

Inferência bayesiana em Modelos Lineares

Prof. Caio Azevedo

Introdução

- Apresentaremos uma introdução à análise bayesiana do modelo de regressão normal linear homocedástico (MRNLH): [aqui](#).
- O MRNLH faz parte da família de modelos lineares.
- Precisaremos de resultados sobre as distribuições: normal multivariada e t de Student multivariada.
- Referências adicionais:
 - MRNLH e alguns outros modelos de regressão: [aqui](#).
 - Modelos lineares: [aqui](#).
 - Normal multivariada: [aqui](#).

Breve revisão de cálculo de probabilidades multivariado

- Como usual, denotaremos por uma letra maiúscula, e.g. Y , uma variável aleatória (va) e por uma letra minúscula, y , um valor observado (realização de um experimento aleatório) desta va.
- Um vetor aleatório (vea) $\mathbf{Y} = (Y_1, \dots, Y_p)'$ é uma coleção (arranjo) de variáveis aleatórias.
- As va's que compõem um vea podem apresentar alguma estrutura de dependência e/ou serem de diferentes tipos (discretas, contínuas ou mistas).

Cont.

- Função densidade de probabilidade ou função de probabilidade: $p(\mathbf{y}|\theta)$, θ são parâmetros desconhecidos (pode ser um vetor).

- Função de distribuição acumulada

$$F(\mathbf{y}|\theta) = P(Y_1 \leq y_1, \dots, Y_p \leq y_p|\theta).$$

- Vetor de médias: $\boldsymbol{\mu} = \mathcal{E}(\mathbf{Y}|\theta) = \begin{bmatrix} \mathcal{E}(Y_1|\theta) \\ \mathcal{E}(Y_2|\theta) \\ \vdots \\ \mathcal{E}(Y_p|\theta) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$

Cont.

- Matriz de covariâncias: $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{Y}|\theta) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_p^2 \end{bmatrix}$
- Sejam \mathbf{A} e \mathbf{B} matrizes não aleatórias, então
 - $\mathcal{E}(\mathbf{A}\mathbf{Y}|\theta) = \mathbf{A}\mathcal{E}(\mathbf{Y}|\theta)$.
 - $\text{Cov}(\mathbf{A}\mathbf{Y}, \mathbf{B}\mathbf{X}|\theta) = \mathbf{A}\text{Cov}(\mathbf{Y}, \mathbf{X}|\theta)\mathbf{B}'$.
 - $\text{Cov}(\mathbf{A}\mathbf{Y}|\theta) = \mathbf{A}\text{Cov}(\mathbf{Y}|\theta)\mathbf{A}'$.

Distribuição Normal multivariada

- Dizemos que $\mathbf{Y} = (Y_1, \dots, Y_p)' | \boldsymbol{\theta} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ se sua fdp é dada por

$$p(\mathbf{y} | \boldsymbol{\theta}) = |\boldsymbol{\Sigma}|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \mathbb{1}_{\mathcal{R}^p}(\mathbf{y})$$

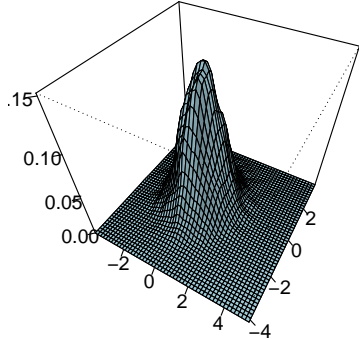
- $\boldsymbol{\mu}$ é o vetor de médias, $\boldsymbol{\Sigma}$ é a matriz de covariâncias e $\boldsymbol{\theta} = (\boldsymbol{\mu}', \text{diags}(\boldsymbol{\Sigma}))'$, $\text{diags}(\boldsymbol{\Sigma})$ representa os elementos distintos de $\boldsymbol{\Sigma}$.

- Fechada sob marginalização: $Y_i | \boldsymbol{\theta} \sim N_1(\mu_i, \sigma_i^2)$. Se $p = 1$, então $Y \sim N_1(\mu, \sigma^2)$.

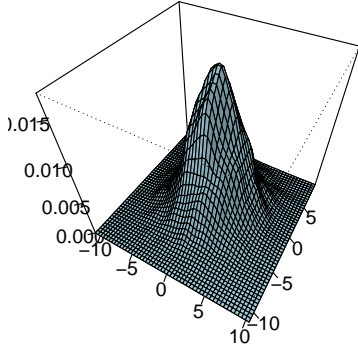
- Fechada por combinações lineares:

$$\mathbf{A}_{(q \times p)} \mathbf{Y}_{(p \times 1)} | \boldsymbol{\theta} \sim N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

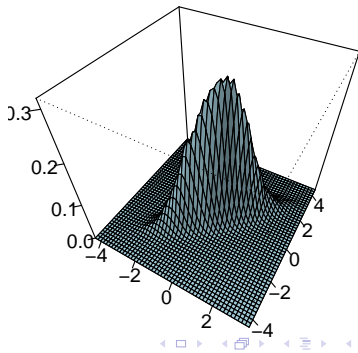
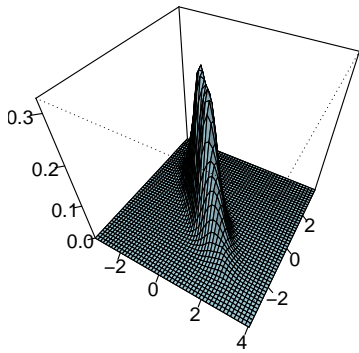
- $Y_i \perp Y_j, \forall i \neq j \Leftrightarrow \sigma_{ij} = 0$.



