

# Dados aninhados: introdução a *linear mixed models* (LMM)

Pedro S.R. Martins

Grupo de estudos em estatística 2023



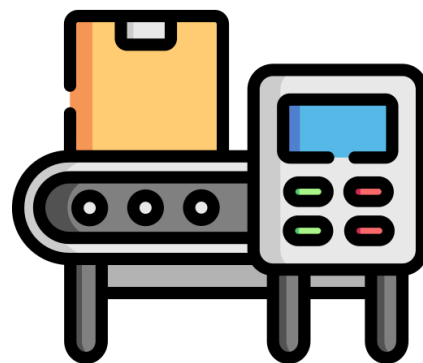
# Parte teórica

- Máquina da confusão
- Modelo linear geral
- Pressupostos regressão
- O que é multinível? (nested data)
- Correlação intraclasses
- Ignorando dados nested
- Efeitos fixos e efeitos aleatórios, uma introdução
- Modelo multinível

# Máquina da confusão

- Primeiro passo para o dia de hoje é entender a terminologia

Linear mixed model



- modelo de coeficiente aleatório
- modelo linear misto
- modelo linear hierárquico
- modelo multinível
- multilevel model
- random coefficient model
- hierarchical linear model



# Um passo para trás

## Modelo linear geral

- O primeiro passo vai ser uma breve revisão da regressão

$$Y = b_0 + b_1X_1 + e$$

### Intercepto

- *Ponto onde a reta toca o zero*
- *Ponto estimado quando todos os preditores são zero*
- *Média condicional de Y na ausência de outros preditores*

### Slope/inclinação

- *Taxa de crescimento de Y em função do X*
- *A cada um ponto em X, Y cresce slope (b) pontos (em modelos **lineares**)*

# Pressupostos regressão

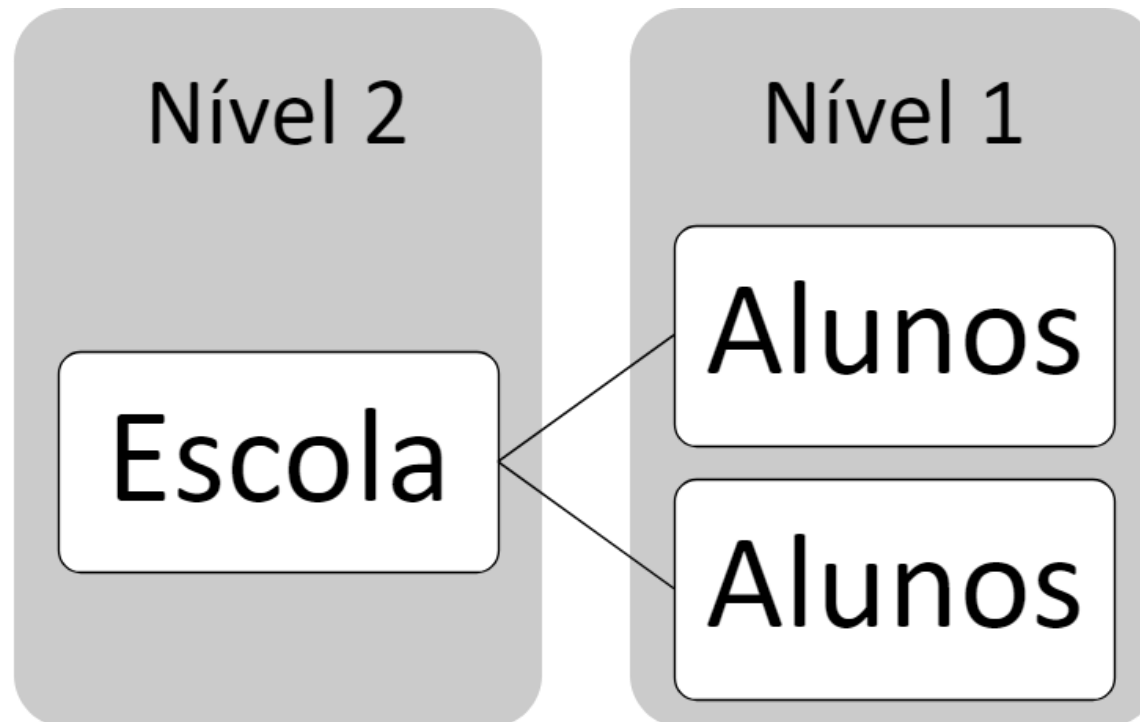
- Um modelo de regressão é extremamente restritivo.
- Para que um modelo linear simples funcione precisamos cumprir vários requisitos
- Um dos principais pressupostos é:
  - Erros independentes (os resíduos dos preditores não podem estar correlacionados entre si)

