Modelos Lineares Generalizados - Estimação em Modelos Lineares Generalizados

Erica Castilho Rodrigues

23 de Maio de 2017

Introdução

- Vimos como encontrar o EMV usando algoritmos numéricos.
- Duas possibilidades:
  - Método de Newton Raphson;
  - Método Escore de Fisher.
- Iremos aplicar esses algoritmos para o caso do MLG.
- Em Regressão Linear o EMV tem forma fechada.
- Pode-se mostrar que é equivalente ao estimador de mínimos quadrados.

- Já nos MGL's os estimadores dos coeficientes β não tem forma fechada.
- Esses estimadores serão obtidos usando os algoritmos que vimos anteriormente.
- Considere uma amostra aleatória

$$Y_1, Y_2, \ldots, Y_n$$

que pertence à família exponencial em sua forma canônica, ou seja,

$$f(y_i, \theta_i) = \exp\{y_i b(\theta_i) + c(\theta_i) + d(\theta_i)\}\ .$$

- Considere que o preditor linear está relacionado com a média da seguinte maneira

$$E(Y_i) = \mu_i \quad \mu_i = g(\mathbf{x}_i^t \boldsymbol{\beta})$$

- Objetivo: estima o vetor $\boldsymbol{\beta}$.
- Temos que a log-verossimilhança é dada por

$$l(y_i, \theta_i) = \log(f(y_i, \theta_i)) = y_i b(\theta_i) + c(\theta_i) + d(\theta_i) \, .$$

- A log-verossimilhança conjunta é dada por

$$l(\mathbf{y}, \boldsymbol{\theta}) = \sum_i l(y_i, \theta_i) = \sum_{i=1}^{n} (y_i b(\theta_i) + c(\theta_i) + d(\theta_i))$$

$$= \sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(\theta_i) \, .$$

- Queremos encontrar o EMV para cada um dos β_j (j=1,2,...,p).
- Devemos então derivar a log-verossimilhança e igualar a zero.
- Vamos primeiro obter a função escore

$$U_j(\boldsymbol{\theta}) = \frac{dl(\mathbf{y}, \boldsymbol{\theta})}{d\beta_j} = \frac{d\sum_i l(y_i, \theta_i)}{d\beta_j} .$$

- Pode-se mostrar que a função escore é dada por

$$U_j(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\frac{(y_i - \mu_i)}{Var(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] .$$

- Queremos então encontrar a a solução de

$$U_j(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\frac{(y_i - \mu_i)}{Var(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] = 0$$

para $j = 1, 2, \ldots, p$.
- Usaremos o método Escore de Fisher.
- Não entraremos em detalhes sobre a construção do algoritmo.

- Vamos chamar **b** o estimador do vetor $\boldsymbol{\beta}$.
- Pode-se mostrar que a equação de atualização do algoritmo é dada por

$$\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{b}^{(m)} = \mathbf{X}^T\mathbf{W}\mathbf{z}$$

  - **W** é uma matriz diagonal $n \times n$, cuja i-ésima entrada é dada por
$$[\mathbf{W}]_{ii} = \frac{1}{Var(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2$$
  - e **z** é um vetor cuja i-ésiman é dada por

$$\mathbf{z}_i = \sum_{k=1}^{p} x_{ik}b_k^{(m-1)} + (y_i - \mu_i)\left(\frac{\partial\eta_i}{\partial\mu_i}\right) .$$

- A equação
$$\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{b}^{(m)} = \mathbf{X}^T\mathbf{W}\mathbf{z}$$
tem o mesmo formato das equações normais no modelo linear quando usamos Mínimos Quadrados Ponderados.
- A diferença é que aqui ela é resolvida iterativamente.
- Esse método é chamado Método de Mínimos Quadrados Ponderados Iterativo.

- A maioria dos softwares utiliza esse método de estimação.
- Escolhem um valor inicial $\mathbf{b}^{(0)}$.
- Utilizam esse valor para encontrar $\mathbf{W}$ e $\mathbf{z}$.

$$[\mathbf{W}]_{ii} = \frac{1}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

$$\mathbf{z}_i = \sum_{k=1}^{p} x_{ik} b_k^{(m-1)} + (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) .$$

- Utilizam $\mathbf{z}$ e $\mathbf{W}$ para atualizar $\mathbf{b}^{(1)}$

$$\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{b}^{(1)} = \mathbf{X}^T\mathbf{W}\mathbf{z}$$

- $\mathbf{b}^{(1)}$ é então usado para obter melhores aproximações de $\mathbf{W}$ e $\mathbf{z}$.
- O algoritmo continua até que

$$|\mathbf{b}^{(m)} - \mathbf{b}^{(m-1)}|$$

seja bem pequeno.

