Análise de Dados Longitudinais
Aula 15.10.2018

José Luiz Padilha da Silva - UFPR
www.docs.ufpr.br/∼jlpadilha

Sumário

Respostas Longitudinal Não-Gaussiana

1. $Y_{ij}, \; i = 1, \ldots, N; \; j = 1, \ldots, n_i$: binária, contagem, etc.

2. Modelos Estatísticos:

   - Modelos Lineares Generalizados Mistos.
   - Modelos Marginais: GEE

Exemplos

1. Mecanismo Evacuatório de Récem-Nascidos
2. Fatores de Risco Coronariano: MCRF, (FLW, pag. 364)

Mecanismo Evacuatório de Récem-Nascidos

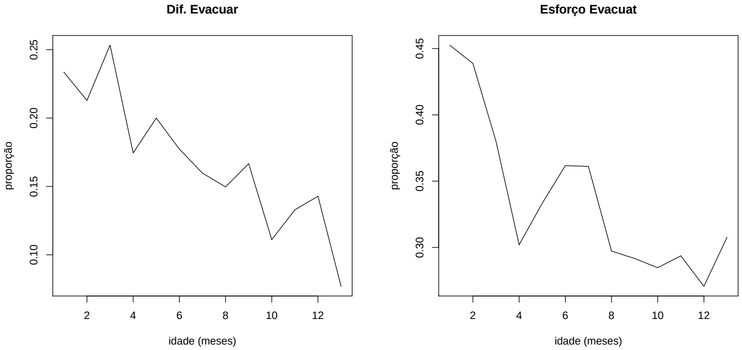
- 151 recém-nascidos acompanhados nos primeiros 12 meses de vida no Hospital das Clínicas da UFMG em 2010 e 2011.
- Acompanhamento mensal totalizando 1751 medidas (61 perdas)
- Respostas:
  1. Binárias: Dificuldade para evacuar; esforço evacuatório; dor ao evacuar.
  2. Contagem: Frequência evacuatória/semana.

Mecanismo Evacuatório de Récem-Nascidos

- Variável temporal: idade (em dias ou meses).
- Covariáveis:
  1. Fixa: sexo.
  2. Dependentes do tempo: aleitamento materno; dieta (0/1): cereais; frutas; vegetais, carnes, etc.
- Objetivo: avaliar o comportamento temporal das respostas e seus respectivos indicadores.

Resposta: Dificuldade e Esforço para Evacuar

Obs.: idade foi arrendondada para mês (um único dígito).



"Muscatine Coronary Risk Factor Study"

- Estudo longitudinal de crianças em idade escolar realizado em Muscatine, Iowa, Estados Unidos na década de 80.
- Cinco coortes de crianças, inicialmente com idades em 5-7, 7-9, 9-11, 11-13 e 13-15 foram acompanhadas bianualmente de 1977 a 1981 (3 medidas).
- Respostas binária: obesidade.
- Variável temporal: idade (em dias ou meses).
- Covariável: sexo.
- Objetivo: avaliar (1) se o risco de obesidade aumenta com a idade e (2) se os padrões são os mesmos para meninos e meninas.

“Muscatine Coronary Risk Factor Study”

		Obesidade (%)		
Gênero	Coorte Idade	1977	1979	1981
Meninos				
	5-7	7.9	15.4	21.2
	7-9	18.8	20.5	23.7
	9-11	21.2	22.7	22.5
	11-13	24.3	21.8	19.4
	13-15	19.2	21.1	18.2
Meninas				
	5-7	14.0	17.2	25.1
	7-9	16.5	24.0	24.9
	9-11	25.4	26.2	22.2
	11-13	23.8	22.1	19.9
	13-15	22.9	25.8	20.9

Revisão: Modelos Lineares Generalizados

Modelos Lineares Generalizados (MLG) é uma classe unificada de modelos de regressão.

1. Considere $Y_1, \ldots, Y_N$ uma amostra aleatória de respostas univariadas (desenho transversal).

2. Um vetor de p-covariáveis associados a cada resposta Y_i. Ou seja

$$X_i = \begin{pmatrix} X_{i0} \\ X_{i1} \\ \vdots \\ X_{ip} \end{pmatrix}$$

em que $X_{i0} = 1$.

Modelos Lineares Generalizados (MLG)

3 O MLG é definido por três componentes:

- Distribuição de Y_i.
- Componente Sistemático (preditor linear).

$$\eta_i = X_i'\beta = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}$$

- Função de Ligação.

MLG - Família Exponencial

A distribuição de Y_i pertence à família exponencial que inclui os principais modelos estatísticos: normal, binomial, poisson, exponencial, etc.

Ou seja, Y_i tem densidade $f(Y_i|\theta,\phi)$ pertencente à família exponencial.

$$f(y_i|\theta_i,\phi) = \exp\{\phi^{-1}(y_i\theta_i - \psi(\theta_i)) + c(y_i,\phi)\}$$

em que θ_i é parâmetro natural, ϕ é o de escala e específicas funções $\psi(.)$ e $c(.)$.

