

Análise de Dados Longitudinais Medidas Repetidas e Dados Perdidos

Enrico A. Colosimo/UFMG

- 1 Medidas Repetidas / Modelos Multiníveis.
- 2 Dados Perdidos / Imputação

Medidas Repetidas / Modelos Multiníveis

- Mesmo indivíduo/unidade amostral é medido várias vezes em diferentes condições.
- Modelo Multinível ou Hierárquico.
- Exemplo (três níveis): criança/moradia/bairro: resposta é a presença de asma. Covariáveis: idade da criança, presença de mofo em casa e nível de poluição da região, etc.
- Dados longitudinais é um caso particular que envolve um nível temporal em uma escala ordenada.
- As duas classes de modelos podem ser utilizadas com uma preferência para o de efeitos aleatórios. Sem esquecer que os objetivos do estudo devem nortear a escolha do modelo.
- Existem dificuldades operacionais no ajuste do GEE.

Exemplo: Avulsão/Trauma Dentário

- Clínica de Traumatismo Dentário da Faculdade de Odontologia da UFMG (CTD-FO-UFMG), coordenada pela Profa. Juliana Vilela Bastos.
- Situação típica envolve a perda do dente devido a um acidente ou trauma.
- Os dentes são reimplantados no serviço de urgência do Hospital Odilon Behrens.
- Inicia-se um processo de reabsorção da raiz. Ou seja, o osso vai destruindo a raiz do dente reimplantado.
- O paciente é encaminhado para a CTD-FO-UFMG para tratamento de canal (tratamento endodôntico radical).
- Um índice, que avalia a gravidade da reabsorção, é medido na chegada do paciente a FO.
- O objetivo do estudo é identificar fatores que aceleram o processo de Reabsorção Radicular Externa (RRE).

Banco de Dados: Avulsão/Trauma Dentário

- O banco de dados compreende o período de 1994 a 2011.
- O banco de dados é formado por 145 pacientes.
- Alguns pacientes perderam mais de um dente: Medidas Repetidas, em um total de 177 dentes.
- Covariáveis incluem marcadores demográficos e clínicos: idade, gênero, tempo decorrido desde o reimplante, período extra oral, meio de armazenamento, dente e uso de antibiótico.
- Resposta: A reabsorção é medida através de um índice que varia de 0 a 12 e agrupada em duas categorias: baixa (≤ 4) e severo (> 4) que representa 76.8% e 23.2% dos casos, respectivamente.
- Objetivo específico: investigar a associação destas variáveis com a gravidade da reabsorção.

Descritiva das covariáveis

Tabela: Meio de armazenamento do dente

Meio	Frequência
Meios úmidos	72 (40,6%)
Leite	38 (21,4%)
Seco	67 (38%)
Total	177 (100%)

Tabela: Prescrição de antibióticos

Antibióticos	Frequência
Sim	20 (13,8%)
Não	40 (27,6%)
Sem informação	85 (58,6 %)
Total	145 (100%)

Descritiva das covariáveis

Tabela: Tipo de dente

Dente	Frequência
Inc lateral superior	20 (11,2%)
Inc central superior	76 (42,9%)
Inc central superior direito	72 (40,6%)
Inc inferior	2 (2,5%)
Outros	7 (2,8%)
Total	177 (100%)

Tabela: Splint

Splint	Frequência
1	74 (41,8%)
2	98 (55,3%)
Sem informação	5 (2,9 %)
Total	145 (100%)

