

Análise de dados e métodos de diagnóstico em modelos de regressão normais lineares (parte 2)

Prof. Caio Azevedo

Exemplo 1: 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: ($i = 4$).
- $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- x_{ij} : carga à que o paciente j que apresenta a etiologia cardíaca i foi submetido (conhecido e não aleatório).
- β_{0i} : consumo esperado para pacientes da i -ésima etiologia submetidos à uma carga igual a 0.
- β_{1i} : incremento (positivo ou negativo) no consumo esperado, de pacientes da i -ésima etiologia, para o aumento em uma unidade da carga.

Gráficos de resíduos

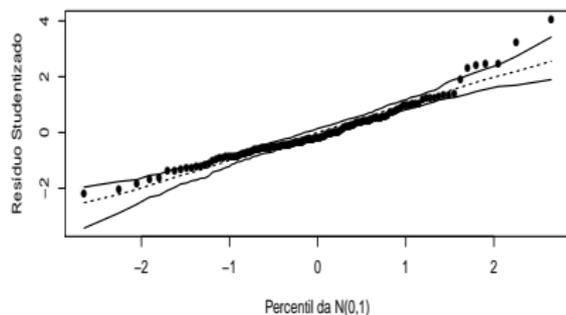
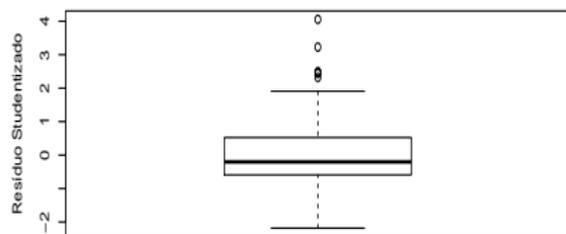
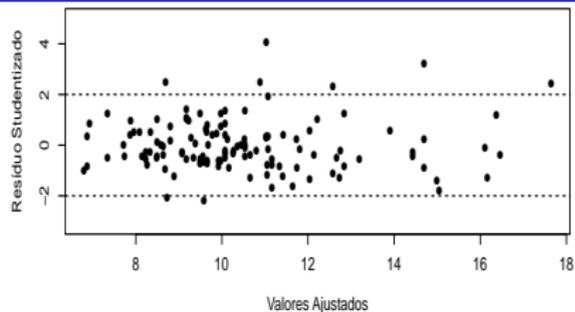
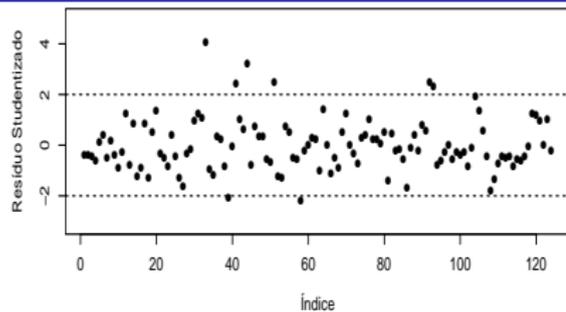
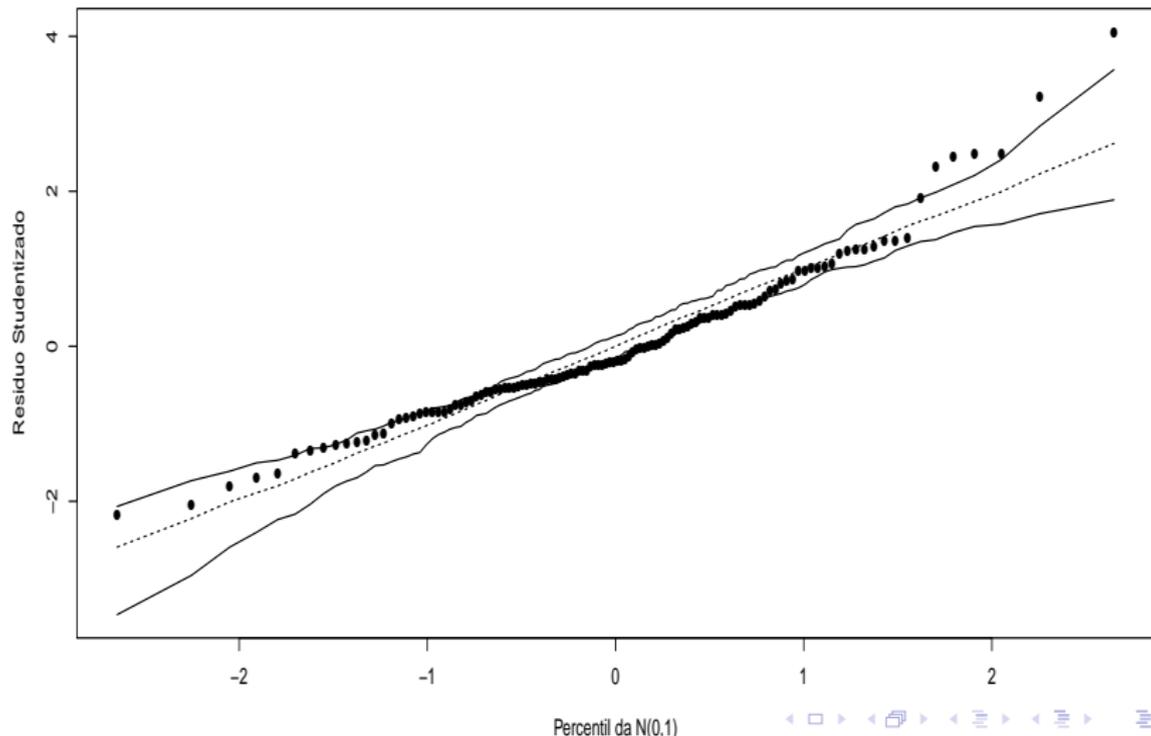