# Pressupostos regressão

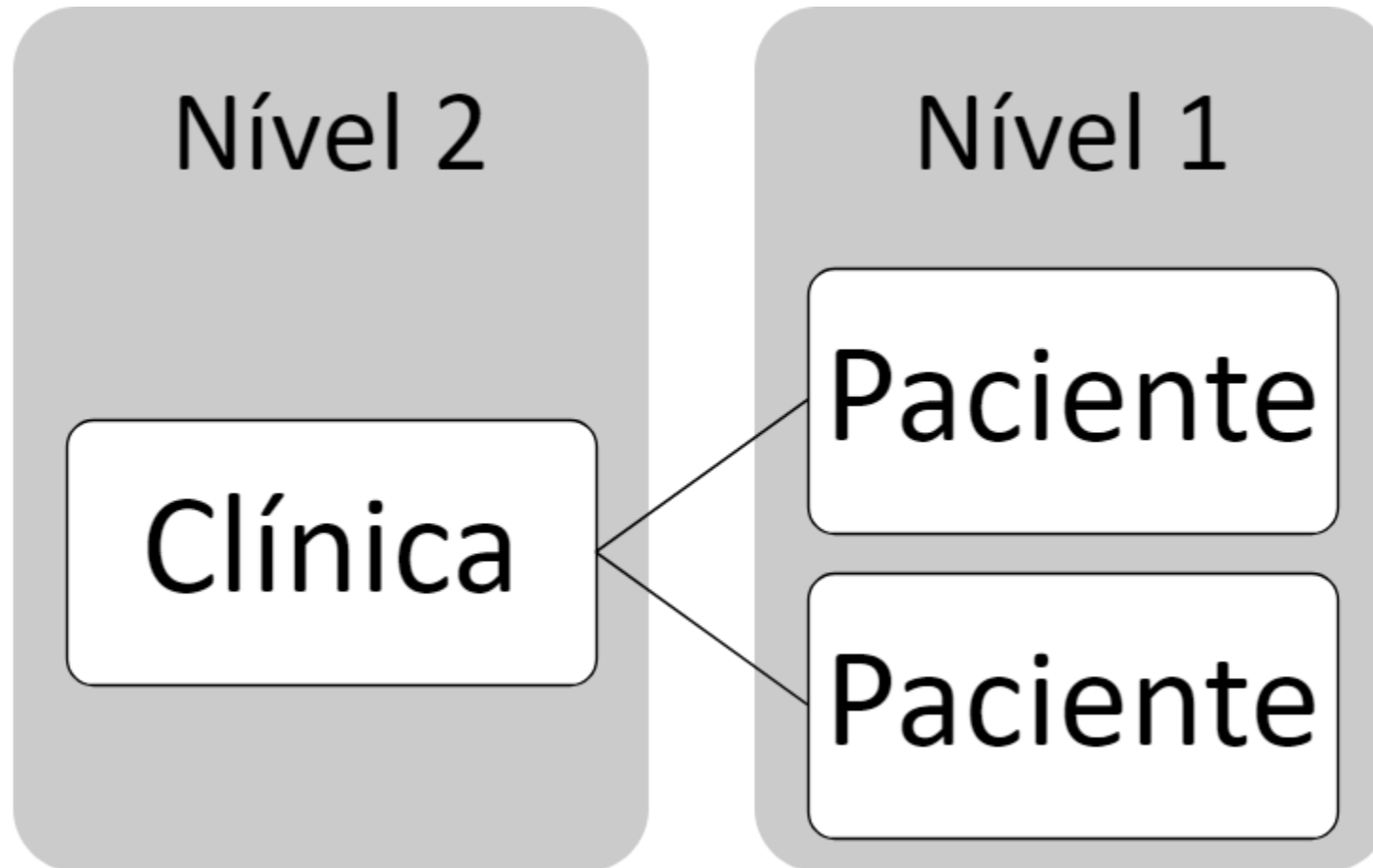
- Uma coisa que sempre pode dar pau é o pressuposto dos erros não correlacionados
- **O que pode fazer com que erros sejam correlacionados?**
  - Participantes mais similares entre si do que com os demais da amostra
  - O fato de um sujeito participar da pesquisa interfere no fato de outro sujeito participar
  - **Medidas repetidas**

