

MAE5763 - Modelos Lineares Generalizados - Resolução da Lista 1

Guilherme Marthe - 8661962

9/21/2020

1 Exercício 1

1.1 Forma matricial do modelo

O modelo na sua forma matricial $y = X\beta + \epsilon$ depende de subcomponentes para cada $i = 1, 2, 3$. A da parametrização sugerida é a seguinte:

$$\begin{aligned}y_{1,j} &= \alpha + \epsilon_{1,j} \\y_{2,j} &= \alpha + \Delta + \epsilon_{2,j} \\y_{3,j} &= \alpha - \Delta + \epsilon_{3,j}\end{aligned}$$

Com $j = 1, \dots, r$ e $\epsilon_{i,j} \sim N(0, \sigma^2)$. Assim, a matriz modelo, vetor de resposta e o componente de erro de cada $i = 1, 2, 3$ é:

$$\begin{aligned}X_1 &= \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix}_{r \times 2} & y_1 &= \begin{bmatrix} y_{1,1} \\ y_{1,2} \\ \vdots \\ y_{1,r} \end{bmatrix}_{r \times 1} & \epsilon_1 &= \begin{bmatrix} \epsilon_{1,1} \\ \epsilon_{1,2} \\ \vdots \\ \epsilon_{1,r} \end{bmatrix}_{r \times 1} \\ \\ X_2 &= \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}_{r \times 2} & y_2 &= \begin{bmatrix} y_{2,1} \\ y_{2,2} \\ \vdots \\ y_{2,r} \end{bmatrix}_{r \times 1} & \epsilon_2 &= \begin{bmatrix} \epsilon_{2,1} \\ \epsilon_{2,2} \\ \vdots \\ \epsilon_{2,r} \end{bmatrix}_{r \times 1} \\ \\ X_3 &= \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \end{bmatrix}_{r \times 2} & y_3 &= \begin{bmatrix} y_{3,1} \\ y_{3,2} \\ \vdots \\ y_{3,r} \end{bmatrix}_{r \times 1} & \epsilon_3 &= \begin{bmatrix} \epsilon_{3,1} \\ \epsilon_{3,2} \\ \vdots \\ \epsilon_{3,r} \end{bmatrix}_{r \times 1}\end{aligned}$$

Utilizando os subcomponentes apresentados em (1), (2) e (3), podemos construir X , y e ϵ da seguinte maneira:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}_{3r \times 2} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_{3r \times 1} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}_{3r \times 1}$$

E por fim o vetor de parâmetros β é fica com a seguinte forma:

$$\beta = \begin{bmatrix} \alpha \\ \Delta \end{bmatrix}_{2 \times 1}$$

1.2 Estimativa de parâmetros

Como a estimativa de modelos lineares por MQO tem solução analítica $\hat{\beta} = (X^T X)^{-1} X^T y$, A estimativa de MQO para o vetor coluna $\beta = [\alpha, \Delta]^T$ é desenvolvida nos a partir dos componentes contruídos anteriormente:

$$X^T X = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & \dots & 1 & 1 & \dots & -1 & -1 \end{bmatrix}_{2 \times 3r} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & -1 \\ 1 & -1 \end{bmatrix}_{3r \times 2} = \begin{bmatrix} 3r & 0 \\ 0 & 2r \end{bmatrix}_{2 \times 2}$$

Como $\det(X^T X) = 6r^2$ podemos calcular $(X^T X)^{-1}$

$$(X^T X)^{-1} = \frac{1}{6r^2} \begin{bmatrix} 2r & 0 \\ 0 & 3r \end{bmatrix}_{2 \times 2} = \begin{bmatrix} \frac{1}{3r} & 0 \\ 0 & \frac{1}{2r} \end{bmatrix}_{2 \times 2}$$

O próximo fator da estimativa é dado por:

$$X^T y = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & \dots & 1 & 1 & \dots & -1 & -1 \end{bmatrix}_{2 \times 3r} \begin{bmatrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ y_{3,r} \end{bmatrix}_{3r \times 1} = \begin{bmatrix} \sum_{i=1}^3 \sum_{j=1}^r y_{i,j} \\ \sum_{j=1}^r y_{2,j} - \sum_{j=1}^r y_{3,j} \end{bmatrix}_{2 \times 1}$$

Por fim, chegamos nas estimativas requisitadas:

$$\begin{aligned} \hat{\beta} &= \begin{bmatrix} \hat{\alpha} \\ \hat{\Delta} \end{bmatrix}_{2 \times 1} \\ &= (X^T X)^{-1} X^T y \\ &= \begin{bmatrix} \frac{1}{3r} & 0 \\ 0 & \frac{1}{2r} \end{bmatrix}_{2 \times 2} \begin{bmatrix} \sum_{i=1}^3 \sum_{j=1}^r y_{i,j} \\ \sum_{j=1}^r y_{2,j} - \sum_{j=1}^r y_{3,j} \end{bmatrix}_{2 \times 1} \\ &= \begin{bmatrix} \frac{1}{3r} \sum_{i=1}^3 \sum_{j=1}^r y_{i,j} \\ \frac{1}{2r} (\sum_{j=1}^r y_{2,j} - \sum_{j=1}^r y_{3,j}) \end{bmatrix}_{2 \times 1} \\ &= \begin{bmatrix} \bar{y} \\ \frac{\bar{y}_2 - \bar{y}_3}{2} \end{bmatrix}_{2 \times 1} \end{aligned}$$

Onde \bar{y} é a média amostral de todos os $y_{i,j}$ e \bar{y}_i é a média amostral das r observações do grupo $i = 1, 2, 3$.

1.3 Variância e covariância das estimativas

As variâncias e covariância das estimativas $\hat{\alpha}$ e $\hat{\Delta}$ são obtidas por meio do seguinte resultado:

$$Var(\hat{\beta}) = \sigma^2(X^T X)^{-1}$$

Consequentemente, temos:

$$\begin{aligned} Var(\hat{\alpha}) &= \frac{\sigma^2}{3r} \\ Var(\hat{\Delta}) &= \frac{\sigma^2}{2r} \\ Cov(\hat{\alpha}, \hat{\Delta}) &= 0 \end{aligned}$$

1.4 Estatística F para $H_0 : \Delta = 0$ contra $H_1 : \Delta \neq 0$

Para mostrar que a estatística F para testar $H_0 : \Delta = 0$ contra $H_1 : \Delta \neq 0$ pode ser expressa como no enunciado, o utilizaremos o teste geral de hipóteses lineares. Neste procedimento, o test $H_0 : \Delta = 0$ contra $H_1 : \Delta \neq 0$ pode ser escrito como:

$$\begin{aligned} H_0 : R\beta &= 0 \\ H_1 : R\beta &\neq 0 \end{aligned}$$

Onde $\beta = [\alpha \ \Delta]^T$ e $R = [0 \ 1]$. O acréscimo à soma dos quadrados dos resíduos devido à H_0 é escrito como:

$$ASQ(R\beta = 0) = (R\hat{\beta})^T [R(X^T X)^{-1} R^T]^{-1} R\hat{\beta}$$

Calculando os componentes temos:

$$\begin{aligned} R\hat{\beta} &= [0 \ 1] \begin{bmatrix} \hat{\alpha} \\ \hat{\Delta} \end{bmatrix} = \hat{\Delta} \\ R(X^T X)^{-1} R^T &= [0 \ 1] \begin{bmatrix} \frac{1}{3r} & 0 \\ 0 & \frac{1}{2r} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= [0 \ \frac{1}{2r}] \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \frac{1}{2r} \end{aligned}$$

E, então, $ASQ(R\beta = 0) = \hat{\Delta}(\frac{1}{2r})^{-1} \hat{\Delta} = 2r(\hat{\Delta})^2$. A estatística F para o teste geral de hipóteses lineares é:

$$F = \frac{ASQ(R\beta = 0)/k}{SQRes/(n - p)}$$

Onde k é o número de linhas de R, p é o número de parâmetros do modelo não restrito pelas hipóteses testadas e n é o tamanho da amostra. No nosso exercício, $k = 1$, $p = 2$ (α e Δ) e $n = 3r$. Com as seguintes manipulações algébricas, conseguimos chegar à forma apresentada no enunciado.

$$\begin{aligned}
F &= \frac{2r(\hat{\Delta})^2/1}{[\sum_{i=1}^3 \sum_{j=1}^r (y_{i,j} - \hat{y}_{i,j})^2]/(3r-2)} \\
&= \frac{2r(3r-2) \frac{(\bar{y}_2 - \bar{y}_3)^2}{4}}{\sum_{i=1}^3 \sum_{j=1}^r (y_{i,j} - \hat{y}_{i,j})^2} \quad (\text{pois } \hat{\Delta} = \frac{\bar{y}_2 - \bar{y}_3}{2}) \\
&= \frac{r(3r-2)}{2} \cdot \frac{(\bar{y}_2 - \bar{y}_3)^2}{\sum_{i=1}^3 \sum_{j=1}^r (y_{i,j} - \hat{y}_{i,j})^2}
\end{aligned}$$

2 Exercício 2

2.1 Família exponencial

As distribuições pertencentes à família exponencial podem ter sua função de densidade de probabilidade, no caso contínuo, ou sua função de massa de probabilidade, no caso discreto, escritas na seguinte forma:

$$f_Y(y; \theta, \phi) = \exp\{\phi[y\theta - b(\theta)] + c(y; \phi)\}$$

Onde θ é o parâmetro de localidade, ϕ é o parâmetro de precisão. Se for possível expressar a função de densidade (ou de massa) de uma variável aleatória Y em termos de θ , ϕ , definindo $b(\theta)$ e $c(y, \phi)$ então Y pertence à família exponencial.

Mostramos com a densidade a seguir, com algumas manipulações algébricas, que ela de fato pertence à família exponencial:

$$\begin{aligned}
f(y; \theta, \phi) &= \frac{\phi a(y; \phi)}{\pi(1+y^2)^{1/2}} \exp[\phi(y\theta + (1-\theta^2)^{1/2})] \\
&= \exp[\phi(y\theta + (1-\theta^2)^{1/2})] \cdot \exp[\log(\frac{\phi a(y; \phi)}{\pi(1+y^2)^{1/2}})] \\
&= \exp[\phi(y\theta + (1-\theta^2)^{1/2}) + \log(\phi) + \log(a(y; \phi)) - \log(\pi) - \frac{1}{2}\log(1+y^2)]
\end{aligned}$$

Onde os componentes são:

- $\theta = \theta$
- $\phi = \phi$
- $b(\theta) = -(1-\theta^2)^{1/2}$
- $c(y; \phi) = \log(\phi) + \log(a(y; \phi)) - \log(\pi) - \frac{1}{2}\log(1+y^2)$.

2.2 Função de variância

Para encontrarmos a função de variância, é necessário manipular a função $b(\theta)$. A teoria da família exponencial nos indica que a função de variância $V(\mu)$ é aquela tal que $V(\mu) = \frac{d\mu}{d\theta}$ onde $\mu = b'(\theta) = \frac{db}{d\theta}$. Realizando essas manipulações temos:

$$\begin{aligned}
\mu &= \frac{d}{d\theta} b(\theta) \\
&= \frac{d}{d\theta} -(1-\theta^2)^{1/2} \\
&= -\frac{1}{2}(1-\theta^2)^{-1/2}(-1)2\theta \\
&= \frac{\theta}{(1-\theta^2)^{1/2}}
\end{aligned} \tag{1}$$

E para achar a função de variância em termos de θ , basta derivar $b'(\theta) = \mu$ com relação à θ mais uma vez:

$$\begin{aligned}
b''(\theta) &= \frac{d}{d\theta} \mu = \frac{d}{d\theta} \frac{\theta}{(1-\theta^2)^{1/2}} \\
&= [(1-\theta^2) - \theta \frac{1}{2}(1-\theta^2)^{-1/2}(-1)2\theta] \cdot \frac{1}{(1-\theta^2)^2} \\
&= \frac{(1-\theta^2)^{1/2}}{(1-\theta^2)^2} + \frac{\theta^2}{(1-\theta^2)^{1/2}(1-\theta^2)} \\
&= \frac{1}{(1-\theta^2)^{1/2}} + \frac{1}{(1-\theta^2)^{1/2}} \cdot \frac{\theta^2}{1-\theta^2} \\
&= \frac{1}{(1-\theta^2)^{1/2}} \left[1 + \frac{\theta^2}{1-\theta^2} \right] \tag{2}
\end{aligned}$$

Para expressar (2) em termos de μ , precisamos identificar duas relações. Partindo de (1) e elevando ao quadrado ambos os lados da igualdade, chegamos em:

$$\mu^2 = \frac{\theta^2}{(1-\theta^2)}$$

Em seguida, partindo da mesma equação (1), vamos isolar o θ , deixando-o em função de μ :

$$\begin{aligned}
\mu &= \frac{\theta}{(1-\theta^2)^{1/2}} \\
\left(\frac{\mu}{\theta}\right)^2 &= \left(\frac{1}{(1-\theta^2)^{1/2}}\right)^2 \\
\frac{\theta^2}{\mu^2} &= (1-\theta^2) \\
0 &= \mu^2 - \theta^2 \mu^2 - \theta^2 \\
0 &= \theta^2 + \theta^2 \mu^2 - \mu^2 \\
0 &= (\mu^2 + 1)(\theta^2) - \mu^2 \\
\theta^2 &= \frac{\mu^2}{\mu^2 + 1} \tag{3}
\end{aligned}$$

$$\theta = \frac{\mu}{(\mu^2 + 1)^{1/2}} \tag{4}$$

E por fim, usando a equação (3) e o primeiro fator na equação (2) a última manipulação necessária é para se chegar na função de variância em termos de μ :

$$(1-\theta^2)^{-1/2} = \left(1 - \frac{\mu^2}{\mu^2 + 1}\right)^{-1/2} = \left(\frac{1}{\mu^2 + 1}\right)^{-1/2} = (\mu^2 + 1)^{1/2} \tag{5}$$

Usando, (5) e (4) em (2), temos:

$$\begin{aligned}
V(\mu) &= \frac{1}{(1-\theta^2)^{1/2}} \left[1 + \frac{\theta^2}{1-\theta^2} \right] \\
&= (\mu^2 + 1)^{1/2} (\mu^2 + 1) \\
&= (\mu^2 + 1)^{3/2}
\end{aligned}$$

