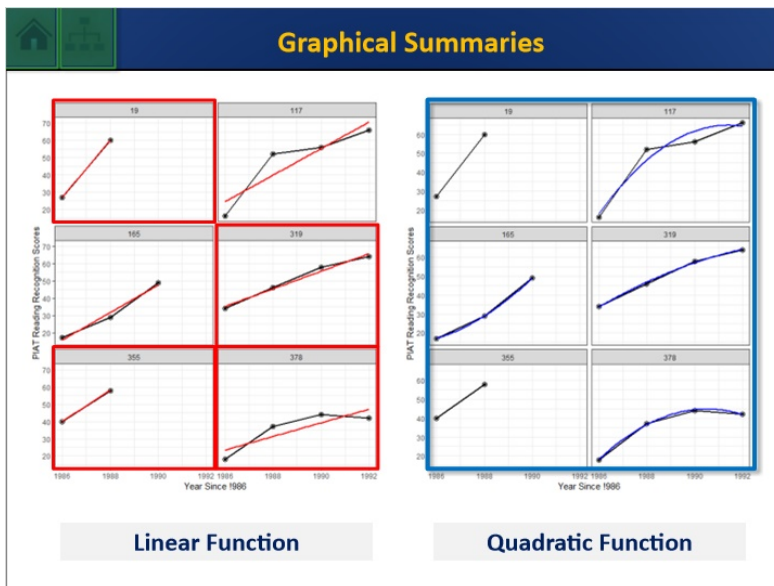
5.8 Missing Data



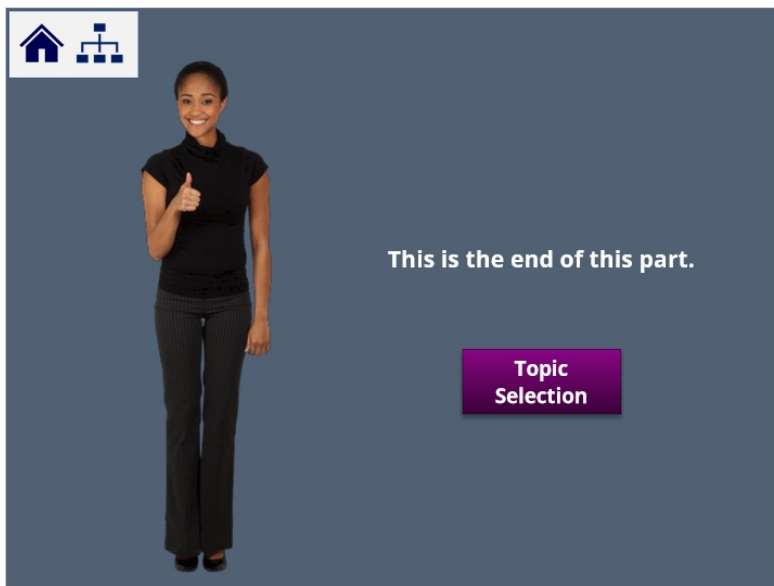
5.9 Graphical Summaries I



5.10 Graphical Summaries II



5.11 Bookend: Exploratory Data Analysis



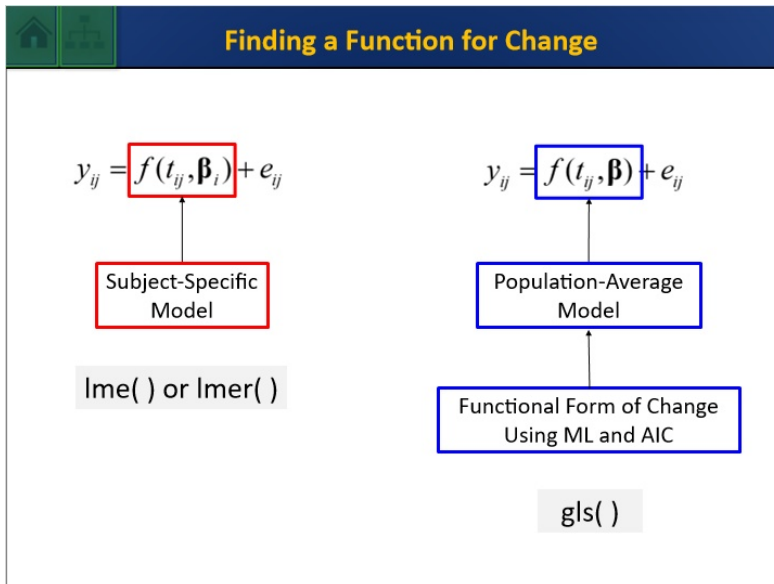
5.12 Bookmark: Defining Growth Functions



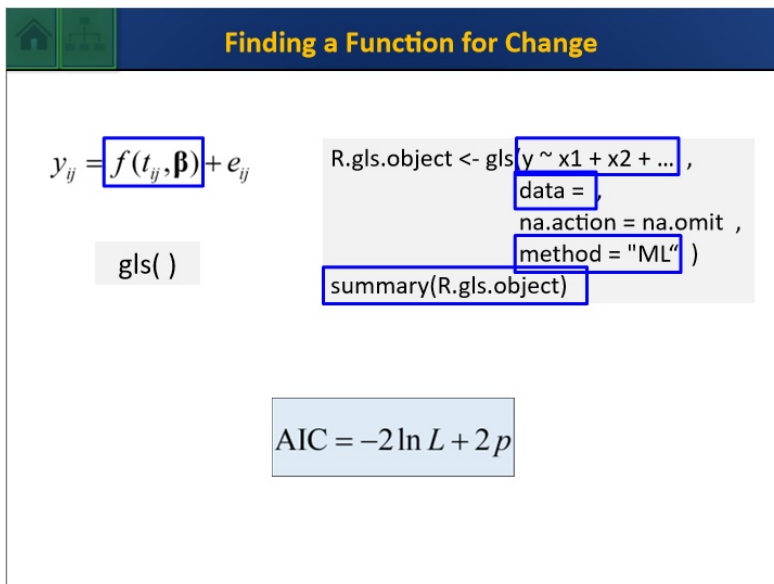
5.13 Bookmark: Exploratory Data Analysis





5.14 Finding a Function for Change I



5.15 Finding a Function for Change II

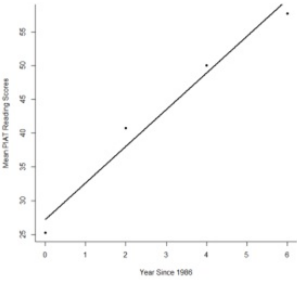


5.16 Finding a Function for Change III



Finding a Function for Change



Linear Function


$$f(t_{ij}, \beta) = \beta_0 + \beta_1 t_{ij}$$

```
p.1 <- gls( read ~ 1 + year ,  
  data = read.long ,  
  na.action = na.omit ,  
  method = "ML" )  
  
summary(p.1)
```

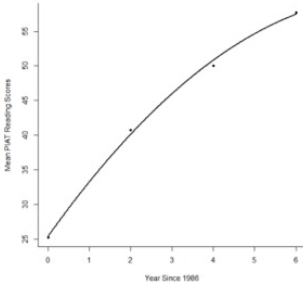
AIC = 10143.32

5.17 Finding a Function for Change IV



Finding a Function for Change



Quadratic Function


$$f(t_{ij}, \beta) = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2$$

```
p.2 <- gls( read ~ 1 + year + I(year^2) ,  
  data = read.long ,  
  na.action = na.omit ,  
  method = "ML" )  
  
summary(p.2)
```

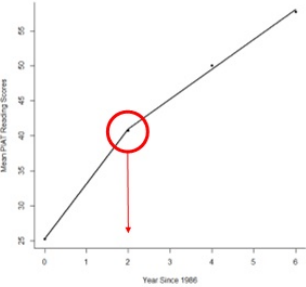
AIC = 10100.36

5.18 Finding a Function for Change V



Finding a Function for Change



Piecewise Function


$$f(t_{ij}, \beta) = \beta_0 + \beta_1 t_{ij} + \beta_2 (t_{ij} - 2)_+$$


AIC = 10098.51

```
read.long$nyear <-  
pmax(read.long$year-2, 0)  
  
p.3 <- gls( read ~ 1 + year + nyear,  
  data = read.long ,  
  na.action = na.omit ,  
  method = "ML" )  
  
summary(p.3)
```

5.19 Bookend: Defining Growth Functions







This is the end of this part.

