Outros modelos de regressão linear e métodos de estimação

Prof. Caio Azevedo

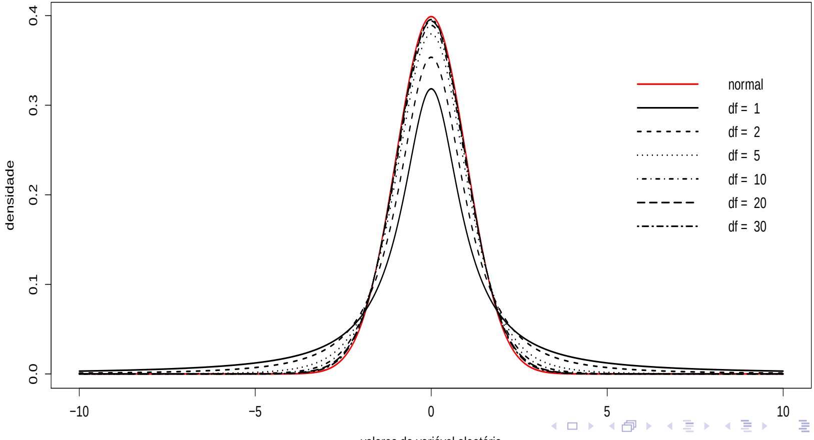
Introdução

- Para os exemplo 2 vimos que considerar uma distribuição com caudas mais pesadas do que a normal para os erros, pode ser mais apropriado para modelar os dados em questão através de um modelo de regressão linear.
- Uma alternativa é o modelo de regressão t de Student.
- Vamos considerar a família t de Student de três parâmetros em que os graus de liberdade são conhecidos.

Distribuição t de Student de 3 parâmetros

- Dizemos que $X \sim t_{(\mu,\psi,\nu)}$ se sua fdp é da forma
$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\psi\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{(x-\mu)^2}{\psi\nu}\right)^{-\frac{\nu+1}{2}} \mathbb{1}_{(-\infty,\infty)}(x)$$
- Temos que $\mathcal{E}(X) = \mu$ e $\mathcal{V}(X) = \psi\frac{\nu}{\nu-2}$. Se $\mu = 0$ e $\psi = 1$, então $X \sim t_{(\nu)}$.
- Se $\nu \to \infty$ então $X \xrightarrow{D} N(\mu,\psi)$
- Representação estocástica: se $X \sim t_{(0,\psi,\nu)}$, então $X|W = w \sim N(0,\psi w^{-1})$ e $W \sim gama(\nu/2,\nu/2)$, ou seja $f(x) = \int_{\Re^+} f(x|w)f(w)dw$, $f(w) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}w^{\nu/2-1}e^{-\nu w/2}\mathbb{1}_{(0,\infty)}(w)$
- Se $X \sim t_{(\nu)}$ e $Y = \mu + \sqrt{\psi}X$, temos que $Y \sim t_{(\mu,\psi,\nu)}$

Densidades da distribuição $t_{(\nu)}$ e da distribuição normal



Exemplo 2 (modelo reduzido): desconsiderando o sexo

$$Y_{ij} = \beta_{1i} x_{ij} + \xi_{ij}, i = 1(\text{convencional}), 2(\text{hugger})(\text{tipo de escova});$$

$$j = 1, ..., 26(\text{criança}).$$

- $\xi_{ij} \overset{i.i.d.}{\sim} t_{(0,\psi,\nu)}$. Assim $Y_{ij} \overset{ind.}{\sim} t_{(\beta_{1i} x_{ij},\psi,\nu)}$, $\mathcal{E}(Y_{ij}) = \beta_{1i} x_{ij}$ e $\mathcal{V}(Y_{ij}) = \psi \frac{\nu}{\nu-2}$.
- x_{ij}: IPB pré-escovação da criança j utilizando a escova do tipo i.
- Y_{ij}: IPB pós-escovação da criança j utilizando a escova do tipo i.
- β_{1i} : diminuição (se $\beta_{i1} \in (0,1)$) ou aumento (se $\beta_{i1} > 1$), no IPB quando se usa a escova do tipo i.