2.3 Função desvio

A função desvio para um MLG em sua forma escalonada é definida por:

$$D^*(y, \hat{\mu}) = \phi D(y, \hat{\mu}) = 2\{L(y, y) - L(y, \hat{\mu})\}$$

Onde $L(y, y)$ é o log da verossimilhança é dada pelo modelo saturado (estimado com n parâmetros, um para cada observação). É conveniente também trabalhar com o desvio em sua forma de somatória, que nos permite para cada observação calcular um componente de desvio não escalonado $d^2(y_i; \hat{\mu}_i)$:

$$D(y, \hat{\mu}) = 2 \sum_{i=1}^n d^2(y_i, \hat{\mu}_i)$$

Com sua versão escalonada $d^{*2}(y_i, \hat{\mu}_i) = \phi d^2(y_i, \hat{\mu}_i)$. Na família exponencial, $d^2(y_i, \hat{\mu}_i)$ pode ser escrito por meio dos parâmetros que definem a família:

$$d^2(y_i, \hat{\mu}_i) = y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i))$$

Onde $\theta_i = \theta(\mu_i)$. Outro resultado importante é, para a família exponencial, sob o modelo saturado $\tilde{\mu}_i = y_i$. Em outras palavras, na família exponencial, a estimativa do modelo saturado para a observação i é próprio valor da variável resposta para aquela observação. No caso da distribuição que estamos estudando, temos que $\theta(\mu) = \frac{\mu}{(\mu^2+1)^{1/2}}$. Assim, as funções necessárias para calcular o componente do desvio são, para o modelo saturado:

$$\tilde{\theta}_i = \frac{y_i}{(y_i^2 + 1)^{1/2}} \quad \text{e} \quad b(\tilde{\theta}_i) = -\frac{1}{(y_i^2 + 1)^{1/2}}$$

E para o modelo estimado:

$$\hat{\theta}_i = \frac{\hat{\mu}}{(\hat{\mu}^2 + 1)^{1/2}} \quad \text{e} \quad b(\hat{\theta}_i) = -\frac{1}{(\hat{\mu}^2 + 1)^{1/2}}$$

A seguir segue o cálculo do componente do desvio não escalonado:

$$\begin{aligned} d^2(y_i, \hat{\mu}_i) &= y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i)) \\ &= y_i \left(\frac{y_i}{(y_i^2 + 1)^{1/2}} - \frac{\hat{\mu}}{(\hat{\mu}^2 + 1)^{1/2}} \right) + \left(-\frac{1}{(\hat{\mu}^2 + 1)^{1/2}} + \frac{1}{(y_i^2 + 1)^{1/2}} \right) \\ &= \frac{y_i^2}{(y_i^2 + 1)^{1/2}} - \frac{y_i \hat{\mu}}{(\hat{\mu}^2 + 1)^{1/2}} - \frac{1}{(\hat{\mu}^2 + 1)^{1/2}} + \frac{1}{(y_i^2 + 1)^{1/2}} \\ &= \frac{y_i^2 + 1}{(y_i^2 + 1)^{1/2}} - \frac{y_i \hat{\mu} + 1}{(\hat{\mu}^2 + 1)^{1/2}} \\ &= (y_i^2 + 1)^{1/2} - \frac{y_i \hat{\mu} + 1}{(\hat{\mu}^2 + 1)^{1/2}} \end{aligned}$$

Assim, para se chegar ao componente escalonado $d^{*2}(y_i, \hat{\mu}_i)$ do desvio, basta multiplicar $d^2(y_i, \hat{\mu}_i)$ por ϕ :

$$d^{*2}(y_i, \hat{\mu}_i) = \phi \left((y_i^2 + 1)^{1/2} - \frac{y_i \hat{\mu} + 1}{(\hat{\mu}^2 + 1)^{1/2}} \right)$$

3 Exercício 3

3.1 Pertencimento à família exponencial

Se uma variável aleatória Y é modelada por uma distribuição de Pascal, ela possui a seguinte função de massa de probabilidades:

$$Y \sim \text{Pascal}(r, \pi) f(y; r, \pi) = \binom{y-1}{r-1} \pi^r (1-\pi)^{(y-r)}$$

Com $y = r, r+1, r+2, \dots$

É pedido então mostrar que a variável $Y^* = \frac{Y}{r}$ pertence à família exponencial. Em primeiro lugar, note que se $y = r, r+1, r+2, \dots$ então $\frac{y}{r} = 1, 1 + \frac{1}{r}, 1 + \frac{2}{r}, \dots$. A função de massa de probabilidade será modificada então para a seguinte forma:

$$f(y^*; r, \pi) = \binom{ry^* - 1}{r-1} \pi^r (1-\pi)^{(ry^* - r)}$$

A forma da função de massa de probabilidade de é a seguinte:

$$f_Y(y; \theta, \phi) = \exp\{\phi[y\theta - b(\theta)] + c(y; \phi)\}$$

No caso de Y^* , a função de massa de probabilidade é manipulada conforme os seguintes passos:

$$\begin{aligned} f(y^*; r, \pi) &= \binom{ry^* - 1}{r-1} \pi^r (1-\pi)^{(ry^* - r)} \\ &= \exp \left[\log \left(\binom{ry^* - 1}{r-1} \pi^r (1-\pi)^{(ry^* - r)} \right) \right] \\ &= \exp \left[\log \binom{ry^* - 1}{r-1} + r \log \pi + (ry^* - r) \log(1-\pi) \right] \\ &= \exp \left[r \log \pi + r(y^* - 1) \log(1-\pi) + \log \binom{ry^* - 1}{r-1} \right] \\ &= \exp \left[r(\log \pi + (y^* - 1) \log(1-\pi)) + \log \binom{ry^* - 1}{r-1} \right] \\ &= \exp \left[r \left[y^* \log(1-\pi) + \log \frac{\pi}{1-\pi} \right] + \log \binom{ry^* - 1}{r-1} \right] \end{aligned}$$

Logo, se tomarmos as definições as seguir:

- $\phi = r$
- $\theta = \log(1-\pi)$
- $b(\theta) = -\log\left(\frac{\pi}{1-\pi}\right) = \theta - \log(1-e^\theta)$
- $c(y^*, \phi) = \log \binom{\phi y^* - 1}{\phi - 1}$

é possível ver que Y^* pertence à família exponencial.

3.2 Função de variância

Para chegarmos na função de variância, é necessário antes encontrar sua média. Na família exponencial, vale o procedimento à seguir:

$$\begin{aligned}
 E(Y^*) &= \mu \\
 &= b'(\theta) \\
 &= \frac{d}{d\theta}(\theta - \log(1 - e^\theta)) \\
 &= 1 - \frac{1}{1 - e^\theta}(-1)e^\theta \\
 &= 1 + \frac{e^\theta}{1 - e^\theta} \\
 &= \frac{1}{1 - e^\theta} \\
 &= \frac{1}{\pi}
 \end{aligned}$$

Então a função de variância $V(\mu)$ é:

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

Em termos de μ , expressamos a função de variância como:

$$\begin{aligned}
 V(\mu) &= \frac{e^\theta}{(1 - e^\theta)^2} = \frac{e^\theta}{(1 - e^\theta)} \frac{1}{(1 - e^\theta)} \\
 &= \left(\frac{\mu - 1}{\mu} \right) \mu^2 \\
 &= (\mu - 1)\mu
 \end{aligned}$$

3.3 Componente da função desvio

A função desvio de um MLG é escrita desta forma:

$$D^*(y, \hat{\mu}) = \phi D(y, \hat{\mu}) = \sum_{i=1}^n d^{*2}(y_i, \hat{\mu}_i) = \sum_{i=1}^n \phi d^2(y_i, \hat{\mu}_i)$$

O componente do desvio requisitado pelo enunciado $d^{*2}(y_i; \hat{\pi}_i)$ tem duas particularidades. Primeiro é o componente escalonado, tal que $d^{*2}(y_i; \hat{\pi}_i) = \phi d^2(y_i; \hat{\pi}_i)$. Segundo, ele é pedido em função da estimativa do parâmetro π_i pelo modelo ajustado (isto é $\hat{\pi}_i$) no lugar do usual μ_i . Isso é importante pois, na família exponencial, o modelo saturado tem o seguinte resultado conhecido $\tilde{\mu}_i = y_i$. Na família exponencial, como no caso de Y^* , pode ser decomposto como:

$$d^2(y_i, \hat{\pi}_i) = 2 \left[y_i^* (\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i)) \right] \quad (1)$$

Tomemos as seguintes funções que partem das partes anteriores dessa resolução:

- $\theta(\pi) = \log(1 - \pi)$
- $\theta(\mu) = \log\left(\frac{\mu-1}{\mu}\right)$, uma vez que $\mu = \frac{1}{\pi}$
- $b(\pi) = -\log\left(\frac{\pi}{1-\pi}\right) = \log\left(\frac{1-\pi}{\pi}\right)$

E por último $b(\mu)$:

$$\begin{aligned}
 b(\mu) &= b(\theta(\mu)) \\
 &= \theta - \log(1 - e^\theta) \\
 &= \log\left(\frac{\mu-1}{\mu}\right) - \log\left(1 - \frac{\mu-1}{\mu}\right) \\
 &= \log\left(\frac{\mu-1}{\mu}\right) - \log\left(\frac{1}{\mu}\right) \\
 &= \log(\mu-1) - \log(\mu) - \log(1) + \log(\mu) \\
 &= \log(\mu-1)
 \end{aligned}$$

Agora, munidos dessas quatro funções, e dos fatos que $y_i^* = \frac{y_i}{r}$ e, sob modelo saturado, $\tilde{\mu}_i = y_i^*$, podemos calcular o componente da função desvio conforme descrito em (1).

$$\begin{aligned}
 \tilde{\theta}_i(\tilde{\mu}_i) &= \log\left(\frac{\tilde{\mu}_i-1}{\tilde{\mu}_i}\right) = \log\left(\frac{y_i^*-1}{y_i^*}\right) = \log\left(\frac{y_i/r-1}{y_i/r}\right) = \log\left(\frac{y_i-r}{y_i}\right) \\
 \hat{\theta}_i(\hat{\pi}_i) &= \log(1 - \hat{\pi}_i) \\
 b(\tilde{\theta}_i(\tilde{\mu}_i)) &= \log(\tilde{\mu}_i-1) = \log(y_i^*-1) = \log\left(\frac{y_i}{r}-1\right) = \log\left(\frac{y_i-r}{r}\right) \\
 b(\hat{\theta}_i(\hat{\pi}_i)) &= \log\left(\frac{1-\hat{\pi}_i}{\hat{\pi}_i}\right)
 \end{aligned}$$

E finalmente, temos o componente não escalonado do desvio:

$$\begin{aligned}
 d^2(y_i, \hat{\pi}_i) &= 2 \left[y_i^* (\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i)) \right] \\
 &= 2 \left[\frac{y_i}{r} \left(\log\left(\frac{y_i-r}{y_i}\right) - \log(1 - \hat{\pi}_i) \right) + \log\left(\frac{1-\hat{\pi}_i}{\hat{\pi}_i}\right) - \log\left(\frac{y_i-r}{r}\right) \right]
 \end{aligned}$$

Que para chegar na sua versão escalonada, basta multiplicar por $\phi = r$:

$$d^{*2}(y_i, \hat{\pi}_i) = \phi d^2(y_i, \hat{\pi}_i) = 2r \left[\frac{y_i}{r} \left(\log\left(\frac{y_i-r}{y_i}\right) - \log(1 - \hat{\pi}_i) \right) + \log\left(\frac{1-\hat{\pi}_i}{\hat{\pi}_i}\right) - \log\left(\frac{y_i-r}{r}\right) \right]$$

Esta expressão, todavia, não está definida caso $y_i = r$ (ou $y_i^* = 1$), uma quantidade com probabilidade não nula de ocorrência. Neste caso temos:

$$\begin{aligned}
 f(y_i = r; r, \pi_i) &= \pi_i^r = \left(\frac{1}{\mu_i}\right)^r \\
 \log(f(y_i = r; r, \pi_i)) &= r \log \pi_i = r \log 1 - r \log \mu
 \end{aligned}$$

Com o resultado da verossimilhança deste caso de fronteira da variável Y , podemos calcular o componente do desvio pela definição:

$$D^*(y, \hat{\mu}) = \phi D(y, \hat{\mu}) = 2\{L(y, y) - L(y, \hat{\mu})\}$$

Onde $L(y, y)$ é a verossimilhança do modelo saturado e $L(y, \hat{\mu})$ do modelo ajustado. Lembrando que se $y_i = r$, então $y_i^* = 1$ temos:

$$d^{*2}(y_i, \hat{\pi}_i) = 2(-r \log y_i^* - r \log \hat{\pi}_i) = -2r \log \hat{\pi}_i$$

Por fim, o componente de desvio escalonado fica dado por:

$$d^{*2}(y_i, \hat{\pi}_i) = \begin{cases} 2r \left[\frac{y_i}{r} \left(\log \left(\frac{y_i - r}{y_i} \right) - \log(1 - \hat{\pi}_i) \right) + \log \left(\frac{1 - \hat{\pi}_i}{\hat{\pi}_i} \right) - \log \left(\frac{y_i - r}{r} \right) \right] & \text{se } y_i > r \\ -2r \log \hat{\pi}_i & \text{se } y_i = r \end{cases}$$

para $i = 1, 2, \dots, n$

4 Exercício 4

4.1 Matriz X e variâncias assintóticas

A matriz X para o modelo proposto no enunciado é:

$$X = \begin{bmatrix} x_1 - \bar{x} & z_1 - \bar{z} \\ x_2 - \bar{x} & z_2 - \bar{z} \\ \vdots & \vdots \\ x_n - \bar{x} & z_n - \bar{z} \end{bmatrix}$$

As variâncias assintóticas são obtidas por meio da matriz de informação de Fisher esperada, uma vez que:

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} \sim N \left(\begin{bmatrix} \beta \\ \gamma \end{bmatrix}, K_{\theta\theta}^{-1} \right)$$

Onde $\theta = [\beta, \gamma]^T$. Num MLG, a matriz de informação de Fisher esperada é dada por:

$$K_{\theta\theta} = \phi X^T W X$$

No nosso exemplo, $\phi = 1$ pois se $Bernoulli(\mu_i) = Binomial(\mu_i, n_i = 1)$ e num modelo binomial, $\phi = n$. A matriz W é uma matriz diagonal de pesos $diag\{\omega_1, \dots, \omega_n\}$ e $\omega_i = \left(\frac{d\mu_i}{d\eta_i} \right)^2 \cdot \frac{1}{V_i}$ onde V_i é a função de variância $V_i = V(\mu_i)$. No nosso exercício:

$$\begin{aligned} \frac{d\mu_i}{d\eta_i} &= \\ \frac{d\eta_i}{d\mu_i}^{-1} &= \left[\frac{d}{d\mu_i} \arcsen(\sqrt{\mu_i}) \right]^{-1} \\ &= \left[\frac{1}{\sqrt{1 - \mu_i^{\frac{1}{2}} \cdot 2}} \cdot \frac{1}{2\mu_i^{1/2}} \right]^{-1} \\ &= \left[\frac{1}{2\sqrt{\mu_i(1 - \mu_i)}} \right]^{-1} \\ &= 2\sqrt{\mu_i(1 - \mu_i)} \end{aligned}$$