# O que é multinível?

- **Sinônimos:** nested data, dados hierárquicos
- Os dados de pesquisa são, normalmente, hierárquicos ou *nested*

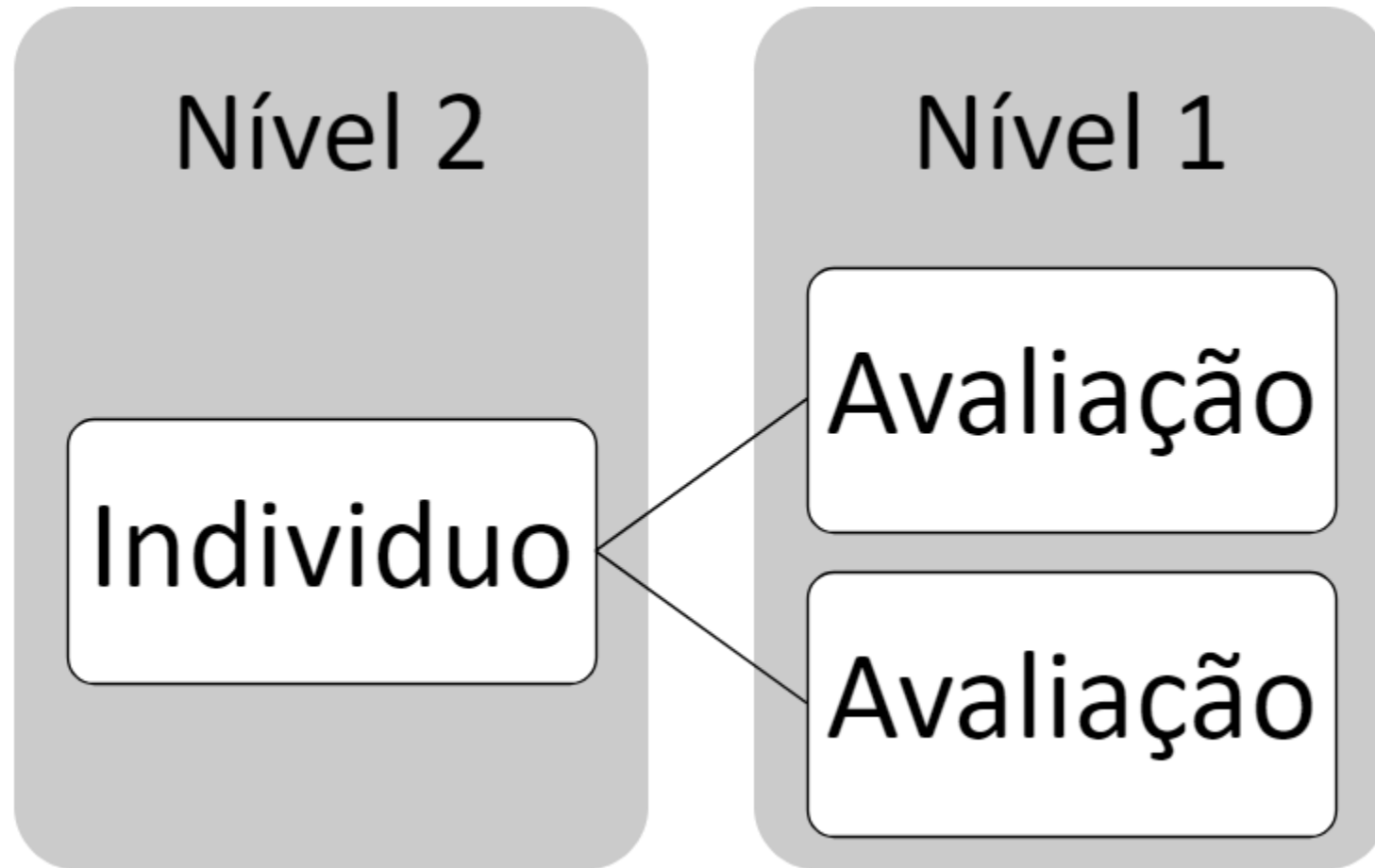


# O que é multinível?





# O que é multinível



# Correlação intra classe (ICC)

- Como saber se uma variável de agrupamento é importante?
- Nem sempre um dado *nested* é tão importante que precisa ser tratado como tal
- A importância do agrupamento é medida por meio da correlação intra-classe (ICC)