Estimação dos parâmetros

- Temos que estimar $\boldsymbol{\beta} = (\beta_{i1}, \beta_{i2})'$ e ψ. Algumas opções são:
    - Estimar $\boldsymbol{\beta}$ por MQO ($\widehat{\boldsymbol{\beta}}_{MQO}$) e usar

$$\widehat{\psi} = \frac{1}{n-2}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{MQO})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{MQO})$$

      . Note que para estimar os parâmetros por MQO não se faz suposições acerca da distribuição de Y_{ij}.
    - Estimar $\boldsymbol{\beta}$ e ψ por MV (usando a verossimilhança original).
    - Estimar $\boldsymbol{\beta}$ e ψ por MV via algoritmo EM (usando a verossimilhança completa).

Verossimilhança original

- Verossimilhança (no nosso caso, n=26)

  $$L(\beta_{11},\beta_{21},\psi) \propto \psi^{-n} \prod_{i=1}^{2} \prod_{j=1}^{n} \left(1 + \frac{(y_{ij} - \beta_{1i}x_{ij})^2}{\psi\nu}\right)^{-\frac{\nu+1}{2}}$$

- Logverossimilhança

  $$l(\beta_{11},\beta_{21},\psi) =$$
  $$-n\ln\psi - \frac{\nu+1}{2} \sum_{i=1}^{2} \sum_{j=1}^{n} \ln\left(1 + \frac{(y_{ij} - \beta_{1i}x_{ij})^2}{\psi\nu}\right) + \text{constante}$$

Sistema de equações de verossimlhança

$$
\begin{cases}
S(\widetilde{\beta}_{11}) = 0 \\
S(\widetilde{\beta}_{12}) = 0 \\
S(\widetilde{\psi}) = 0
\end{cases} \rightarrow
$$

$$
\begin{cases}
(\nu+1)\sum_{j=1}^{n} h_{1j}(\widetilde{\beta}_{i1},\widetilde{\psi})\left(\dfrac{(y_{1j}-\widetilde{\beta}_{11}x_{1j})x_{1j}}{\widetilde{\psi}\nu}\right) = 0 \\
(\nu+1)\sum_{j=1}^{n} h_{2j}(\widetilde{\beta}_{i1},\widetilde{\psi})\left(\dfrac{(y_{2j}-\widetilde{\beta}_{12}x_{2j})x_{2j}}{\widetilde{\psi}\nu}\right) = 0 \\
-\frac{n}{\widetilde{\psi}} + \frac{\nu+1}{2}\sum_{i=1}^{2}\sum_{j=1}^{n} h_{ij}(\widetilde{\beta}_{i1},\widetilde{\psi})\frac{(y_{ij}-\beta_{1i}x_{ij})^2}{\psi^2\nu} = 0
\end{cases}
$$

em que $h_{ij}(\widetilde{\beta}_{i1},\widetilde{\psi}) = \left(1 + \dfrac{(y_{ij}-\widetilde{\beta}_{1i}x_{ij})^2}{\widetilde{\psi}\nu}\right)^{-1}, i = 1, 2$

Obtenção dos estimadores de máxima verossimilhaça e inferência

- As equações de verossimilhança não têm solução explícita e, assim, algum método numérico tem de ser empregado para obtê-las como: Algoritmo Newton-Raphson, Escore de Fisher, Nelder-Med, etc.
- Em nosso caso, a obtenção da informação de Fisher (esperada) é bem complicada.
- Defina $\widehat{\boldsymbol{\theta}}_{MV} = (\widehat{\beta}_{11}, \widehat{\beta}_{12}, \widehat{\psi})'$. Nesse caso, podemos usar o seguinte resultado: para n (tamanho da amostra) suficiente grande $\widehat{\theta}_{MV} \approx N_3(\boldsymbol{\theta}, \mathbf{I}^o(\boldsymbol{\theta})^{-1})$ em que $\mathbf{I}^o = -\mathbf{H}(\boldsymbol{\theta})$ e $\mathbf{H}(\boldsymbol{\theta})$ é a matriz hessiana.