Gráfico de envelopes para os resíduos



Comentários

- Aparentemente os resíduos apresentam uma distribuição assimétrica positiva.
- Talvez exista uma “leve” heterocedasticidade nos resíduos.
- Ausência de correlação entre os resíduos.
- $R^2 = 0,97$ e $\bar{R}^2 = 0,97$.

Exemplo 1: modelo reduzido (desconsiderando as etiologias cardíacas)

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

- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- $(\beta_0, \beta_1, \sigma^2)'$: parâmetros desconhecidos.
- x_i : carga à que o paciente i foi submetido (conhecida e não aleatória).
- Parte sistemática: $\mathcal{E}(Y_i) = \beta_0 + \beta_1 x_i$.
- Parte aleatória: ξ_i .
- O modelo acima implica que $Y_i \stackrel{ind.}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$, Y_i : valor do consumo de oxigênio do paciente i .

Gráficos de resíduos

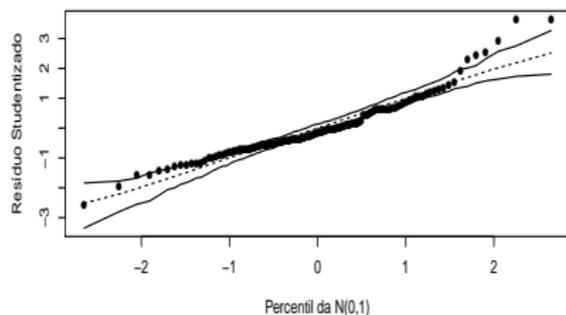
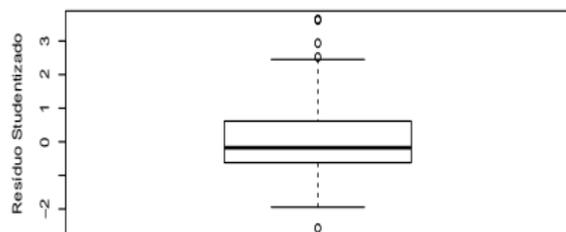
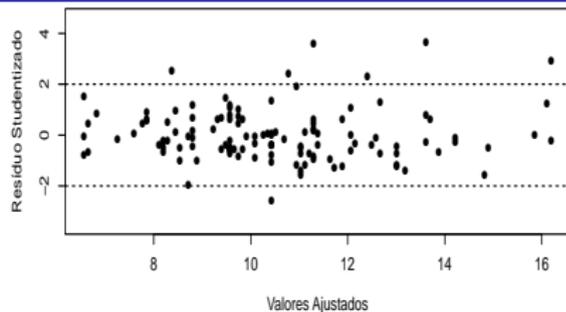
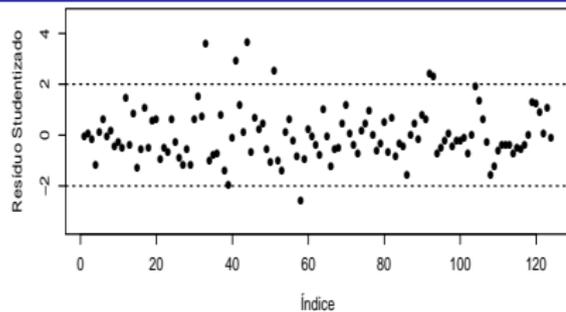