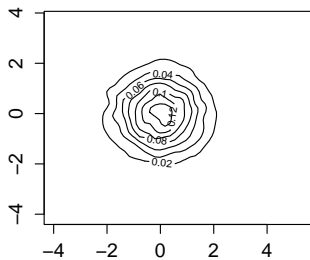
correlação = -0.9



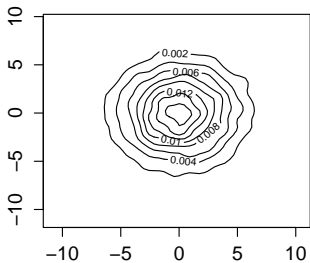
correlação = 0.9



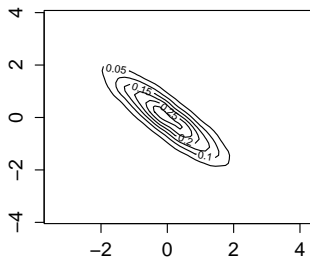
correlação = 0



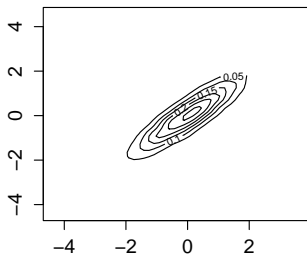
correlação = 0



correlação = -0.9



correlação = 0.9



Distribuições condicionais

- Seja $\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2)'$, $\boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2)'$ e

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

em que $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}'_{12}$.

- Então $\mathbf{Y}_1 | (\boldsymbol{\theta}, \mathbf{Y}_2 = \mathbf{y}_2) \sim N(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$, em que

$$\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{y}_2 - \boldsymbol{\mu}_2); \bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}.$$

- Estimadores de máxima verossimilhança (dada uma amostra aleatória) $\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}_i$ e $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Y}_i - \bar{\mathbf{Y}})' (\mathbf{Y}_i - \bar{\mathbf{Y}})$.

Distribuição t de Student multivariada

- Dizemos que $\mathbf{Y} = (Y_1, \dots, Y_p)' | \boldsymbol{\theta} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, k)$ se sua fdp é dada por

$$p(\mathbf{y}|\boldsymbol{\theta}) = \frac{\Gamma(\frac{p+k}{2})}{(\sqrt{\pi})^p \Gamma(k/2)} k^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \left[1 + \frac{(\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})}{k} \right]^{-(p+k)/2}$$

- Se $k > 1$, $\mathcal{E}(\mathbf{Y}|\boldsymbol{\theta}) = \boldsymbol{\mu}$, se $k > 2$, $\text{Cov}(\mathbf{Y}|\boldsymbol{\theta}) = \frac{k}{k-2} \boldsymbol{\Sigma}$. Além disso, $\boldsymbol{\theta} = (\boldsymbol{\mu}', \text{diags}(\boldsymbol{\Sigma}), k)$, $\text{diags}(\boldsymbol{\Sigma})$ representa os elementos distintos de $\boldsymbol{\Sigma}$.
- Fechada sob marginalização: $Y_i | \boldsymbol{\theta} \sim t(\mu_i, \sigma_i^2, k)$. Se $p = 1$, então $Y \sim t(\mu, \sigma^2, k)$ ([aqui](#)).
- Fechada por comb. lineares: $\mathbf{A}_{(q \times p)} \mathbf{Y}_{(p \times 1)} | \boldsymbol{\theta} \sim t_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}', k)$.

Distribuição Normal multivariada inversa gama

- Seja $\mathbf{Y} = (Y_1, \dots, Y_p)'$ e W , tais que

$$\mathbf{Y} | (\boldsymbol{\mu}, \boldsymbol{\Sigma}, W = w) \sim N_p(\boldsymbol{\mu}, w\boldsymbol{\Sigma}) \text{ e } W | (a, b) \sim IG(a, b),$$

então $(\mathbf{Y}', W)' | (\boldsymbol{\mu}, \boldsymbol{\Sigma}, a, b) \sim N_p IG(\boldsymbol{\mu}, \boldsymbol{\Sigma}, a, b)$

$$\begin{aligned} p(\mathbf{y}, w | \boldsymbol{\theta}) &= w^{-1/2} |\boldsymbol{\Sigma}|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2w} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \\ &\times \mathbb{1}_{\mathcal{R}^p}(\mathbf{y}) \frac{b^a}{\Gamma(a)} w^{-(a+1)} e^{b/w} \mathbb{1}_{\mathcal{R}^+}(w) \\ &= |\boldsymbol{\Sigma}|^{-1/2} (2\pi)^{-p/2} w^{-(a+1/2+1)} \\ &\times \exp \left\{ -\frac{1}{2w} [2b + (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})] \right\} \mathbb{1}_{\mathcal{R}^p}(\mathbf{y}) \mathbb{1}_{\mathcal{R}^+}(w) \end{aligned}$$

- Além disso, $\mathbf{Y} | (\boldsymbol{\mu}, \boldsymbol{\Sigma}, a, b) \sim t_p \left(\boldsymbol{\mu}, \frac{a}{b} \boldsymbol{\Sigma}, 2a \right)$.