Matriz Hessiana

$$H(\beta_{1i},\beta_{1i}) = -(\nu+1)\sum_{j=1}^{n}\left[\frac{2(y_{ij}-\beta_{1i}x_{ij})^2x_{ij}^2}{\psi^2\nu^2} + \frac{x_{ij}^2}{\psi\nu h_{ij}}\right]$$

$$H(\beta_{11},\beta_{12}) = 0$$

$$H(\beta_{i1},\psi) = -(\nu+1)\sum_{j=1}^{n}\left[\frac{(y_{ij}-\beta_{11}x_{ij})x_{ij}}{\psi^2\nu h_{ij}} + \frac{(y_{ij}-\beta_{11}x_{ij})^3x_{ij}}{\psi^3\nu^2}\right]$$

$$H(\psi,\psi) = -\frac{n}{\psi^2} - \frac{\nu+1}{2}\sum_{i=1}^{2}\sum_{j=1}^{n}\left[\frac{2(y_{ij}-\beta_{1i}x_{is})^2}{\psi^3\nu h_{ij}} + \frac{(y_{ij}-\beta_{1i}x_{ij})^4}{\psi^4\nu}\right]$$

em que $h_{ij} = h_{ij}(\widetilde{\beta}_{i1},\widetilde{\psi})$

Logverossimilhança completa

- Verossimilhança completa (distribuição conjunta de $(\mathbf{y}', \mathbf{w}')'$)

$$
\begin{aligned}
L(\boldsymbol{\beta}, \psi, \mathbf{w}) &\propto \psi^{-n} \prod_{i=1}^{2} \prod_{j=1}^{n} w_{ij} \exp\left\{ - \sum_{i=1}^{2} \sum_{j=1}^{n} \frac{w_{ij}(y_{ij} - \beta_{1i} x_{ij})^2}{2\psi} \right\} \\
&\times \prod_{i=1}^{2} \prod_{j=1}^{n} w_{ij}^{\nu/2-1} e^{-\sum_{i=1}^{2} \sum_{j=1}^{n} \nu w_{ij}/2}
\end{aligned}
$$

- Logverossimilhança completa

$$
\begin{aligned}
l(\boldsymbol{\beta}, \psi, \mathbf{w}) &= -n \ln \psi + \sum_{i=1}^{2} \sum_{j=1}^{n} \ln w_{ij} - \sum_{i=1}^{2} \sum_{j=1}^{n} \frac{w_{ij}(y_{ij} - \beta_{1i} x_{ij})^2}{2\psi} \\
&\times (\nu/2 - 1) \sum_{i=1}^{2} \sum_{j=1}^{n} \ln w_{ij} - \sum_{i=1}^{2} \sum_{j=1}^{n} \nu w_{ij}/2 + \text{constante}
\end{aligned}
$$

em que $\mathbf{w} = (w_{11}, ..., w_{1(26)}, w_{21}, w_{2(26)})'$.

Algorito EM

- Sejam $\mathbf{Y}$ o conjunto de variáveis observadas e $\mathbf{W}$ o conjunto de variáveis não observadas (não observáveis, dados faltantes).
- Em geral, $\mathbf{Y}$ é chamado de dados incompletos e $(\mathbf{Y}, \mathbf{W})$ são os dados completos.
- Seja $l(\boldsymbol{\theta})$ a logverossimilhança em que $\boldsymbol{\theta} \in \Theta \subset \mathcal{R}^p$ é o conjunto de parâmetros a ser estimado.

Estrutura do algoritmo EM

- Sejam $l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{w})$ a log-verossimilhança aumentada e $\boldsymbol{\theta}^{(t)}$ estimativas provisórias para $\boldsymbol{\theta}$. O algoritmo EM pode ser resumido nos seguintes passos

Passo E: Calcule a esperança condicional (na log-verossimilhança) dos dados faltantes condicionado as variáveis observadas e à estimativas provisórias de $\boldsymbol{\theta}^{(t)}$, ou seja

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathcal{E}\left[l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{w})|\mathbf{y}, \boldsymbol{\theta}\right]$$

Passo M: Maximizar a esperança acima em relação à $\boldsymbol{\theta}$ ou seja, obter

$$\boldsymbol{\theta}^{(t+1)} = \text{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$$

até que que algum critério de convergência seja alcançado.

