Lista 1

Modelos lineares generalizados

Paulo Ricardo Seganfredo Campana

16 de dezembro de 2023

Questão 1. Mostre que a distribuição Binomial, pertence à família exponencial. Calcule μ e $V(\mu)$.

$$\begin{aligned} f(y;n,\mu) &= \binom{n}{y}\mu^y(1-\mu)^{n-y} \\ &= \exp\left\{\ln\binom{n}{y} + y\ln\mu + (n-y)\ln(1-\mu)\right\} \\ &= \exp\left\{y\ln\left(\frac{\mu}{1-\mu}\right) + n\ln(1-\mu) + \ln\binom{n}{y}\right\} \end{aligned}$$

$$\phi = 1, \quad \theta = \ln\left(\frac{\mu}{1-\mu}\right), \quad c(y,\phi) = \ln\binom{n}{y}$$

$$b(\theta) = -n\ln(1-\mu) = n\ln(1+e^\theta)$$

$$\begin{aligned} E(X) &= b'(\theta) \\ &= \frac{\partial}{\partial\theta} n\ln(1+e^\theta) \\ &= \frac{ne^\theta}{1+e^\theta} \\ &= \frac{n\frac{\mu}{1-\mu}}{1+\frac{\mu}{1-\mu}} \\ &= n\mu \end{aligned}$$

$$\begin{aligned} V(\mu) &= b''(\theta) \\ &= \frac{\partial}{\partial\theta}\frac{ne^\theta}{1+e^\theta} \\ &= \frac{ne^\theta}{(1+e^\theta)^2} \\ &= \frac{n\frac{\mu}{1-\mu}}{(1+\frac{\mu}{1-\mu})^2} \\ &= n\mu(1-\mu) \end{aligned}$$

Questão 2. Mostre que a distribuição Normal inversa pertence à família exponencial. Calcule μ e $V(\mu)$.

$$
\begin{aligned}
f(y;\mu,\phi) &= \sqrt{\frac{\phi}{2\pi y^3}} \exp\left\{\frac{-\phi(y-\mu)^2}{2y\mu^2}\right\} \\
&= \exp\left\{\frac{-\phi(y-\mu)^2}{2y\mu^2} + \frac{1}{2}\ln\left(\frac{\phi}{2\pi y^3}\right)\right\} \\
&= \exp\left\{\phi\left(\frac{y}{2\mu^2} - \frac{1}{\mu} + \frac{1}{2y}\right) + \frac{1}{2}\ln\left(\frac{\phi}{2\pi y^3}\right)\right\} \\
&= \exp\left\{\phi\left(\frac{y}{2\mu^2} - \frac{1}{\mu}\right) + \frac{\phi}{2y} + \frac{1}{2}\ln\left(\frac{\phi}{2\pi y^3}\right)\right\}
\end{aligned}
$$

$$
\phi = \phi, \quad \theta = \frac{1}{2\mu^2}, \quad c(y,\phi) = \frac{\phi}{2y} + \frac{1}{2}\ln\left(\frac{\phi}{2\pi y^3}\right)
$$

$$
b(\theta) = \frac{1}{\mu} = \sqrt{2\theta}
$$

$$
\begin{aligned}
E(X) &= b'(\theta) \\
&= \frac{\partial}{\partial\theta}\sqrt{2\theta} \\
&= \frac{1}{\sqrt{2\theta}} \\
&= \mu
\end{aligned}
$$

$$
\begin{aligned}
V(\mu) &= b''(\theta) \\
&= \frac{\partial}{\partial\theta}\frac{1}{\sqrt{2\theta}} \\
&= \frac{-\theta^{-3/2}}{2\sqrt{2}} \\
&= \frac{-(2\mu^2)^{3/2}}{2\sqrt{2}} \\
&= -\mu^3
\end{aligned}
$$

Questão 3. Mostre que a distribuição logarítmica pertence à família exponencial. Calcule μ e $V(\mu)$.

$$
\begin{aligned}
f(y;\rho) &= \frac{\rho^y}{-y\ln(1-\rho)} \\
&= \exp\left\{y\ln\rho - \ln y - \ln(-\ln(1-\rho))\right\} \\
&= \exp\left\{y\ln\rho - \ln(-\ln(1-\rho)) - \ln y\right\}
\end{aligned}
$$