Exemplo 12: Teste de esforço cardiopulmonar

- Considere o estudo sobre teste de esforço cardiopulmonar em pacientes com insuficiência cardíaca realizado no InCor da Faculdade de Medicina da USP pela Dra. Ana Fonseca Braga.
- Um dos objetivos do estudo é comparar os grupos formados pelas diferentes etiologias cardíacas quanto às respostas respiratórias e metabólicas obtidas do teste de esforço cardiopulmonar.
- Outro objetivo do estudo é saber se alguma das características observadas (ou combinação delas) pode ser utilizada como fator prognóstico de óbito.

cont.

- Os dados podem ser encontrados [aqui](#) e também no site do curso.
- Etiologias = CH: chagásicos, ID: idiopáticos, IS: isquêmicos, C: controle.
- Considere que o objetivo é o de explicar a variação do consumo de oxigênio no limiar anaeróbio ($ml/(kg.min)$) em função da carga utilizada na esteira ergométrica para pacientes com diferentes etiologias (causas) de insuficiência cardíaca.

cont.

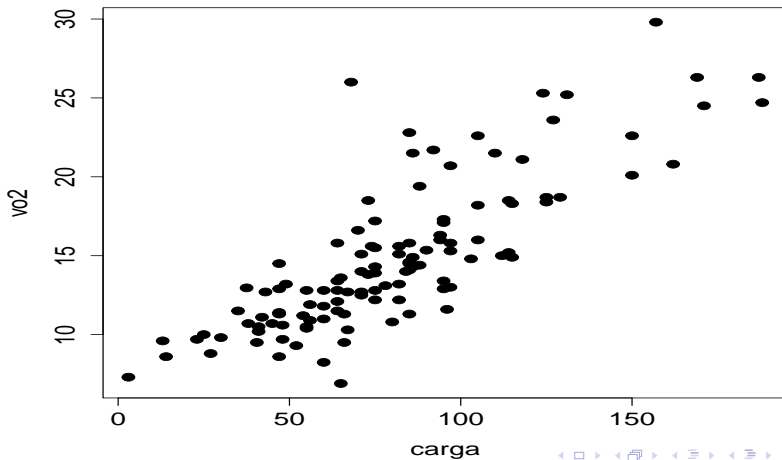
- A grosso modo o Limiar Anaeróbio é um ponto (limite), de divisão entre metabolismo essencialmente aeróbio e metabolismo essencialmente anaeróbio.
- Aeróbio (com a utilização de oxigênio) ; anaeróbio (sem a utilização de oxigênio).
- Como responder à pergunta de interesse (ignorando as etiologias cardíacas, num primeiro momento)?.

Dados

ID	Etiologia	Carga	VO2
1	CH	41	10,0
2	CH	56	11,5
3	ID	8	7,0
4	ID	53	8,9
⋮	⋮	⋮	
7	ID	0	6,5
⋮	⋮	⋮	
123	C	64	14,1
124	C	70	12,4

Dispersão sem considerar os grupos

Consumo de oxigênio em função da carga



Exemplo 12: desconsiderando as etiologias cardíacas

$$Y_i = \beta_0 + \beta_1 x_i + \xi_i, i = 1, \dots, 124$$

- $\xi_i | \boldsymbol{\theta} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2)'$ parâmetros desconhecidos.
- x_i : carga à que o paciente i foi submetido (conhecido e não aleatório).
- Parte sistemática: $\mathcal{E}(Y_i | \boldsymbol{\theta}) = \beta_0 + \beta_1 x_i$.
- Parte aleatória: ξ_i .
- Assim, temos que: $Y_i | \boldsymbol{\theta} \stackrel{ind.}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$.

Cont.

- β_1 : é o aumento esperado no consumo de oxigênio para o aumento de uma unidade na carga imposta.
- Se for possível observar $x_i = 0$, carga igual à 0, temos que:
 - β_0 : valor esperado do consumo de oxigênio para pacientes submetidos à uma carga igual à 0.
- Caso contrário, podemos considerar o seguinte modelo:

$$Y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \xi_i, i = 1, \dots, 124, \bar{x} = \frac{1}{124} \sum_{i=1}^n x_i.$$

- Neste caso, β_0 é o valor esperado do consumo de oxigênio para pacientes submetidos à uma carga igual à média amostral.

Exemplo 12: considerando as etiologias cardíacas

$$Y_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + \xi_{ij}, i = 1, \dots, ; j = 1, \dots, n_i$$

- Etiologias = CH ($i = 1$), ID ($i = 2$), IS ($i = 3$), C: ($i = 4$).
- $\xi_{ij}|\boldsymbol{\theta} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- $\boldsymbol{\theta} = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \sigma^2)'$ parâmetros desconhecidos.
- x_{ij} : carga à que o paciente j que apresenta a etiologia cardíaca i foi submetido (conhecido e não aleatório).
- Parte sistemática: $\mathcal{E}(Y_{ij}|\boldsymbol{\theta}) = \beta_{0i} + \beta_{1i}x_{ij}$.
- Parte aleatória: ξ_{ij} .
- Assim, temos que: $Y_{ij}|\boldsymbol{\theta} \stackrel{ind.}{\sim} N(\beta_{0i} + \beta_{1i}x_{ij}, \sigma^2)$.

Modelo de regressão normal linear

- Modelo geral:

$$Y_i = \beta_0 + \sum_{j=1}^{p-1} X_{ij}\beta_j + \xi_i, i = 1, 2, \dots, n,$$
$$\xi_i | \boldsymbol{\theta} \stackrel{iid}{\sim} N(0, \sigma^2).$$

em que $\boldsymbol{\theta} = (\beta_0, \beta_1, \dots, \beta_{p-1})'$.