Observações

- O cálculo das esperanças necessárias podem ser complicadas, exigindo o emprego de aproximações analítivas ou numéricas.
- Para a maximização pode ser necessário o emprego de métodos numéricos.
- Na família exponencial as contas ficam mais simples.
- A utilização de valores iniciais apropriados para $\boldsymbol{\theta}$ auxilia na convergência do algoritmo EM.
- A introdução de variáveis "aumentadas" apropriadas, é outro fator de importância na convergência e obtenção de estimativas acuradas.

Estimando os erros-padrão

- Usar a informação de Fisher observada ou esperada (verossimilhança original).
- Identidade de Louis (para se calcular tanto a Informação de Fisher observada ou a esperada). A informação observada, é dada por:

$$\mathcal{E}\left(-H(\widetilde{\boldsymbol{\theta}})^{*}|\mathbf{y},\boldsymbol{\theta}\right) - Cov\left(S(\widetilde{\boldsymbol{\theta}})^{*}|\mathbf{y},\boldsymbol{\theta}\right)$$

- Reamostragem.

Voltando ao exemplo

- Passo E: Calcular

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= -n\ln\psi + \sum_{i=1}^{2}\sum_{j=1}^{n}\mathcal{E}(\ln\widetilde{w}_{ij}|y_{ij},\boldsymbol{\theta}^{(k)}) - \sum_{i=1}^{2}\sum_{j=1}^{n}\frac{\widetilde{w}_{ij}^{(k)}(y_{ij}-\beta_{1i}x_{ij})^2}{2\psi} \\
&\times \quad (\nu/2-1)\sum_{i=1}^{2}\sum_{j=1}^{n}\mathcal{E}(\ln\widetilde{w}_{ij}|y_{ij},\boldsymbol{\theta}^{(k)}) - \sum_{i=1}^{2}\sum_{j=1}^{n}\nu\widetilde{w}_{ij}^{(k)}/2 + \text{constante}
\end{aligned}
$$

em que $\widetilde{w}_{ij}^{(k)} = \mathcal{E}(W_{ij}|y_{ij},\widetilde{\boldsymbol{\theta}}^{(k)}) =$
$(\frac{\nu}{2}+1)\left(\frac{1}{2}(\nu+(y_{ij}-\beta_{i1}^{(k-1)}x_{ij})^2/\psi^{(k-1)})\right)^{-1}$, pois
$W_{ij}|y_{ij},\widetilde{\boldsymbol{\theta}}^{(k)} \sim gama(\nu/2+1,\nu/2+(y_{ij}-\beta_{i1}x_{ij})^2/2\psi)$

Voltando ao exemplo

- Passo M: Calcular

$$\widetilde{\boldsymbol{\beta}}^{(k)} = \left(\mathbf{X}'\left(\mathbf{W}^{(k)}\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{W}^{(k)}\right)^{-1}\mathbf{y}$$

$$\widetilde{\psi}^{(k)} = \frac{1}{2n}\sum_{i=1}^{2}\sum_{j=1}^{n}\widetilde{w}_{ij}^{(k)}(y_{ij} - \widetilde{\beta}_{1i}x_{ij})^2$$

em que $\mathbf{X}$ é matriz de planejamento do modelo, $\mathbf{y}$ o vetor com as respostas observadas e

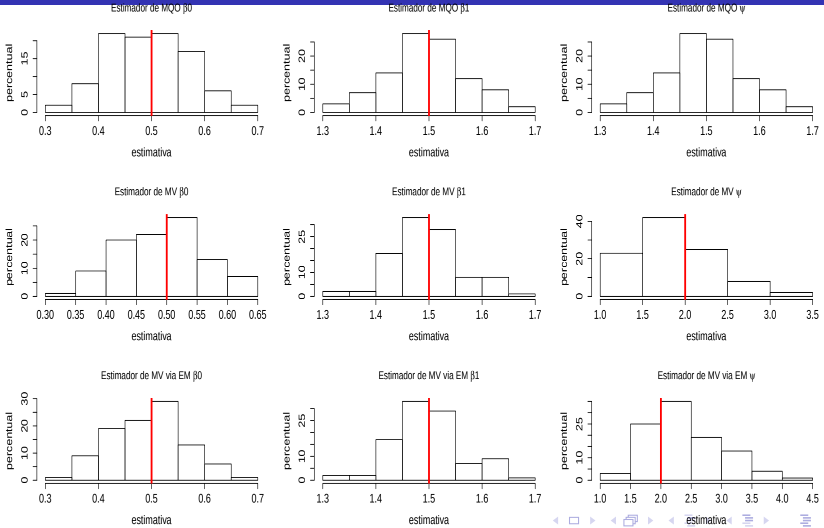
$\mathbf{W}^{(k)} = \psi^{(k-1)}\mathrm{diag}(w_{11}^{(k)}, ..., w_{1(26)}^{(k)}, w_{21}^{(k)}, ..., w_{2(26)}^{(k)})^{-1}$

