

# Análise de Dados Longitudinais

## Introdução à Disciplina

Enrico A. Colosimo-UFMG  
[www.est.ufmg.br/~enricoc](http://www.est.ufmg.br/~enricoc)

# Medidas Repetidas/Dados Longitudinais

## Medidas Repetidas

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

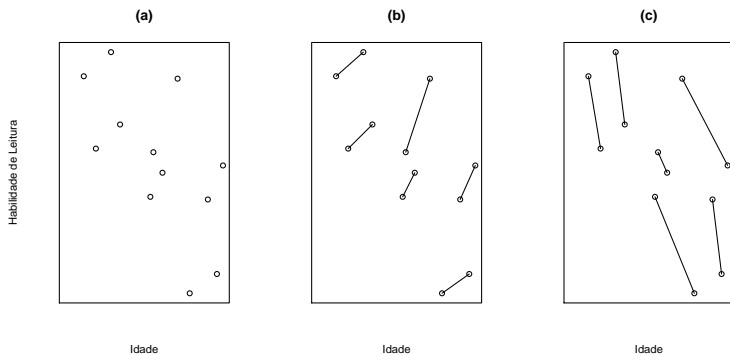
- Unidades: indivíduos, pacientes, animais, plantas, conglomerados (família, empresa, cidade, etc), etc.
- Caso especial: Dados Longitudinais: são medidas ao longo do tempo em uma mesma unidade/indivíduo.

## 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 à estudos transversais. Unidades podem estar agrupadas em conglomerados.

# Estudo Transversal vs Longitudinal



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

## Vantagens do Desenho Longitudinal

- Permite avaliar o comportamento da resposta ao longo do tempo.
- Permite avaliar a variação intra-indivíduo.
- No exemplo, o efeito de idade pode ser confundido com um possível efeito de coorte.
- Se quisermos comparar o depois com o antes, cada indivíduo atua como seu próprio controle evitando o efeito de variação entre indivíduo (fatores de confusão).

## Medidas Repetidas/Dados Longitudinais

### Característica Principal

Medidas Repetidas na mesma unidade são  
**CORRELACIONADAS.**

## Exemplos: 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 Sobrevida)

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

(Séries Temporais)

# Análise de Dados Longitudinais - Medidas Repetidas

## 1 Características:

- Dados de mesma natureza;
- grande número de pequenas séries;
- os tempos de medição são fixos (\*\*balanceado ou não balanceado\*\*);
- covariáveis: fixas ou dependentes do tempo;
- áreas de aplicação: saúde, economia, engenharia, etc.

## 2 Vantagens:

- \*\*avaliar mudança no tempo\*\*;
- eficiência no custo das observações;
- homogeneidade nas comparações.

## 3 Dificuldades:

- \*\*observações correlacionadas\*\*;
- usualmente mais demorados;
- fonte de vício: dados perdidos (viés de seleção).

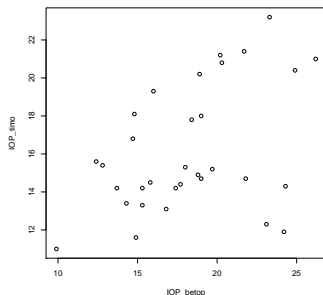
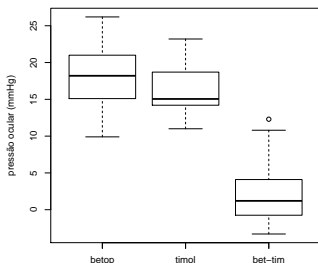


### Estudo "cross-over"

- O Dr. Emílio Suzuki quer comparar o efeito de dois colírios (B, T) redutores da pressão ocular com relação ao fluxo sanguíneo.
- Para tal ele submeteu 32 pacientes aos dois colírios por um período de dois meses com um descanso de igual tamanho. Foram 17 pacientes submetidos a sequência BT e 15 à TB.
- A ordem da aplicação dos colírios foi aleatória
- Duas medidas de pressão (colírios B e T) foram tomadas ao fim do estudo para cada paciente.
- Eventualmente, a medida de linha de base pode ser útil na análise estatística.

## Exemplo: Colírio A: Timoptol (timo) e Colírio B: Betoptic

- Ponto Principal: Existe diferença entre os colírios?
- Existe efeito da ordem?
- Existe efeito de período?
- A medida de linha de base é útil?
- O descanso ("washout") de dois meses foi suficiente?