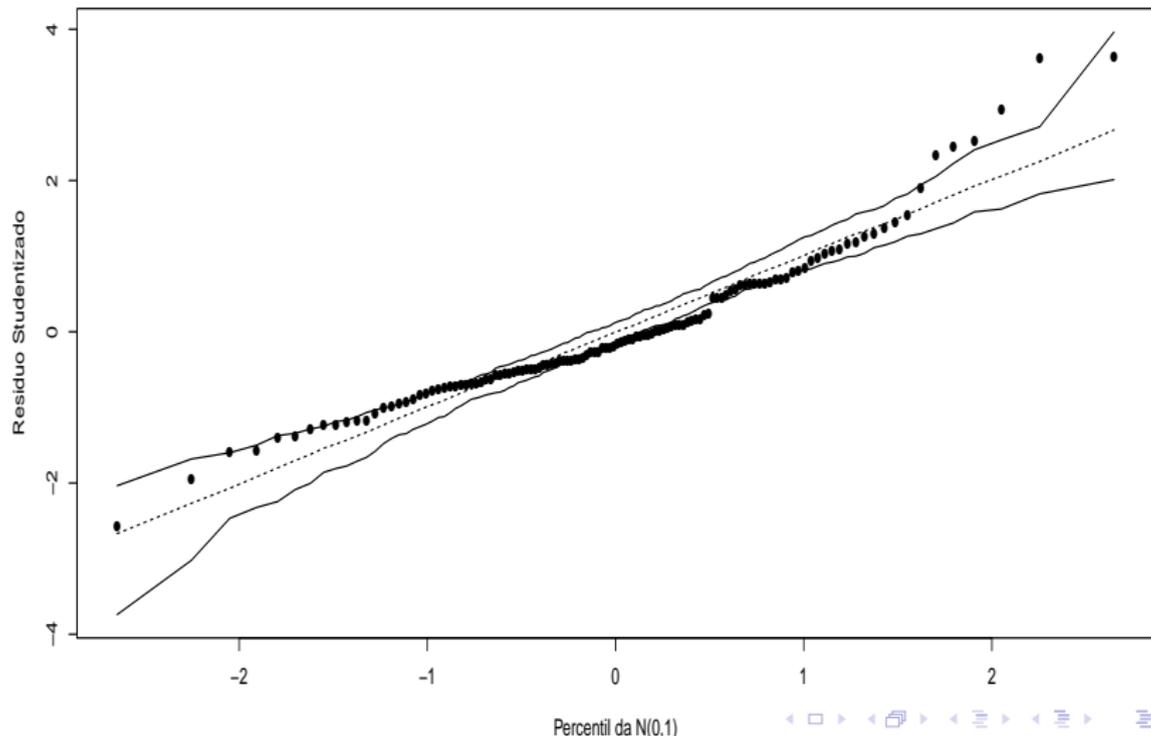


Gráfico de envelopes para os resíduos



Comentários

- Aparentemente os resíduos apresentam uma distribuição assimétrica positiva.
- Talvez exista uma “leve” heterocedasticidade nos resíduos.
- Ausência de correlação entre os resíduos.
- $R^2 = 0,57$ e $\bar{R}^2 = 0,56$.

Exemplo 2: desconsiderando o sexo

$$Y_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + \xi_{ij}, i = 1(\text{convencional}), 2(\text{hugger})(\text{tipo de escova}), \\ j = 1, \dots, 26(\text{criança}).$$

- $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- x_{ij} : IPB pré-escovação da criança j utilizando a escova do tipo i .
- Y_{ij} : IPB pós-escovação da criança j utilizando a escova do tipo i .
- β_{0i} : IPB pós-escovação esperado quando se utiliza a escova do tipo i para um IPB pré-escovação igual a 0.
- β_{1i} : incremento (positivo ou negativo) no IPB pós-escovação esperado quando se utiliza a escova do tipo i , para o aumento em uma unidade no IPB pré-escovação.