Modelos Lineares Generalizados

- $\psi(.)$ é a função geradora de momentos
  - $\mu = E(Y) = \psi'(\theta)$ e
  - $Var(Y) = \phi\psi''(\theta)$
- Em geral, média e variância são relacionadas.

$$Var(Y) = \phi\psi'' \ \ (\psi'^{-1}(\mu) = \phi\nu(\mu)\ )$$

- A função $\nu(\mu)$ é chamada de função de variância.
- ψ'^{-1} que relaciona θ com μ é chamada de função de ligação.

Exemplos

1. Modelo Normal (μ, σ^2)

    - $\theta = \mu$
    - $\phi = \sigma^2$
    - $\psi(\theta) = \theta^2/2$
    - Média: $\mu = \theta$ e $\nu(\mu) = 1$
    - Observe que no modelo normal, média e variância não são relacionadas

    $$\phi\nu(\mu) = \sigma^2$$

    - Função de ligação natural: $\theta = \mu$.

Exemplos

2 Modelo Bernoulli (π)

- $\theta = \log(\pi/(1-\pi))$
- $\phi = 1$
- $\psi(\theta) = -\log(1-\pi) = log(1+\exp(\theta))$
- Média: $\mu = \pi = \frac{\exp(\theta)}{1+\exp(\theta)}$ e $\nu(\mu) = \pi(1-\pi) = \frac{\exp(\theta)}{(1+\exp(\theta))^2}$
- Observe que no modelo bernoulli, média e variância são relacionadas
$$\phi\nu(\mu) = \mu(1-\mu)$$
- Função de ligação natural: $\theta = \log(\mu/(1-\mu))$.

Função de Ligação Natural ou Canônica

$$g(\mu_i) = \eta_i = X_i'\beta = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}$$

- Gaussiano: $g(\mu_i) = \eta_i$ (identidade)
- Bernoulli: $g(\mu_i) = logit(\eta_i)$.
- Poisson: $g(\mu_i) = \log(\eta_i)$

Inferência por MV

- Função de log-verossimilhança $logL(.) = l(.)$

$$L(\beta) = \prod_{i=1}^{N} f(y_i|\theta_i, \phi) = \prod_{i=1}^{N} \exp\{\phi^{-1}(y_i\theta_i - \psi(\theta_i)) + c(y_i, \phi)\}$$

- Equações escore: derivada de $l(.)$.

- Inferência baseada na teoria assintótica de MV.

Exemplo - Regressão Binária

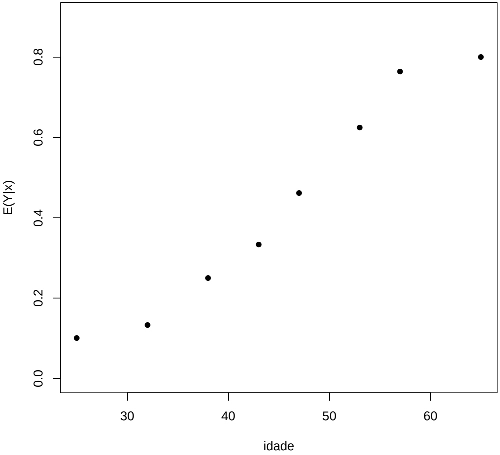
- Uma amostra de 100 indivíduos acompanhados por um período de cinco anos.
- Resposta: ocorrência de doença coronariana.
- Resposta para cada indivíduo foi sim (1) ou não (0).
- Covariável de interesse: 8 faixas etárias (idade): 20-29, ..., 60-69.
- Aconteceram 43 ocorrências de doença coronariana.

Ref: Giolo (2010) pg. 98- Introdução à Análise de Dados Categóricos.

Tabela Resumo

	Doença coronária			
Idade ($X = x$)	Não ($Y = 0$)	Sim ($Y = 1$)	Totais	$E(Y \mid x)$
20-29	9	1	10	0,10
30-34	13	2	15	0,13
35-39	9	3	12	0,25
40-44	10	5	15	0,33
45-49	7	6	13	0,46
50-54	3	5	8	0,63
55-59	4	13	17	0,76
60-69	2	8	10	0,80
Totais	57	43	100	0,43

Descrição Gráfica por Faixa Etária



MLG

$$logit(idade_i) = \log\{\mu_i/(1-\mu_i)\} = \beta_0 + \beta_1 idade_i$$

e

$$E(Y_i|idade_i) = P(Y_i = 1|idade_i)$$

O modelo logístico pode ser escrito como:

$$P(Y_i = 1|idade_i) = \frac{\exp(\beta_0 + \beta_1 idade_i)}{1 + \exp(\beta_0 + \beta_1 idade_i)}$$

Resultados do Ajuste MV

```
> summary(ajust1)

Coefficients:
             Estimate    Std. Error z value Pr(>|z|)
  (Intercept) -5.12300    1.11111  -4.611 4.01e-06 ***
   idade       0.10578    0.02337   4.527 5.99e-06 ***

Number of Fisher Scoring iterations: 4


> anova(ajust1,test="Chisq")
Terms added sequentially (first to last)

       Df Deviance Resid. Df Resid. Dev P(>|Chi|)
 NULL                      7    28.7015
 idade  1   28.118         6     0.5838 1.142e-07 ***
```

Resultados do Ajuste

Y: presença ou não de doença coronariana;
X: idade (em anos);
$n = 100$.

