

Métodos Estatísticos Avançados em Epidemiologia

Modelo de Poisson e Análise de Dados Longitudinais

Enrico A. Colosimo

Departamento de Estatística
Universidade Federal de Minas Gerais
<http://www.est.ufmg.br/~enricoc>

2011

Modelos Estatísticos

- ▶ Regressão Linear (resposta quantitativa).
- ▶ Regressão Logística (resposta binária/categórica).
- ▶ Análise de Sobrevida (resposta: tempo até evento).
- ▶ Regressão de Poisson (resposta contagem).
- ▶ Dados Longitudinais/Medidas Repetidas (mais de uma resposta por indivíduo)

Modelo de Poisson

1. Resposta: contagem

2. Exemplos:

- ▶ No. chamadas telefônicas/dia em um call center (dia).
- ▶ No. de surtos epiléticos em dois anos (paciente).
- ▶ No. de casos de câncer em 2009 em cidades do Estado de MG (cidade).
- ▶ No. de partos por hospital em um ano (hospital).

Modelo de Poisson: Observações

1. Resposta: contagem ou taxa.
2. Tempo de exposição pode ser diferente para cada indivíduo da amostra.
3. Taxa de Incidência.

$$TI = \frac{\text{no. de casos novos}}{\text{quantidade de pessoa tempo}}$$

4. Presença de covariáveis.

Modelo de Poisson

1. Objetivo: explicar a variação da contagem/taxa através das covariáveis.
2. Como modelar a contagem/taxa?
 - ▶ Aproximar pela distribuição gaussiana (regressão linear): válido para contagens grandes.
 - ▶ Usar o modelo de Poisson.

$$P(Y = y) = \frac{e^{-\lambda(x\beta)} \lambda(x\beta)}{y!}$$

- ▶ Usual é tomar

$$\lambda(x\beta) = \exp(x\beta).$$

Modelo de Poisson

1. A partir de uma amostra, a inferência para β é realizada utilizando o método de máxima verossimilhança.
2. A interpretação das quantidades estimadas é razão de taxas de incidência.
3. Existem todas as quantidades necessárias para validar a adequação do modelo ajustado.
4. Na presença de exposição diferente utilizamos um termo extra no modelo (offset) para acomodar este fato.

Dados Longitudinais: Tipos de Estudos

- ▶ Transversal: uma única resposta é medida para cada indivíduo em um instante de tempo
- ▶ ****Longitudinal**:** indivíduos são medidos repetidamente ao longo do tempo.

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 ou Medidas Repetidas **)

2. Indivíduos foram divididos aleatoriamente em dois grupos (A e B) e foi registrado o tempo até a pressão sistólica atingir um certo 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: fixa ou dependente 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**;
- ▶ fonte de vício: dados perdidos.

Exemplos Reais

- ▶ Estudo "cross-over"

O Dr. Emílio Suzuki quer comparar o efeito de dois colírios (A, B) redutores da pressão ocular com relação ao fluxo sanguíneo. Para tal ele submeteu cada paciente aos dois colírios por um período de dois meses com um intervalo de igual tamanho. A ordem da aplicação dos colírios foi aleatória e duas medidas de pressão foi tomada ao fim do estudo para cada paciente.

Exemplos Reais

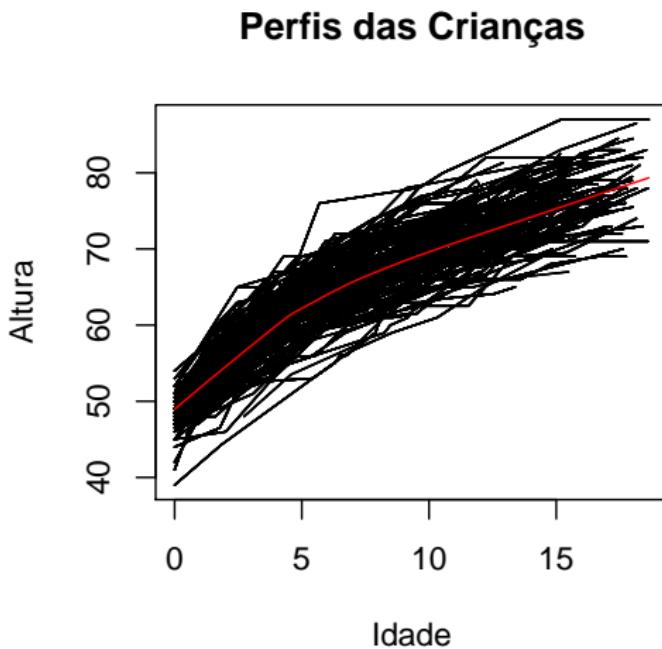
Estudo Longitudinal Desbalanceado: Avaliação longitudinal do crescimento de lactentes nascidos de mães infectadas com o HIV-1.