Exemplo:

- Vamos ajutar uma Regressão de Poisson.
- Considere uma amostra aleatória

$$Y_1, \ldots, Y_n$$

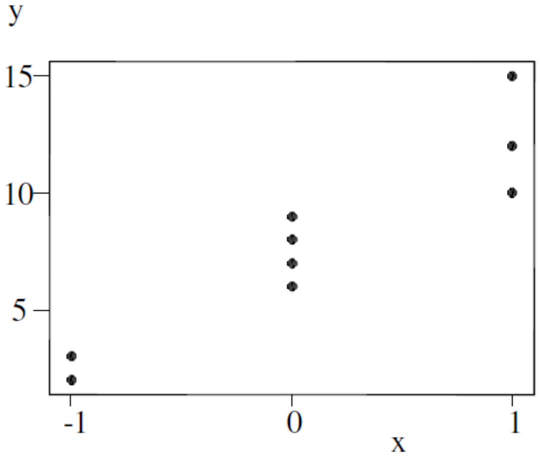
tal que $Y_i \sim^{iid} Poisson(\mu_i)$.

- A Tabela a seguir apresenta os dados observados

y_i	2	3	6	7	8	9	10	12	15
x_i	-1	-1	0	0	0	0	1	1	1

Exemplo: (continuação)

- A figura a seguir mostra o gráfico de disperão entre as duas variáveis.



Exemplo: **(continuação)**

- Como a variância se comporta?
- Ela aumenta quando Y aumenta.
- Isso faz sentido no modelo Poisson?
- Sim pois
$$E(Y_i) = Var(Y_i) .$$
- Vamos modelar a média dessas variáveis da seguinte maneira
$$E(Y_i) = \mu_i = \beta_0 + \beta_1 x_i .$$
- Precisamo somar o erro ϵ_i nesse caso?
- Não.

Exemplo: **(continuação)**

- Então o modelo é perfeito? Não comete erro nenhum?
- O erro vem do fato de

$$Y_i \sim Poisson(\mu_i) \ .$$

- Os valores observamos de Y_i estarão em torno de μ_i.
- Não serão exatamente μ_i.
- Assim como acontece no modelo de regressão

$$Y_i \sim N(\mu_i, \sigma^2) \ .$$

- A diferença é que o modelo de regressão pode ser escrito de duas maneiras equivalentes

$$Y_i \sim N(\mu_i, \sigma^2) \quad Y_i = \mu_i + \epsilon_i \text{ com } \epsilon_i \sim N(0, \sigma^2) \ .$$

- Os MLG's podem ser escritos apenas do primeiro modo.

Exemplo: (continuação)

- Temos então que

$$Y_i \sim Poisson(\mu_i)$$

com

$$E(Y_i) = \mu_i = \beta_0 + \beta_1 x_i = \mathbf{x}_i^T \boldsymbol{\beta}$$

onde

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \mathbf{x}_i = \begin{bmatrix} 1 \\ x_1 \end{bmatrix}$$

- Qual função de ligação estamos usando aqui?
- Identidade

$$\mu_i = g(\mathbf{x}_i^T \boldsymbol{\beta}) = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i$$

Exemplo: (continuação)

- Vejamos como usar o Método Escore de Fisher para estimar $\boldsymbol{\beta}$.
- Vamos obter a matriz **W**.
- Vimos que

$$[\mathbf{W}]_{ii} = \frac{1}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 .$$

- Temos que

$$Var(Y_i) = \mu_i = \beta_0 + \beta_1 x_i .$$

- Além disso

$$\mu_i = \eta_i \Rightarrow \frac{\partial \mu_i}{\partial \eta_i} = 1.$$

Exemplo: (continuação)

- Portanto

$$[\mathbf{W}]_{ii} = \frac{1}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = \frac{1}{\beta_0 + \beta_1 x_i}\ .$$

- Como não temos o valor de β_0 e β_1, substituimos pelos seus estimadores b_0 e b_1

$$[\mathbf{W}]_{ii} = \frac{1}{b_0 + b_1 x_i}\ .$$

- Veremos agora como fica **z.**
- Vimos que

$$\mathbf{z}_i = \sum_{k=1}^{p} x_{ik} b_k^{(m-1)} + (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i}\right)\ .$$

- Temos que

$$\eta_i = \mu_i \Rightarrow \frac{\partial \eta_i}{\partial \mu_i} = 1.$$

Exemplo: (continuação)

- Portanto

$$\mathbf{z}_i = \sum_{k=1}^{p} x_{ik} b_k^{(m-1)} + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)$$

$$= b_0 + b_1 x_i + (y_i - b_0 - b_1 x_i) = y_i$$

  onde

  - b_0 é estimador de β_0;
  - b_1 é estimador de β_1.