$$
\phi = 1, \quad \theta = \ln\rho, \quad c(y,\phi) = -\ln y
$$

$$
b(\theta) = \ln(-\ln(1-\rho)) = \ln(-\ln(1-e^\theta))
$$

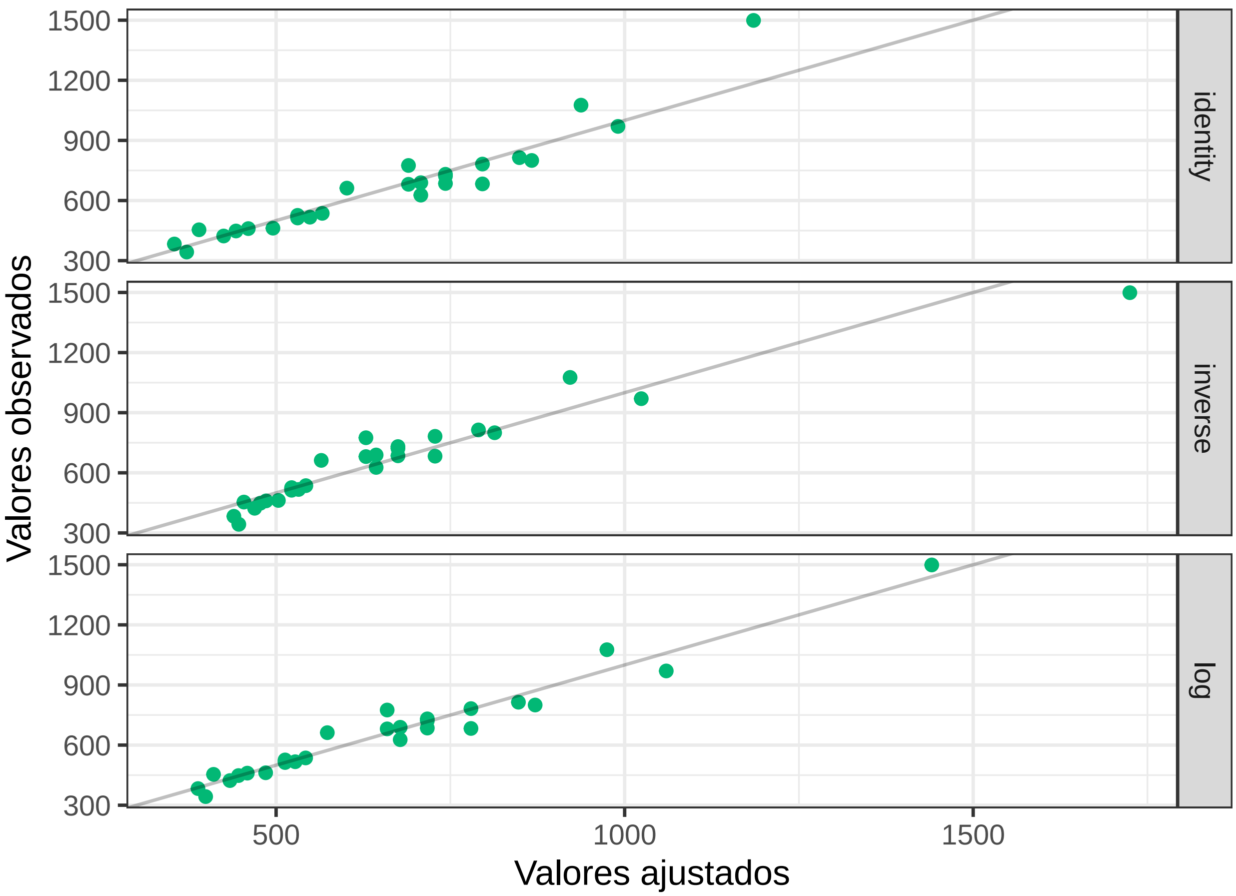
$$
\begin{aligned}
E(X) &= b'(\theta) \\
&= \frac{\partial}{\partial\theta}\ln(-\ln(1-e^\theta)) \\
&= \frac{e^\theta}{-\ln(1-e^\theta)(1-e^\theta)} \\
&= \frac{\rho}{-\ln(1-\rho)(1-\rho)}
\end{aligned}
$$

$$
\begin{aligned}
V(\mu) &= b''(\theta) \\
&= \frac{\partial}{\partial\theta}\frac{e^\theta}{-\ln(1-e^\theta)(1-e^\theta)} \\
&= \frac{-e^\theta\ln(1-e^\theta) - e^{2\theta}}{(\ln(1-e^\theta)(1-e^\theta))^2} \\
&= \frac{-\rho\ln(1-\rho) - \rho^2}{(\ln(1-\rho)(1-\rho))^2}
\end{aligned}
$$

Questão 4. O conjunto de dados descrito no arquivo censo.txt, extraído do censo do IBGE de 2000, apresenta para cada unidade da federação o número médio de anos de estudo e a renda média mensal (em reais) do chefe ou chefes do domicílio. Suponha que a renda tenha distribuição gama.