Assim, como para uma variável aleatória Bernoulli $V(\mu) = \mu(1 - \mu)$

$$\begin{aligned}\omega_i &= \left(\frac{d\mu_i}{d\eta_i}\right)^2 \cdot \frac{1}{V_i} \\ &= \frac{4\mu_i(1 - \mu_i)}{\mu_i(1 - \mu_i)} \\ &= 4\end{aligned}$$

E assim, a matriz $W = \text{diag}\{4, \dots, 4\} = 4 \cdot I_n$ onde I_n é a matriz identidade n -dimensional. Então:

$$K_{\theta\theta} = \phi X^T W X = 4X^T X$$

Calculando o último componente de $K_{\theta\theta}$:

$$\begin{aligned}X^T X &= \begin{bmatrix} x_1 - \bar{x} & \dots & x_n - \bar{x} \\ z_1 - \bar{z} & \dots & z_n - \bar{z} \end{bmatrix} \begin{bmatrix} x_1 - \bar{x} & z_1 - \bar{z} \\ x_2 - \bar{x} & z_2 - \bar{z} \\ \vdots & \vdots \\ x_n - \bar{x} & z_n - \bar{z} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^n (x_i - \bar{x})^2 & \sum_{i=1}^n (x_i - \bar{x})(z_i - \bar{z}) \\ \sum_{i=1}^n (x_i - \bar{x})(z_i - \bar{z}) & \sum_{i=1}^n (z_i - \bar{z})^2 \end{bmatrix} \\ &= \begin{bmatrix} S_{xx} & S_{xz} \\ S_{xz} & S_{zz} \end{bmatrix}\end{aligned}$$

Assim colocando todos os componentes calculados até então,

$$K_{\theta\theta} = 4 \begin{bmatrix} S_{xx} & S_{xz} \\ S_{xz} & S_{zz} \end{bmatrix}$$

Então, para obter as variâncias assintóticas, basta calcular $K_{\theta\theta}^{-1}$:

$$K_{\theta\theta}^{-1} = \frac{1}{4\det(X^T X)} \begin{bmatrix} S_{zz} & -S_{xz} \\ -S_{xz} & S_{xx} \end{bmatrix}$$

Calculando o determinante temos $\det(X^T X) = S_{xx}S_{zz} - S_{xz}^2$. Multiplicando o determinante por $1 = \frac{S_{xx}S_{zz}}{S_{xx}S_{zz}}$, $\det(X^T X) = S_{xx}S_{zz} - S_{xz}^2 = (1 - r_{xz}^2)S_{xx}S_{zz}$. Então podemos terminar o cálculo do inverso da informação de Fisher esperada:

$$K_{\theta\theta}^{-1} = \begin{bmatrix} \frac{1}{(1-r_{xz}^2)S_{xx}} & -\frac{S_{xz}}{(1-r_{xz}^2)S_{xx}S_{zz}} \\ -\frac{S_{xz}}{(1-r_{xz}^2)S_{xx}S_{zz}} & \frac{1}{(1-r_{xz}^2)S_{zz}} \end{bmatrix}$$

Por fim, as variâncias e covariância assintóticas de $\hat{\theta} = [\hat{\beta}, \hat{\gamma}]^T$ são:

$$\text{Var}(\hat{\beta}) = \frac{1}{(1 - r_{xz}^2)S_{xx}}$$

$$\text{Var}(\hat{\gamma}) = \frac{1}{(1 - r_{xz}^2)S_{zz}}$$

$$Cov(\hat{\beta}, \hat{\gamma}) = -\frac{S_{xz}}{(1 - r_{xz}^2)S_{xx}S_{zz}}$$

4.2 Comparação entre correlações lineares

O cálculo da correlação linear $\rho(\hat{\beta}, \hat{\gamma})$, e por consequência, sua relação com a correlação amostral $r_{xz} = S_{xz}(S_{xx}S_{zz})^{-1/2}$ é a seguinte:

$$\begin{aligned}\rho(\hat{\beta}, \hat{\gamma}) &= \frac{Cov(\hat{\beta}, \hat{\gamma})}{\sqrt{Var(\hat{\beta})Var(\hat{\gamma})}} \\ &= -\frac{S_{xz}}{(1 - r_{xz}^2)S_{xx}S_{zz}} \cdot (1 - r_{xz}^2)^{1/2} S_{zz}^{1/2} (1 - r_{xz}^2)^{1/2} S_{xx}^{1/2} \\ &= -\frac{S_{xz}}{S_{xx}^{1/2} S_{zz}^{1/2}} \\ &= -r_{xz}\end{aligned}$$

Alguns comentários sobre a igualdade $\rho(\hat{\beta}, \hat{\gamma}) = -r_{xz}$:

- Se a correlação linear entre x e z for nula, $r_{xz} = 0$, as estimativas $\hat{\beta}$ e $\hat{\gamma}$ são ortogonais.
- Quanto mais positiva é r_{xz} mais negativa será a correlação entre $\hat{\beta}$ e $\hat{\gamma}$.

5 Exercício 5

Quando uma variável aleatória é dita distribuída por uma distribuição binomial negativa ajustada para os zeros, ela possui a seguinte função de massa de probabilidade:

$$f_Z(z_i; \mu, \nu, \pi) = \begin{cases} \pi & \text{se } z_i = 0 \\ (1 - \pi) \frac{f_Y(z_i; \mu, \nu)}{1 - f_Y(0; \mu, \nu)} & \text{se } z_i > 0 \end{cases}$$

onde $z_i = 0, 1, 2, \dots$, $i = 1, 2, \dots, n$ e f_y é:

$$f_Y(y; \mu, \nu) = \frac{\Gamma(\nu + y)}{\Gamma(y + 1)\Gamma(\nu)} \cdot \left(\frac{\mu}{\mu + \nu}\right)^y \left(\frac{\nu}{\mu + \nu}\right)^\nu$$

A estatística do teste de razão de verossimilhança para testar $H_0 : \mu = 1$ contra $H_1 : \mu \neq 1$ é:

$$\xi_{RV} = 2[L(z_i; \hat{\mu}, \nu, \hat{\pi}) - L(z_i; 1, \nu, \hat{\pi})]$$

No caso de $Z \sim ZANBI(\mu, \nu, \pi)$ o logarítmo da sua função de massa de probabilidade é:

$$\log(f_Z(z_i; \mu, \nu, \pi)) = \begin{cases} \log(\pi) & \text{se } z_i = 0 \\ \log(1 - \pi) + \log(f_Y(z_i; \mu, \nu)) - \log(1 - f_Y(0; \mu, \nu)) & \text{se } z_i > 0 \end{cases}$$

Definindo k como o número de observações dentre as n que são zero, isto é os i que $z_i = 0$, ao somarmos as log verossimilhanças temos:

$$L(z_i; \mu, \nu, \pi) = \sum_{i=1}^n f_Z(z_i; \mu, \nu, \pi) =$$

$$k \cdot \log(\pi) + (n - k) \cdot \log(1 - \pi) - (n - k) \cdot \log\left(1 - \left(\frac{\nu}{\nu + \mu}\right)^\nu\right) + \\ (n - k) \log(\delta(\nu, y)) + \log(\mu) \sum_{i=1}^{n-k} z_i + (n - k) \nu \log(\nu) - (n - k) \nu \log(\nu + \mu)$$

Onde definimos $\delta(\nu, y) = \frac{\Gamma(\nu+y)}{\Gamma(y+1)\Gamma(\nu)}$, e $f_Y(0; \mu, \nu) = \left(\frac{\nu}{\nu+\mu}\right)^\nu$.

Com o objetivo de calcular ξ_{RV} realizamos a diferença $2\{L(z_i; \hat{\mu}, \nu, \hat{\pi}) - L(z_i; 1, \nu, \hat{\pi})\}$. Nesta conta, todos os termos que não dependem de μ são cancelados. Além disso, os termos restantes tem como um de seus fatores o tamanho da amostra não nula $(n - k)$ que é deixado em evidência. Outro aspecto importante é a definição da média amostra não nula $\bar{z}_{z>0} = \frac{1}{(n-k)} \sum_{i=1}^{n-k} z_i$ com $z_i \neq 0$. Então:

$$\begin{aligned} \xi_{RV} &= 2\{L(z_i; \hat{\mu}, \nu, \hat{\pi}) - L(z_i; 1, \nu, \hat{\pi})\} \\ &= 2(n - k) \left[\log\left(\frac{1 - \left(\frac{\nu}{1+\nu}\right)^\nu}{1 - \left(\frac{\nu}{\hat{\mu}+\nu}\right)^\nu}\right) + \bar{z}_{z>0} \cdot \log(\hat{\mu}) + \nu \cdot \log\left(\frac{1 + \nu}{\hat{\mu} + \nu}\right) \right] \end{aligned}$$

Que é a estatística do teste de razão de verossimilhança proposto para testar $H_0 : \mu = 1$ contra $H_1 : \mu \neq 1$.

6 Exercício 6

A seguir iremos analisar a base **fuel2001.txt**. Conforme o enunciado, são descritas as seguintes variáveis referentes aos 50 estados norte-americanos mais o Distrito de Columbia no ano de 2001:

- i) UF, unidade da federação
- ii) Drivers, número de motoristas licenciados
- iii) FuelC, total de gasolina vendida (em mil galões)
- iv) Income, renda per capita em 2000 (em mil USD)
- v) Miles, total de milhas em estradas federais
- vi) MPC, milhas per capita percorridas
- vii) Pop, população ≥ 16 anos
- viii) Tax, taxa da gasolina (em cents por galão)

Abaixo mostro as 5 primeiras linhas da base.

State	Drivers	FuelC	Income	Miles	MPC	Pop	Tax
AL	3559897	2382507	23471	94440	12737.00	3451586	18.0
AK	472211	235400	30064	13628	7639.16	457728	8.0
AZ	3550367	2428430	25578	55245	9411.55	3907526	18.0
AR	1961883	1358174	22257	98132	11268.40	2072622	21.7
CA	21623793	14691753	32275	168771	8923.89	25599275	18.0

Conforme sugerido, realizaremos as transformações $\text{Fuel} = 1000 \text{FuelC} / \text{Pop}$ e $\text{Dlic} = 1000 \text{Drivers} / \text{Pop}$ para possibilitar a comparação entre estados com diferentes populações. Em seguida criamos a variável $\text{IMiles} = \log(\text{Miles})$.

Com essas transformações, partimos para uma análise descritiva da base, antes de partirmos para a modelagem.

6.1 Análise univariada

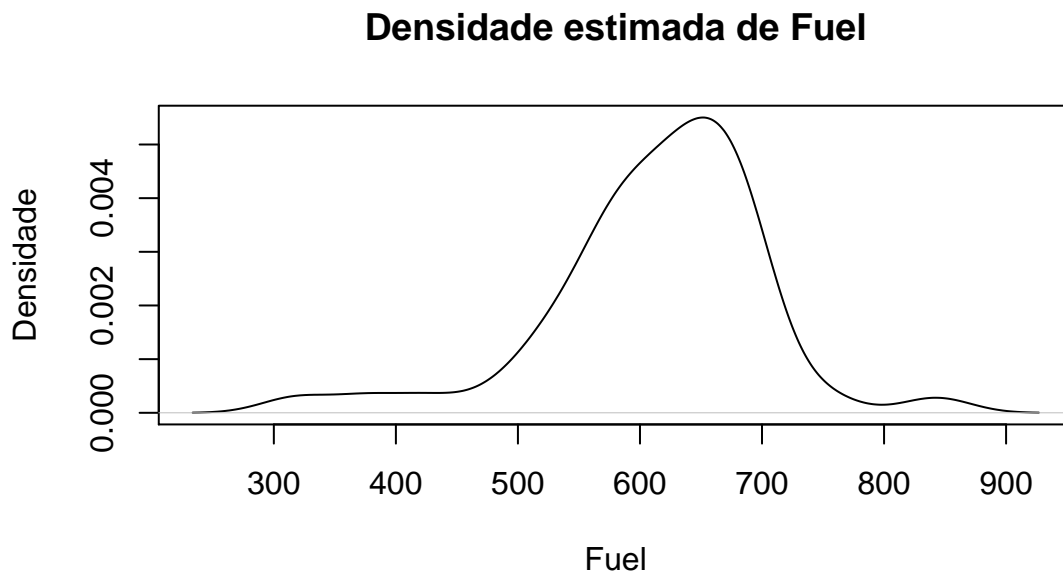
Primeiramente, a partir de algumas estatísticas descritivas de cada uma das variáveis estudadas que mostro abaixo, podemos postular algumas características das variáveis:

- a variável *fuel* mostra uma leve assimetria à esquerda uma vez que sua média é menor que a mediana
- a variável *income* mostra uma assimetria à direita uma vez que sua média é maior que a mediana

variável	média	desvio padrão	quartil 1	mediana	quartil 3
dlic	904	73	864	909	943
fuel	613	89	575	626	667
income	28404	4452	25323	27871	31208
lmiles	11	1	11	11	12
tax	20	5	18	20	23

Neste exercício, a variável *fuel* é a variável resposta. Por isso à inspecionaremos primeiro. o gráfico abaixo mostra a densidade estimada dela. Podemos ver que ela possui uma tendência gaussiana, apesar de possuir caudas ligeiramente pesadas.