## Exemplos Reais

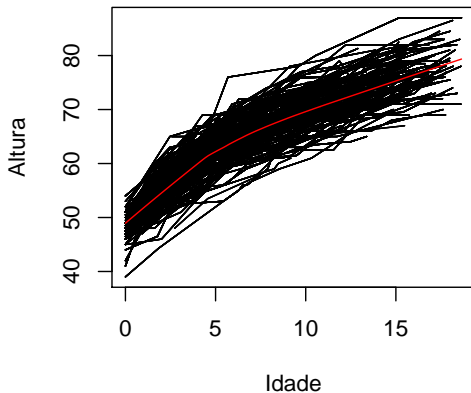
Estudo Longitudinal Desbalanceado: Avaliação longitudinal do crescimento de lactentes nascidos de mães infectadas com o HIV-1.

- Comparar longitudinalmente altura de lactentes infectados e não-infectados nascidos de mães infectadas pelo HIV.
- Uma coorte aberta acompanhada no ambulatório de AIDS pediátrica do Hospital das Clínicas da Universidade Federal de Minas Gerais.
- Período: 1995 a 2003.
- Inclusão: primeiros três meses de vida.
- Grupos: (1) não-infectados: 97; (2) infectados: 42.
- Controlado por sexo.

## Estrutura Longitudinal

- Visitas regulares ao pediatra.
- Planejado para acompanhamento de 18 meses.
- Tempo: idade da criança.
- Tempo mediano de acompanhamento foi 15 meses (7 a 18).
- Número total de medidas: Não-infectados: 907; Infectados: 411.
- Número médio de visitas por criança: 9,5.
- Delineamento não-balanceado.

**Perfis das Crianças**



# Perfis médio por grupo

Gráfico para os Grupos

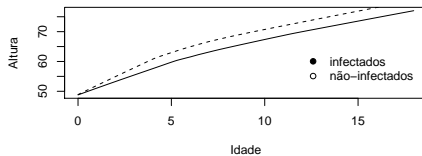
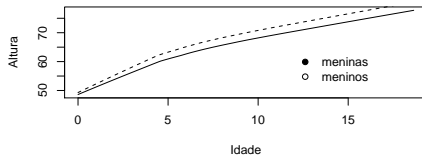


Gráfico para Meninos e Meninas



### Marcadores Psicofisiológicos de Proteção e Vulnerabilidade ao Estresse Psicosocial

Os objetivos gerais deste estudo são:

- Investigar as reações cardíacas a uma situação de estresse social.
- Investigar a capacidade de regulação dessas respostas em função da afetiva individual (fatores internos) e da indução prévia de um estado de afeto positivo ou negativo (fator externo).

## Exemplos: Estresse Psicosocial

- Participaram do experimento 72 estudantes universitários da Universidade de Granada (Espanha) de ambos os sexos, com idade entre 18 a 30 anos.
- Foram utilizadas 40 fotos agradáveis (famílias e bebês) e 40 fotos desagradáveis (pessoas com mutilações) selecionadas do catálogo International Affective Picture System - IAPS.
- O objetivo das fotos é induzir um estado de humor positivo ou negativo, respectivamente.
- Resposta: período cardíaco médio avaliado em 12 momentos.



## Perfis individuais e médio

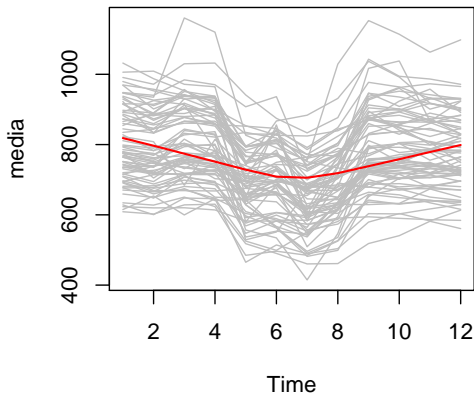


Figura: Perfis dos estudantes e uma curva alisda

## Exemplo - Medidas Repetidas

- ❶ Trauma Odontológico - Profa. Juliana Bastos
  - O indivíduo perde o(s) dente(s) por acidente.
  - O dente é reimplantado em um serviço de urgência (Odilon Behrens).
  - Em seguida ele é encaminhado ao serviço de trauma da Faculdade de Odontologia da UFMG para tratamento de canal.
  - No período entre o reimplante e o canal, existe um processo de reabsorção óssea.
  - Este processo de reabsorção é medido por um índice (raio-x).
  - A resposta de interesse é a avaliação deste índice, em especial se ele ultrapassou ou não o valor 4.
- ❷ O objetivo é identificar fatores que aceleram ou desaceleram o crescimento do índice.
- ❸ Fatores: período extra-oral, meio de armazenamento, idade, etc.
- ❹ Alguns pacientes contribuem com mais de um dente (componente transversal) e o índice é avaliado em mais de um momento (componente longitudinal).