Gráficos de resíduos

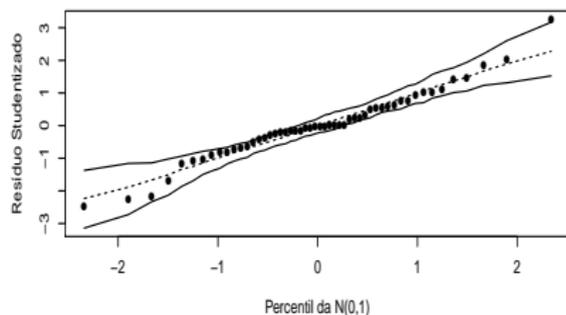
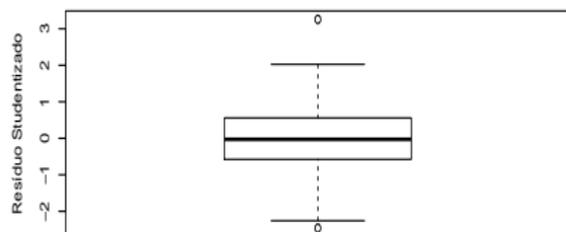
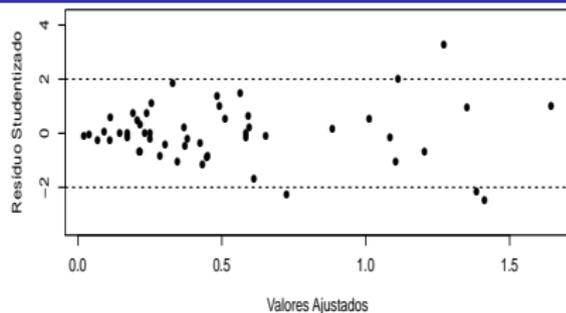
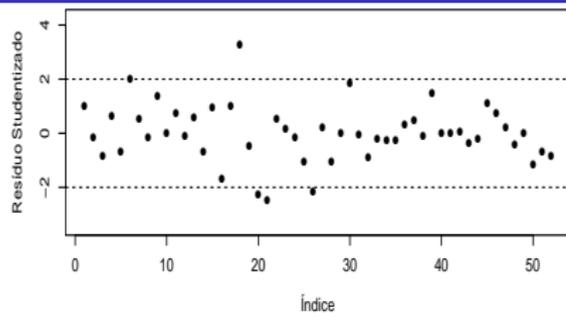
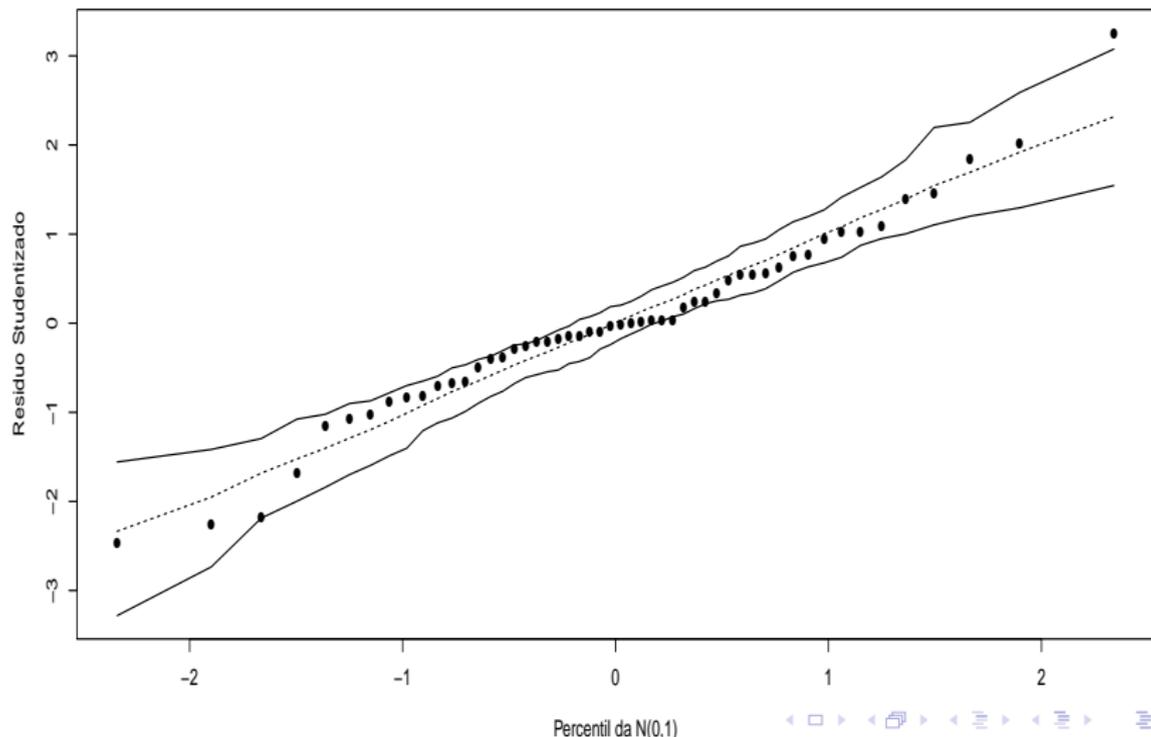


