

# CE075 - Análise de Dados Longitudinais

Silva, J.L.P.

07 de agosto, 2019

# Introdução aos Estudos Longitudinais

# Estudos Longitudinais

- 1 Indivíduos foram divididos aleatoriamente em dois grupos (A e B) e a pressão sistólica foi medida em 5 tempos distintos.

(\*\*Dados Longitudinais\*\*)

- 2 Indivíduos foram divididos aleatoriamente em dois grupos (A e B) e foi registrado o tempo até a pressão sistólica atingir um certo valor/patamar.

(Análise de Sobrevivência)

- 3 Uma série histórica (200 valores) de medidas de pressão sistólica foi registrada para o Sr. João.

(Séries Temporais)

# Medidas Repetidas/Dados Longitudinais

## Medidas Repetidas

Medidas Repetidas são obtidas quando uma resposta é medida repetidamente em um grupo de unidades.

- Unidades: indivíduos, pacientes, animais, plantas, conglomerados (família, empresa, cidade, etc), etc.
- Caso especial: Dados Longitudinais: são medidas ao longo do tempo em uma mesma unidade/indivíduo.

# Tipos de Estudos

- Transversal: uma única resposta é medida em cada unidade em um certo instante de tempo.
- Longitudinal: unidades são, geralmente, medidas repetidamente ao longo do tempo.

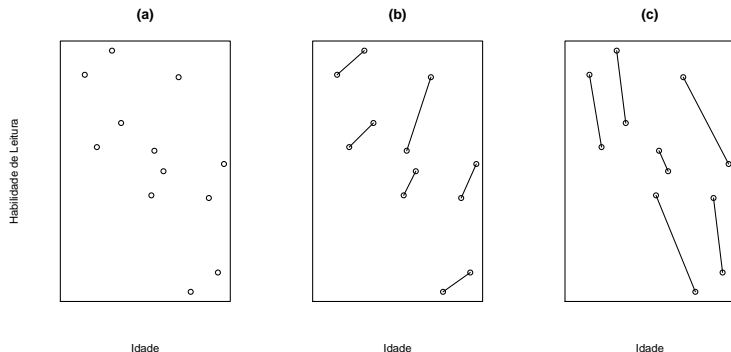
Obs.: Medidas repetidas podem estar associadas a estudos transversais. Unidades podem estar agrupadas em conglomerados.

# Estudo Transversal vs Longitudinal

Embora frequentemente seja possível responder as mesmas questões científicas com um estudo longitudinal ou transversal, a grande vantagem de um estudo longitudinal é sua capacidade de separar os chamados efeitos de *coorte* e *idade*.

A ideia é ilustrada na figura a seguir que mostra a habilidade de leitura em função da idade.

# Estudo Transversal vs Longitudinal



**Figura 1:** Habilidade de Leitura: transversal (a) vs longitudinal (b) e (c). Fonte: Diggle et al. (2002).

# Estudo Transversal vs Longitudinal

- Fig. 1(a): os dados foram obtidos de um estudo transversal. A habilidade de leitura parece ser pior entre as crianças mais velhas.
- Fig. 1(b): os mesmos dados foram obtidos de um estudo longitudinal no qual cada indivíduo foi medido duas vezes. Agora fica claro que, enquanto crianças mais novas começam em níveis mais altos de leitura, todas melhoram com o tempo. Tal padrão poderia ser resultante da introdução de educação elementar em uma comunidade rural pobre começando pelas crianças mais novas.
- Fig. 1(c). Os padrões longitudinal e transversal agora contam a mesma história - que a habilidade de leitura piora com a idade.



# Estudo Transversal vs Longitudinal

Este exemplo nos diz que estudos longitudinais (Fig. 1.(b) e (c)) conseguem distinguir mudanças no tempo dentro dentro dos indivíduos (efeito da idade) de diferenças entre as pessoas em seus valores *baseline* (efeitos de coorte).

Os estudos transversais, por sua vez, não têm esta propriedade.

Trabalharemos com estes conceitos posteriormente na disciplina.

# Estudos Longitudinais

Dados longitudinais podem ser coletados ou prospectivamente (seguindo os indivíduos no tempo) ou retrospectivamente (extraíndo múltiplos registros do indivíduo de registros históricos). Os métodos que veremos são aplicáveis a ambas situações.

A coleta prospectiva é mais comum já que a qualidade das medidas repetidas de registros passados ou da memória do indivíduo pode ser inferior.

Dados longitudinais requerem métodos estatísticos especiais porque o conjunto de observações dentro de um mesmo indivíduo tende a ser correlacionados. Esta correlação deve ser levada em conta para obtermos inferências válidas.