- Assim, temos que $Y_i | \boldsymbol{\theta} \stackrel{ind.}{\sim} N(\beta_0 + \sum_{j=1}^{p-1} X_{ij}\beta_j, \sigma^2)$.
- Consideremos, por enquanto, σ^2 conhecido.

Aplicação no modelo de regressão normal linear

■ Modelo

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1(p-1)} \\ 1 & X_{21} & \dots & X_{2(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{n(p-1)} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}$$

- Suposição $\boldsymbol{\xi}|\boldsymbol{\theta} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Assim, temos que

$$\mathbf{Y}|\boldsymbol{\theta} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n).$$

- Estimador de mínimos quadrados de $\boldsymbol{\beta}$, minimizar

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \text{ equivale a resolver } (\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}.$$

Estimação por mínimos quadrados

- Solução $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.
- Sob normalidade $\hat{\beta} \sim N_p(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$.
- Consequentemente, $\hat{\beta}_j \sim N_1(\beta_j, \Psi_{jj}), j = 1, \dots, p$ em que $\Psi_{jj} = \sigma^2[(\mathbf{X}'\mathbf{X})^{-1}]_{jj}$.
- IC exato: $\hat{\beta}_j \pm z_{\frac{1-\gamma}{2}} \sqrt{\Psi_{jj}}$. $P(Z > z_{\frac{1-\gamma}{2}}) = \frac{1-\gamma}{2}$, $Z \sim N(0, 1)$.
- Testes de hipótese: $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$. Teste exato, rejeitar H_0 se $|Z_c| > z_{\frac{1-\gamma}{2}}$, em que

$$Z_c = \frac{\hat{\beta}_j}{\sqrt{\Psi_{jj}}}$$

Inferência Bayesiana

- Considere $Y_i | \boldsymbol{\theta} \stackrel{ind.}{\sim} N(\mathbf{x}_i' \boldsymbol{\beta}, \sigma^2)$, em que \mathbf{x}_i é a i -ésima linha da matriz \mathbf{X} e $\boldsymbol{\theta} = (\beta_0, \beta_1, \dots, \beta_{p-1})'$.
- Verossimilhança:

$$p(\mathbf{y} | \boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \quad (2)$$

$$\propto \exp \left\{ -\frac{1}{2\sigma^2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\}, \quad (3)$$

que corresponde ao núcleo de uma distribuição $N_p(\hat{\boldsymbol{\beta}}, (\mathbf{X}' \mathbf{X})^{-1})$.

Assim, a família conjugadas é a distribuição $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (família conjugada).

Esboço da demonstração

- Basta desenvolver as duas formas quadráticas (fórmulas (2) e (3)) do slide anterior e ver que elas são semelhantes (\propto) (próximo slide).
- Ou seja, devemos provar que

$$\begin{aligned} & \exp \left\{ -\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}) \right]' \left[(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}) \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{X}' \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right] \right\}, \end{aligned}$$

notando que $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$

Cont.

- Vamos assumir a seguinte priori: $\beta \sim N_p(\mu_\beta, \Psi_\beta)$.
- Pode-se provar que $\beta|\mathbf{y} \sim N_p(\mu_\beta^*, \Psi_\beta^*)$, em que

$$\begin{aligned}\Psi_\beta^* &= \left((\sigma^2)^{-1} (\mathbf{X}'\mathbf{X}) + \Psi_\beta^{-1} \right)^{-1} \\ \mu_\beta^* &= \Psi_\beta^* \left(\Psi_\beta^{-1} \mu_\beta + (\sigma^2)^{-1} \mathbf{X}'\mathbf{y} \right)\end{aligned}$$

- Assim, $\beta_j|\mathbf{y} \sim N_1(\mu_{\beta_j}^*, \Psi_{\beta_{jj}}^*)$, em que $\mu_{\beta_j}^* = (\mu_\beta^*)_j$ e $\Psi_{\beta_{jj}}^* = (\Psi_\beta^*)_{jj}$.

Esboço da demonstração

$$\begin{aligned} p(\beta|\mathbf{y}) &\propto p(\mathbf{y}|\beta)p(\beta) \\ &\propto \exp \left\{ -\frac{1}{2} \left(\beta' \frac{\mathbf{X}'\mathbf{X}}{\sigma^2} \beta - 2\hat{\beta}' \mathbf{X}'\mathbf{X} \beta + \beta' \boldsymbol{\Psi}_\beta^{-1} \beta - 2\boldsymbol{\mu}'_\beta \boldsymbol{\Psi}_\beta^{-1} \beta \right) \right\} \mathbb{1}_{\mathcal{R}^p}(\beta) \\ &= \exp \left\{ -\frac{1}{2} \left[\beta' \left(\frac{\mathbf{X}'\mathbf{X}}{\sigma^2} + \boldsymbol{\Psi}_\beta^{-1} \right) \beta - 2 \left(\boldsymbol{\mu}'_\beta \boldsymbol{\Psi}_\beta^{-1} + \hat{\beta}' \frac{\mathbf{X}'\mathbf{X}}{\sigma^2} \right) \beta \right] \right\} \mathbb{1}_{\mathcal{R}^p}(\beta) \\ &\propto \exp \left\{ -\frac{1}{2} (\beta - \boldsymbol{\mu}_\beta^*)' (\boldsymbol{\Psi}_\beta^*)^{-1} (\beta - \boldsymbol{\mu}_\beta^*) \right\} \mathbb{1}_{\mathcal{R}^p}(\beta) \end{aligned}$$

Cont.