- ▶ 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 individuais e médio



Perfis médio por grupo

Gráfico para os Grupos

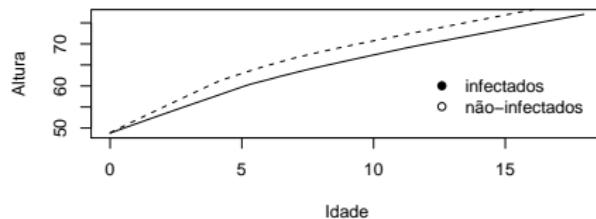
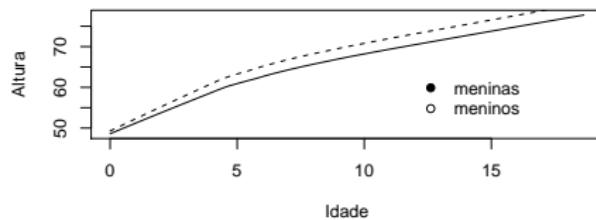


Gráfico para Meninos e Meninas



Exemplos Reais

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 recrutados durante as aulas de graduação dos professores do laboratório de Psicofisiologia Humana desta universidade.

Foram utilizadas 40 fotos agradáveis (famílias e bebês) e 40 fotos desagradáveis (pessoas com mutilações) selecionadas do catálogo International Affective Picture System - IAPS (Center for the Study of Emotion and Attention [CSEA-NIMH], 1999; Lang et al., 1999) para induzir um estado de humor positivo ou negativo, respectivamente.

Resposta: período cardíaco médio avaliado em 12 momentos.

Perfis individuais e médio

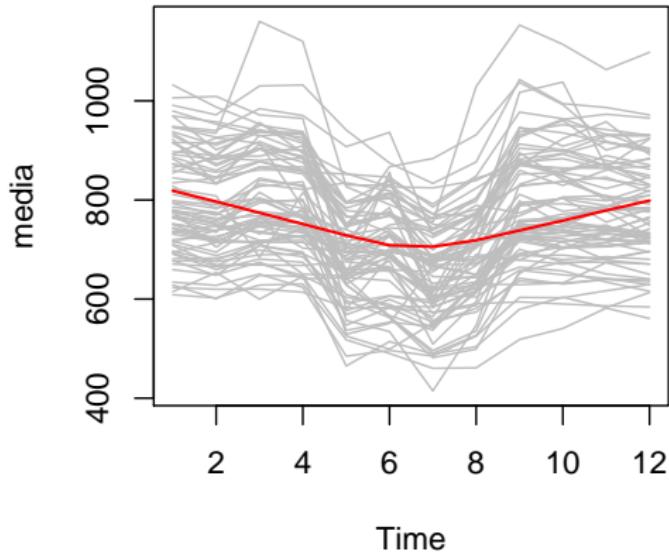


Figura: Perfis dos estudantes e uma curva alisada

Análise de Dados Longitudinais

1. Características:

- ▶ As respostas de diferentes indivíduos são independentes;
- ▶ As respostas para o mesmo indivíduo 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.

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.