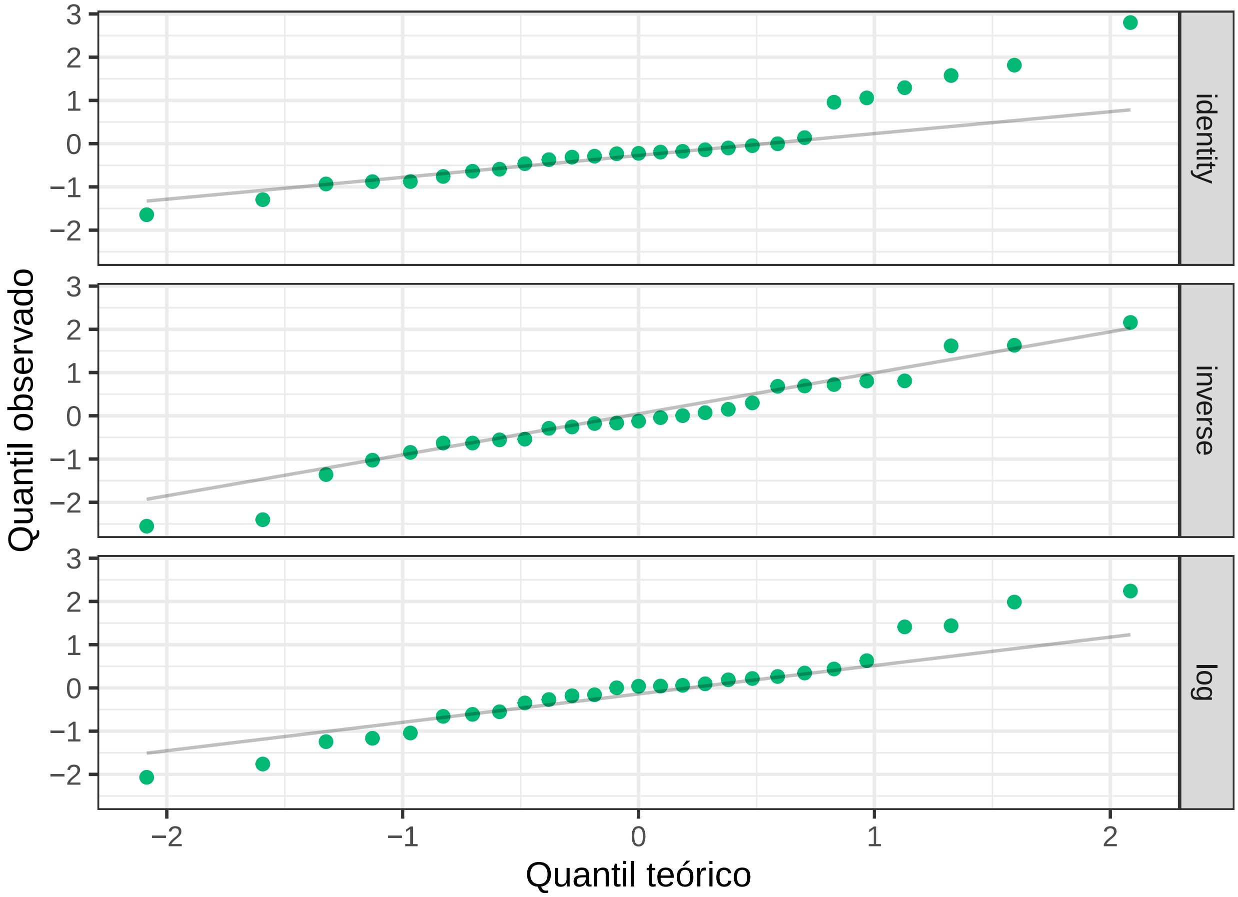
a) Realize o ajuste da gama com as possíveis funções de ligação e decida, entre elas, qual a função de ligação é melhor. Justifique sua escolha.

Eu escolho a função de ligação log, por estimar melhor as observações tanto para valores altos e baixos de renda.



b) Selecione as variáveis. Realize uma análise residual para o melhor modelo obtido. O modelo é adequado? Por quê?

Os dados possuem apenas uma variável independente, a seleção não é necessária. O modelo com função de ligação inversa obteve melhor normalidade nos resíduos, podemos aceitar os três pois os desvios normalizados são menores que o quantil da Chi-quadrado para estes dados.



Função de ligação	$D(y;\mu)/\phi$	Quantil χ^2
Identidade	23.592	37.652
Inversa	25.056	37.652
Log	24.625	37.652

c) Realize uma análise de diagnóstico para o melhor modelo obtido em (a). Analise os resultados obtidos.

Os coeficientes são altamentes significativos atráves o teste t, os testes de Lilliefors e Shapiro-Wilk contribuem para a hipótese de normalidade dos resíduos (p-valores 0.390 e 0.202). Apenas uma observação dos dados possui distância de Cook e h_{ii} fora do padrão.

Coeficiente	Estimativa	Erro padrão	Estatística	p-valor
(Intercepto)	4.984	0.0679	73.4	1.03×10^{-30}
ano	0.279	0.0128	21.9	7.98×10^{-18}