- Intervalos de Credibilidade: $IC_B = \mu_{\beta_j}^* \pm z_{\frac{1-\gamma}{2}} \sqrt{\Psi_{\beta_{jj}}^*}$.
- Teste de hipótese $\frac{P(\beta_j \in \Theta_1 | \mathbf{y})}{P(\beta_j \in \Theta_0 | \mathbf{y})}$
- Estimação (empírica) dos hiperparâmetros $(\mu_{\beta}, \Sigma_{\beta})$: por exemplo, usar a distribuição preditiva à priori de toda a amostra $(Y_1, Y_2, \dots, Y_n)'$.
- Dist. preditiva a priori de uma nova observação:
 $Y_{i+1} | \mathbf{y} \sim N(\mathbf{x}'_i \mu_{\beta}, \mathbf{x}'_i \Sigma_{\beta} \mathbf{x}_i)$.
- Dist. preditiva a posteriori de uma nova observação:
 $Y_{i+1} | \mathbf{y} \sim N(\mathbf{x}'_i \Sigma_{\beta}^* \mu_{\beta}^*, \mathbf{x}'_i \Sigma_{\beta}^* \mathbf{x}_i)$.

σ^2 desconhecido

- Dizemos que $(\mathbf{X}', y)' | (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha, \beta)$ tem uma distribuição normal multivariada inversa gama

$$(\mathbf{X}', y)' | (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha, \beta) \sim N_p IG(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha, \beta)$$

se, $\mathbf{X} | (\boldsymbol{\mu}, \boldsymbol{\Sigma}, y) \sim N_p(\boldsymbol{\mu}, y\boldsymbol{\Sigma}^{-1})$ e $y | (\alpha, \beta) \sim IG(\alpha, \beta)$, ou seja:

$$\begin{aligned} p(\mathbf{x}, y | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha, \beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} (y)^{-(\alpha+1)} \frac{|\boldsymbol{\Sigma}|^{p/2}}{\sqrt{y}\sqrt{2\pi}} \\ &\times \exp \left\{ -\frac{1}{2y} [2\beta + (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma} (\mathbf{x} - \boldsymbol{\mu})] \right\} \\ &\times \mathbb{1}_{\mathcal{R}^p}(\mathbf{x}) \mathbb{1}_{\mathcal{R}^+}(y) \end{aligned}$$

σ^2 desconhecido

- Nesse caso, $\boldsymbol{\theta} = (\beta_0, \beta_1, \dots, \beta_{p-1}, \sigma^2)'$ e a verossimilhança pode ser escrita como:

$$p(\mathbf{y}|\boldsymbol{\theta}) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \quad (4)$$

$$\propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[ks^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right] \right\}, \quad (5)$$

em que $ks^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, $k = n - p$.

- A expressão anterior corresponde ao núcleo de uma distribuição $N_p IG(\hat{\boldsymbol{\beta}}, (\mathbf{X}' \mathbf{X}), ks^2/2, n - 1)$

Esboço da demonstração

- Assim, a família conjugada de prioris é a normal multivariada inversa-gama, ou seja $\beta|\sigma^2 \sim N_p(\mu_\beta, \sigma^2 \Sigma_\beta^{-1})$, $\sigma^2 \sim IG(c/2, d/2)$.
- Com relação as expressões (4) e (5), para mostrar sua equivalência tem-se que mostrar que:

$$\begin{aligned} & \exp \left\{ -\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\hat{\beta}) + (\mathbf{X}\hat{\beta} - \mathbf{X}\beta) \right]' \left[(\mathbf{y} - \mathbf{X}\hat{\beta}) + (\mathbf{X}\hat{\beta} - \mathbf{X}\beta) \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\hat{\beta})' (\mathbf{y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \beta)' \mathbf{X}' \mathbf{X} (\hat{\beta} - \beta) \right] \right\}, \end{aligned}$$

notando que $(\mathbf{y} - \mathbf{X}\hat{\beta})' (\mathbf{X}\hat{\beta} - \mathbf{X}\beta) = \mathbf{0}$.

Posteriori

- Tem-se, finalmente, que

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) &\propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}\left[d + ks^2 + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)' \mathbf{C} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)\right]\right. \\ &\quad \left.+ (\boldsymbol{\beta} - \mathbf{c})' (\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_\beta^{-1})^{-1} (\boldsymbol{\beta} - \mathbf{c})\right\} (\sigma^2)^{-\left(\frac{n+c}{2}+1\right)} (\sigma^2)^{-1/2} \\ &\quad \times \mathbb{1}_{\mathcal{R}^p}(\boldsymbol{\beta})\mathbb{1}_{(0,\infty)}(\sigma^2) \end{aligned}$$

em que $\mathbf{C} = \left((\mathbf{X}'\mathbf{X})^{-1} + \boldsymbol{\Sigma}_\beta\right)^{-1}$ e

$$\mathbf{c} = \left(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_\beta^{-1}\right)^{-1} \left(\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta + \mathbf{X}'\mathbf{y}\right)$$

Posteriori

- Assim,

$$\beta|\sigma^2, \mathbf{y} \sim N_p(\boldsymbol{\mu}_\beta^*, \boldsymbol{\Sigma}_\beta^*) \text{ e } \sigma^2|\mathbf{y} \sim IG(c_1/1, d_1/2).$$

- Portanto,

$$\beta|\mathbf{y} \sim t_p(\boldsymbol{\mu}_\beta^*, \frac{d_1}{c_1}\boldsymbol{\Sigma}_\beta^*, 2c_1),$$

em que $\boldsymbol{\mu}_\beta^* = \mathbf{c}$, $\boldsymbol{\Sigma}_\beta^* = (\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_\beta^{-1})^{-1}$, $c_1 = n + c$ e $d_1 = d + ks^2 + (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)' \mathbf{C} (\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)$.