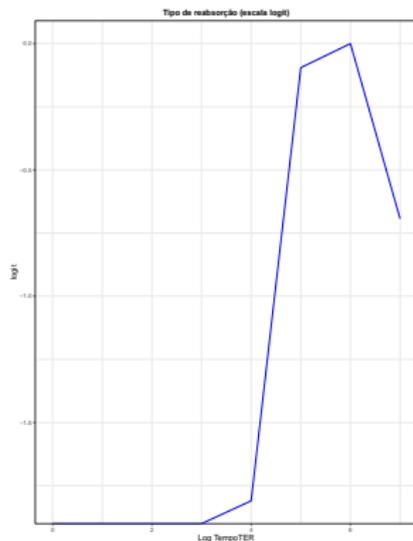
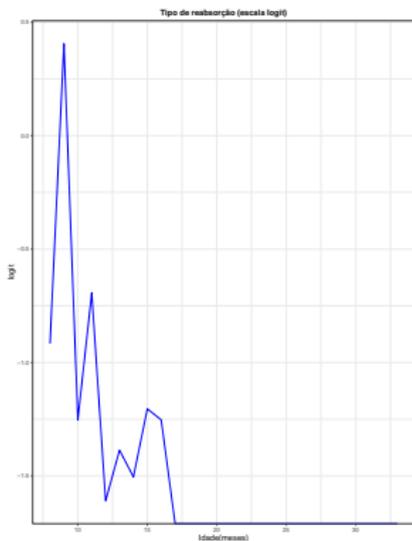
Descritiva das covariáveis

Tabela: Covariáveis contínuas

Covariável	Min	1Q	Med	Média	3Q	Máx	DP
Idade	7,8	10,7	12,8	13,3	14,6	33,1	4,10
PerEO	5,0	60,0	120,0	236,6	240,0	4320,0	442,9
Log PerEO	1,6	4,1	4,8	4,7	5,5	8,4	1,2
TER	1,0	31,0	67,0	127,3	135,0	1206,0	174,7
Log TER	0,0	3,4	4,2	4,2	4,9	7,1	1,3

Gráfico na escala do Logit

Obs.: idade e Log do TempoTER foram arredondados para um único dígito. Avaliar a possibilidade de incluir duas inclinações para Log do TempoTER no valor 6.



Resultados do Ajuste GEE - Resposta binária

```
Call:
geeglm(formula = resposta ~ TempoTERdR_log + Idade + Log_tempoter_2,
        family = binomial(link = "logit"), data = dados, id = Paciente,
        corstr = "exchangeable", std.err = "san.se")
```

Coefficients:

	Estimate	Std.err	Wald	Pr(> W)	
(Intercept)	-6.2552	1.6339	14.66	0.00013	***
TempoTERdR_log	1.7371	0.3108	31.24	2.3e-08	***
Idade	-0.2137	0.0723	8.74	0.00311	**
Log_tempoter_2	-0.3466	0.1447	5.74	0.01658	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation structure = exchangeable

Estimated Scale Parameters:

	Estimate	Std.err
(Intercept)	0.7	0.0695

Link = identity

Estimated Correlation Parameters:

	Estimate	Std.err
alpha	0.371	0.154

Number of clusters: 146 Maximum cluster size: 4

Resultados do Ajuste

Y: ($0 \leq 4/1 \rightarrow 4$) Reaboração.

X: Covariáveis: idade em meses:(2) Antibiótico (1-Sim e 2-Não), Meio (1-úmido, 2-Leite e 3-Seco), Dente (1-Inc lateral superior, 2-Inc Central Superior e 3-Inc central superior direito), TempoTER com duas inclinações e Período Extra-oral

$n = 145$ pacientes e 177 dentes.

Variável	Estimativa	E.P.	Wald
Idade	-0,21	0,07	8,74 ($p = 0,003$)
Log (TempoTER)	1,73	0,31	31,24 ($p < 0,001$)
Log (TempoTER) ₂	-0,34	0,14	5,74 ($p = 0,01$)

Interpretação dos Coeficientes

- Idade: Razão de chances = $\exp(-0,21) = 0,81$ (0,67;0,95), isto significa que com o aumento de um ano na idade a chance de reabsorção grave (maior que 4), reduz em 19%;
- TempoTER: No modelo foi considerado na escala logarítmica. É necessário realizar a interpretação na escala original. Uma possibilidade é fazer $1.2^{1.73} = 1.37$. (1,23; 2,34) Isso significa que com o aumento de 20% no TempoTER, a chance de reabsorção grave aumenta em 37%.
- Antibiótico não foi significativo. Dificuldade: Mais de 60% da amostra não tem esse informação ("missing data").