- O próximo passo é encontrar a equação para atualização do estimador $\mathbf{b}^{(m)}$.

Exemplo: (continuação)

- Essa atualização é obtida através de

$$\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{b}^{(m)} = \mathbf{X}^T\mathbf{W}\mathbf{z}$$

- Vejamos como fica o termo $\mathbf{X}^T\mathbf{W}\mathbf{X}$.
- Temos que

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} \frac{1}{b_0+b_1x_1} & 0 & \dots & 0 \\ 0 & \frac{1}{b_0+b_1x_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1}{b_0+b_1x_n} \end{bmatrix}$$

logo

$$\mathbf{X}^T\mathbf{W} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_3 & \dots & x_n \end{bmatrix} \begin{bmatrix} \frac{1}{b_0+b_1x_1} & 0 & \dots & 0 \\ 0 & \frac{1}{b_0+b_1x_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1}{b_0+b_1x_n} \end{bmatrix}$$

Exemplo: (continuação)

$$\mathbf{X}^T\mathbf{W} = \begin{bmatrix} \frac{1}{b_0+b_1x_1} & \frac{1}{b_0+b_1x_2} & \cdots & \frac{1}{b_0+b_1x_n} \\ \frac{x_1}{b_0+b_1x_1} & \frac{x_2}{b_0+b_1x_2} & \cdots & \frac{x_n}{b_0+b_1x_n} \end{bmatrix}$$

então

$$\mathbf{X}^T\mathbf{W}\mathbf{X} = \begin{bmatrix} \frac{1}{b_0+b_1x_1} & \frac{1}{b_0+b_1x_2} & \cdots & \frac{1}{b_0+b_1x_n} \\ \frac{x_1}{b_0+b_1x_1} & \frac{x_2}{b_0+b_1x_2} & \cdots & \frac{x_n}{b_0+b_1x_n} \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^n \frac{1}{b_0+b_1x_i} & \sum_{i=1}^n \frac{x_i}{b_0+b_1x_i} \\ \sum_{i=1}^n \frac{x_i}{b_0+b_1x_i} & \sum_{i=1}^n \frac{x_i^2}{b_0+b_1x_i} \end{bmatrix}$$

Exemplo: (continuação)

- Veremos agora como fica o termo $\mathbf{X}^T\mathbf{W}\mathbf{z}$.
- Temos que

$$\mathbf{z} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} \qquad \mathbf{X}^T\mathbf{W} = \begin{bmatrix} \frac{1}{b_0+b_1x_1} & \frac{1}{b_0+b_1x_2} & \cdots & \frac{1}{b_0+b_1x_n} \\ \frac{x_1}{b_0+b_1x_1} & \frac{x_2}{b_0+b_1x_2} & \cdots & \frac{x_n}{b_0+b_1x_n} \end{bmatrix}$$

- Logo

$$\mathbf{X}^T\mathbf{W}\mathbf{z} = \begin{bmatrix} \frac{1}{b_0+b_1x_1} & \frac{1}{b_0+b_1x_2} & \cdots & \frac{1}{b_0+b_1x_n} \\ \frac{x_1}{b_0+b_1x_1} & \frac{x_2}{b_0+b_1x_2} & \cdots & \frac{x_n}{b_0+b_1x_n} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$$

Exemplo: **(continuação)**

$$\mathbf{X}^T\mathbf{W}\mathbf{z} = \begin{bmatrix} \sum_{i=1}^n \frac{y_i}{b_0+b_1x_i} \\ \sum_{i=1}^n \frac{x_iy_i}{b_0+b_1x_i} \end{bmatrix}$$