Topic Selection

5.20 Bookmark: Are Random Effects Necessary



5.21 Are Random Effects Necessary I



Are Random Effects Necessary?

Random Intercept Model to Compute the ICC

$$y_{ij} = \beta_{0i} + e_{ij} \quad e_{ij} \sim N(0, \sigma^2)$$
$$\beta_{0i} = \beta_0 + b_{0i} \quad b_{0i} \sim N(0, \varphi_{00})$$

```
r0.out <- lme (read ~ 1,
               data = read.long,
               na.action = na.omit,
               method = "ML",
               random = ~ 1 | id,
               control=list(maxIter=100) )

summary(r0.out)
```

$$ICC = \frac{\varphi_{00}}{\sqrt{\varphi_{00} \cdot \sigma^2}} = \frac{29.79}{\sqrt{29.79 \cdot 239.03}} \approx 0.111$$

5.22 Are Random Effects Necessary II

🏠 📄
Are Random Effects Necessary?

$$y_{ij} = f(t_{ij}, \beta_i) + e_{ij}$$

$$= \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+ + e_{ij}$$

$$\beta_i = \begin{pmatrix} \beta_{0i} \\ \beta_{1i} \\ \beta_{2i} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{pmatrix}$$

$$\mathbf{b}_i \sim MVN(\mathbf{0}, \Phi)$$

$$\begin{pmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{pmatrix} \sim MVN \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \varphi_{00} & \varphi_{10} & \varphi_{20} \\ \varphi_{10} & \varphi_{11} & \varphi_{21} \\ \varphi_{20} & \varphi_{21} & \varphi_{22} \end{pmatrix} \right]$$

5.23 Testing Variances of Random Effects I

🏠 📄
Testing Variances of Random Effects

$$\mathbf{b}_i \sim MVN(\mathbf{0}, \Phi)$$

$$\Phi = \begin{pmatrix} \varphi_{00} & & \\ \varphi_{10} & \varphi_{11} & \\ \varphi_{20} & \varphi_{21} & \varphi_{22} \end{pmatrix}$$

3 Random Effects

$$f(t_{ij}, \beta_i) = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+$$

$$\Phi = \begin{pmatrix} \varphi_{00} & & \\ \varphi_{10} & \varphi_{11} & \\ \varphi_{20} & \varphi_{21} & \varphi_{22} \end{pmatrix}$$

2 Random Effects

$$f(t_{ij}, \beta_i) = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+$$

5.24 Testing Variances of Random Effects II

Testing Variances of Random Effects

$f(t_{ij}, \beta_i) = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+$

$$\Phi = \begin{pmatrix} \varphi_{00} & & \\ \varphi_{10} & \varphi_{11} & \\ \varphi_{20} & \varphi_{21} & \varphi_{22} \end{pmatrix}$$

3 Random Effects

$$\Phi = \begin{pmatrix} \varphi_{00} & & \\ \varphi_{10} & \varphi_{11} & \\ \varphi_{20} & \varphi_{21} & \varphi_{22} \end{pmatrix} = \begin{pmatrix} \varphi_{00} & & \\ \varphi_{10} & \varphi_{11} & \\ 0 & 0 & 0 \end{pmatrix}$$

2 Random Effects

$f(t_{ij}, \beta_i) = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+$

$$\chi^2_{LRT} = 2 \ln L(\text{reduced}) - 2 \ln L(\text{full})$$

$$v = \# \text{par}(\text{full}) - \# \text{par}(\text{reduced})$$

$$\chi^2 = \frac{1}{2} \chi^2_{2df} + \frac{1}{2} \chi^2_{3df}$$

5.25 Testing Variances of Random Effects III

Testing Variances of Random Effects

$f(t_{ij}, \beta_i) = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+$

$$\Phi = \begin{pmatrix} \varphi_{00} & & \\ \varphi_{10} & \varphi_{11} & \\ \varphi_{20} & \varphi_{21} & \varphi_{22} \end{pmatrix}$$

3 Random Effects

$$\Phi = \begin{pmatrix} \varphi_{00} & & \\ \varphi_{10} & \varphi_{11} & \\ \varphi_{20} & \varphi_{21} & \varphi_{22} \end{pmatrix} = \begin{pmatrix} \varphi_{00} & & \\ 0 & 0 & \\ 0 & 0 & 0 \end{pmatrix}$$

1 Random Effects

$f(t_{ij}, \beta_i) = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+$

$\chi^2_{LRT} \sim \chi^2_{5df}$
➔
 $\chi^2_{LRT} \sim (w_1 \cdot \chi^2_{3df} + w_2 \cdot \chi^2_{4df} + w_3 \cdot \chi^2_{5df})$

$$\sum_{k=1}^3 w_k = 1$$

Reference

Reference (Slide Layer)

Reference

On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints.

EXPORT Add To My List Request Permissions

Database: APA PsycArticles Journal Article

Stoel, Reinoud D., Gams, Francisco Salgado, Dolan, Connor, van den Wittenboer, Godfried

Citation

Stoel, R. D., Gams, F. G., Dolan, C., & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*, 11(4), 439-456. <https://doi.org/10.1037/1082-989X.11.4.439>

Abstract

The authors show how the use of inequality constraints on parameters in structural equation models may affect the distribution of the likelihood ratio test. Inequality constraints are implicitly used in the testing of commonly applied structural equation models, such as the common factor model, the autoregressive model, and the latent growth curve model, although this is not commonly acknowledged. Such constraints are the result of the null hypothesis in which the parameter value or values are placed on the boundary of the parameter space. For instance, this occurs in testing whether the variance of a growth parameter is significantly different from 0. It is shown that in these cases, the asymptotic distribution of the chi-square difference cannot be treated as that of a central chi-square-distributed random variable with degrees of freedom equal to the number of constraints. The correct distribution for testing 1 or a few parameters at a time is inferred for the 3 structural equation models.

Psychological Methods

Journal TOC

Get Access



Related Content


Revised structural equation models: Noncentrality and power of restriction test. Raykov, Sorku, Pateris, Spasidou, 1998

A note on the use of missing auxiliary variables in full information maximum

Back

5.26 Bookend: Random Effects





This is the end of this part.