Intra class correlation

# ICC

- O cálculo da ICC pode ser feito com diversas equações e em diferentes contextos
- O termo não é exclusivo da aplicação de LMM

Journal of Chiropractic Medicine (2016) 15, 155–163



JCM  
JOURNAL of  
CHIROPRACTIC MEDICINE  
www.journalchirod.com

Cracking the Code: Providing Insight Into the Fundamentals  
of Research and Evidence-Based Practice

## A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research



Terry K. Koo, PhD<sup>a,\*</sup>, Mae Y. Li, BPS<sup>b</sup>

<sup>a</sup> Director & Associate Professor, Foot Levelers Biomechanics Research Laboratory, New York Chiropractic College, Seneca Falls, NY

<sup>b</sup> DC Candidate, Foot Levelers Biomechanics Research Laboratory, New York Chiropractic College, Seneca Falls, NY

Received 30 July 2015; received in revised form 3 November 2015; accepted 9 November 2015

**Key indexing terms:**  
Reliability and validity;  
Research;  
Statistics

### Abstract

**Objective:** Intraclass correlation coefficient (ICC) is a widely used reliability index in test-retest, intrarater, and interrater reliability analyses. This article introduces the basic concept of ICC in the content of reliability analysis.

**Discussion for Researchers:** There are 10 forms of ICCs. Because each form involves distinct assumptions in their calculation and will lead to different interpretations, researchers should explicitly specify the ICC form they used in their calculation. A thorough review of the research design is needed in selecting the appropriate form of ICC to evaluate reliability. The best practice of reporting ICC should include software information, “model,” “type,” and “definition” selections. **Discussion for Readers:** When coming across an article that includes ICC, readers should first check whether information about the ICC form has been reported and if an appropriate ICC form was used. Based on the 95% confident interval of the ICC estimate, values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively.

**Conclusion:** This article provides a practical guideline for clinical researchers to choose the correct form of ICC and suggests the best practice of reporting ICC parameters in scientific publications. This article

# ICC

- No contexto de modelos multinível o ICC assume uma forma genérica de:

$$ICC = \frac{var_x}{var_x + var_{erro}}$$

- Valores de ICC são restritos para estarem entre 0 e 1 ICC [0,1]
- O ICC indica quanto do desfecho total pode ser atribuído à variável de agrupamento
- ICC acima de 0,10 podem ser considerados como relevantes
  - Caso uma variável **tenha que ser tratada como variável de agrupamento** (i.e., em casos de dados longitudinais) essa regra de bolso para valores relevantes de ICC deixa de ser tão importante
  - A teoria é importante.

# Ignorando dados nested

O que ocorre quando ignoramos fonte de correlação entre os erros dos nossos participantes?

- A correlação vira uma estimativa inapropriada dos erros padrão



estimativas enviesadas dos valores de  $p$

- Não conseguimos nunca cumprir os pressupostos do modelo

# Introdução a efeitos fixos e efeitos aleatórios

- As estatísticas tradicionais trabalham com efeitos fixos
- Ou seja, até agora, tudo que vimos foram os efeitos fixos

# Introdução a efeitos fixos e efeitos aleatórios

## Efeito fixo

- Efeitos de variáveis em que todos os níveis foram incluídos no experimento (i.e., sexo)
- O efeito não varia significativamente entre os participantes