# Estudos Longitudinais

Dados correlacionados também ocorrem em uma série temporal longa de medidas repetidas. A análise de dados longitudinais tende a ser mais simples porque os indivíduos podem (geralmente) ser considerados independentes.

Inferências válidas podem ser feitas tomando emprestado informação entre os indivíduos. Ou seja, a consistência de um padrão entre os indivíduos é a base das conclusões.

Por isso, as inferências de estudos longitudinais podem ser mais robustas às suposições do modelo do que aquelas das séries temporais, particularmente quanto à natureza das correlações.

# Estudos Longitudinais

Sociólogos e economistas frequentemente se referem a dados longitudinais como *dados em painel*.

Embora os métodos estatísticos que veremos sejam aplicáveis à análise de dados das ciências sociais, a ênfase será motivada pela experiência em estudos longitudinais provenientes das ciências da saúde.

# Vantagens de Estudos Longitudinais

## Em Comparação com Estudos Transversais

Primeira vantagem: como as medidas na mesma unidade são correlacionadas, os estudos longitudinais são mais poderosos que os estudos transversais para um número fixo de indivíduos.

Ou seja, menos indivíduos são necessários para alcançar um nível similar de poder estatístico.

Medidas repetidas de um mesmo indivíduo fornece mais informação independente que uma única medida obtida de um único indivíduo.

## Em Comparação com Estudos Transversais

Segunda vantagem: cada indivíduo pode servir como seu próprio controle.

Por exemplo, em um estudo cross-over, cada indivíduo recebe ambas as condições experimental e controle.

Em geral, a variabilidade intra-indivíduo é consideravelmente menor que a variabilidade entre-indivíduos, o que resulta em testes estatísticos mais sensíveis ou poderosos.

## Em Comparação com Estudos Transversais

A intervenção de interesse pode ser tempo-dependente, assim as mudanças intra-indivíduo na intervenção podem estar relacionadas com mudanças na variável de interesse dentro dos indivíduos.

Temos a exclusão da variabilidade entre-indivíduos do termo de erro que resulta em estimadores mais eficientes dos efeitos do tratamento comparado com desenhos transversais com o mesmo número e padrão de observações.



## Em Comparação com Estudos Transversais

Terceira vantagem: permite separar os efeitos de idade dos efeitos de coorte. Sem dados longitudinais, não podemos fazer esta distinção.

Por fim, dados longitudinais podem fornecer informação sobre mudanças individuais, o que não ocorre em estudos transversais.

Estimativas de tendências individuais podem ser usadas para melhor compreender a heterogeneidade na população e os determinantes de crescimento e mudança em nível de indivíduo.

# Desafios de Estudos Longitudinais

# Desafios de Estudos Longitudinais

As observações não são independentes e devemos levar em conta essa dependência usando métodos estatísticos mais sofisticados.

Em certos casos, por exemplo para modelos não lineares para dados binários, ordinais ou nominais, a estimação dos parâmetros pode ser computacionalmente intensiva devido à necessidade de métodos numéricos ou de simulação Monte Carlo para avaliação da verossimilhança dos modelos de regressão.

# Desafios de Estudos Longitudinais

Uma complicação adicional é a invariável presença de *missing data*.

Em alguns casos, a medida em um indivíduo pode ser ausente em uma ou várias ocasiões de medida. A perda pode ser *intermitente* embora comumente ela seja do tipo *dropout* (abandono do estudo).

Além da perda de poder, a análise com os dados disponível pode apresentar sérios vieses dependendo do mecanismo de dados ausentes.

Uma forma de tratar o problema é a imputação múltipla.

# Desafios de Estudos Longitudinais

Outro desafio é que não apenas a resposta muda ao longo do tempo mas também os preditores podem mudar.

O tratamento de covariáveis tempo-dependente na análise permite mais fortes inferências sobre a dinâmica das relações em comparação com dados transversais.

O preço, contudo, é a complexidade extra do modelo estatístico.

# Desafios de Estudos Longitudinais

Finalmente, em alguns casos, as medidas repetidas envolvem condições diferentes às quais os indivíduos são expostos.

Um exemplo clássico é um estudo crossover no qual dois ou mais tratamentos são administrados ao mesmo indivíduo em diferentes ordens.

Neste caso, as inferências podem ser comprometidas pela ordem ou efeitos *cross-over* no qual a resposta a um tratamento pode ser condicional na exposição ao tratamento anterior.

Lidar com efeitos *cross-over* não é trivial e é o preço pago para obtermos inferências estatísticas mais fortes permitidas por este tipo de desenho.