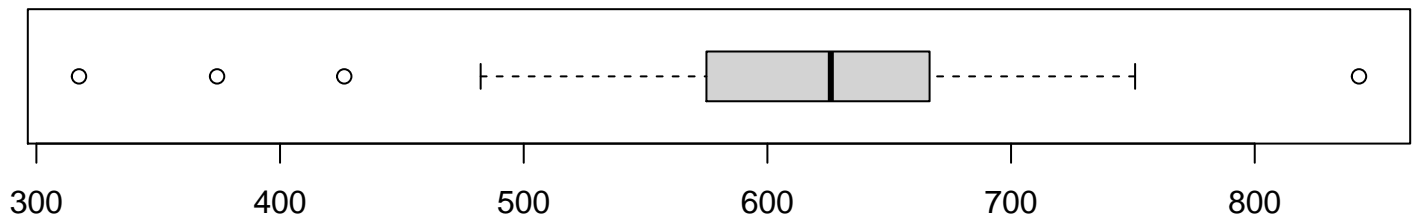
```
plot(density(mdata$fuel), xlab="Fuel", ylab="Densidade", main="Densidade estimada de Fuel")
```



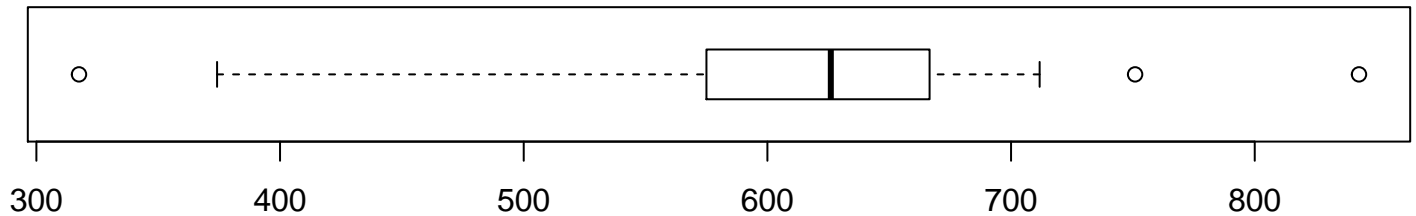
Abaixo mostramos os boxplots comum (criado através das estatísticas usuais como mediana e quartis) e robusto para a variável *fuel*. No boxplot comum estão 3 pontos que poderiam ser considerados extremos à esquerda, sendo mais baixo que o esperado para essa variável, e a extremo à direita. Com o boxplot robusto os pontos extremos à esquerda não são tão evidenciados, porém à direita um ponto novo aparece.

```
par(mfrow=c(2, 1), mar=c(2, 2, 2, 2))
boxplot(mdata$fuel, horizontal = T, main='Fuel - boxplot comum')
adjbox(mdata$fuel, horizontal = T, main='Fuel - boxplot robusto')
```

Fuel – boxplot comum

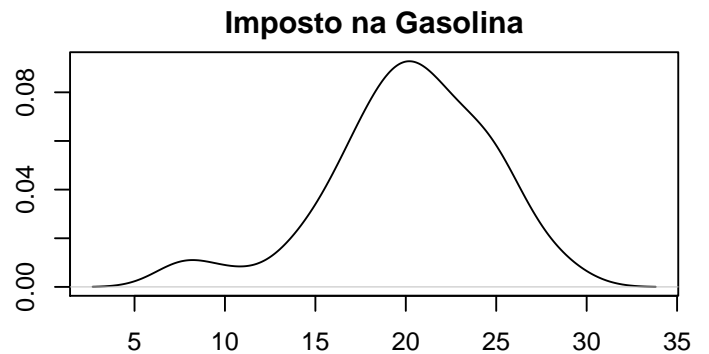
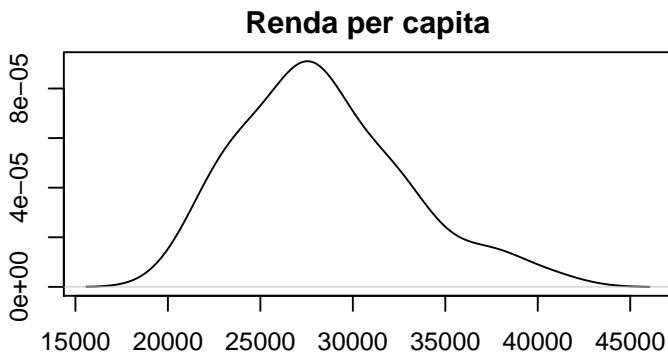
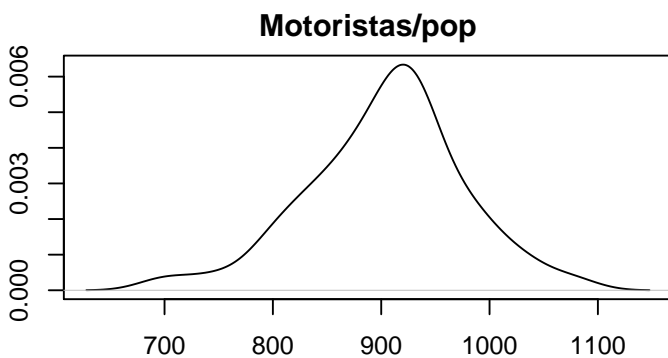


Fuel – boxplot robusto



Por completude, mostro as densidades empíricas das variáveis explicativas do estudo. Como sob a construção de modelos elas são consideradas fixas, suas distribuições não são importantes para o ajuste de modelos de regressão linear múltipla.

```
par(mfrow=c(2,2), mar=c(2, 2, 2, 2))
plot(density(mdata$dlic), xlab="motoristas/pop", ylab="Densidade", main='Motoristas/pop')
plot(density(mdata$lmliles), xlab="milhas (log)", ylab="Densidade", main='Milhas de estradas')
plot(density(mdata$income), xlab="Renda (1000 usd)", ylab="Densidade", main='Renda per capita')
plot(density(mdata$tax), xlab="imposto (centavos/galão) ", ylab="Densidade",
      main='Imposto na Gasolina')
```



6.2 Análise bivariada

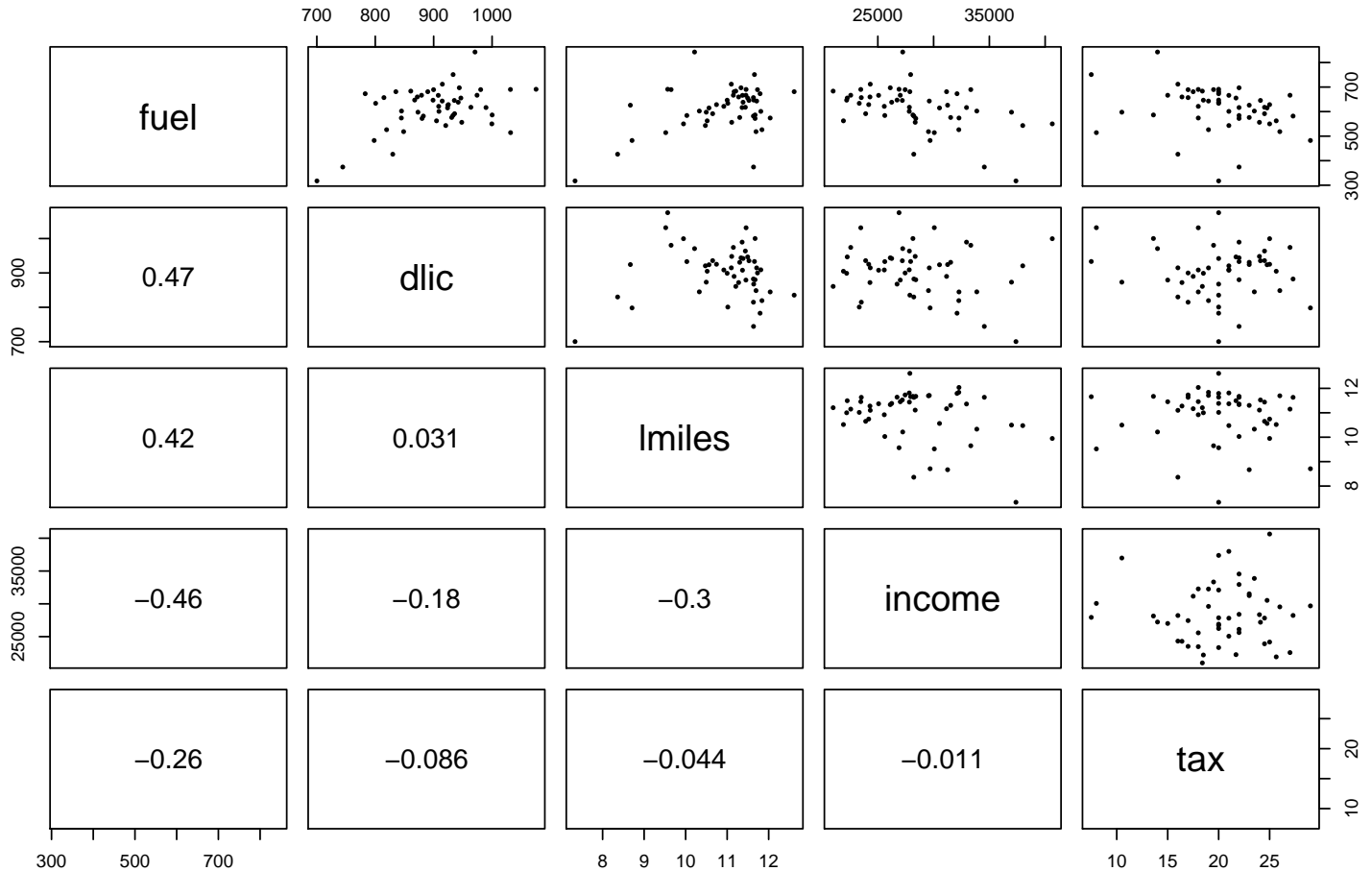
Como ponto de partida para análise bivariada, apresentamos uma matriz de painéis de gráficos de dispersão para as variáveis que estamos estudando junto com o índice de correlação linear entre as variáveis em questão.

```

panel.cor <- function(x, y, ...)
{
  par(usr = c(0, 1, 0, 1))
  txt <- as.character(format(cor(x, y), digits=2))
  text(0.5, 0.5, txt, cex = 1.57)
}

pairs(mdata, pch=16, cex=0.6, lower.panel = panel.cor)

```



A inspeção dessa matriz levanta alguns pontos:

- 3 variáveis explicativas (*dlic*, *lmiles*, *income*) apresentam uma correlação linear relativamente forte com a variável resposta *fuel*.
- em geral a correlação entre as variáveis explicativas não é grande, com exceção do par *lmiles* e *income*. É importante notar isso para que não haja problemas de multicolinearidade.

6.3 Ajuste de modelos

6.3.1 Seleção de variáveis

Devido à tendência gaussiana da variável resposta *fuel* iremos trabalhar com regressões lineares. A primeira etapa será aplicar o procedimento de seleção de variáveis stepwise avaliando o critério de Akaike do modelo resultante de cada iteração.

O método stepwise consiste em, partindo do modelo nulo (somente com o intercepto) adicionar variáveis a partir do critério de entrada avaliando todas as possíveis candidatas. Assim que houver ao menos duas variáveis

inseridas no modelo, avaliar a partir de um critério de saída a exclusão de alguma delas do modelo. Nesse estudo, o critério de entrada e saída é a minimização o critério de informação de Akaike.

A seguir está a tabela que resultou da aplicação do método setpwise usando o AIC para o modelo apenas com os efeitos principais.

```
modelo_nulo = lm(fuel ~ 1 , data = mdata)

aic <- stepAIC(modelo_nulo,
               scope=list(lower= ~ 1,
                           upper= ~ dlic + income + lmiles + tax),
               direction = 'both', trace = F)

interactions <- combn(
  names(aic$coefficients)[-1], m = 2,
  function(x) paste0(x[1], '*', x[2])
)
```

No caso das variáveis e amostra em questão o método stepwise inseriu todas as variáveis no modelo. Como vemos na tabela a seguir, a inclusão de cada um dos efeitos principais culminou em uma redução do AIC.

Seleção de variáveis via stepwise com critério de AIC

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
	NA	NA	50	395694.1	458.7851
+ dlic	1	86854.11	49	308840.0	448.1464
+ lmiles	1	65833.72	48	243006.2	437.9199
+ income	1	31042.28	47	211964.0	432.9496
+ tax	1	18263.94	46	193700.0	430.3543

Por enquanto não tentaremos incluir ajustes polinomiais dos efeitos principais, somente após uma análise de inclusão de interações e análise de pontos aberrantes e influentes, pois sua inclusão pode melhorar o modelo nesses aspectos.

6.3.2 Inclusão de interações

Iremos realizar o teste de interações entre cada uma das variáveis 2 a 2. O modelo com efeitos principais resultante do procedimento de seleção de variáveis tem a seguinte forma. Nos refereremos à esse modelo como “modelo cheio” daqui em diante.

$$\text{fuel} = \alpha + \beta_1(\text{dlic}) + \beta_2(\text{lmiles}) + \beta_3(\text{income}) + \beta_4(\text{tax}) + \epsilon$$

Iremos então realizar testes de hipóteses para a inclusão das seguintes interações *dliclmiles*, *dlicincome*, *dlictax*, *lmilesincome*, *lmilestax*, *incometax*. Conforme mostra a tabela abaixo. Note que as interações foram ordenadas de maneira decrescente com base no valor da estatística F.

```
modelo_cheio <- lm(fuel ~ ., data=mdata)

int1 <- update(modelo_cheio, ~ . + dlic*lmiles)
int2 <- update(modelo_cheio, ~ . + dlic*income)
int3 <- update(modelo_cheio, ~ . + dlic*tax)
int4 <- update(modelo_cheio, ~ . + lmiles*income)
int5 <- update(modelo_cheio, ~ . + lmiles*tax)
```

```

int6 <- update(modelo_cheio, ~ . + income*tax)

int_models <- list(
  'dlic*lmiles' = int1,
  'dlic*income' = int2,
  'dlic*tax' = int3,
  'lmiles*income' = int4,
  'lmiles*tax' = int5,
  'income*tax' = int6
)

do_f_test <- function(int_model) {
  teste <- anova(modelo_cheio, int_model)
  f <- teste[['F']][2]
  pvalor <- teste[['Pr(>F)']][2]
  c('F' = f, 'pvalor' = pvalor)
}

sapply(int_models, do_f_test, simplify = T) %>%
  t %>%
  as.data.frame() %>%
  rename('Valor F' = 'F', 'Valor P' = 'pvalor') %>%
  arrange(-`Valor F`) %>%
  rownames_to_column('interação testada') %>%
  janitor::adorn_rounding(digits = 2) %>%
  kbl(booktabs = T,
      format = 'latex',
      caption='Testes F de inclusão de interação no modelo cheio',
      align = 'c',
      centering = T
  ) %>%
  kable_styling(full_width = F,
                position = 'center', latex_options = "HOLD_position")

```

Testes F de inclusão de interação no modelo cheio

interação testada	Valor F	Valor P
lmiles*tax	5.71	0.02
dlic*tax	3.15	0.08
dlic*income	2.12	0.15
dlic*lmiles	1.49	0.23
income*tax	0.82	0.37
lmiles*income	0.15	0.70

Assim decidimos adicionar a interação `lmiles*tax` por ser a única que proveu um teste F com nível descritivo menor que 5%. Devido à pequena amostra, ou à natureza do estudo, poderíamos aceitar níveis de significância mais permissivos (10% por exemplo), mas nos reservaremos de tomar essa decisão no presente exercício. Por fim, o modelo final nesta etapa é possui a seguinte fórmula:

$$\text{fuel} = \alpha + \beta_1(\text{dlic}) + \beta_2(\text{lmiles}) + \beta_3(\text{income}) + \beta_4(\text{tax}) + \beta_5(\text{lmiles} \times \text{tax}) + \epsilon$$

6.4 Interpretação de parâmetros estimados

Na tabela abaixo, apresentamos as estimativas do modelo de efeitos principais e do modelo incluindo a interação.