## Efeito aleatório

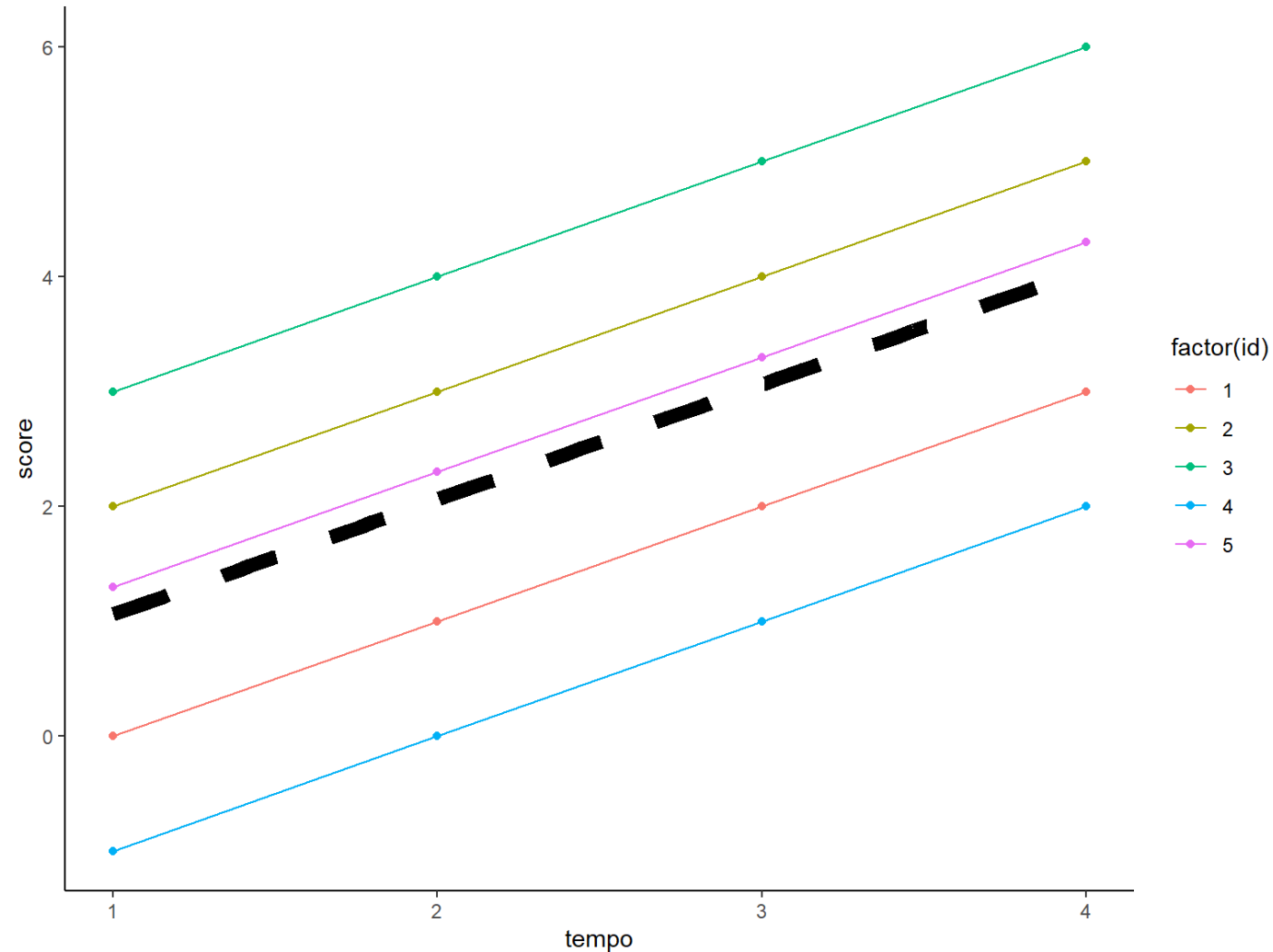
- Efeitos de variáveis de "**agrupamento**" e que incluem **apenas uma parcela** de todos os níveis possíveis.
  - *Por exemplo, normalmente os pesquisadores querem investigar todos os humanos, mas incluem apenas uma parcela da população nos experimentos. Portanto, os participantes são fatores aleatórios*
- O efeito varia significativamente entre os participantes



