

MAT02035 - Modelos para dados correlacionados

Introdução aos dados longitudinais e agrupados

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Porto Alegre, 2019

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Introdução

A pesquisa estatística em métodos para o delineamento e análise de investigações humanas expandiu explosivamente na segunda metade do século 20.

- Nos EUA (metade do século 20): mudança de suporte de pesquisa da área militar para a área biomédica.
- Nos EUA (1944): aprovação da Lei do Serviço de Saúde Pública ⇒ crescimento do *National Institutes of Health* (NIH¹).
 - Orçamento do NIH: 1947 - \$8 milhões ⇒ 1966 - \$1 bilhão.

¹Um pouco da história do NIH: www.nih.gov/about-nih/who-we-are/history

- O NIH patrocinou vários dos estudos epidemiológicos importantes e ensaios clínicos daquele período, incluindo o ***Framingham Heart Study***² (FHS).
- O foco destes primeiros estudos foi a **morbidade** e, especialmente, a **mortalidade**.
 - Pesquisadores procuravam **identificar as causas** da morte prematura e avaliar a efetividade dos tratamentos para atrasar a morte morbidade.
 - **Regressão logística** (1960s).
 - A análise de dados de **tempo até o evento** foi revolucionada pelo artigo de 1972 de **D. R. Cox**³, descrevendo **modelo de riscos proporcionais**⁴.

²Um pouco da história do NIH: www.ncbi.nlm.nih.gov/pmc/articles/PMC4159698/

³Reid, N. A Conversation with Sir David Cox. *Statistical Science*, 9:439-455, 1994.

⁴**Leituras para o fim do dia:** Cox, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B*, 34:187-220, 1972.

- Embora o delineamento do FHS exigisse a **medição periódica** das características do paciente, consideradas determinantes de doenças crônicas, o interesse pelos níveis e **padrões de mudança** dessas características **ao longo do tempo** foi inicialmente limitado.
- Com avanços da pesquisa, investigadores começaram a fazer perguntas sobre o **comportamento** desses **fatores de risco**.
 - No FHS os pesquisadores começaram a perguntar se os **níveis de pressão arterial** na infância eram preditivos de hipertensão na vida adulta.
 - No *Coronary Artery Risk Development in Young Adults (CARDIA) Study*⁵, os investigadores procuraram identificar os determinantes da transição do estado normotenso⁶ ou normocolesterolêmico⁷ no início da vida adulta para hipertensão e hipercolesterolemia na meia-idade.

⁵Friedman, G. D *et al.* CARDIA: study design, recruitment, and some characteristics of the examined subjects. *Journal of Clinical Epidemiology*, 41:1105-1116, 1988.

⁶“pressão arterial normal”

⁷“colesterol normal”

- No tratamento de artrite, asma e outras doenças que normalmente não ameaçam a vida, os pesquisadores começaram a estudar os efeitos dos tratamentos na mudança ao longo do tempo em medidas de gravidade da doença.
 - Questões semelhantes estavam sendo colocadas em todos os contextos de doenças.
- Os pesquisadores começaram a **acompanhar populações** de todas as idades ao longo do tempo, tanto em estudos observacionais quanto em ensaios clínicos, para **entender o desenvolvimento e a persistência da doença e para identificar fatores que alteram o curso do desenvolvimento da doença.**

Introdução

- Esse interesse nos padrões temporais de mudança nas características humanas ocorreu em um período em que os avanços no poder da computação tornaram novas e mais intensivas abordagens computacionais para a análise estatística disponíveis no desktop.



- **Laird e Ware** propuseram o uso do algoritmo EM para ajustar uma classe de **modelos lineares de efeitos mistos** apropriados para a análise de medidas repetidas⁸;
- Mais tarde na década, **Liang e Zeger** introduziram as **equações de estimativas generalizadas** na literatura bioestatística e propuseram uma família de modelos lineares generalizados para ajustar observações repetidas de dados binários e contados^{9,10}.

⁸Laird, N. M., Ware, J. H. Random-effects models for longitudinal data. *Biometrics*, 38:963-974, 1982.

⁹Liang, K. Y., Zeger, S. L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73:13-22, 1986.

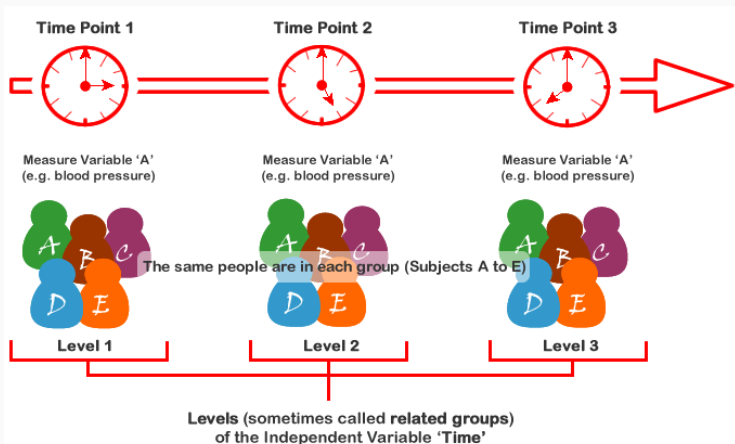
¹⁰Zeger, S. L., Liang, K. Y. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, 42:121-130, 1986.

- Muitos outros pesquisadores que escrevem na literatura biomédica, educacional e psicométrica contribuíram para o rápido desenvolvimento de metodologia para a análise desses dados “longitudinais”.
- Os últimos 40 anos assistiram a progressos consideráveis no desenvolvimento de métodos estatísticos para a análise de dados longitudinais.
- Este curso apresentará parte desta literatura de maneira rigorosa apontando para as possibilidades de aplicação destas técnicas.