Modelos construídos		
	efeitos principais	+ interação
(Intercept)	154.193 (194.906)	-877.018* (469.654)
dlic	0.472*** (0.129)	0.512*** (0.124)
lmiles	26.755*** (9.337)	119.603*** (39.848)
income	-0.006*** (0.002)	-0.006*** (0.002)
tax	-4.228** (2.030)	45.070** (20.714)
lmiles \times tax		-4.616** (1.931)
Num.Obs.	51	51
R2	0.510	0.566
R2 Adj.	0.468	0.517
Log.Lik.	-282.543	-279.495
F	11.992	11.720
AIC	430.354	426.258

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Para apresentar a interpretação dos parâmetros estimados, vamos focar apenas no modelo com interação + efeitos principais, uma vez que esse é o melhor modelo em termos de R2 Ajustado e AIC. As interpretações são:

- **intercepto:** neste modelo o intercepto não tem uma interpretação direta, uma vez que possuir as outras covariáveis nulas não faz muito sentido (o que seria um estado com 0 milhas de estradas? ou um estado que não tem renda per capita?).
- **dlic:** aqui vemos que para cada unidade a mais de 1000 motoristas per capita é esperado que o valor de vendas de combustível per capita aumente em 530 dólares per capita em média em um determinado estado.
- **income:** um aumento de mil dólares na renda per capita de um determinado estado é esperado uma diminuição de 6 dólares na venda de combustíveis de um determinado estado. Este efeito parece praticamente pequeno.
- **lmiles:** esta variável tem duas peculiaridades. Primeiro, passou por uma transformação em log. Sua interpretação, então, é baseada em mudanças percentuais no total de milhas de estradas em estradas federais no estado e seu efeito na variável respota. Segundo, ela está no modelo em interação com a variável *tax*, ou seja seu efeito depende do nível da variável que interage. No caso, quanto mais alto é nível de taxa da gasolina (a variável *tax*) menor é o efeito de um aumento percentual em *lmiles*. Assim, todo o resto constante, um aumento de 1% de malha rodoviária federal no estado leva a um aumento de $119 - 4.61 \times \text{tax}$ dólares no valor de vendas de combustível per capita. A taxa média de imposto sob a gasolina dentre os 51 estados é 20 centavos de dolar por galão. Nesse caso o efeito de 1% a mais de malha rodoviária federal na venda média de gasolina é $119 - 4.61 \times 20 = 26.8$ mil dólares per capita.

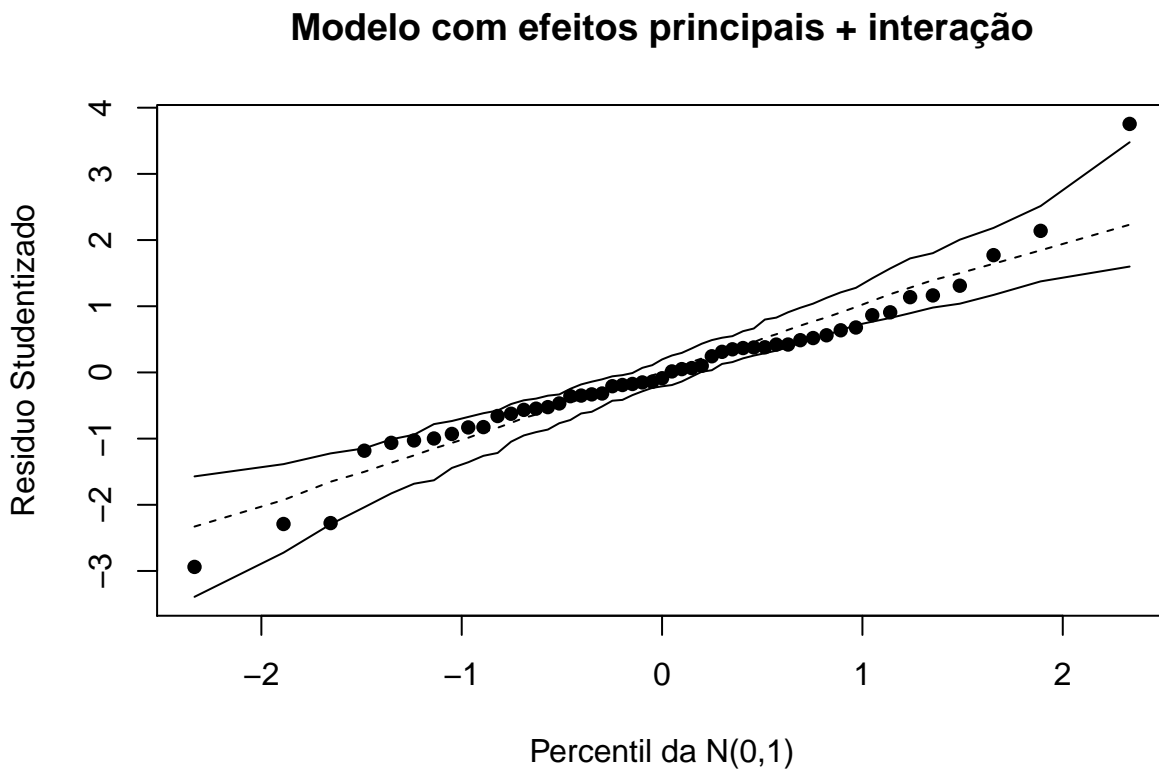
- **tax**: o efeito do aumento dessa variável é similar ao da variável *lmiles*, porém sem a interpretação em aumento percentual, uma vez que a variável *tax* não foi transformada em log. O seu efeito nas vendas de combustível médio depende da variável *lmiles* devido à interação incluída no modelo. Um aumento de um centavo na taxa de gasolina num estado que tem $e^{10.91} = 54885$ milhas de malha rodoviária associa-se com uma diminuição na média de galões de gasolina vendidos per capita em 5116 dólares.

6.5 Diagnósticos de modelo

6.5.1 Análise de resíduos

O gráfico abaixo contrasta os percentis do resíduo studentizado com os da distribuição normal com uma banda empírica de confiança. Nas extremidades podemos ver pontos que caem fora da banda de confiança, indicando que esse modelo não atende a suposição de normalidade dos erros. Isso pode melhorar se detectarmos pontos aberrantes e com alta influência posteriormente.

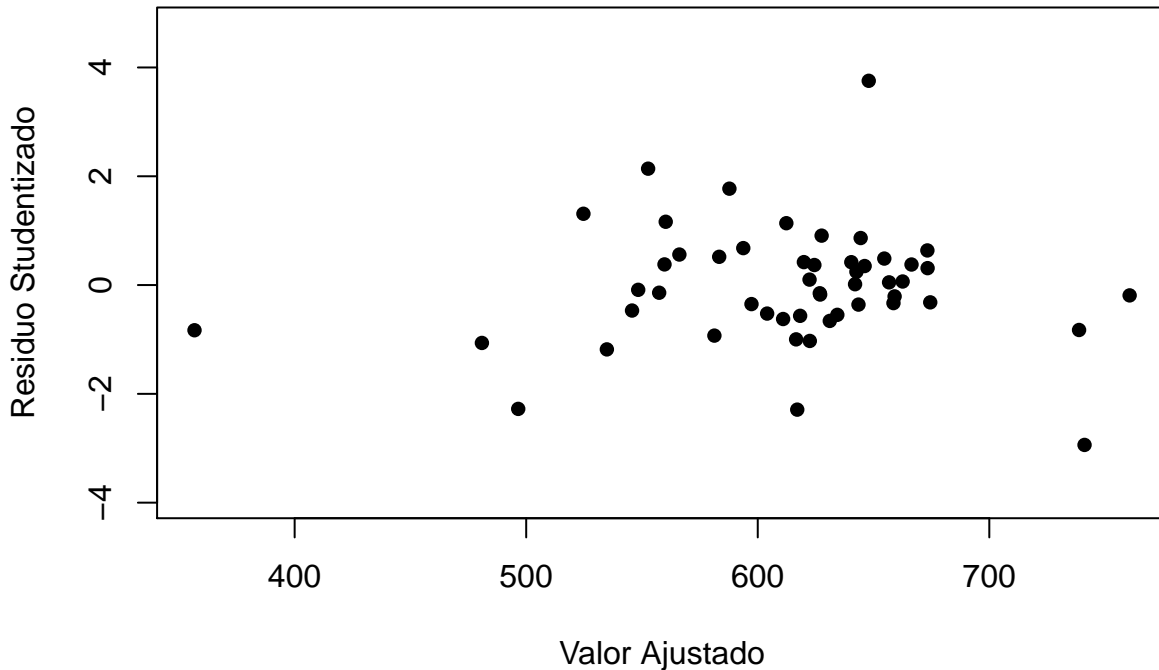
```
set.seed(410)
fit.model <- int5
title <- 'Modelo com efeitos principais + interação'
source('../envel_norm.R')
```



Na inspeção do valor ajustado contra o resíduo, não encontramos padrões de variância heterocedástica. Apenas um ponto parece ter o seu valor ajustado muito menor que os outros, mas nada que indique uma variância não constante.

```
title <- 'Valor ajustado e resíduos para o modelo com interação'
source('../diag_resid_norm.R')
```

Valor ajustado e resíduos para o modelo com interação

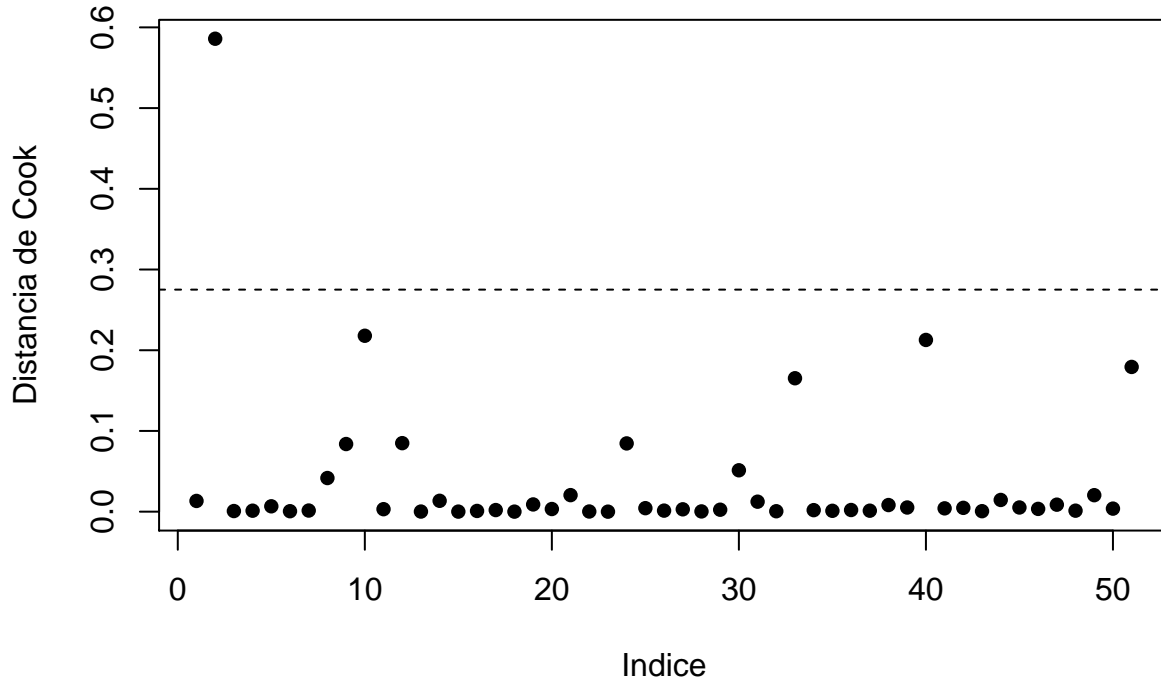


6.5.2 Análise de influência

Em seguida, faremos uma análise de observações influentes para identificar quais pontos têm um grande impacto nas estimativas dos parâmetros. Como só estamos aqui preocupados com os coeficientes da regressão, e não com a estimativa do erro sigma, vamos nos basear na distância de cook para avaliar a influência das observações. Para avaliação da distância de Cook, vamos compará-las com um fator k do erro padrão das distâncias calculadas. No nosso caso, como a amostra é pequena, iremos comparar as distâncias de Cook adotando $k = 2.5$. Assim se a distância $D_i \geq 2.5\hat{e}p(D_i)$. Esse fator de comparação está apontado em linha pontilhada no gráfico abaixo.