# Variabilidade no slope e intercepto

Intercepto aleatório  
(random intercept):

considera variação no ponto de partida



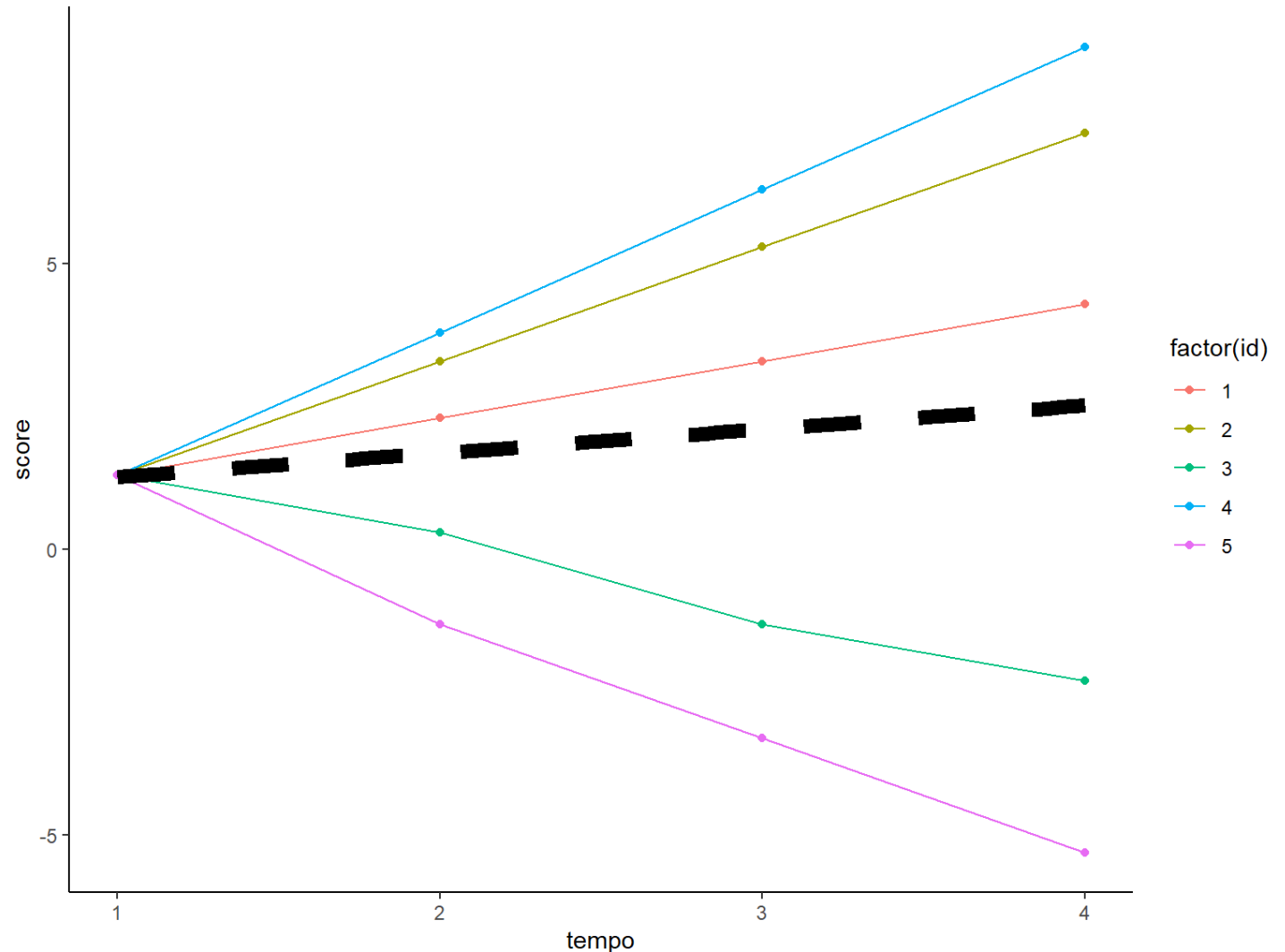


# Variabilidade no slope e intercepto

Slope aleatório

(Random slope)