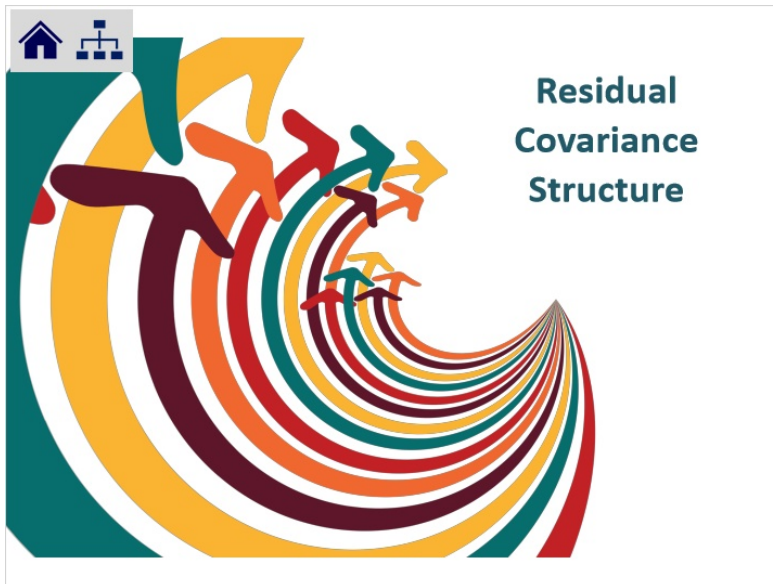
Topic Selection

DM16 SLIDES

74 / 105

8/1/2020

5.27 Bookmark: Residual Covariance Structure



5.28 Residual Variances and Covariances I

Residual Variances and Covariances

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+ + e_{ij} \quad \text{var}(\mathbf{e}_i) = \mathbf{\Theta}_i$$

$$\mathbf{\Theta}_i = \begin{pmatrix} \sigma^2 & & & \\ 0 & \sigma^2 & & \\ 0 & 0 & \sigma^2 & \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

Default in most statistical software

Homogeneous, independent Variances

$$\mathbf{\Theta}_i = \begin{pmatrix} \sigma_{r=1}^2 & & & \\ 0 & \sigma_{r=2}^2 & & \\ 0 & 0 & \sigma_{r=3}^2 & \\ 0 & 0 & 0 & \sigma_{r=4}^2 \end{pmatrix}$$

Heterogeneous Variances

$$\mathbf{\Theta}_i = \sigma^2 \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho & \rho & 1 & \\ \rho & \rho & \rho & 1 \end{pmatrix}$$

Compound Symmetry

$$\mathbf{\Theta}_i = \sigma^2 \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho^2 & \rho & 1 & \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

First-Order Autoregressive

5.29 Residual Variances and Covariances II

Residual Variances and Covariances

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+ + e_{ij} \quad \text{var}(e_i) = \Theta_i$$

$\Theta_i =$

weights()

correlation()

```
r.4.out <- lme (read ~ 1 + year + nyear ,
  data = read.long ,
  na.action = na.omit ,
  method = "ML" ,
  weights = ( ) ,
  correlation = ( ) ,
  random = ~ 1 + year + nyear | id ,
  control=list(maxIter=100))

summary(r.4.out)
```

AIC = $-2\ln L + 2p$

5.30 Residual Variances and Covariances III

Residual Variances and Covariances

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+ + e_{ij} \quad \text{var}(e_i) = \Theta_i$$

$$\begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho^2 & \rho & 1 & \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

First-Order Autoregressive

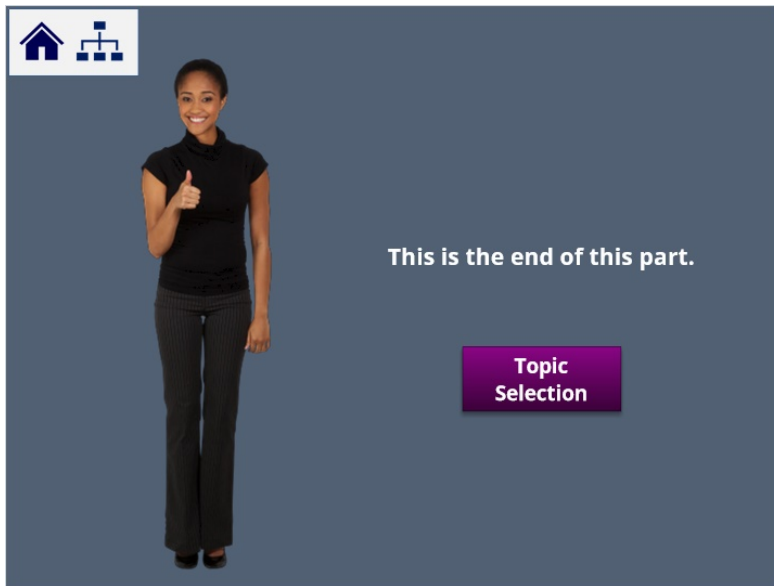
```
r.4.out <- lme (read ~ 1 + year + nyear ,
  data = read.long ,
  na.action = na.omit ,
  method = "ML" ,
  weights=varIdent(form = ~ 1 | id) ,
  correlation=corCAR1(value=.3,form = ~ 1 | id) ,
  random = ~ 1 + year + nyear | id ,
  control=list(maxIter=100))

summary(r.4.out)
```

Heterogeneous Variances

$$\Theta_i = \begin{pmatrix} \sigma^2_{t=1} & & & \\ 0 & \sigma^2_{t=2} & & \\ 0 & 0 & \sigma^2_{t=3} & \\ 0 & 0 & 0 & \sigma^2_{t=4} \end{pmatrix}$$

5.31 Bookend: Residual Covariances



5.32 Bookmark: Determinants of Change



5.33 Determinants of Change I

Determinants of Change

Determinants of change are **time-invariant covariates** (i.e., gender, treatment condition) that help explain why **individuals differ in growth parameters**

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+ + e_{ij} \quad \beta_i = g(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i)$$

$$\begin{aligned}\beta_{0i} &= \beta_0 + \beta_3 z_{1i} + b_{0i} \\ \beta_{1i} &= \beta_1 + \beta_4 z_{1i} + b_{1i} \\ \beta_{2i} &= \beta_2 + \beta_5 z_{1i} + b_{2i}\end{aligned}$$

Home Cognition
Scores

5.34 Determinants of Change II

Determinants of Change

$$y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}(t_{ij} - 2)_+ + e_{ij}$$

$$\begin{aligned}\beta_{0i} &= \beta_0 + \beta_3 z_{1i} + b_{0i} \\ \beta_{1i} &= \beta_1 + \beta_4 z_{1i} + b_{1i} \\ \beta_{2i} &= \beta_2 + \beta_5 z_{1i} + b_{2i}\end{aligned}$$

$\boldsymbol{\beta}' = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$

$\mathbf{b}'_i = (b_{0i}, b_{1i}, b_{2i})$

\mathbf{X}_i
 $n_i \times p$

\mathbf{Z}_i
 $n_i \times q$

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

5.35 Determinants of Change III

Determinants of Change

$$y_i = X_i\beta + Z_i b_i + e_i$$

$$\beta_{0i} = \beta_0 + \beta_3 z_{1i} + b_{0i}$$

$$\beta_{1i} = \beta_1 + \beta_4 z_{1i} + b_{1i}$$

$$\beta_{2i} = \beta_2 + \beta_5 z_{1i} + b_{2i}$$

```

r.9.out <- lme (read ~ 1 + year + nyear +
  homecog + year:homecog + nyear:homecog ,
  data = read.long ,
  na.action = na.omit ,
  method = "ML" ,
  random = ~ 1 + year + nyear | id ,
  control=list(maxIter=100))

summary(r.9.out)
```