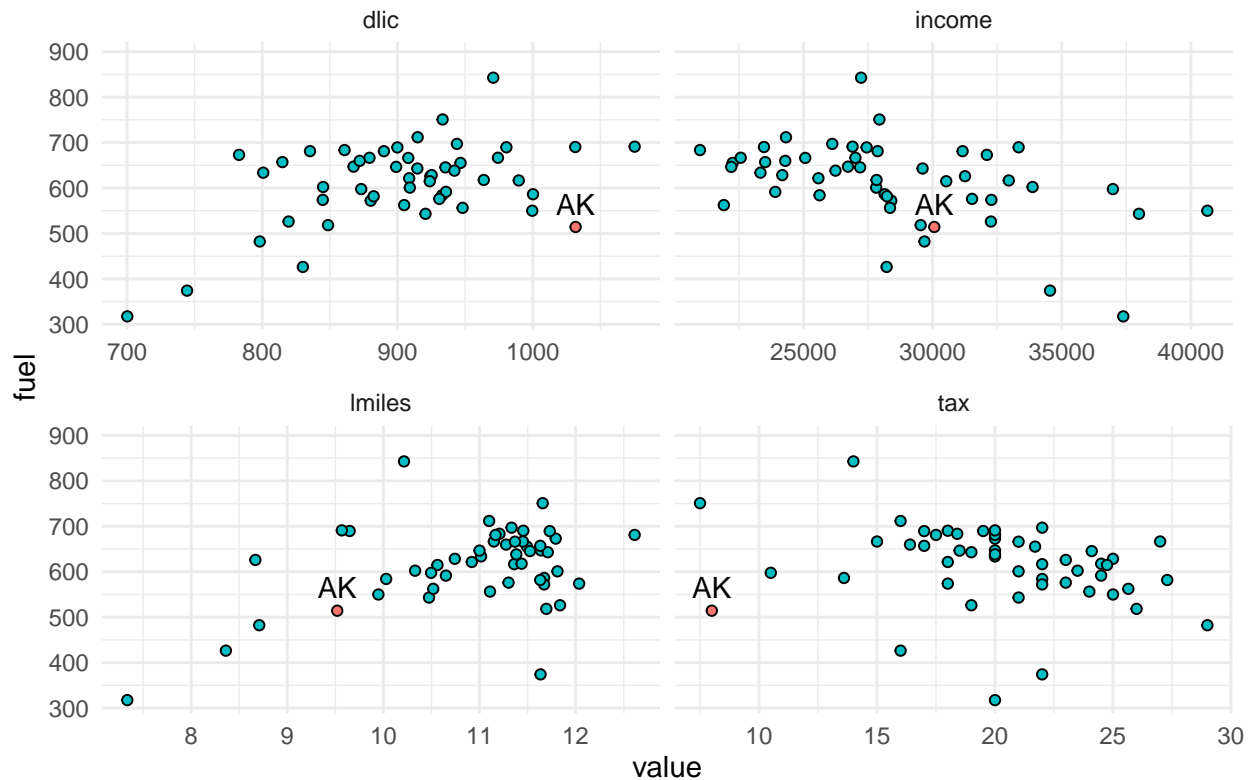
```
title <- 'Distâncias de Cook e linha de comparação para o\nmodelo com interação'
source('../diag_cook_norm.R')
```

Distâncias de Cook e linha de comparação para o modelo com interação



Existe então uma observação influente ao utilizarmos o critério proposto acima. Esta observação corresponde ao estado do Alasca (AK), que possui um nível de taxação da gasolina baixo, para a quantidade de gasolina vendida.

Posição do estado de Arkansas nos relacionamentos marginais no modelo



Para avaliar o impacto da exclusão da observação correspondente ao estado do Alasca nas estimativas e inferências resultantes da modelagem iremos apresentar 2 ajustes com formas funcionais diferentes, e cada um com as duas amostras.

```

full_data_int_md <- lm(formula = fuel ~ dlic + lmiles + income + tax + lmiles:tax,
  data = mdata)

full_data_ep_md <- lm(formula = fuel ~ dlic + lmiles + income + tax,
  data = mdata)

excl_ak_int_md <- lm(formula = fuel ~ dlic + lmiles + income + tax + lmiles:tax,
  data = mdata[di < cut, ])

excl_ak_ep_md <- lm(formula = fuel ~ dlic + lmiles + income + tax,
  data = mdata[di < cut, ])

```

Modelos construídos

	Efeitos principais + interação		Apenas Efeitos principais	
	Amostra completa (I)	S/ Alasca (II)	Amostra completa (III)	S/ Alasca (IV)
(Intercept)	-877.018* (469.654)	-345.088 (505.460)	154.193 (194.906)	183.425 (178.167)
dlic	0.512*** (0.124)	0.571*** (0.121)	0.472*** (0.129)	0.569*** (0.121)
lmiles	119.603*** (39.848)	68.810 (44.077)	26.755*** (9.337)	20.554** (8.743)
income	-0.006*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)
tax	45.070** (20.714)	18.739 (22.896)	-4.228** (2.030)	-6.735*** (2.013)
lmiles × tax	-4.616** (1.931)	-2.341 (2.096)		
Num.Obs.	51	50	51	50
R2	0.566	0.602	0.510	0.591
R2 Adj.	0.517	0.557	0.468	0.554
Log.Lik.	-279.495	-271.694	-282.543	-272.393
F	11.720	13.304	11.992	16.229
AIC	426.258	413.493	430.354	412.891

* p < 0.1, ** p < 0.05, *** p < 0.01

O modelo com interação e sem a observação do Alasca (II) tem as estimativas muito similares ao modelo I, porém os parâmetros de interação deixaram de ser significantes à 5%. Junto com isso, o tamanho da estimativa do coeficiente *lmiles* diminuiu bastante e ficou insignificante à 5%. Conjuntamente com esse apontamento, é possível ver que nos modelos sem a interação, a exclusão da observação do Alasca não troca os sinais das estimativas, não muda bruscamente suas magnitudes ou altera a significância dos parâmetros. Assim, concluímos que a observação do Alasca, por suas características relativamente atípicas, induz a aparição da interação entre *lmiles* e *tax*. Nesse sentido, o modelo preferível é o modelo apenas com os efeitos principais.

Vamos realizar uma curta análise de diagnóstico para checar se o modelo apenas com os efeitos principais é apropriado.

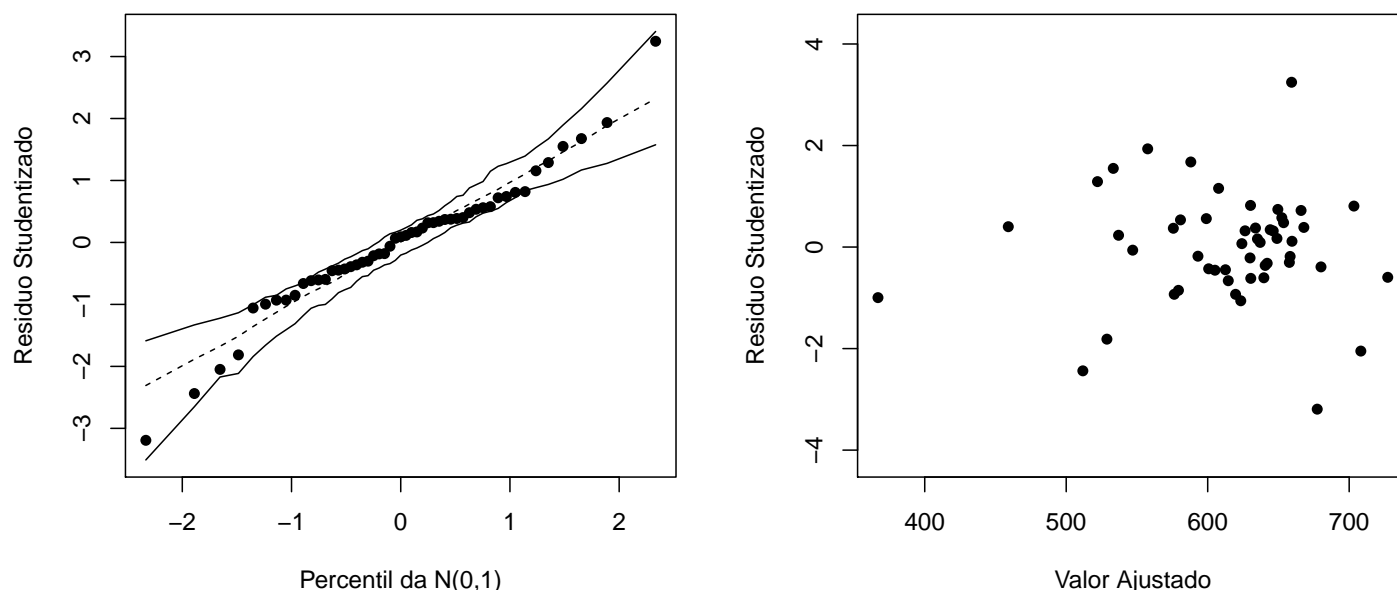
```

fit.model <- full_data_ep_md
title1 = ''
title2 = ''

```

```
set.seed(4155)
source('../diag_resid_envel.R')
title("Análise de resíduos para o modelo com efeitos principais", line = -1, outer = TRUE)
```

Análise de resíduos para o modelo com efeitos principais



Como podemos checar, o ajuste desse modelo é melhor, uma vez que todos os pontos estão dentro do envelope, dando indícios de que a normalidade dos resíduos é atendida. Além disso, inspecionando o gráfico de valor ajustado contra o resíduo studentizado, nenhum padrão na dispersão é observado indicando uma variância constante.

6.6 Ajuste por gamlss

Por fim, de maneira ilustrativa, ajustamos o modelo de efeitos principais com o gamlss, usando a família normal. Os gráficos de análise de resíduos apresentados a seguir sugerem um bom ajuste do modelo. A dispersão de resíduos quantílicos contra os valores ajustados está relativamente uniforme no intervalo, sugerindo uma variância constante. O gráfico de probabilidades dos quantis do resíduo versus os da normal mostram que, tirando uma observação extrema a direita, os pontos se alinham à reta $y = x$, indicando uma possível normalidade dos resíduos.

```
ajuste = gamlss(fuel ~ dlic + lmiles + income + tax, family=NO(), data = mdata)
```

```
## GAMLSS-RS iteration 1: Global Deviance = 565.086
```

```
## GAMLSS-RS iteration 2: Global Deviance = 565.086
```

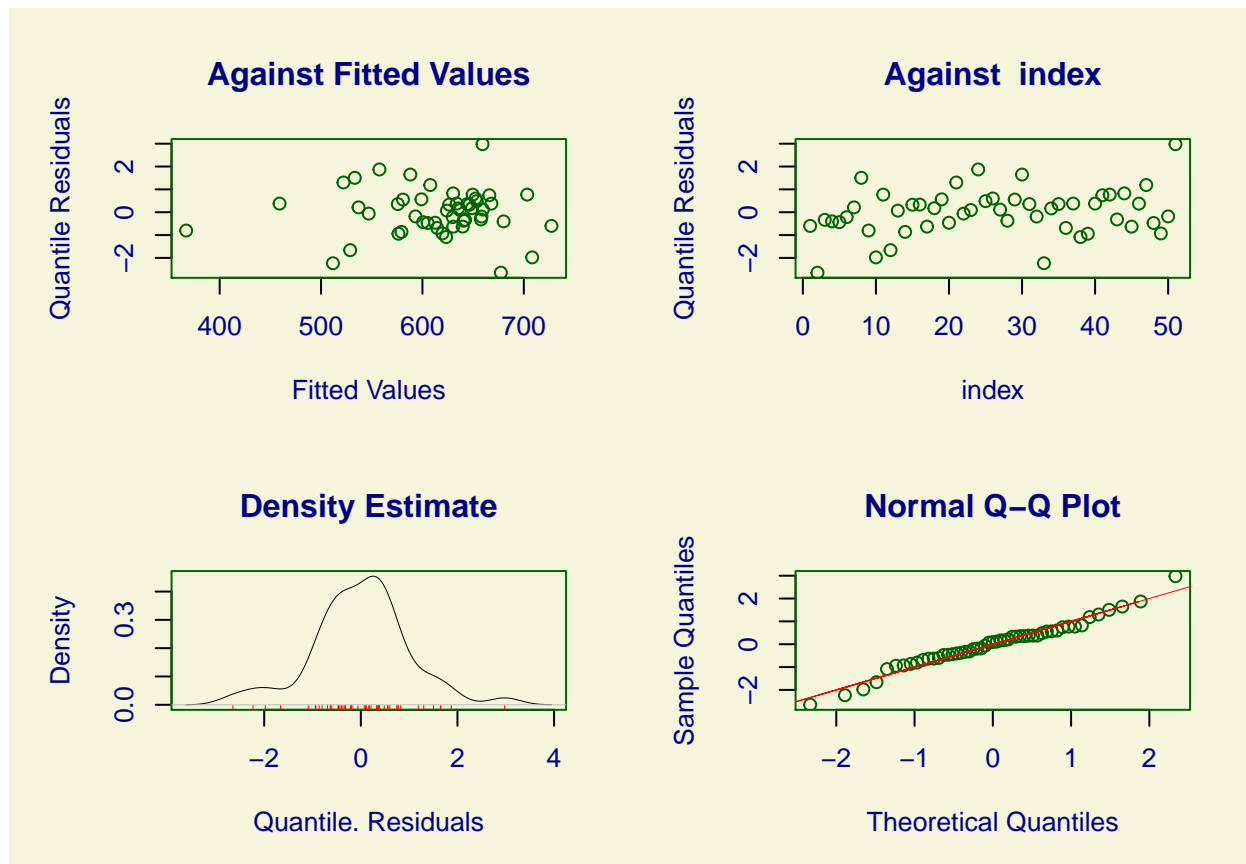
```
summary(ajuste)
```

```
## *****
## Family:  c("NO", "Normal")
##
## Call:   gamlss(formula = fuel ~ dlic + lmiles + income + tax,
##               family = NO(), data = mdata)
##
## Fitting method: RS()
##
```



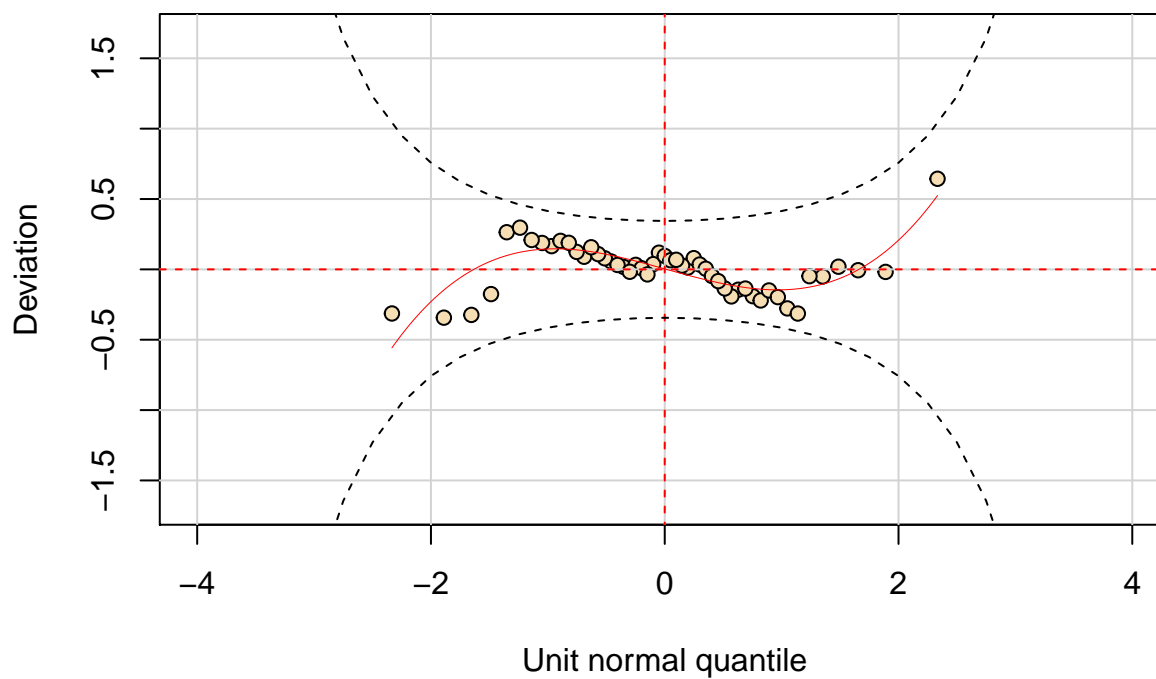
```
## -----
## Mu link function:  identity
## Mu Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 154.192845 185.109822   0.833 0.409255
## dlic        0.471871   0.122053   3.866 0.000353 ***
## lmiles      26.755176   8.867917   3.017 0.004189 **
## income      -0.006135   0.002083  -2.945 0.005097 **
## tax         -4.227983   1.928043  -2.193 0.033526 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## Sigma link function:  log
## Sigma Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.12112   0.09901  41.62  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## No. of observations in the fit:  51
## Degrees of Freedom for the fit:  6
##      Residual Deg. of Freedom:  45
##                      at cycle:  2
##
## Global Deviance:      565.086
##           AIC:        577.086
##           SBC:        588.6769
## *****
```

```
plot(ajuste, summaries=F)
```



A visualização do *worm plot* também indica um bom ajuste, uma vez que os resíduos não sobrepõem as bandas de confiança geradas.