## Comparação de duas Médias

### Retomar a Comparação dos Colírios A e B

- Pacientes com pressão intra-ocular (PIO) elevada irão participar do estudo.
- A pressão será medida após dois meses de uso do colírio.
- O objetivo é comparar a redução média de PIO dos dois colírios (A e B).

Então, queremos o seguinte:

$$\delta = \mu_A - \mu_B.$$

O interesse é então testar a hipótese:

$$H_0 : \delta = 0$$

## Comparação de duas Médias

- Existem duas formas de conduzir o estudo:
  - 50 pacientes são submetidos ao colírio A e ao colírio B (medidas repetidas). Considera-se um período de descanso de dois meses entre a aplicação dos colírios. É importante aleatorizar a ordem de aplicação de A e B.
  - 100 pacientes são selecionados e 50 são sorteados para receber o colírio A e os demais recebem o B.
- Ambos estudos são experimentais
  - Pareado: Estudo Cross-over
  - Amostras Independentes: Estudo Clínico Aleatorizado.
- Qual forma você utilizaria?

## Amostra Pareada ou Independente?

### 1 Vantagens em Parear Pacientes

- Controlar por possíveis fatores de confusão.
- Menos pacientes/unidades na amostra.
- Teste mais preciso com menos suposições.
- Controla pelo efeito de coorte.

### 2 Vantagens de Amostras Independentes

- Dados não Correlacionados.
- Dados são obtidos de forma mais rápida.

## Amostras Pareada ou Independente?

Quando devemos parear?

SEMPRE (que for possível).

- Caso típico: antes e depois.
- Situações não típicas: comparar fumantes e não-fumantes pareado por sexo e idade.

## Teste-t pareado

- Comparar duas  $n$  respostas pareadas.



$$\delta = \mu_A - \mu_B.$$

Uma estimativa natural para  $\delta$  é a diferença das médias. Ou seja

$$\hat{\delta} = \hat{\mu}_A - \hat{\mu}_B.$$

A variância de  $\hat{\delta}$  é

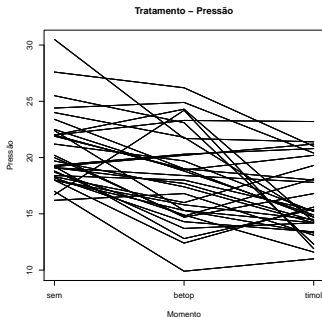
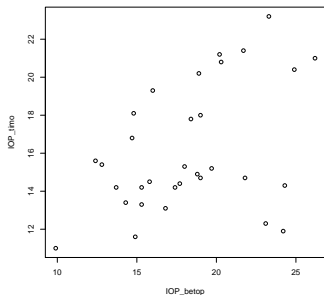
$$\text{Var}(\hat{\delta}) = \frac{1}{n}(\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB})$$

## Teste-t pareado

Usualmente dados longitudinais têm correlação positiva. Ou seja  $\sigma_{AB} > 0$ .

Isto significa que a estatística a ser utilizada tem menor variância do que aquela com dados independentes.

Exemplo: Colírio A: Timoptol (timo) e Colírio B: Betoptic (cor=0,43).





## Teste-t pareado

Considere as diferenças:

$$d_i = y_{i1} - y_{i2} \quad i = 1, \dots, n.$$

A estatística teste é:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

que sob  $H_0$  tem uma distribuição t com  $n-1$  graus de liberdade.

Teste para uma única amostra.

**Suposição:**  $d_i$  vem de uma distribuição normal.

Exemplo Colírio:  $t = 2.9934$ ,  $df = 31$ ,  $p\text{-value} = 0.005378$ .

Neste exemplo o teste somente é válido se não houver efeito de período e de "carry-over".

## Em resumo: Análise de Dados Longitudinais

### 1 Características:

- As respostas de diferentes unidades são independentes;
- As respostas para a mesma unidade são correlacionadas. De uma forma geral, as respostas próximas no tempo devem ser mais correlacionadas.

### 2 Medida Temporal

- Idade;
- Calendário medido a partir de um certo evento. Evento: aplicação de um tratamento, diagnóstico de doença, etc.