5.36 Determinants of Change IV

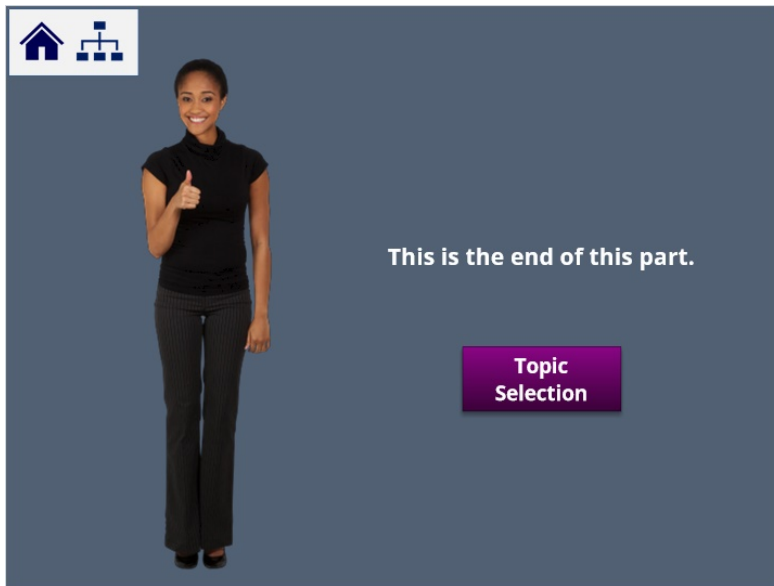
Determinants of Change

Parameter	Unconditional Model	Conditional Model
β_0	25.23	25.23
β_1	7.81	7.81
β_2	-3.41	-3.43
β_3	n/a	0.31
β_4	n/a	0.24
β_5	n/a	-0.13
σ_{ω}	60.90	60.23
σ_{10}	1.96	1.50
σ_{11}	5.96	5.55
σ_{20}	-3.18	-2.93
σ_{21}	-3.43	-3.18
σ_{22}	2.32	2.16
σ^2	24.72	24.73

$$\text{var}_{\text{slope diff}} = 100 \times \left(\frac{2.32 - 2.16}{2.32} \right) = 6.90\%$$

**Explained Variance
Intercept, Slope, and
Slope Difference**

5.37 Bookend: Determinants of Change



5.38 Bookmark: Assumptions and Diagnostics



5.39 LME Assumptions & Diagnostics

LME Assumptions & Diagnostics



Assumptions	Diagnostics
<ol style="list-style-type: none"> 1. Distribution of the level 1 residuals is multivariate normal 2. Distribution of the level 2 random effects is multivariate normal 3. Predicted random effects and residuals are often used to assess these assumptions 	<p>Identification of influential subjects uses residuals</p> <ul style="list-style-type: none"> Outlier detection Leverage Influence
$\hat{\mathbf{b}}_i = \hat{\Phi} \mathbf{Z}_i' \hat{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})$	$\hat{\mathbf{e}}_{(c)i} = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{Z}_i \hat{\mathbf{b}}_i$
<div style="background-color: #e6e6fa; padding: 2px 10px; display: inline-block;">Random Effects</div>	<div style="background-color: #e6e6fa; padding: 2px 10px; display: inline-block;">Conditional</div>
	$\hat{\mathbf{e}}_{(m)i} = \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}$
	<div style="background-color: #e6e6fa; padding: 2px 10px; display: inline-block;">Marginal</div>

5.40 Bookend: Assumptions & Diagnostics

This is the end of this part.

Topic Selection

5.41 Summary I



Summary



Research scenarios frequently arise where...

- Repeated measurements are gathered on each of several individuals
- There is variability in the relationship between response and time
- There is a scientifically-relevant model available for individual behavior in terms of meaningful parameters that vary across individuals and dictate variation in patterns of time-response

Common research objectives are to understand...

- Typical behavior of a phenomenon
- Extent to which phenomena vary across individuals
- Whether some variation associated with individual attributes

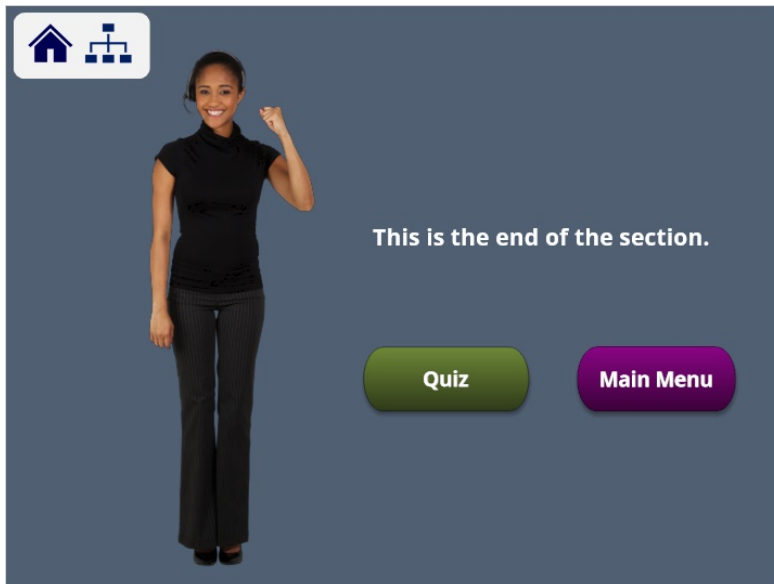
5.42 Summary II



Summary

- LME models are well-suited to attend to the **goals of many longitudinal analyses** and flexible enough to handle the **myriad data and design nuances** encountered in real-world scenarios
- LME analyses involve **data exploration, model investigation and refinement**, and **checking assumptions and diagnosing problems** that may hinder making valid inferences
- The **guided data analysis** in the last section of this module will provide an opportunity to **gain practice** with the many data-analytic activities outlined earlier in this module

5.43 Bookend: Section 4





6. Section 5: Data Activity


6.1 Cover: Section 5



6.2 Learning Objectives





Learning Objectives

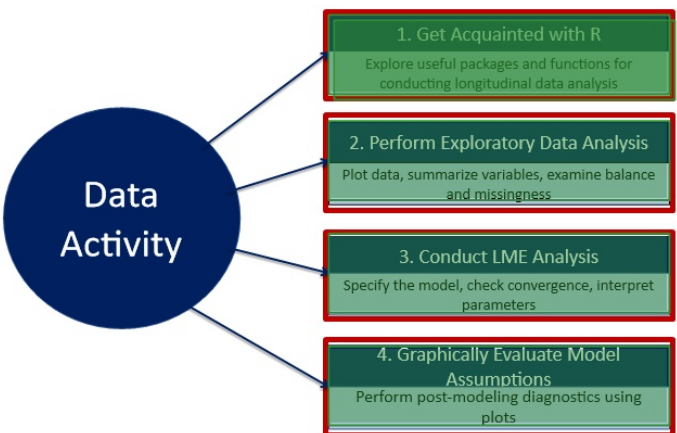



1. Identify the software packages available for conducting a mixed-effects analysis in R
2. Conduct exploratory data analysis, reshaping data and calculating descriptive statistics
3. Specify the model in R using the appropriate function arguments
4. Assess the model fit by examining output, model assumptions, and parameter estimates

6.3 Topic Selection



Section Overview





Click on each button to learn more 

Real Data Analysis

6.4 Bookmark: Get Acquainted with R



6.5 Many Sides of R





Many Sides of R

- Two primary options exist for working with data in R:
“**Base R**” and “**Tidyverse**”
- In this module, we will demonstrate how users can leverage **tidyverse** packages and principles for each stage of a longitudinal data analysis.