`wp(ajuste)`



7 Exercício 7

A base de dados *heart.txt* contém 100 observações de pacientes que foram diagnosticados com doença arterial coronariana. As variáveis dessa base são:

- HD: ausência de doença (0), evidência de doença (1)
- Age: idade do paciente
- FE: faixa etária do paciente

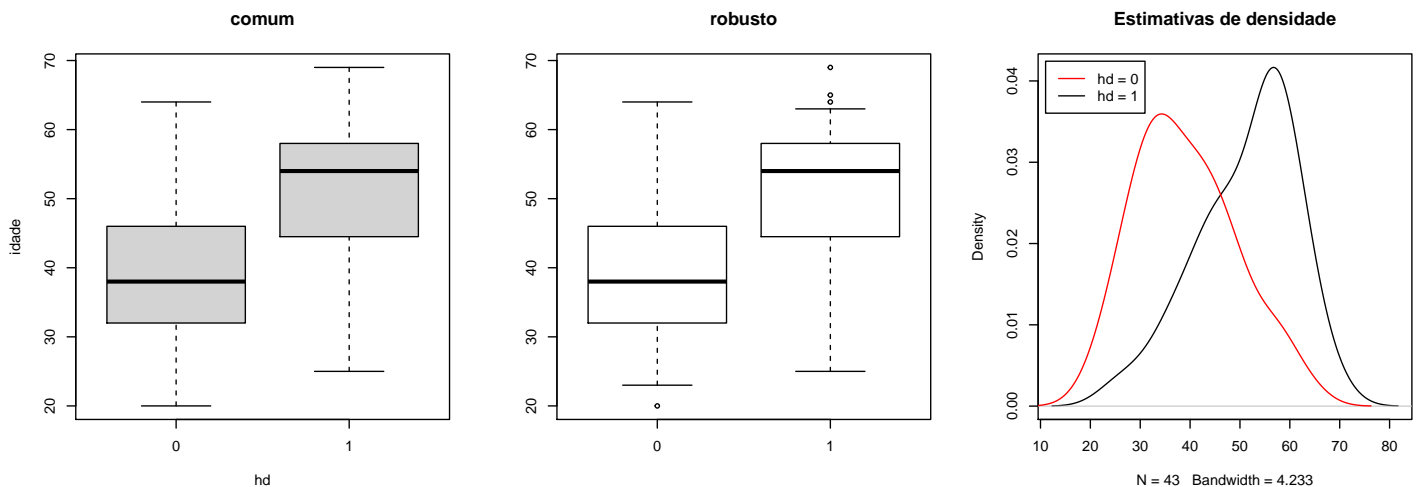
Abaixo mostro as 5 primeiras linhas dessa base de dados.

idade	fe	hd
20	1	0
23	1	0
24	1	0
25	1	0
25	1	1

7.1 Análise descritiva

Primeiro mostraremos dois conjuntos de boxplots, o comum e o robusto para os grupos da variável resposta *hd* junto com as estimativas de densidade de da variável idade para cada um desses grupos.

```
par(mfrow = c(1, 3))
boxplot(idade ~ hd, data = heart, main='comum')
adjbox(idade ~ hd, data = heart, main='robusto')
plot(density(heart[heart$hd == 1, 'idade']), col = 'black', main='Estimativas de densidade')
lines(density(heart[heart$hd == 0, 'idade']), col = 'red')
legend('topleft', inset = 0.02, legend=c("hd = 0", "hd = 1"), col=c("red", "black"), lty=c(1, 1))
```



Aguns pontos de discussão surgem da inspeção desses boxpots:

- os pacientes que possuem evidências de doença ($hd = 1$) são mais velhos em geral, possuindo um intervalo interquartil acima pacientes que têm ausência de doença.
- utilizando-se do boxplot comum, aparentemente não há pacientes com idades extremas. Porém, tulizando o boxplot robusto, 3 pacientes que possuem evidências da doença atipicamente mais velhos, quando um paciente dos que não possuem doença é atipicamente mais novo.

- parece existir uma certa tendência de assimetria à direita nas idades do grupo $hd = 0$ e assimetria à esquerda para o $hd = 1$.

7.2 Tabelas de contingência

A tabela de contingência a seguir mostra as proporções de eventos $hd = 1$ e $hd = 0$ para cada faixa etária presente nos dados.

Tabela de contingência para faixa etária e diagnóstico de doença arterial coronariana

faixa etária	hd = 0	hd = 1	total
1	9 (90%)	1 (10%)	10
2	13 (86%)	2 (14%)	15
3	9 (76%)	3 (24%)	12
4	10 (66%)	5 (34%)	15
5	7 (54%)	6 (46%)	13
6	3 (38%)	5 (62%)	8
7	4 (24%)	13 (76%)	17
8	2 (20%)	8 (80%)	10

É possível observar a tendência crescente de da proporção de evidência de doença coronariana. O gráfico abaixo confirma essa afirmação de maneira mais visual.



7.3 Ajuste de modelo logístico

Ajustaremos um modelo logístico com a ligação canônica $\log\left(\frac{\mu}{1-\mu}\right)$.

```
md <- glm(hd ~ idade, data = heart, family = binomial(link = 'logit'))
```

	Modelo proposto
(Intercept)	0.005*** (1.134)
idade	1.117*** (0.024)
Num.Obs.	100
AIC	111.4
BIC	116.6
Log.Lik.	-53.677
* p < 0.1, ** p < 0.05, *** p < 0.01	
Nota: Coeficientes exponenciados	

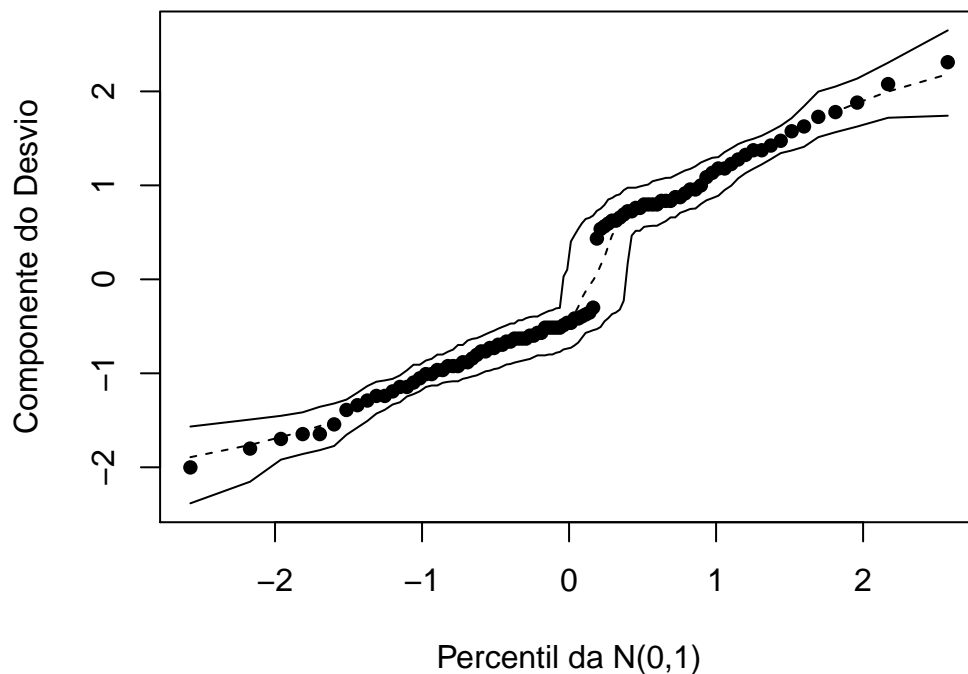
O modelo tem algumas características em concordância com a análise descritiva já apresentada. O intercepto não têm uma interpretação satisfatória. Poderia significar a estimativa do modelo quando a idade = 0, mas não temos nenhuma observação com esse valor de idade. Tampouco me parece razoável pensar em doença cardíacas como as estudadas num recém nascido. Poderíamos deslocar a variável idade com relação ao mínimo da amostra, e então o intercepto ficaria com uma interpretação mais apropriada. Outro ponto importante é o sinal da variável *idade* que está no mesmo sentido que o encontrado durante a análise descritiva.

7.4 Diagnósticos de modelo

Observando o gráfico de probabilidades do resíduo contra a distribuição normal padrão, com uma ênfase para as bandas empíricas calculadas, vemos que não há um desvio do resíduos para fora dessa banda, indicando um ajuste satisfatório.

```
fit.model <- md
title <- 'Gráfico de envelope para os resíduos do ajuste'
source('envel_gamma.R')
```

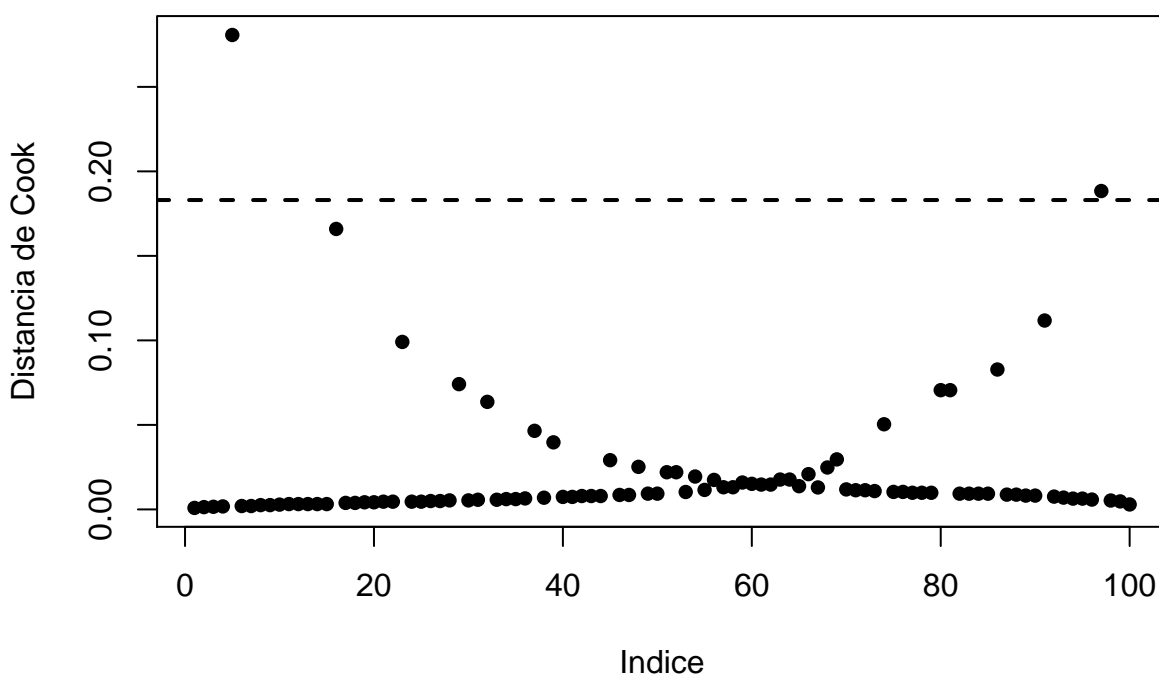
Gráfico de envelope para os resíduos do ajuste



Para avaliação da distância de Cook, vamos compará-las com um fator k do erro padrão das distâncias calculadas. No nosso caso, escolhemos $k = 4$, de tal forma que se $D_i \geq 4\hat{e}p(D_i)$, a observação é considerada influente. Esse fator de comparação está apontado em linha pontilhada no gráfico abaixo.

```
title = 'Distância de Cook para o modelo ajustado'
source('diag_cook_bino.R')
```

Distância de Cook para o modelo ajustado



Como podemos ver, duas observações pode ser consideradas influentes. Iremos-re ajustar o modelo retirando essas observações e ver se a nossa inferência com relação aos parâmetros muda de maneira brusca (traca de

sinais ou mudança forte na magnitude da estimativa. As duas observações, são os indivíduos 5, e 97, e abaixo podemos checar seus valores da variável resposta e explicativa:

```
heart[di > mean(di) + 4*sd(di),]
```

	idade	fe	hd
5	25	1	1
97	64	8	0

O resultado dos modelos reajustados é promissor, pois não altera as estimativas, e nem os níveis de significância dos testes de Wald. Todos os parâmetros continuam sendo significativos à 1%.

```
md_all_samples <- glm(hd ~ idade, data = heart, family = binomial(link = 'logit'))
md_drop <- glm(hd ~ idade, data = heart[di < mean(di) + 4*sd(di),],
               family = binomial(link = 'logit'))
```

	todas as observações	excl. [5, 97]
(Intercept)	-5.309*** (1.134)	-6.397*** (1.295)
idade	0.111*** (0.024)	0.134*** (0.027)
Num.Obs.	100	98
AIC	111.4	101.5
BIC	116.6	106.6
Log.Lik.	-53.677	-48.726

* p < 0.1, ** p < 0.05, *** p < 0.01

Nota: Coeficientes exponenciados

7.5 Interpretação de coeficientes bandas de confiança e estimativas intervalares

Como já mencionado anteriormente, a interpretação do intercepto para o modelo ajustado é problemática, uma vez que não um paciente com zero anos de idade não é razoável. Todavia uma interpretação para o coeficiente da variável idade é possível:

- interpretação da coeficiente da idade: cada 1 ano a mais de idade, é esperado que sua chance de ter doença cardíaca em 11%, como podemos checar na comparação abaixo, conforme a tabelas de coeficientes do modelo que contém os parâmetros exponenciados abaixo.