Gráfico de envelopes para os resíduos



Comentários

- Aparentemente os resíduos apresentam uma distribuição simétrica com caudas um pouco mais pesadas do que as da normal.
- Presença de heterocedasticidade nos resíduos.
- Ausência de correlação entre os resíduos.
- $R^2 = 0,87$ e $\bar{R}^2 = 0,87$.

Exemplo 2 (modelo reduzido): desconsiderando o sexo

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

$$j = 1, \dots, 26(\text{criança}).$$

- $\xi_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- x_{ij} : IPB pré-escovação da criança j utilizando a escova do tipo i .
- Y_{ij} : IPB pós-escovação da criança j utilizando a escova do tipo i .
- β_{1i} : diminuição (se $\beta_{1i} \in (0, 1)$) ou aumento (se $\beta_{1i} > 1$), no IPB quando se usa a escova do tipo i .

Gráficos de resíduos

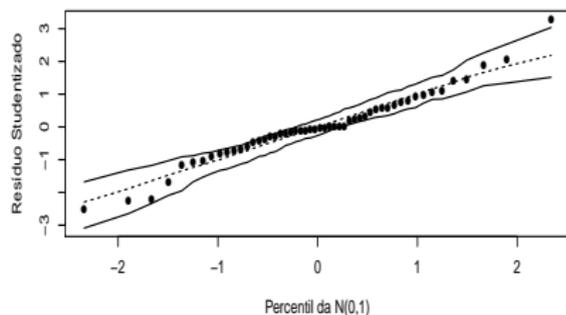
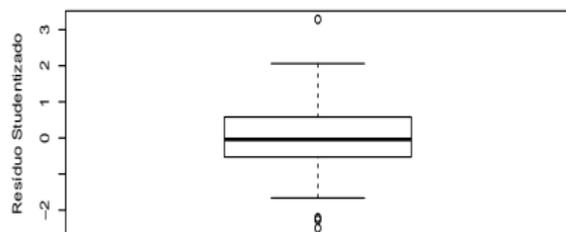
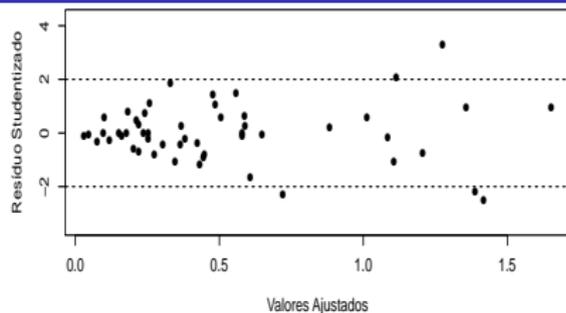