Base R:	Tidyverse:
<ul style="list-style-type: none">Access to variety – hundreds of great statistical packages exist on CRAN and most are not part of TidyverseMay have advantages in computational speed	<ul style="list-style-type: none">Data-science orientedConsistent logic across packages and functionsActive development & emerging resources for complex statistical models

6.6 Welcome to Tidyverse



Welcome to Tidyverse!



What makes data “tidy”?

- **Variables** are in **columns**
- **Observations** are in **rows**
- **Tables** contain a **single observational unit**


What is Tidyverse?

- A set of R packages built for **simplified data management**
- Tidyverse packages are all share a **coherent structure and logic**

6.7 Tidyverse for Longitudinal Data



Tidyverse for Longitudinal Data



Manage

- Convert data from wide to long
- Assess balance and missingness



Describe

- Calculate descriptive statistics across time variables and levels of predictors
- Assessing balance in missing data

Plot

- Plot individual and aggregated trajectories
- Create custom residual and random effects graphics



6.8 Selected Tidy Functions for Longitudinal Data




Selected Tidy Functions for Longitudinal Data

Read/Write	Manipulate/ Summarize	Visualize
readr::read_csv <ul style="list-style-type: none">• Takes a .CSV file (other file format available)• Outputs a tibble, which is based on the R dataframe class• Recode missing data from within the function	tidyr::pivot_longer <ul style="list-style-type: none">• Covert wide data to the long format dplyr::mutate <ul style="list-style-type: none">• Create new variables, manipulate existing variables dplyr::group_by & dplyr::summarize <ul style="list-style-type: none">• produce descriptive statistics for different levels of grouping variables	ggplot2::geom_histogram <ul style="list-style-type: none">• Create a histogram ggplot2::geom_point <ul style="list-style-type: none">• Plot a scatterplot ggplot2::geom_line <ul style="list-style-type: none">• Generate spaghetti/ individual trajectory plot ggplot2::geom_qq <ul style="list-style-type: none">• Assess normality of data points

6.9 Bookend: Get Acquainted with R







This is the end of this part.

Topic Selection

6.10 Bookmark: Perform Exploratory Data Analysis



6.11 EDA1: Load the Data



Load the Data



id	gen	momage	homecog	homeemo	an_1	an_2	an_3	an_4	r_1	r_2	r_3	r_4	ag_1	ag_2	ag_3	ag_4
<dbl>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
9	female	24	3	7	5	3	2	0	29	NA	35	38	8	10	12	14
10	female	28	9	11	2	3	6	5	45	58	76	80	8	10	12	15

Loading NLSY Data

- Child's age, reading skill, and anti-social behavior were measured at 4 waves
- Mom's age, child's sex, home cognitive stimulation, and home emotional support were measured at Wave 1
- Model children's anti-social behavior over time

Code

Code (Slide Layer)



Example Code



```
# read data from .CSV file  
path_dat <- "inst/extdata/anti-read.csv"  
dat <- readr::read_csv(path_dat, na = "-99")
```

NOTE: R will start looking for the "inst" folder in the current working directory

NOTE: na = "-99" tells R that any value of -99 should be recoded to missing

Back

6.12 EDA 2: Determine Necessary Manipulations





Determine Necessary Manipulations

Converting from Wide to Long Format

- Repeated observations are stored **in the wide format**, with data from each wave contained in a separate column
- Following tidy data principles, "pivot" the wide data into the **long format** using the {tidyr} package

Code

Code (Slide Layer)



Example Code

NOTE: R will look for all of the columns that start with "an", which are the anti-social scores at each wave, and combine them into one column

```
# pivot wide to long  
dat_long <- dat %>%  
  tidyr::pivot_longer(data = .,  
    cols = dplyr::starts_with("an"),  
    names_to = "wave")
```



NOTE: This symbol is called a "pipe" and tells R to feed whatever comes before it into the function after it

Wherever you see a "%%" in the function following the pipe, that's essentially a placeholder for the object that came before the pipe

NOTE: An additional column, "wave", will be created to indicate the wave in which the anti-social score was measured

Back

6.13 EDA 3: Summarize and Describe the Data



Summarize and Describe the Data

Calculating Descriptive Statistics



- **Mean and variance of the outcome** at each time point (in this case we have two time points to choose from: wave or child's age)
- **Correlation/covariance** of the outcome measured at time 1 with the outcome measured at time 2, etc.

Each of statistics can be grouped, for example:

- Describe **average change** in **anti-social behavior** among boys and girls over time (group by the sex variable, "gen")
- Explore anti-social behavior sample statistics across **levels of home cognitive stimulation** (group by the cognition variable, "homecog")

Code

Code (Slide Layer)



Example Code

```
dat_long %>%  
  dplyr::group_by(., ag) %>%  
  dplyr::summarize(  
    mean = round(mean(anti_score, na.rm = TRUE), 2),  
    var = round(var(anti_score, na.rm = TRUE), 2),  
    n = dplyr::n())
```



NOTE: We use the `dplyr::group_by` function to tell R to calculate the statistics we request in the next step within each level of the grouping variable. Here, we choose to group by “ag”, or the age of the child, rather than the “wave” of data collection.

To also group by a predictor, e.g., child’s sex, add the grouping variable of interest after “ag”, like so: `dplyr::group_by(., ag, gen)`.

NOTE: Using `dplyr::n()` tells R to calculate the sample size in each group. This is especially important in our case because we have chosen to group by age rather than wave.

Back



6.14 EDA 3: Example Table



Example Table

Child Age	Sample Size	Mean	Variance
6	122	1.57	2.78
7	168	1.55	2.42
8	146	1.97	3.23
9	192	1.89	3.82
10	151	2.14	4.56
11	174	1.79	3.80
12	135	1.84	3.17
13	173	2.24	5.16
14	101	1.96	4.36
15	8	1.12	3.55
NA	250	—	—

6.15 EDA 4: Visualize Change Over Time





Visualize Change Over Time

Graphical Analysis: Spaghetti Plots

- Spaghetti plots portray **the trajectories of the outcome** for each individual over time
- To improve readability, plot **a small random subset of individuals** rather than the whole sample
- Plots can be **split by group** to help visualize potential sources of covariate effects

Code

Code (Slide Layer)



Example Code

```
# randomly select individuals (ids) from dat_final
sample_id <- dat_final %>%
  dplyr::distinct(., person_id) %>%
  dplyr::sample_n(., size = 12) %>%
  unlist(.)
```

NOTE: This set of code randomly samples 12 individuals from our dataset to be plotted

```
# filter dat_final by sampled ids
dat_final %>%
  dplyr::filter(., person_id %in% sample_id) %>%
  ggplot2::ggplot(.) +
  ggplot2::aes(y = anti_score, x = child_age, group = person_id, color =
    assigned_sex) +
  ggplot2::geom_point() +
  ggplot2::geom_line() +
  ggplot2::facet_wrap(dplyr::vars(person_id))
```