- Portano $\mathbf{b}^{(m)}$ será atualizado a partir da equação

$$\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{b}^{(m)} = \mathbf{X}^T\mathbf{W}\mathbf{z}$$

ou seja

$$\begin{bmatrix} \sum_{i=1}^n \frac{1}{b_0+b_1x_i} & \sum_{i=1}^n \frac{x_i}{b_0+b_1x_i} \\ \sum_{i=1}^n \frac{x_i}{b_0+b_1x_i} & \sum_{i=1}^n \frac{x_i^2}{b_0+b_1x_i} \end{bmatrix} \mathbf{b}^{(m)} = \begin{bmatrix} \sum_{i=1}^n \frac{y_i}{b_0+b_1x_i} \\ \sum_{i=1}^n \frac{x_iy_i}{b_0+b_1x_i} \end{bmatrix}$$

Exemplo: (continuação)

- Para os dados observados temos que

$$\mathbf{y} = \mathbf{z} = \begin{bmatrix} 2 \\ 3 \\ \cdots \\ 15 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

- Vamos escolher os seguintes valores iniciais

$$b_0^{(0)} = 7 \quad b_1^{(0)} = 5\,.$$

- Temos que

$$(\mathbf{X}^T\mathbf{W}\mathbf{X})^{(1)} = \begin{bmatrix} 1.821429 & -0.75 \\ -0.75 & 1.25 \end{bmatrix}$$

Exemplo: (continuação)

$$(\mathbf{X}^T\mathbf{W}\mathbf{z})^{(1)} = \begin{bmatrix} 9.869048 \\ 0.5833 \end{bmatrix}$$

- Portanto

$$\begin{bmatrix} 1.821429 & -0.75 \\ -0.75 & 1.25 \end{bmatrix} \mathbf{b}^{(1)} = \begin{bmatrix} 9.869048 \\ 0.5833 \end{bmatrix}$$

o que implica que

$$\mathbf{b}^{(1)} = \begin{bmatrix} 1.821429 & -0.75 \\ -0.75 & 1.25 \end{bmatrix}^{-1} \begin{bmatrix} 9.869048 \\ 0.5833 \end{bmatrix} = \begin{bmatrix} 7.4514 \\ 4.9375 \end{bmatrix}$$

Exemplo: (continuação)

- A tabela a seguir mostra os valores de **b** para alguns passos do algoritmo

m	1	2	3	4
$b_1^{(m)}$	7	7.45139	7.45163	7.45163
$b_2^{(m)}$	5	4.93750	4.93531	4.93530

- O algoritmo parece ter convergido para os seguintes valores de **b**

$$b_0^{(m)} = \hat{\beta}_0 = 7.45 \quad b_1^{(m)} = \hat{\beta}_1 = 4.93\,.$$

Propriedades do Estimador Máxima Verossimilhança

- Assintoticamente não viesados

$$\lim_{n \to \infty} E(\hat{\theta}) = \theta \,.$$

- Consistente

$$\hat{\theta} \xrightarrow{\mathcal{P}} \theta$$

  ($\hat{\theta}$) converge em probabilidade para θ).
- Variância mínima:
  - $\hat{\theta}$ possui variância mínima para amostras grandes.
- Distribuição assintótica:
  - a distribuição de $\hat{\theta}$ converge para uma normal.

- Veremos agora as propriedades do EMV para o caso dos Modelos Lineares Generalizados.

- Duas principais ferramentas para inferência estatística:
  - intervalos de confiança e testes de hipóteses.
- Veremos como obtê-los no caso nos Modelos Lineares Generalizados.

Intervalo de Confiança

- Fornece uma estimativa intervalar para o parâmetro.
- A amplitude do intervalo dá uma ideia da precisão dessa estimativa.

Testes de Hipóteses

- Podemos comparar dois modelos.
- Queremos identificar qual deles se ajusta melhor aos dados.
- Um modelo é mais simples que o outro - tem menos parâmetros.
- Esse modelo é chamado modelo nulo e corresponde a H_0.
- Deve ser um caso particular do modelo H_1.
- Se os dois modelos ajustam bem os dados:
  - devemos escolher H_0, pelo critério da parcimônia.