Exemplo: Intensidade Pixel em Tomografia Computadorizada (Pinheiro e Bates, 2000, p.40)

- Experimento com cães envolvendo a medida de intensidade de pixel após injeção de contraste.
- A medida é feita via Tomografia Computadorizada.
- Medidas feitas em nódulos linfomas nos lados direito e esquerdo da cabeça em vários dias após a injeção do contraste.
- Estrutura hierárquica em três níveis: cão, lado e ocasião. Combinação de medida repetida (lados) com estrutura longitudinal.

Gráfico de Perfis

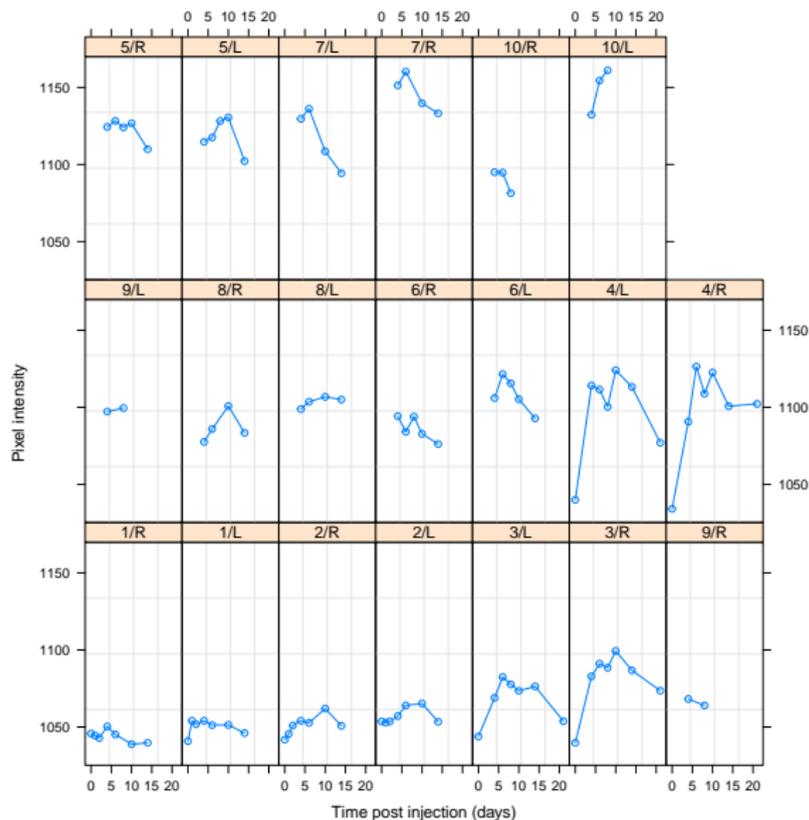


Gráfico de Perfis (Pinheiro e Bates, 2000, p.40)

- O perfil cresce e descreve, com pico em torno de 10 dias.
- Muita variabilidade entre cães, e pouca diferença entre os lados.

●

$$y_{ijk} = \beta_0 + \beta_1 t_{ik} + \beta_2 t_{ik}^2 + b_{0i} + b_{1i} t_{ik} + b_{ij} + i_{jk}$$

$$i = 1, \dots, 10; j = 1, 2, k = 1, \dots, n_{ij}$$

- R script

```
>require(nlme)
```

```
>Pixel
```

```
>names(Pixel)
```

```
>plot(Pixel)
```

```
#
```

```
>Pixel.lme2<-lme(pixel ~ day + I(day^2), data=Pixel,  
+random= list(Dog = ~day, Side=~1), m="REML") # ignore
```

```
>summary(Pixel.lme2)
```

```
>plot(Pixel.lme2)
```

```
>intervals(Pixel.lme2)
```

```
#
```