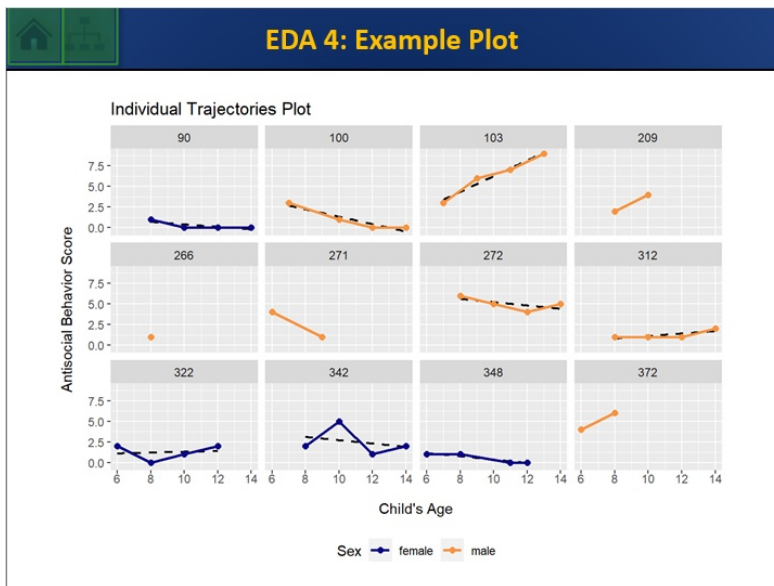
NOTE: dplyr::filter keeps only those individuals selected in "sample_id" for the plot

NOTE: ggplot2::geom_point plots the x and y values for each person; ggplot2::geom_line creates the "spaghetti" (lines for each individual) connecting each of their points

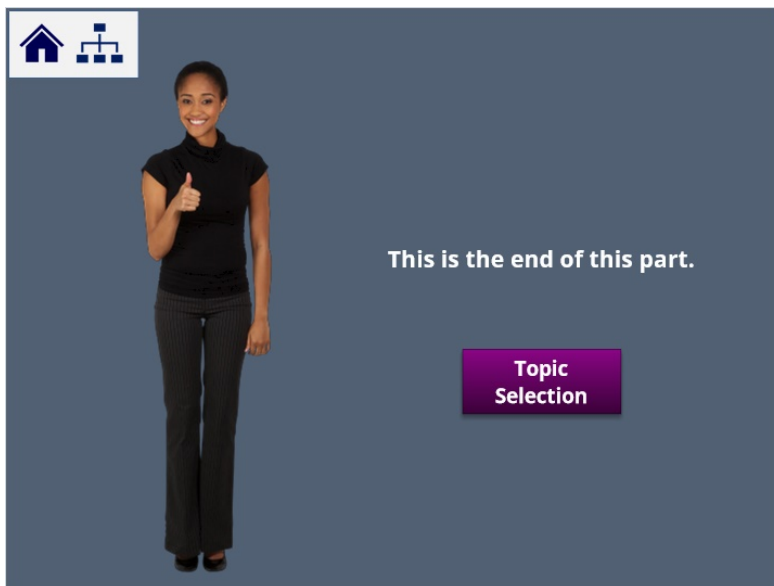
NOTE: ggplot2::facet_wrap splits the plot into frames for each individual

Back

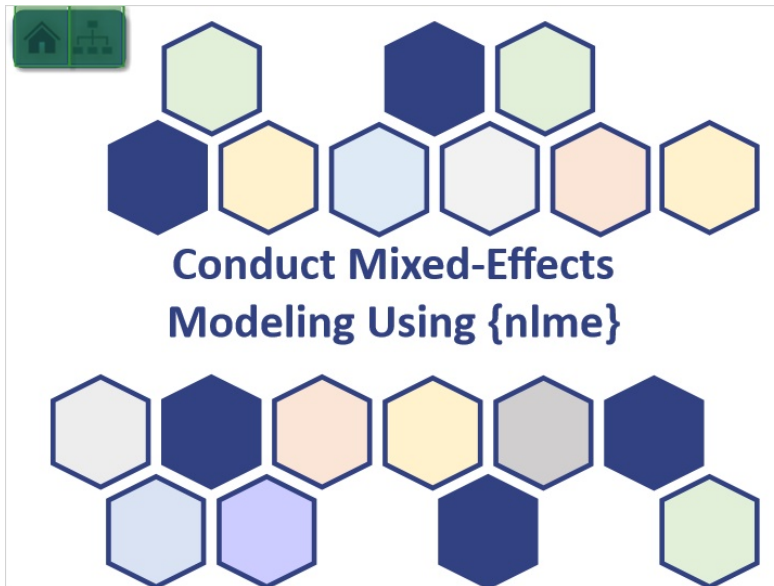
6.16 EDA 4: Example Plot





6.17 Bookend: Perform Exploratory Data Analysis



6.18 Bookmark: Conduct Mixed-Effects Modeling Using {nlme}



6.19 Graphical Eval: Model Assumptions





Graphical Evaluation: Model Assumptions

Primary Assumptions of Linear Mixed-Effects Model

- The relation between the outcome variable and time is **linear**
- The model residuals and random effects are centered around 0 and **normally distributed** with **constant variance**
- The level-1 residuals are **independent** of the level-2 random effects

The mixed-effects model **does not assume** that observations are independent, just that **residuals and random effects are independent**

6.20 Model Fit 1: Calculate the ICC





Model Fit 1: Calculate the ICC

What is an ICC for Longitudinal Data?

- The **intraclass correlation coefficient (ICC)** tells us what **proportion of variance** in a fully unconditional model exists at level-2 relative to the total variance
- A large ICC, **closer to 1**, implies **more variability** in average anti-social behavior across persons
- A smaller ICC, **closer to 0**, implies **persons are similar to one another** in anti-social behavior

Code

Code (Slide Layer)



Model Fit 1: Example Code

```
null_icc <- dat_final %>%  
  nlme::lme(  
    anti_score ~ 1,  
    data = .,  
    method = "ML",  
    random = ~ 1 | person_id,  
    control = list(maxiter = 100, returnObject = TRUE)  
  )  
# calc icc  
l1_var <- summary(null_icc)$sigma2  
l2_var <- as.numeric(nlme::VarCorr(null_icc)[1,"Intercept"], "Variance")  
icc <- l2_var / (l1_var + l2_var)
```



NOTE: "anti_score ~ 1" gives the model formula. This tells nlme::lme() to model the anti-social behavior scores with an intercept only, where the intercept is represented by the 1.

NOTE: This statement extracts the level-1 residual variance from the null_icc model object.

NOTE: This statement extracts the level-2 intercept variance from the null_icc model.

Back

6.21 Model Fit 2: Identify “Best” Model Structure





Model Fit 2: Identify “Best” Model Structure

Modeling Random Effects, Error Variances, and Error Covariances

- Start with **the least complex model** and **add terms one at a time**, estimating with **maximum likelihood**
- Utilize the **knowledge gained from EDA** to choose which terms to add (e.g., are random slopes necessary? do variances appear homogenous?)
- Compare models using information criteria**, searching for the model that **minimizes AIC/BIC**

Code

Code (Slide Layer)



Model Fit 2: Example Code

```
null_rslp <- dat_final %>%
  nlme::lme(
    anti_score ~ 1 + child_age,
    data = .,
    na.action = na.exclude,
    method = "ML",
    random = ~ 1 + child_age | person_id,
    weights = nlme::varIdent(form = ~ 1 | child_age),
    correlation = nlme::corCompSymm(
      value = -.3, form = ~ 1 + child_age | person_id
    )
    control = list(maxIter = 100, returnObject = TRUE)
  )
```



NOTE: This formula for the model random effects includes “1” for the intercept and “child_age” for the random slope of the effect of time.

NOTE: The “weights” argument in combination with the “nlme::varIdent” function tells nlme that the within-person residual errors should be allowed to be unequal across time.

NOTE: The “correlation” argument with the “nlme::corCompSymm” function tells nlme to allow the within-person covariances to have a compound symmetric structure. Other structures are available.