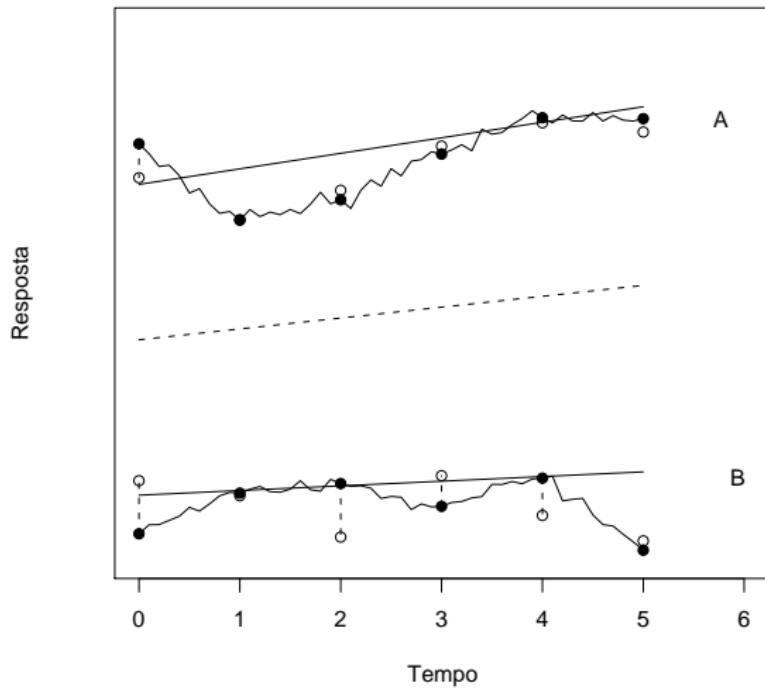
Fontes de Variabilidade em Estudos Longitudinais

- ▶ Variação entre-indivíduos;
- ▶ Variação no indivíduo;
- ▶ Erro de medição.

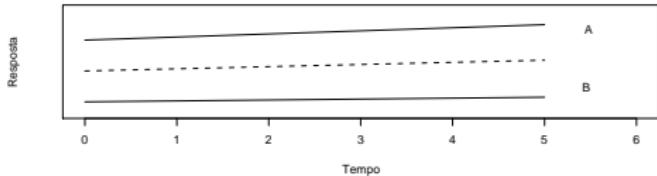
Fontes de Variabilidade em Estudos Longitudinais

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

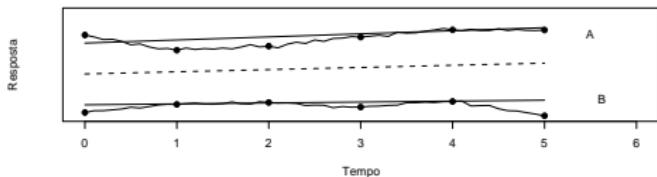
- ▶ pontos pretos são respostas livre de erro de medição;
- ▶ pontos brancos são as respostas observadas;
- ▶ A e B são diferentes indivíduos.



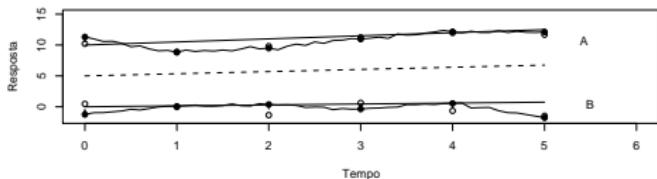
(a)



(b)



(c)



Situação simples

Considere o caso mais simples em que existem somente duas medidas repetidas, digamos nos tempos 1 e 2. O objetivo principal do estudo é determinar se existe mudança da média ao longo do tempo. Ou seja

$$\delta = \mu_1 - \mu_2.$$

Uma estimativa natural para δ é a diferença das médias. Ou seja

$$\hat{\delta} = \hat{\mu}_1 - \hat{\mu}_2.$$

A variância de $\hat{\delta}$ é

$$Var(\hat{\delta}) = \frac{1}{n}(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$$

Situação simples

Usualmente dados longitudinais têm correlação positiva. Ou seja

$$\sigma_{12} > 0$$

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

Outras vantagens:

- ▶ pareamento controla por fatores de confusão;
- ▶ evita efeito coorte.

Exemplo simples de Dados Longitudinais

Deseja-se verificar a eficácia de uma certa droga para reduzir a pressão arterial. 100 pacientes hipertensos participaram do estudo. A pressão sistólica foi medida no início (tempo 1) do estudo e após os pacientes terem sido submetidos a droga de interesse (tempo 2). Então

$$\delta = \mu_1 - \mu_2.$$

O interesse é então testar a hipótese:

$$H_0 : \delta = 0$$

Teste-t pareado

Considere as diferenças:

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

A estatística é:

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

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

Suposição: d_i vem de uma distribuição normal.

Modelos de Regressão para Dados Longitudinais

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