Dados longitudinais e agrupados

Estudos longitudinais

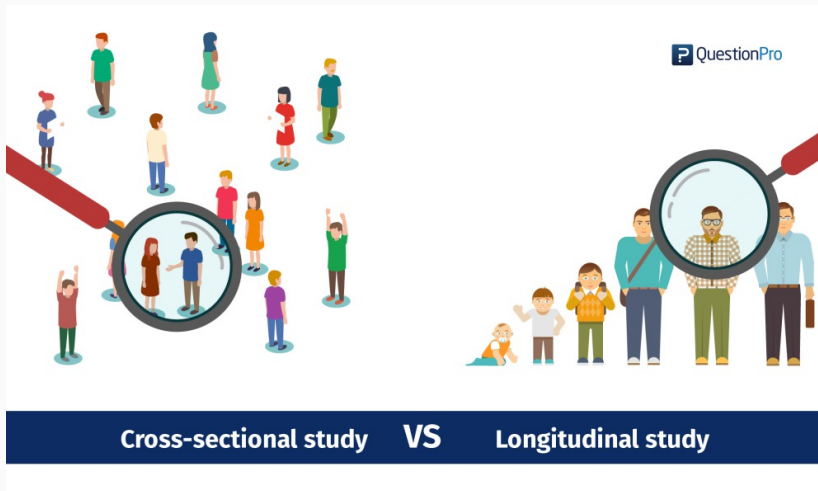
- A característica definidora de estudos longitudinais é que medidas dos mesmos indivíduos são tomadas **repetidamente** através do tempo, assim permitindo o estudo direto da **mudança com o tempo**.



Estudos longitudinais vs. estudos transversais

- Com medidas repetidas nos indivíduos, é possível capturar a **mudança dentro de indivíduo**.
- Em um **estudo transversal**, em que a resposta é medida em apenas uma ocasião, é possível apenas estimar **diferenças** nas respostas **entre indivíduos**.
 - Permite comparações entre subpopulações que diferem em idade, mas não fornece qualquer informação a respeito de como os indivíduos mudam durante o correspondente período.

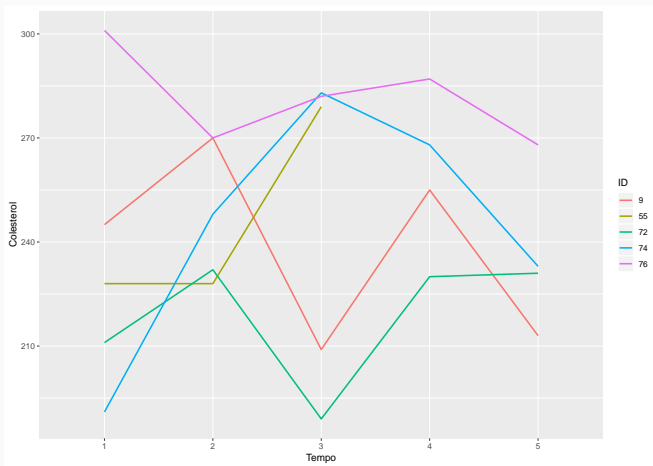
Estudos longitudinais vs. estudos transversais



Um exemplo:

Dados longitudinais são dados agrupados

- Uma característica distintiva de dados longitudinais é que eles são **agrupados**.



- Cada indivíduo forma um grupo de observações repetidas ao longo do tempo.

Dados agrupados

- Observações dentro de um grupo tipicamente irão exibir **correlação positiva**, e esta correlação **deve ser levada em conta** na análise.
- Dados agrupados podem surgir de amostragens aleatórias de grupos que se caracterizam naturalmente na população:
 - Famílias;
 - Domicílios;
 - Enfermarias hospitalares;
 - Práticas médicas;
 - Vizinhanças;
 - Escolas.
- Ainda, dados agrupados podem surgir quando a resposta de interesse é simultaneamente obtida de múltiplos avaliadores (juízes, examinadores) ou de diferentes instrumentos de medição.

- Nestes casos, esperamos que medidas em unidades dentro de um grupo são mais similares que as medidas em unidades de grupos diferentes.
- O grau de agrupamento pode ser expresso em termos da **correlação** entre as medidas em unidades do mesmo grupo.
- Esta correlação **invalida a suposição crucial de independência** que é o pilar de tantas técnicas estatísticas.
 - Modelos estatísticos para dados correlacionados devem explicitamente descrever e levar em conta para esta correlação.

- Dados longitudinais são um caso especial de dados agrupados (com uma ordenação natural das medições dentro de um grupo),
 - Faremos a descrição dos métodos de análise para dados agrupados, mais amplamente definidos.
- Um dos objetivos deste curso é demonstrar que os métodos para a análise de dados longitudinais são, mais ou menos, casos especiais de métodos de regressão mais gerais para dados agrupados.
 - Como resultado, uma compreensão abrangente de métodos para a análise de dados longitudinais fornece a base para uma compreensão mais ampla de métodos para analisar a ampla gama de dados agrupados.

- Os exemplos descritos anteriormente consideram apenas um único nível de agrupamento.
- Mais recentemente, pesquisadores desenvolveram metodologia para a análise de **dados multiníveis**, em que as observações podem ser agrupadas em mais de um nível.
 - Os dados podem consistir em medições repetidas em pacientes agrupados por clínica.
 - Os dados podem consistir em observações sobre crianças aninhadas dentro de salas de aula, aninhadas dentro das escolas.
- Os dados multiníveis serão discutidos no final do curso.

Exemplos

Modelos de regressão para respostas correlacionadas

Por hoje é só! Sejam todos bem-vindos!