Back

6.22 Model Fit 3: Test Covariates of Interest





Model Fit 3: Test Covariates of Interest

Assessing Covariate Effects

- Predictors can be added to a longitudinal mixed effects model **in much the same way as with regular linear regression**, and all the usual caveats about predictor selection apply
- In the same way that we added a random effect, or random slope, to the effect of time (child_age) in the previous example, we can **add random effects to all covariates**
- The more random effects a model estimates, the more complex. Convergence and estimation issues may be **a result of over-parameterized models**

Code

Code (Slide Layer)



Model Fit 3: Example Code

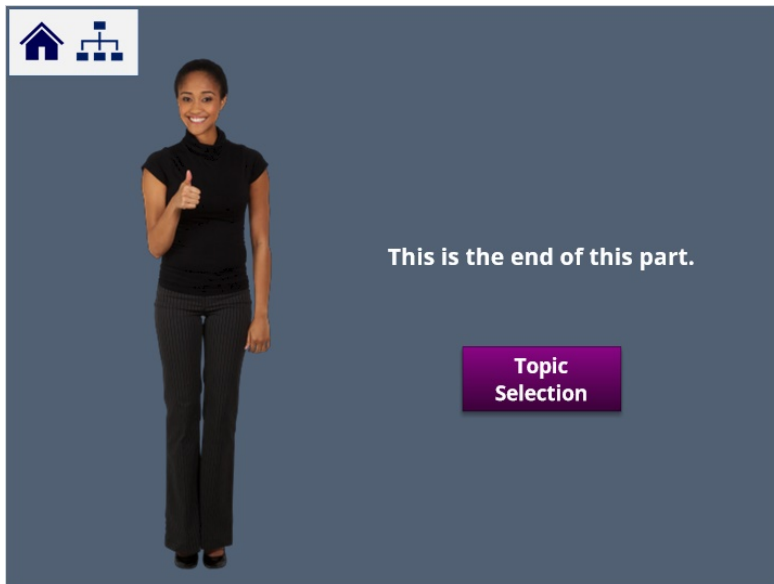
NOTE: Adding assigned_sex to the random effects structure indicates that the model will estimate a random slope for this variable. This choice would make sense if we believed that the relation between assigned sex and anti-social behavior was different across different children.

```
cond_rslp <- dat_final %>%
  nlme::lme(
    anti_score ~ 1 + child_age + assigned_sex,
    data = .,
    na.action = na.exclude,
    method = "ML",
    random = ~ 1 + child_age + assigned_sex | person_id,
    control = list(maxIter = 100, returnObject = TRUE)
  )
```

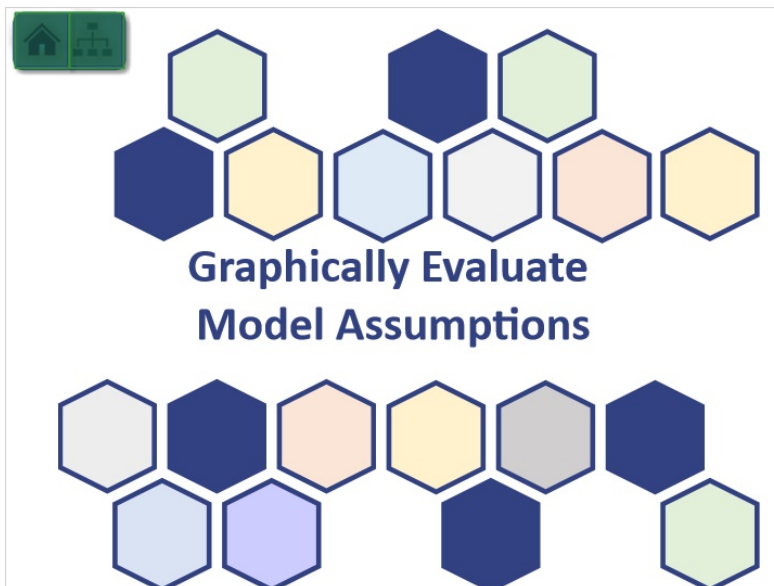
NOTE: Here, we add the predictor "assigned_sex" to our model formula to explore whether anti-social behavior trajectory differs between boys and girls.

Back



6.23 Bookend: Conduct Mixed-Effects Modeling Using {nlme}



6.24 Bookmark: Graphically Evaluate Model Assumptions



6.25 Graphical Eval: Diagnostic Plots in R





Diagnostic Plots in R

Evaluating Assumptions Using ggplot2

- **ggplot2** is a flexible tool for graphically evaluating modeling assumptions
- Plotting is most effective when estimated **residuals and random effects** can be extracted from the model object in R and re-merged with the raw data

Code

Code (Slide Layer)



```
model_eval <- cond_reml %>%  
  nlme::ranef(., condVar = TRUE) %>%  
  tibble::as_tibble(.) %>%  
  dplyr::mutate(  
    person_id = dat_final %>% dplyr::distinct(., person_id) %>% unlist(.,  
    .before = "(Intercept)"  
  ) %>%  
  dplyr::rename(., resid_int = `(Intercept)`, resid_age = child_age) %>%  
  dplyr::right_join(., dat_final, by = "person_id") %>%  
  dplyr::mutate(  
    resid_id = residuals(mod, type = "pearson"),  
    fitted_id = fitted(mod)  
  )
```

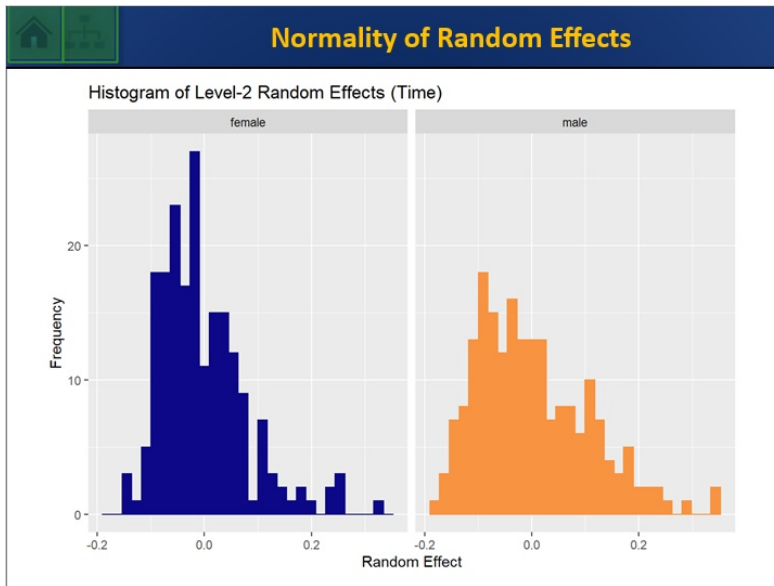
NOTE: The nlme::ranef() function extracts the estimated random effects from the model object "cond_reml".

NOTE: In the last step, level-1 residuals and fitted values are extracted using the "residuals" and "fitted" functions from base R.