Resultados do Ajuste GEE - Resposta binária

Random effects:

Formula: ~day | Dog

Structure: General positive-definite, Log-Cholesky parametrization

StdDev Corr

(Intercept) 28.36990 (Intr)

day 1.84375 -0.555

Formula: ~1 | Side %in% Dog

(Intercept) Residual

StdDev: 16.82431 8.989606

Fixed effects: pixel ~ day + I(day^2)

Value Std.Error DF t-value p-value

(Intercept) 1073.3391 10.171686 80 105.52225 0

day 6.1296 0.879321 80 6.97083 0

I(day^2) -0.3674 0.033945 80 -10.82179 0

Correlation:

(Intr) day

day -0.517

I(day^2) 0.186 -0.668

Standardized Within-Group Residuals:

Min Q1 Med Q3 Max

-2.8290572 -0.4491811 0.0255493 0.5572163 2.7519651

Number of Observations: 102

Number of Groups:

Dog Side %in% Dog

10 20

Resultados do Ajuste GEE - Resposta binária

```
> intervals(Pixel.lme2)
Approximate 95% confidence intervals

Fixed effects:
      lower      est.      upper
(Intercept) 1053.0968388 1073.3391382 1093.5814376
day          4.3796925   6.1295971   7.8795016
I(day^2)     -0.4349038   -0.3673503   -0.2997967
attr(,"label")
[1] "Fixed effects:"

Random Effects:
Level: Dog
      lower      est.      upper
sd((Intercept)) 15.9292099 28.3699038 50.5267650
sd(day)          1.0814073 1.8437505 3.1435110
cor((Intercept),day) -0.8943751 -0.5547222 0.1906591
Level: Side
      lower      est.      upper
sd((Intercept)) 10.41733 16.82431 27.17176

Within-group standard error:
      lower      est.      upper
7.634529  8.989606 10.585199
>
```

Dados Perdidos (Dados Longitudinais)

Dados Perdidos acontecem com frequência em Estudos Longitudinais.

Tipos:

- 1 **"Drop out"**: Perda completa a partir de um certo instante;
- 2 **Intermitente**: Somente uma perda pontual.

Implicações:

- Perda do desbalanciamento;
- Perda de precisão (aumento da variância);
- "Possível" viés de seleção de acordo com o mecanismo de perda dos dados e o método de estimação.

1 "MCAR" Missing Completely at Random:

- Perda completamente aleatória;
- Dados observados podem ser considerados como uma amostra aleatória dos dados completos.

Exemplo: Mudança da cidade na qual estava sendo desenvolvido o estudo (migração)

Impacto:

- Qualquer método de análise gera inferências válidas.
- Perda de precisão.

2 "MAR": Missing at RAndom

- A probabilidade da resposta perdida depende do conjunto de valores observados.

Exemplo: Óbito devido a gravidade da resposta medida.

Impacto:

- As médias amostrais ficam viesadas.
- GEE não produzem estimativas válidas para β .
- EMV produz estimativas válidas se a distribuição normal for corretamente especificada para a resposta.

3 "NMAR": Not Missing at Random

- A probabilidade da resposta perdida depende do conjunto de valores observados e não observados.

Exemplo: gravidade devido a renda/salário.

Impacto:

- As médias amostrais ficam viesadas.
- GEE não produzem estimativas válidas para β .
- EMV não produz estimativas válidas.

Imputação de Dados Perdidos

1 Imputação Simples

- Substituir o dado perdido por uma única estimativa (média, último valor, etc).
- Subestima a variância dos estimadores.

2 Imputação Múltipla.

- Modelos de Regressão.
- Distância de valores observados (o mais próximo segundo alguma métrica).
- Análise Bayesiana:

$$p(y_{\text{perdido}} / \text{dados})$$