- Logo,

$$\beta_j|\mathbf{y} \sim t_1((\boldsymbol{\mu}_\beta^*)_j, \frac{d_1}{c_1}(\boldsymbol{\Sigma}_\beta^*)_{jj}, 2c_1),$$

em que j ou jj significa o j -ésimo elemento do vetor ou da diagonal principal da matriz.

Observações

- 1 As distribuições preditivas (posteriori e priori) podem ser obtidas facilmente com os resultados apresentados anteriormente.
- 2 Testes de hipóteses podem ser implementados utilizando a idéia de prioris com **estrutura de mistura**.
- 3 Hipóteses do tipo $H_0 : \mathbf{C}_{(q \times p)} \boldsymbol{\beta}_{(p \times 1)} = \mathbf{M}_{(q \times 1)}$ vs $H_1 : \mathbf{C} \boldsymbol{\beta} \neq \mathbf{M}$, podem ser testadas utilizando o ponto 2 e a distribuição de $\mathbf{C} \boldsymbol{\beta} | \mathbf{y} \sim t_q(\mathbf{C} \boldsymbol{\mu}_\beta^*, \frac{d_1}{c_1} \mathbf{C} \boldsymbol{\Sigma}_\beta^* \mathbf{C}', 2c_1)$.
- 4 As estimativas bayesianas (EAP, MAP, MdAP) podem ser usadas para gerar os resíduos usuais. Tais resíduos “bayesianos” podem ser analisados semelhantemente ao caso em que se utiliza as estimativas de mínimos quadrados e/ou, do ponto de vista Bayesiano.

Voltando ao conjunto de dados

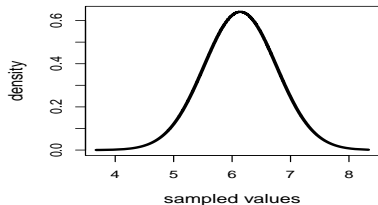
- 1 Vamos assumir que $\beta|\sigma^2 \sim N(\mathbf{0}, \Sigma_\beta)$, em que

$$\Sigma_\beta = \begin{bmatrix} 1000 & 0 \\ 0 & 1000 \end{bmatrix}.$$

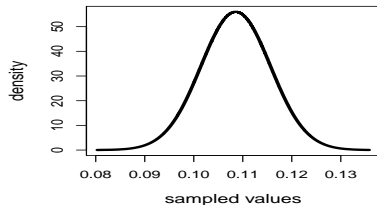
- 2 Para σ^2 , vamos considerar um caso limite da inversa gamma, o qual é obtido fazendo-se $c \rightarrow 0$ e $d \rightarrow 0$. Em termos práticos, basta tomar $c = d = 0$. Isto equivale à $p(\sigma^2) \propto (\sigma^2)^{-1} \mathbb{1}_{(0,\infty)}(\sigma^2)$.

Posterioris

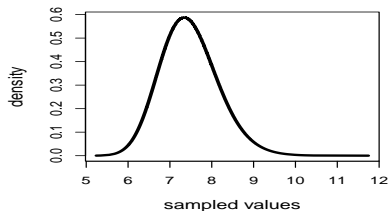
(Intercept)



carga



σ^2



Resultados da análise

- Estimativas dos parâmetros

Parâmetro	EAP	DPAP	IC(95%)	HPD(95%)
β_0	6,14	0,62	[4,91 ; 7,37]	[4,91 ; 7,37]
β_1	0,11	0,01	[0,09 ; 0,12]	[0,09 ; 0,12]
σ^2	7,53	0,69	[6,23 ; 8,94]	[6,16 ; 8,85]

- Pelo respectivo HPD, vemos que há uma relação positiva (e significativa) entre carga e consumo de oxigênio (β_1).

Diagnóstico

- Uma das etapas mais importantes na modelagem de problemas é verificar a adequabilidade das ferramentas estatísticas utilizadas.
- Essa etapa deve preceder a apresentação dos resultados inferenciais (estimativas e posteriores)
- Uma abordagem consiste em analisar o comportamento de resíduos apropriados.
- Materiais adicionais a respeito: [link1](#), [link2](#), [link3](#), [link4](#).

Suposições do MRNLH

- As principais suposições do MRNLH são:
 - Homocedasticidade (dos erros).
 - Independência (correlação nula) dos erros.
 - Normalidade dos erros.
- Como verificar as suposições do modelo?
- Como proceder se uma ou mais suposições não forem (satisfatoriamente) válida(s)?
- Vamos, primeiramente, lembrar a abordagem usual, do ponto de vista frequentista

Cont. (livro do Prof. Gilberto, págs. 48, 49 e 50)

- Defina

$$V_i = \frac{R_i}{\sqrt{S^2(1 - h_{ii})}} = \frac{Y_i - \hat{Y}_i}{\sqrt{S^2(1 - h_{ii})}}, \quad (6)$$

em que $\hat{Y}_i = \mathbf{X}'_i \hat{\boldsymbol{\beta}}$, $S^2 = \frac{1}{n-p} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})' (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})$,

$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}$ e \mathbf{X}'_i , $h_i = \text{diag} (\mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}')$ é a i -ésima linha da matriz \mathbf{X} .