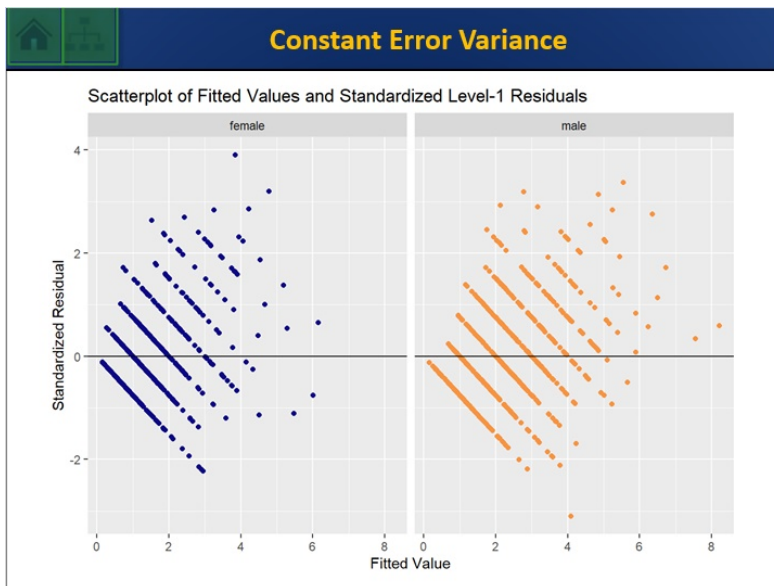
NOTE: Using the dplyr::right_join function, we can merge the level-2 random effects back with the "dat_final" dataset.

Back

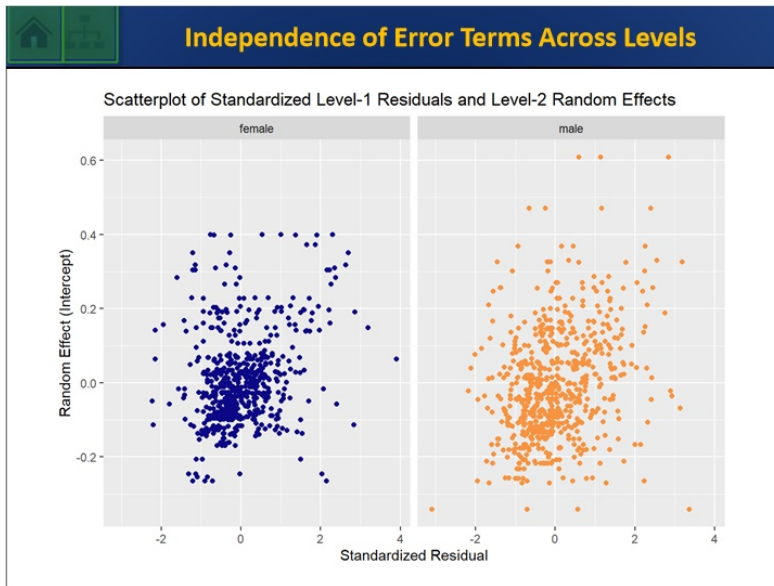
6.26 Graphical Eval: Normality of Random Effects



6.27 Graphical Eval: Constant Error Variance



6.28 Graphical Eval: Independence of Error Terms Across Levels



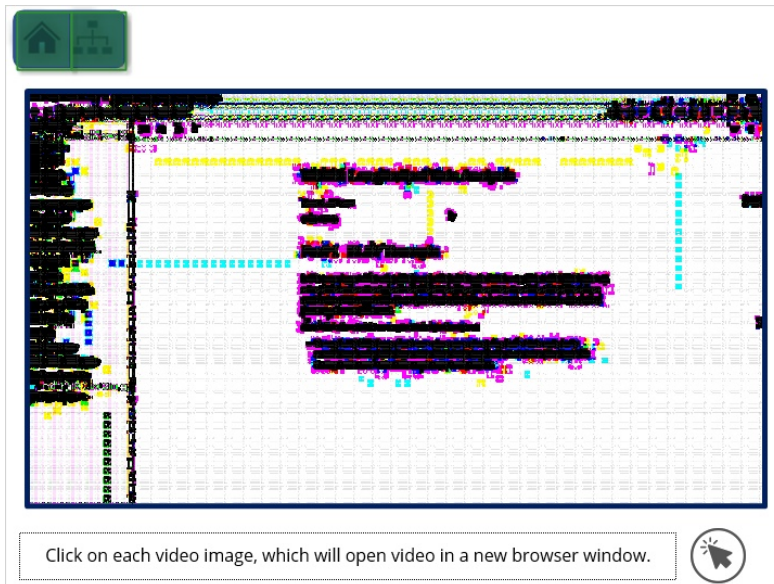
6.29 Bookend: Graphical Evaluation





6.30 Bookmark: Real Data Analysis



6.31 Example 1



6.32 Graphical Evaluation



Graphical Evaluation

- Linear mixed-effects models for longitudinal data are **similar to linear regression models** in that they impose several assumptions
- With increasing model complexity, there are both more assumptions and more opportunities to **“relax” assumptions that are untenable**

6.33 Example 2

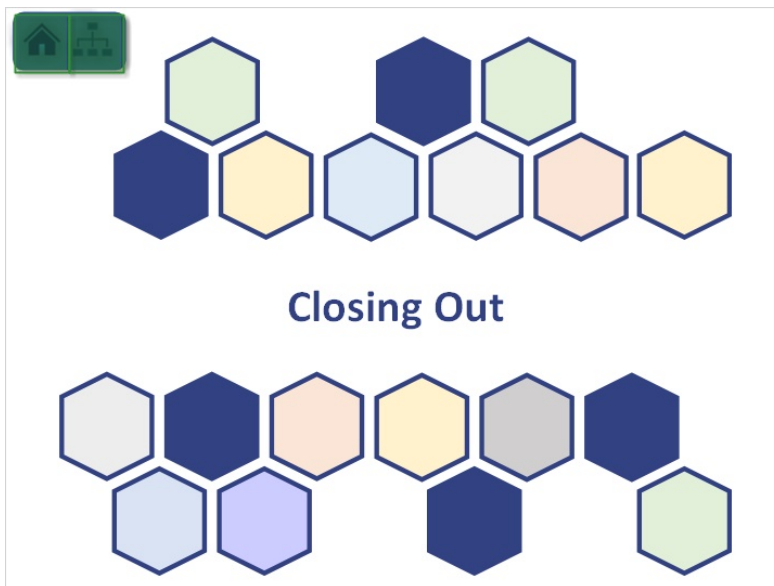




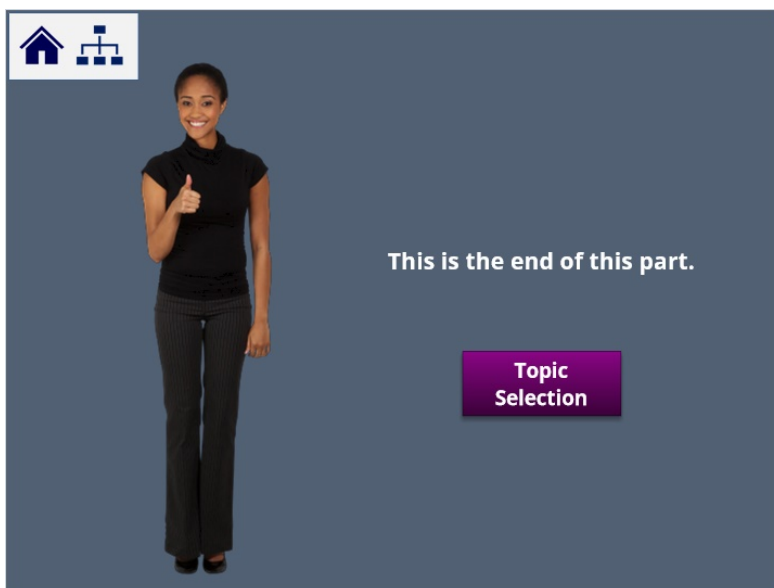
Click on each video image, which will open video in a new browser window.



6.34 End of Section



6.35 Bookend: Data Activity



6.36 Module Cover (END)