Variável	Estimativa	E.P.	Wald
Idade	0,106	0,023	4,53 ($p < 0,001$)
Constante	-5,123	1,11	-4,61 ($p < 0,001$)

$$\widehat{\pi}(x) = \frac{\exp(-5,12 + 0,106 \text{ idade})}{1 + \exp(-5,12 + 0,106 \text{ idade})}$$

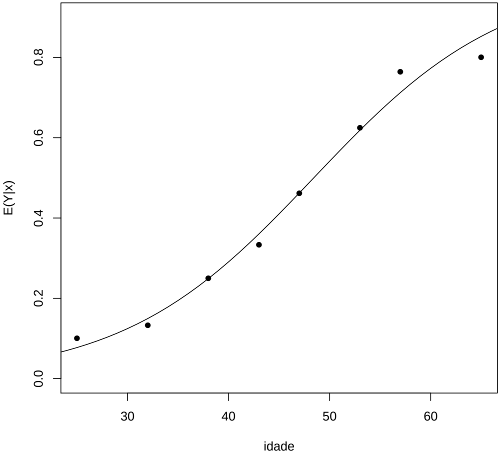
$$\widehat{logit}(x) = -5,12 + 0,106 \text{ idade}$$

$$\log(\text{verossimilhança}) = \log L(\widehat{\beta}_0, \widehat{\beta}_1) = -10,86$$

Sob $H_0 : \beta_1 = 0$, $logL(\widehat{\beta}_0) = -24,92$.

TRV = $2(-10,86 + 24,92) = \text{Null Deviance} - \text{Residual Deviance} = 28,118$.

Modelo Estimado



Interpretação dos Coeficientes

Interpretação: Razão de chances = $\exp(0,1058)$ = 1,11 (1,06;1,16).

Isto significa que para o aumento de um ano na idade a chance de doença coronariana aumenta em 11%.

Outros MLG

- Y tem uma Bernoulli.
- Outras funções de ligação:
  - $\pi(x) = \Phi(x)$ (probit)
  - $\pi(x) = \exp\{-\exp(x)\}$ (complemento log-log)
  - etc (qualquer função de distribuição)

Modelos para Resposta Gaussiana Longitudinal

1. Modelo Marginal

$$Y_{ij} = X_{ij}'\beta + \varepsilon_{ij}$$

e

$$E(Y_{ij}|X_{ij}) = X_{ij}'\beta.$$

2. Modelo Condicional

$$Y_{ij} = X_{ij}'\beta + Z_{ij}'b_i + \varepsilon_{ij}$$

em que:

$(\beta)_{p\times 1}$: efeitos fixos;

$(b_i)_{q\times 1}$: efeitos aletaórios.

e,

$$b_i \sim N_q(0, \Sigma) \text{ e } \varepsilon_{ij} \sim N(0, \sigma^2)$$

Sendo b_i e ε_{ij} independentes.

Modelos para Resposta Gaussiana

- Média Condicional ou Específica por Indivíduo

$$E(Y_{ij}|b_i, X_{ij}) = X_{ij}'\beta + Z_{ij}'b_i.$$

e a Covariância Marginal

$$Var(Y_i) = Z_i \Sigma Z_i' + \sigma^2 I_{n_i}.$$

Modelos para Resposta Não-Gaussiana

1. $\mu_{ij} = E(Y_{ij}|X_{ij})$ (modelo marginal)
   $\mu_{ij} = E(Y_{ij}|b_i, X_{ij})$ (modelo condicional).

2. Modelo Bernoulli

   - Y_{ij} : $0/1$ (Bernoulli)

   - função de ligação: logit (mais comum)

$$logit(\mu_{ij}) = X_{ij}'\beta \quad \text{Modelo Marginal}$$

$$logit(\mu_{ij}) = X_{ij}'\beta + Z_{ij}'b_i \quad \text{Modelo Condicional}$$

Modelos para Resposta Não-Gaussiana

3 Modelo Poisson

- Y_{ij} :contagem (Poisson)
- função de ligação: logarítmica (mais comum)

$$\log(\mu_{ij}) = X'_{ij}\beta \quad \text{Modelo Marginal}$$

$$\log(\mu_{ij}) = X'_{ij}\beta + Z'_{ij}b_i \quad \text{Modelo Condicional}$$