Estudo de simulação

$Y_{ij} = \beta_{1i} x_{ij} + \xi_{ij}, i = 1, 2; j = 1, ..., 26$(criança).

- $\xi \overset{i.i.d.}{\sim} t_{(0,\psi=2,\nu=5)}$
- $\beta_{11} = 0,5$ e $\beta_{12} = 1,5$.
- Foram geradas 100 réplicas de tamanho 52.
- Em cada réplica, os parâmetros β_{11}, β_{12} e ψ foram estimados por MQO, MV e MV via EM (estes dois últimos através do modelo acima).
- Para cada método de estimação e parâmetros, obtivemos a média e variância das estimativas das réplicas, bem como o vício e a raiz quadrática do erro quadrático médio (rqeqm).

Histogramas das estimativas



Mét. de est.	Par.	Valor verd.	Média	Vício	Var.	RQEQM
MQO	β_{11}	0,5	0,495	-0,005	0,005	0,074
	β_{12}	1,5	1,499	-0,001	0,005	0,071
	ψ	2	3,195	1,195	1,275	1,644
MV	β_{11}	0,5	0,494	-0,006	0,005	0,069
	β_{12}	1,5	1,498	-0,002	0,004	0,064
	ψ	2	1,883	-0,117	0,222	0,485
MV via EM	β_{11}	0,5	0,494	-0,006	0,005	0,069
	β_{12}	1,5	1,498	-0,002	0,004	0,064
	ψ	2	2,400	0,400	0,368	0,727

Exemplo 2 (modelo reduzido): desconsiderando o sexo

$$Y_{ij} = \beta_{1i} x_{ij} + \xi_{ij}, i = 1(\text{convencional}), 2(\text{hugger})(\text{tipo de escova});$$

$$j = 1, ..., 26(\text{criança}).$$

- $\xi_{ij} \overset{i.i.d.}{\sim} t_{(0,\psi,\nu)}$. Assim $Y_{ij} \overset{ind.}{\sim} t_{(\beta_{1i}x_{ij},\psi,\nu)}$, $\mathcal{E}(Y_{ij}) = \beta_{1i}x_{ij}$ e $\mathcal{V}(Y_{ij}) = \psi\frac{\nu}{\nu-2}$.
- x_{ij}: IPB pré-escovação da criança j utilizando a escova do tipo i.
- Y_{ij}: IPB pós-escovação da criança j utilizando a escova do tipo i.
- β_{1i} : diminuição (se $\beta_{i1} \in (0,1)$) ou aumento (se $\beta_{i1} > 1$), no IPB quando se usa a escova do tipo i.
- Regressão "gridge": ajustou-se um modelo para determinados valores de $\nu \in \{2,5; 3,5; 8; 10; 20; 30\}$.

Voltando ao exemplo 2

V1	AICn	BICn
Normal	7,90	13,76
t com $\nu = 2{,}5$	6,27	12,12
t com $\nu = 3$	5,93	11,79
t com $\nu = 5$	5,84	11,69
t com ν =8	6,21	12,07
t com ν =10	6,43	12,28
t com ν =20	7,03	12,88
t com ν =30	7,28	13,14

Comentários

- O modelo de regressão mais apropriado é aquele que considera uma distribuição $t_{(5)}$ graus de liberdade.
- No entanto, é necessári verificar se o modelo se ajustou bem aos dados (análise de resíduos e preditiva).
- Resíduos : adaptação dos resíduos vistos para o modelo de regressão normal linear homocedástico (ver a literatura).