- Para fazermos inferências sobre o parâmetro:
  - precisamos da distribuição de probabilidade do estimador.
- Seja S uma estatística de interesse.
- Em geral, pode-se mostrar que, para tamanhos de amostra grandes

$$\frac{S - E(S)}{\sqrt{Var(S)}} \sim N(0,1)$$

$$\frac{(S - E(S))^2}{Var(S)} \sim \chi^2_{(1)}$$

- Para o caso que **S** é um vetor de dimensão p temos que

$$[\mathbf{S} - E(\mathbf{S})]^T \mathbf{V}^{-1} [\mathbf{S} - E(\mathbf{S})] \sim \chi^2_{(p)}$$

onde

$$Cov(\mathbf{S}) = \mathbf{V}\,.$$

- Vamos denotar por **b** o estimador do vetor de parâmetros $\boldsymbol{\beta}$.
- Pode-se mostrar que, se a amostra tem tamanho grande,

$$E(\mathbf{b}) = \boldsymbol{\beta}$$

$$(\mathbf{b} - \boldsymbol{\beta})^T I(\boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta}) \sim \chi^2_{(p)}$$

ou

$$\mathbf{b} \sim N\left(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1}\right)$$

onde $I(\boldsymbol{\beta})$ é matriz de Informação de Fisher de $\boldsymbol{\beta}$ cujo termo jk é dado por

$$[I(\boldsymbol{\beta})]_{jk} = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 .$$

- No caso em que $\boldsymbol{\beta}$ tem apenas um parâmetro temos que, para *n* grande
$$\mathbf{b} \sim N\left(\beta, I(\beta)^{-1}\right)$$
onde
$$I(\beta) = \sum_{i=1}^{n} \frac{x_i^2}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 .$$
- No caso dos MLG's todos esses resultados só valem se a amostra é grande.
- No caso dos modelos de Regressão Linear esses resultados são exatos.

Exemplo:

- Vamos considerar o caso da distribuição Normal.
- Ele não deixa de ser um caso específico dos MLG's.
- Veremos que os resultados coincidem com aqueles obtidos no curso de Regressão Linear.
- Iremos considerar então que

$$Y_i \sim N(\mu_i, \sigma^2) \quad \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$$

onde $\boldsymbol{\beta}$ é um vetor de parâmetros de dimensão p.

Exemplo: (continuação)

- Vejamos como usar o Método Escore de Fisher para estimar $\boldsymbol{\beta}$.
- Vamos obter a matriz **W**.
- Vimos que

$$[\mathbf{W}]_{ii} = \frac{1}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 .$$

- Temos que

$$Var(Y_i) = \sigma^2 .$$

- Além disso

$$\mu_i = \eta_i \Rightarrow \frac{\partial \mu_i}{\partial \eta_i} = 1.$$

Exemplo: (continuação)

- Portanto

$$[\mathbf{W}]_{ii} = \frac{1}{Var(Y_i)} = \frac{1}{\sigma^2}\,.$$

- Além disso

$$\mathbf{z}_i = \sum_{k=1}^{p} x_{ik} b_k^{(m-1)} + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)$$

$$= \sum_{k=1}^{p} x_{ik} b_k^{(m-1)} + (y_i - \mu_i) = y_i$$

Exemplo: (continuação)

- Vimos que a equação de atualização do parâmetro é dada por

$$\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{b}^{(m)} = \mathbf{X}^T\mathbf{W}\mathbf{z} \,.$$

- E temos que

$$\mathbf{W} = \frac{1}{\sigma^2}\mathbf{I}$$

$$\mathbf{z} = \mathbf{y}$$

logo a equação fica

$$\frac{1}{\sigma^2}\mathbf{X}^T\mathbf{X}\mathbf{b}^{(m)} = \frac{1}{\sigma^2}\mathbf{X}^T\mathbf{y}$$

e portanto

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

que é justamente o estimador do mínimos quadrados do modelo de regressão.

Exemplo: (continuação)

- Temos nesse caso que

$$\mathbf{b} \sim N\left(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1}\right)$$

os termos da matriz de Informação de Fisher ficam

$$[I(\boldsymbol{\beta})]_{jk} = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{\sigma^2}.$$

- Temos então que

$$I(\boldsymbol{\beta}) = \frac{1}{\sigma^2}(\mathbf{X}^T\mathbf{X})$$

ou seja

$$I(\boldsymbol{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

- Isso significa que

$$\mathbf{b} \sim N\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right)$$

que foi o mesmo resultado visto no curso de Regressão Linear.

- **Exercício:** Verifique se a distribuição normal inversa pertence à família exponencial. Caso isso seja verdade, encontre o parâmetro canônico e a ligação canônica dessa distriuição.

$$f(y;\mu;\sigma^2) = \left(\frac{1}{2\pi\sigma^2 y^3}\right)^{1/2} \exp\left\{-\frac{(y-\mu)^2}{2\mu^2\sigma^2 y}\right\}$$

$$f(y_i,\theta_i) = \exp\{y_i b(\theta_i) + c(\theta_i) + d(\theta_i)\}\ .$$