considera a variação no crescimento

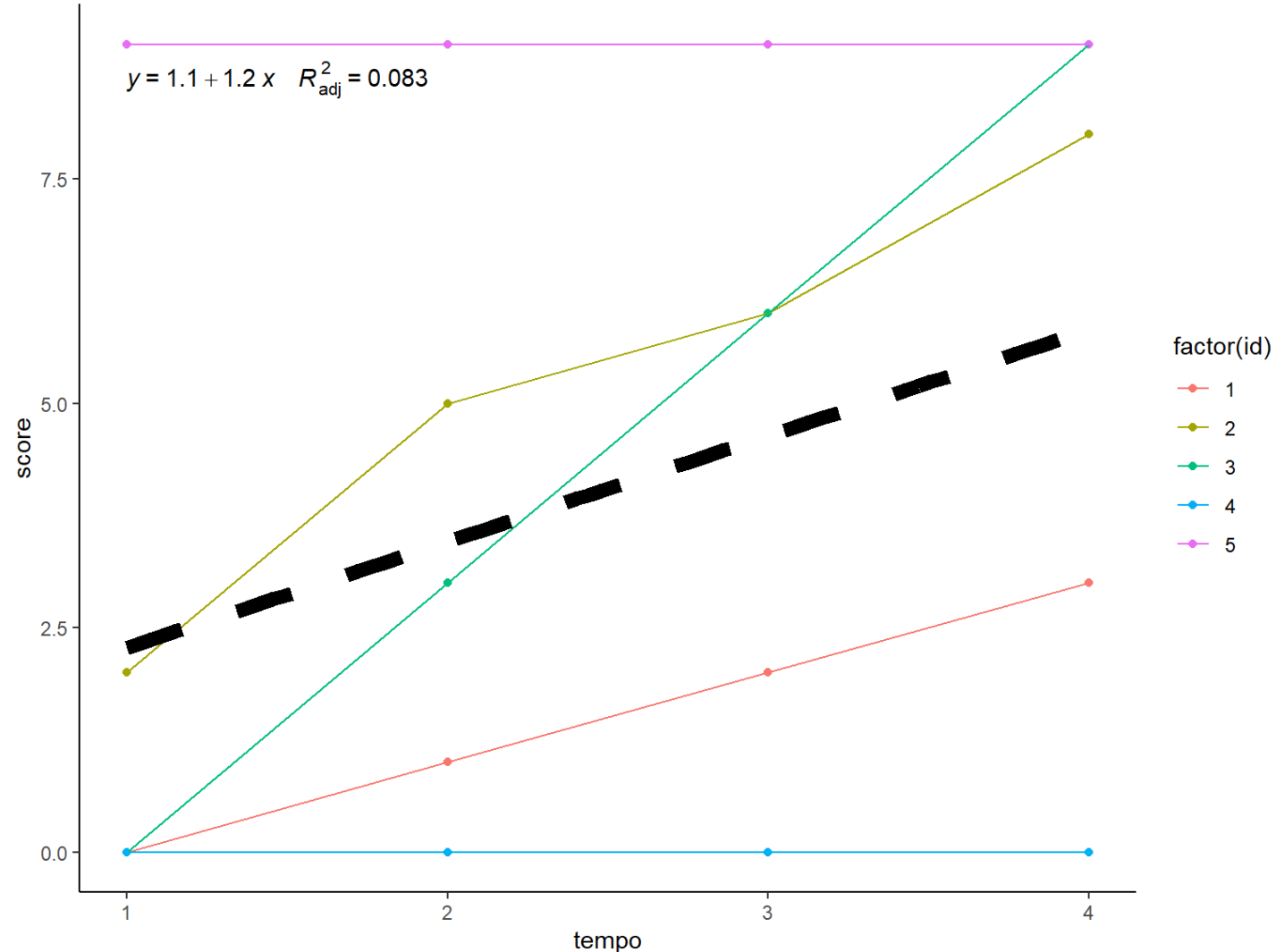


# Variabilidade no slope e intercepto

Slopes e interceptos  
aleatórios

(random slopes and random  
intercepts):

as pessoas começam em  
lugares diferentes e têm taxas  
de crescimento diferentes



# Como interpretar?

- No programa, vamos ver esses padrões nas variâncias do slope e intercepto

## Variância no intercepto

- A variabilidade no ponto de partida
- Alta variância: pessoas começam em lugares muito diferentes
- Baixa variância: as pessoas começam em um lugar parecido

## Variância no slope

- A variabilidade na mudança ao longo do tempo
- Alta variância: pessoas têm taxas de crescimento muito diferentes
- Baixa variância: pessoas têm taxas de crescimento parecidas

# Referências

- Finch, W. H., Bolin, J. E., & Kelley, K. (2019). Multilevel modeling using R. Crc Press.
- Gelman, A. (2006). Multilevel (hierarchical) modeling: What It can and cannot do. *Technometrics*, 48(3), 432–435.  
<https://doi.org/10.1198/004017005000000661>
- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*.  
*Discovering Statistics Using IBM SPSS Statistics*.