- A divisão por $(1 - h_{ii})$ atenua a correlação entre os resíduos.
- Contudo, R_i e S^2 não são independentes.
- Porém, $S^2_{(i)}$ e R_i o são (em que $S^2_{(i)}$ corresponde à S^2 obtido no modelo sem a i -ésima observação e (i) indica que a i -ésima observação foi excluída).

Cont. (livro do Prof. Gilberto, pág. 48, 49 e 50)

- $S_{(i)}^2 = \frac{1}{n-p-1}(\mathbf{Y}_{(i)} - \mathbf{X}_{(i)}\widehat{\boldsymbol{\beta}}_{(i)})'(\mathbf{Y}_{(i)} - \mathbf{X}_{(i)}\widehat{\boldsymbol{\beta}}_{(i)}),$
 $\widehat{\boldsymbol{\beta}}_{(i)} = (\mathbf{X}'_{(i)}\mathbf{X}_{(i)})^{-1}\mathbf{X}'_{(i)}\mathbf{Y}_{(i)}.$

- Pode-se provar, além disso, que

$$S_{(i)}^2 = S^2 \left(\frac{n-p-V_i^2}{n-p-1} \right), \quad (7)$$

(facilita seu cálculo, dispensando o ajuste do modelo “n” vezes).

- Tem-se, então, que

$$T_i = \frac{R_i}{\sqrt{S_{(i)}^2(1-h_{ii})}} \sim t_{(n-p-1)}, \quad (8)$$

sob a validade das hipóteses do modelo (exercício). Lembre-se de que, se $\nu \geq 30$, então $t_{(\nu)} \approx N(0, 1)$.

Análise bayesiana dos resíduos

- Do ponto de vista bayesiano, substituímos $\hat{\beta}$ e S^2 em (6) e em (8), por alguma **estimativa bayesiana pontual** (EAP, MeAP e MAP). No caso de $S_{(i)}^2$ a substituição levará a uma aproximação da expressão (7).
- Devido à esses resultados frequentistas e ao **Teorema de Bernstein-Von-Mises** (que implica, dentre outros fatos, que os estimadores Bayesianos pontuais são consistentes) e, assim, os resíduos (ainda que assintoticamente) terão a distribuição de referência dada pela Equação (8).

Análise bayesiana dos resíduos

- Mesmo que a distribuição de referência dos resíduos não seja conhecida, é possível construir gráficos de quantil quantil com envelopes, com base na distribuição à posteriori dos parâmetros e dos resíduos.
- Uma outra forma, é obter a posteriori de cada um dos resíduos e estudar o comportamento desses, para cada observação, de forma individual. Ou seja, considerar:

$$T_i | \mathbf{y},$$

que em geral tem de ser feito de forma numérica, como a seguir.

Obtenção da posteriori de cada resíduo

- Algoritmo:

- 1 Simular R valores de β e σ^2 , a partir de $\beta|\mathbf{y}$ e $\sigma^2|\mathbf{y}$ (slide 32).
 - 2 Com base nos valores simulados no item 1), obter R valores para cada uma das seguintes quantidades: V_i , $S_{(i)}^2$ e $T_{(i)}$.
 - 3 Avaliar, com base nos R valores simulados no item 2), estudar a posteriori de cada um dos T_j .
- Note que a distribuição de $T_j|\mathbf{y}$ (algoritmo acima), em geral, é diferente da distribuição de T_j (Equação (8)).

O que e como observar nos resíduos?

- No caso dos resíduos (T_i):
 - Gráfico de dispersão dos resíduos versus o índice da observação: identificação de dependência/tendência/correlação.
 - Gráfico de dispersão dos resíduos versus os valores ajustados: homocedasticidade.
 - Boxplot e/ou gráfico de quantis-quantis: simetria, identificação de “out-liers” e multimodalidade.
 - Problema no gráfico de quantis-quantis: Visualmente, muitas vezes, é complicado avaliar a proximidade dos quantis.
 - Solução: criar bandas de confiança ([gráficos de envelope](#)).

O que e como observar nos resíduos?

- No caso dos resíduos ($T_i|\mathbf{y}$):
 - Estudar a densidade (estimar não parametricamente) de cada resíduo.
 - Estudar o box-plot de cada resíduo.
 - Calcular a probabilidade de ser um outlier.
 - Verificar se o modelo não se adequa bem a um determinado conjunto de observações.

Envelopes

- Com relação aos resíduos (T_i), utilizar a abordagem frequentista para a construção de envelopes pode ser muito custoso (computacional e temporalmente), além de exigir um certo cuidado, pois usa o conceito de distribuições amostrais de estatísticas.
- Formas bayesianas de se obter envelopes podem ser encontradas [aqui](#), por exemplo.
- Em nosso curso, como conhecemos a distribuição de referência dos resíduos sob o bom ajuste do modelo, vamos [simular os envelopes](#) como mostrado a seguir.

Procedimento para se gerar o gráfico de envelopes (quantil-quantil)

- 1) Simule n variáveis aleatórias independentes de interesse ($t_{(n-p-1)}$ ou $N(0, 1)$). Repita este processo m vezes.
- 2) Ao final teremos uma matriz com valores simulados dessas variáveis aleatórias, digamos V_{ij} , $i=1, \dots, n$, (tamanho da amostra) $j=1, \dots, m$ (réplica).

$$\mathbf{V} = \begin{bmatrix} V_{11} & V_{12} & \dots & V_{1m} \\ V_{21} & V_{22} & \dots & V_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ V_{n1} & V_{n2} & \dots & V_{nm} \end{bmatrix}$$

Cont.