### 3 Objetivos do Estudo:

- avaliar o comportamento temporal;
- avaliar o efeito de covariáveis sobre a resposta;
- predição.

### 4 Modelos de Regressão

- Modelos marginais (modelar a média e a estrutura de covariância);
- Modelos de efeitos aleatórios.
- Modelos de transição.

## Características da Correlação dos Dados

- As correlações usualmente são positiva;
- as correlações usualmente diminuem a medida que aumenta a separação no tempo;
- as correlações entre medidas repetidas raramente aproximam do zero.
- medidas muito próximas tendem a ter correlação um.

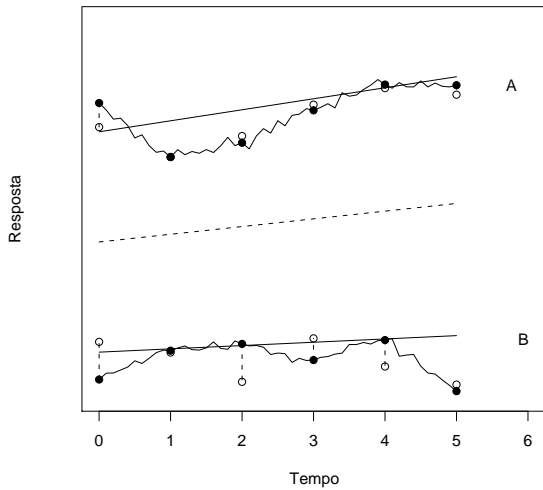
## Fontes de Variabilidade em Estudos Longitudinais

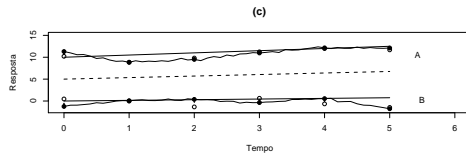
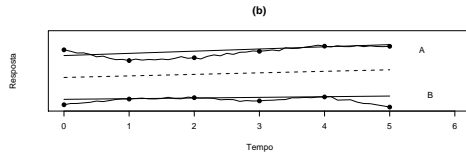
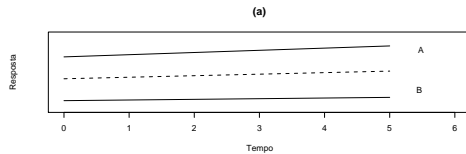
- Variação entre-unidades;
- Variação intra-unidade;
- Erro de medição.

## Fontes de Variabilidade em Estudos Longitudinais

Estas três fontes de variação podem ser visualizadas de forma gráfica.

- pontos pretos são respostas livre de erro de medição;
- pontos brancos são as respostas observadas;
- A e B são diferentes indivíduos.





## Comparação de mais de duas Médias

### Comparação dos Colírios A e B e C

- Pacientes com pressão intra-ocular elevada irão participar do estudo.
- A pressão será medida após dois meses de uso do colírio.
- O objetivo é comparar a redução média dos três colírios.

Então, queremos testar a seguinte hipótese:

$$H_0 : \mu_A = \mu_B = \mu_C.$$

**ANOVA é válido?**



## 1 Tratamento de Dados Correlacionados

- Duas amostras pareadas  $\implies$  solução simples;
  - Teste t pareado (resposta contínua);
  - Teste McNemar (resposta binária).
- Desenhos mais complexos  $\implies$  solução "difícil"(não trivial);
  - Modelos efeitos aleatórios;
  - Modelar a estrutura de correlação;
  - Métodos inferenciais complexos.

## 2 É necessário tratar a correlação dos dados?

## 3 Qual é o impacto de ignorar a correlação dos dados na inferência estatística?

## Simulação: Duas amostras Pareadas

$y_1$ : pressão sistólica de pacientes hipertensos;

$y_2$ : pressão sistólica após o tratamentos com a droga X;

- $y_1 \sim N(\mu = 150, \sigma = 15)$ ;
- $y_2 \sim N(\mu = 110, \sigma = 15)$ ;
- $cor(y_1, y_2) = 0,8$ ;  $cov(y_1, y_2) = 180$ ;
- $n_1 = n_2 = 100$
- $t = 46,56, p < 2,210^{-16}$

## Simulação: Duas amostras Pareadas

Resultados:

- Regressão Linear Simples (suposição de independência):

$$t = \frac{-41,88}{1,889} = -22,17$$

- Regressão Linear Ponderada com a estrutura de correlação (estimada).

$$t = \frac{-41,88}{0,9064} = -46,20$$

**O ERRO-PADRÃO fica muito mal estimado.**