# Recursos extras

- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of personality and social psychology*, 103(1), 54–69.  
<https://doi.org/10.1037/a0028347>
- [http://mfviz.com/hierarchical-models/?fbclid=IwAR2zLJIN5O2QlhpVp8BESPUmicnlfujAGq\\_EKC9e7BYZCKqUyW0yEMeQB7Y](http://mfviz.com/hierarchical-models/?fbclid=IwAR2zLJIN5O2QlhpVp8BESPUmicnlfujAGq_EKC9e7BYZCKqUyW0yEMeQB7Y)
- <https://www.scielo.br/j/pope/a/5FqmXZfgcfMWpyJZ8HvCVLq/?lang=pt>

# Recursos extras



*Annual Review of Psychology*

## Catching Up on Multilevel Modeling

Lesia Hoffman<sup>1</sup> and Ryan W. Walters<sup>2</sup>

<sup>1</sup>Department of Psychological and Quantitative Foundations, University of Iowa, Iowa City, Iowa 52242, USA; email: Lesia-Hoffman@UIowa.edu

<sup>2</sup>Department of Clinical Research, Creighton University, Omaha, Nebraska 68178, USA

<https://doi.org/10.1146/annurev-psych-020821-103525>

Special Section: Using Simulation to Convey Statistical Concepts  
*Tutorial*



Advances in Methods and Practices in Psychological Science  
January-March 2021, Vol. 4, No. 1,  
pp. 1–15  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2515245920965119  
[www.psychologicalscience.org/AMPPS](http://www.psychologicalscience.org/AMPPS)



## Understanding Mixed-Effects Models Through Data Simulation



Lisa M. DeBruine<sup>ID</sup> and Dale J. Barr<sup>ID</sup>  
Institute of Neuroscience & Psychology, University of Glasgow

### Abstract

Experimental designs that sample both subjects and stimuli from a larger population need to account for random effects of both subjects and stimuli using mixed-effects models. However, much of this research is analyzed using analysis of variance on aggregated responses because researchers are not confident specifying and interpreting mixed-effects models. This Tutorial explains how to simulate data with random-effects structure and analyze the data using linear mixed-effects regression (with the *lme4* R package), with a focus on interpreting the output in light of the simulated parameters. Data simulation not only can enhance understanding of how these models work, but also enables researchers to perform power calculations for complex designs. All materials associated with this article can be accessed at <https://osf.io/3cz2e/>.

### Keywords

simulation, mixed-effects models, power, *lme4*, R, open materials

Received 6/2/19; Revision accepted 9/5/20

<https://doi.org/10.1177/2515245920965119>

# Recursos extras

MINI REVIEW article

Front. Psychol., 22 January 2015

Sec. Quantitative Psychology and Measurement

Volume 6 - 2015 | <https://doi.org/10.3389/fpsyg.2015.00002>

## Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui)



David A. Magezi\*

Neurology Unit, Laboratory for Cognitive and Neurological Sciences, Department of Medicine, Faculty of Science, University of Fribourg, Fribourg, Switzerland

<https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00002/full>

# Modelo multinível - adendo

- Como é a equação de um modelo misto?
- Partindo de um modelo linear com todos os efeitos fixos temos:

$$(1) - y_i = \beta_0 + \beta_1 X_1 + e_i$$

- Se temos uma variável de agrupamento, a equação será modificada com a inserção de um subscrito

$$(2) - y_{ij} = \beta_{0j} + \beta_{1j} X_1 + e_{ij}$$



# Modelo multinível - adendo

- Para modelos com intercepto aleatório, o intercepto é expandido para ter um parâmetro extra que indica sua variabilidade

$$\text{Nível1} - y_{ij} = \beta_{0j} + \beta_{1j}X_1 + e_{ij}$$

$$\text{Nível2} - \beta_{0j} = \gamma_{00} + U_{0j}$$

# Modelo multinível - adendo

- Para modelos com slope aleatório, o slope é expandido para ter um parâmetro extra que indica sua variabilidade

$$\text{Nível1} - y_{ij} = \beta_{0j} + \beta_{1j}X_1 + e_{ij}$$

$$\text{Nível2} - \beta_{0j} = \gamma_{00} + U_{0j}$$

$$\beta_{1j}X_1 = \gamma_{10}x_{ij} + U_{1j}x_{ij}$$