As estimativas intervalares de 95% para ambos os parâmetros estão na tabela a seguir:

```
se <- sqrt(diag(vcov(md)))

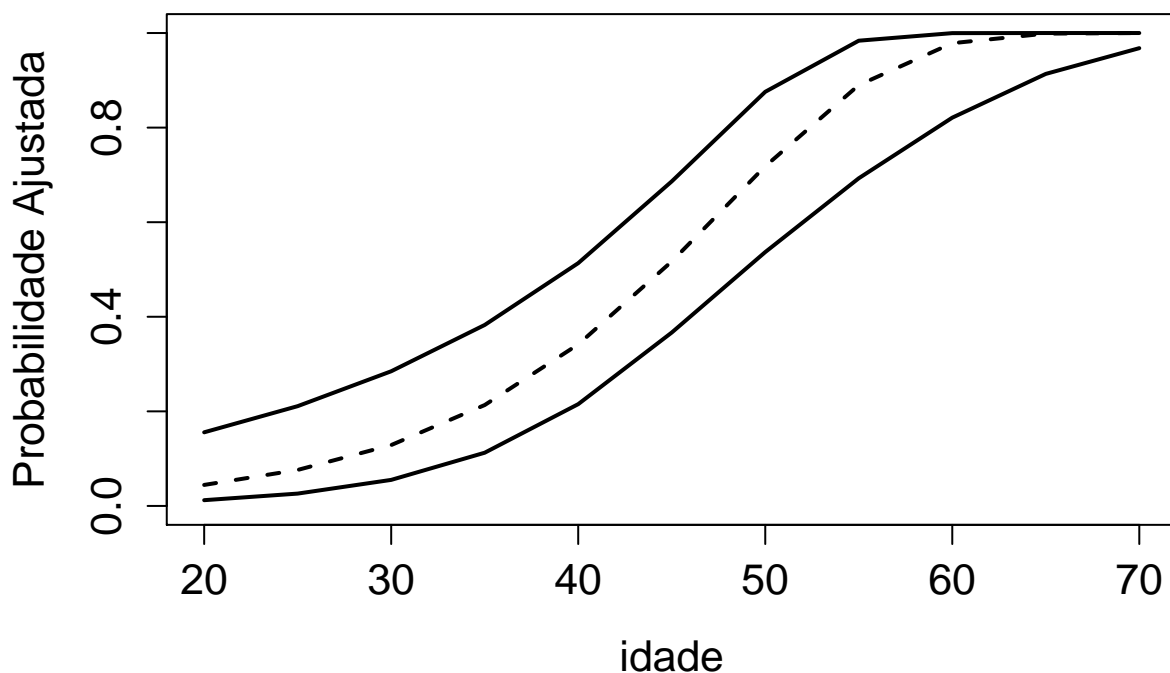
data.frame(coef = coef(md),
            ep = se,
            inferior = coef(md) - 1.96*se,
            superior = coef(md) + 1.96*se,
            exp_coef = exp(coef(md)),
            exp_inferior = exp(coef(md) - 1.96*se),
            exp_superior = exp(coef(md) + 1.96*se)
            ) %>%
  kbl(booktabs = T, format = 'latex') %>%
  kable_styling(full_width = F)
```

	coef	ep	inferior	superior	exp_coef	exp_inferior	exp_superior
(Intercept)	-5.3094534	1.1336536	-7.5314145	-3.0874922	0.0049446	0.000536	0.0456162
idade	0.1109211	0.0240598	0.0637639	0.1580784	1.1173068	1.065841	1.1712580

E por fim, apresentamos as bandas de confiança para a probabilidade ajusta pelo modelo de ser diagnosticado com a doença cardíaca ao longo da variável explicativa *idade*.

```
X=model.matrix(md)
w = md$weight
W = diag(w)
A = t(X)%*%W%*%X
A = solve(A)
z=seq(20, 70, 5)
vz = cbind(1,z)
erro = vz%*%A%*%t(vz)
erro = diag(erro)
erro = 1.96*sqrt(erro)
etaz = predict(md, data.frame(idade = z))
eta1 = etaz + erro
eta2 = etaz - erro
freqz= 1 - exp(-exp(etaz))
freq1= 1 - exp(-exp(eta1))
freq2= 1 - exp(-exp(eta2))
plot(z, freqz, type="n", ylim=c(0,1), xlab="idade",
      ylab="Probabilidade Ajustada", main='Banda de confiança para idade',
      cex=2, cex.axis=1.3, cex.lab=1.3)
lines(z, freqz, lty=2,lwd=2)
lines(z, freq1,lwd=2)
lines(z, freq2,lwd=2)
```

Banda de confiança para idade



7.6 Curva ROC e capacidade discriminativa

A curva ROC, apresentada abaixo, mostra a troca entre a taxa de verdadeiros positivos e a taxa de falsos positivos do modelo ajustado. A área sob essa curva nos dá uma boa noção da discriminação que o modelo consegue fazer. Além disso, pode nos ajudar a encontrar um ponto de corte da probabilidade ajustada pelo modelo que poderia ser utilizada para atribuir uma classificação à observação.

Na curva mostramos o ponto que minimiza a distância entre o caso de perfeita discriminação, ou seja, com 100% dos casos classificados como positivos de maneira correta e nenhum caso errado entre aqueles que foram classificados como positivos.

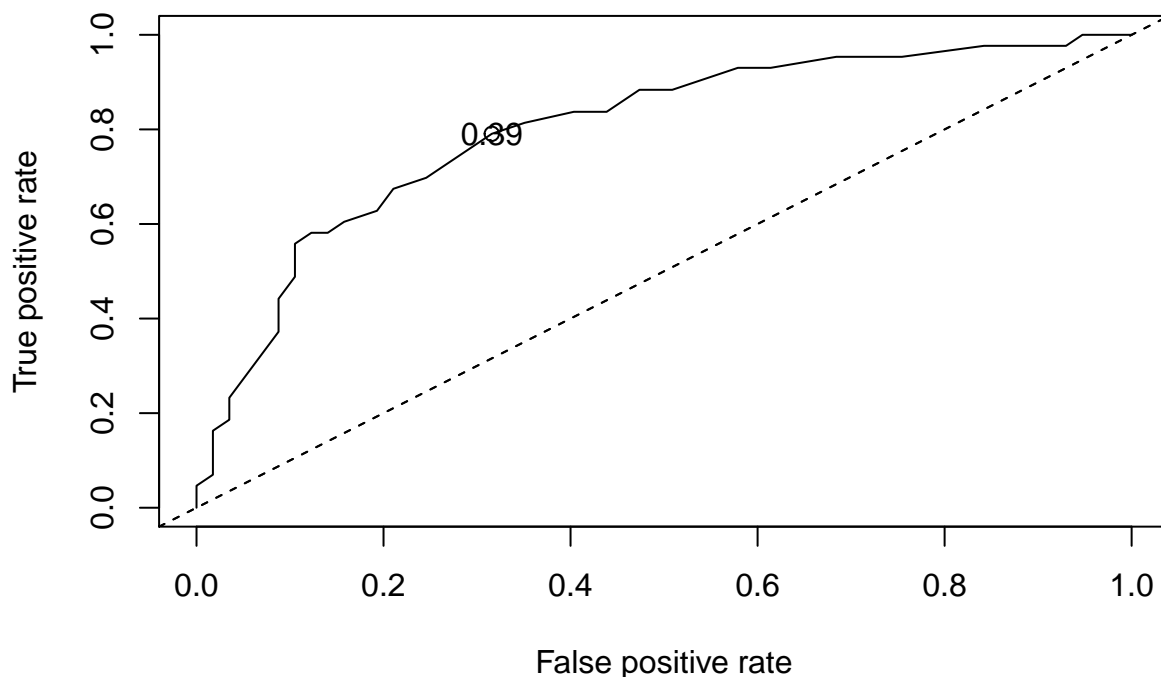
```
pred = prediction(fitted(md), heart$hd)
perf = performance(pred, "tpr", "fpr")
x <- perf@x.values[[1]]
y <- perf@y.values[[1]]

dist <- sqrt((0 - x)^2 + (1 - y)^2)

which_min = min(dist) == dist

min_dist_cut <- perf@alpha.values[[1]][which_min]
plot(perf, print.cutoffs.at=min_dist_cut, main='Curva ROC e o ponto de corte escolhido')
abline(0,1,lty=2)
abline(0,1,lty=2)
```

Curva ROC e o ponto de corte escolhido



Para esse ponto de corte, iremos mostrar a matriz de confusão associada a ele:

Doença cardíaca	prev. modelo = 0	prev. modelo = 1	Total
0	41	16	57
1	11	32	43
Total	52	48	100

Nessa matriz de confusão, essas são algumas métricas de interesse:

- Acurácia: $\frac{41+32}{100} = 0.83$
- Sensibilidade (taxa de verdadeiros positivos): $\frac{32}{48} = 0.6666667$
- Especificidade: $\frac{41}{52} = 0.7884615$

7.7 Ajuste por gamlss

```
ajuste = gamlss(cbind(hd, 1-hd) ~ idade, family=BI, data = heart)
```

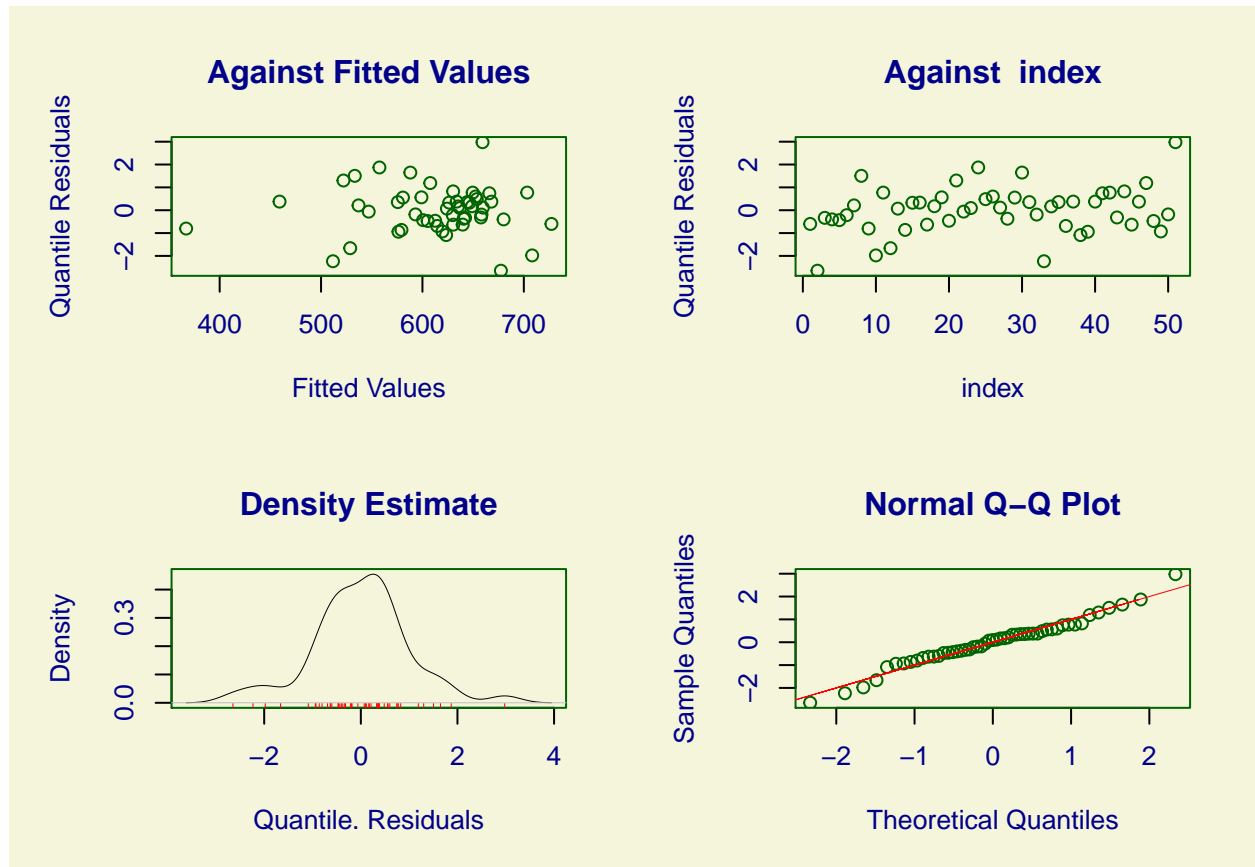
Como podemos comparar, o ajuste do GAMLS resulta em estimativas muito próximas à aquelas geradas pelo glm.

```
summary(ajuste)
```

```
## *****
## Family:  c("NO", "Normal")
##
## Call:    gamlss(formula = fuel ~ dlic + lmiles + income + tax,
##               family = NO(), data = mdata)
##
## Fitting method: RS()
##
## -----
## Mu link function:  identity
## Mu Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 154.192845 185.109822   0.833 0.409255
## dlic         0.471871   0.122053   3.866 0.000353 ***
## lmiles       26.755176   8.867917   3.017 0.004189 **
## income      -0.006135   0.002083  -2.945 0.005097 **
## tax          -4.227983   1.928043  -2.193 0.033526 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## Sigma link function:  log
## Sigma Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.12112    0.09901  41.62  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## No. of observations in the fit:  51
## Degrees of Freedom for the fit:   6
##      Residual Deg. of Freedom:  45
##                               at cycle:  2
##
## Global Deviance:      565.086
##               AIC:      577.086
##               SBC:      588.6769
## *****
```

Os gráficos de resíduos também indicam um bom ajuste, uma vez que não há padrões no gráfico de resíduos quantílicos X valores ajustados, nem um distanciamento da normalidade, como podemos ver no gráfico de probabilidades normal dos mesmos resíduos.

```
plot(ajuste, summaries=F)
```



Por fim, no gráfico de *worm plot*, como não há sobreposição com as regiões circulares superiores e inferiores do desvio, temos indicativos de um bom ajuste.

```
wp(ajuste)
```