- 3) Dentro de cada amostra (coluna), ordena-se, de modo crescente, os valores simulados, obtendo-se $v_{(i)j}^*$ (i -ésima estatística de ordem da j -ésima amostra):

$$\mathbf{V}^* = \begin{bmatrix} v_{(1)1} & v_{(1)2} & \cdots & v_{(1)m} \\ v_{(2)1} & v_{(2)2} & \cdots & v_{(2)m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{(n)1} & v_{(n)2} & \cdots & v_{(n)m} \end{bmatrix}$$

- 4) Pode-se obter os limites $v_{(i)l} = \min_{1 \leq j \leq m} v_{(i)j}$ e $v_{(i)s} = \max_{1 \leq j \leq m} v_{(i)j}$.

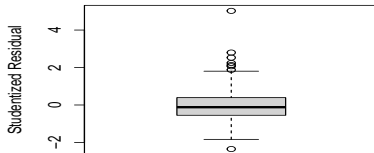
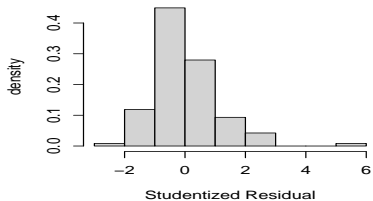
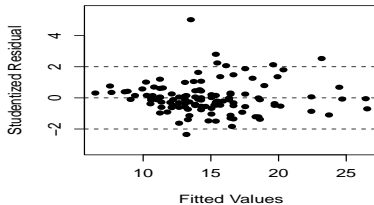
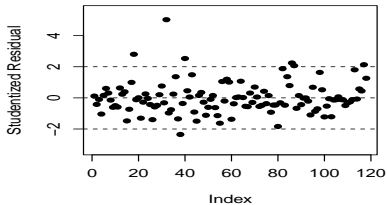
Cont.

- 5) Porém, na prática considera-se $v_{(i)l} = \frac{v_{(i)(2)} + v_{(i)(3)}}{2}$ e $v_{(i)s} = \frac{v_{(i)(m-2)} + v_{(i)(m-1)}}{2}$ (para se gerar limites de confiança), em que $v_{(i)(r)}$ é a r -ésima estatística de ordem dentro de cada linha, $i = 1, 2, \dots, n$.

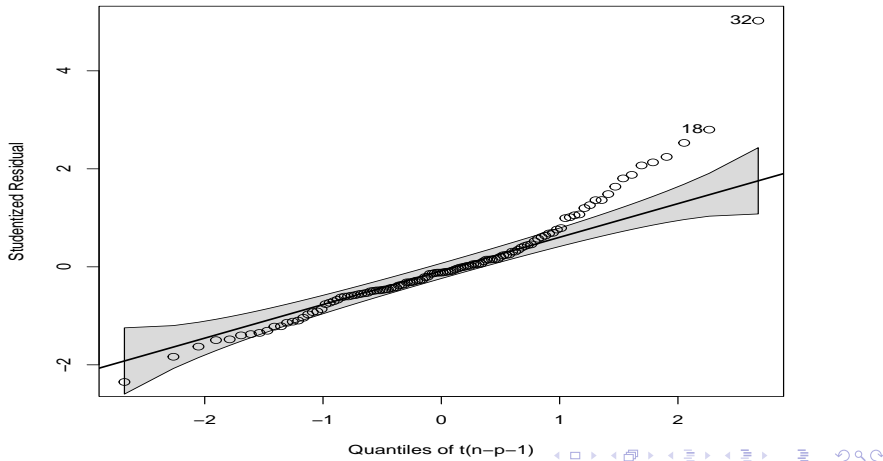
- Além disso, consideramos como a linha de referência

$$v_{(i)} = \frac{1}{m} \sum_{j=1}^m v_{(i)j}, i = 1, 2, \dots, n.$$

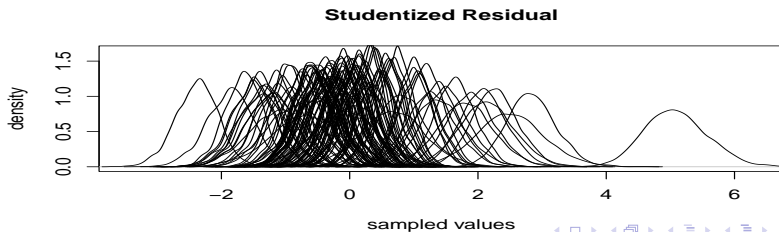
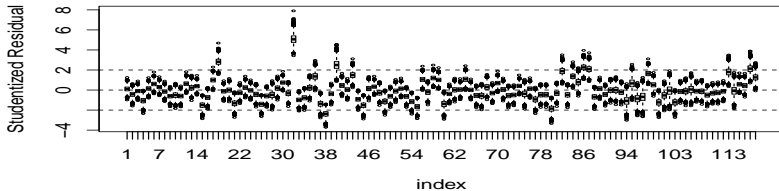
Gráficos de Diagnóstico dos resíduos T_i



Gráficos QQ com envelopes dos resíduos T_i



Distribuições dos resíduos $T_i|y$



Sobre o ajuste do modelo

- Os resíduos (T_i) apresentam não normalidade e heterocedasticidade, além de apresentarem um número maior do que o esperado (sob normalidade $\approx 5\%$).
- Os resíduos (T_i) reforçam a existência de uma quantidade maior do que o esperado de outliers, com pelo menos uma observação com comportamento bastante destacado.
- Alternativa, uso de [MLG- Modelos lineares generalizados](#) com distribuição gama e função de ligação log para a média.