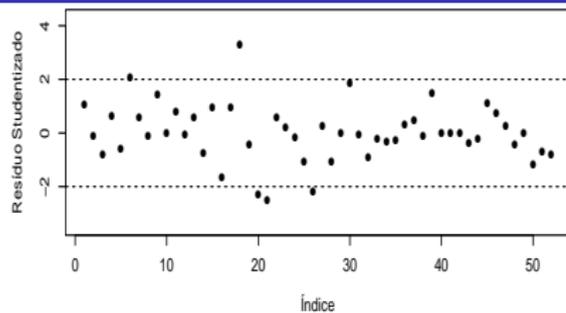
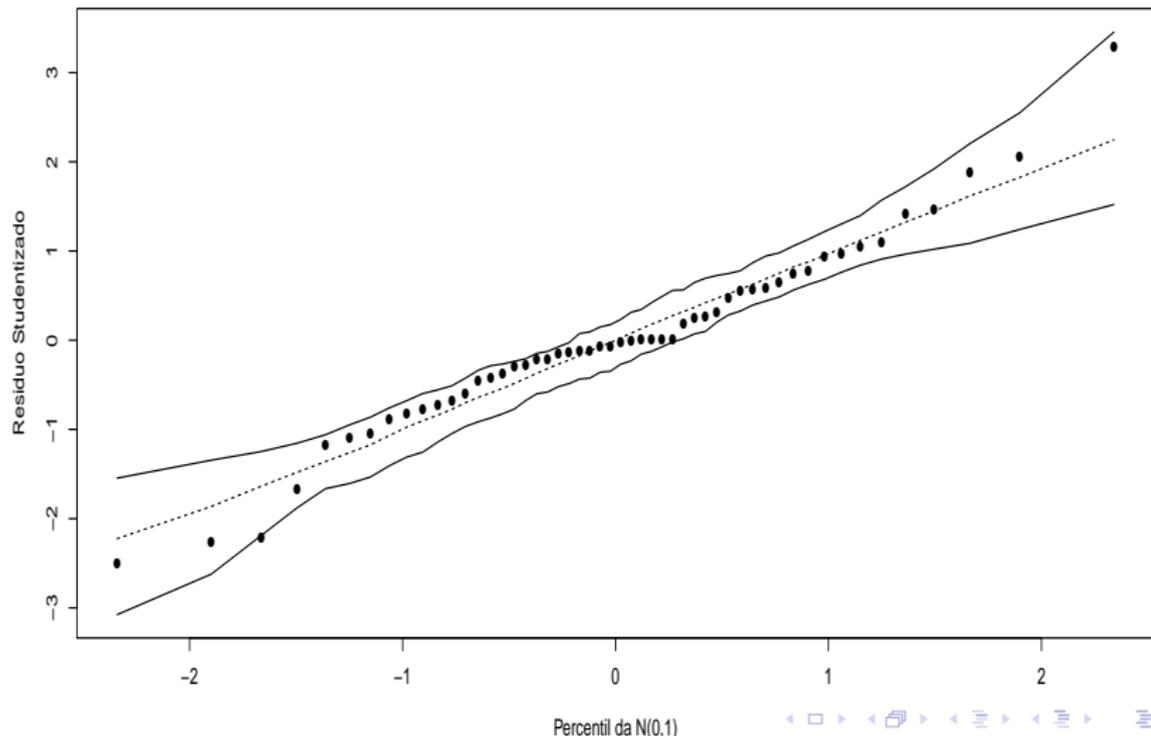


Gráfico de envelopes para os resíduos



Comentários

- Aparentemente os resíduos apresentam uma distribuição simétrica com caudas um pouco mais pesadas do que as da normal.
- Presença de heterocedasticidade nos resíduos.
- Ausência de correlação entre os resíduos.
- $R^2 = 0,87$ e $\bar{R}^2 = 0,87$.

Exemplo 3 (efeito do fósforo na produção de milho)

Modelo linear 1: reta

$$Y_i = \beta_0 + \beta_1 x_i + \xi_i, i = 1, 2, \dots, 20$$

- x_i : quantidade de fósforo ministrada a i -ésima parcela.
- β_0 : valor esperado (média) da produção de milho quando a quantidade de fósforo aplicada é igual à 0.
- β_1 : incremento no valor esperado da produção de milho quando a quantidade de fósforo aplicada aumenta em uma unidade.
- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.

Gráficos de resíduos

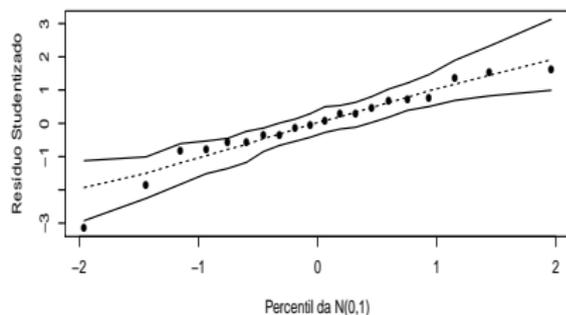
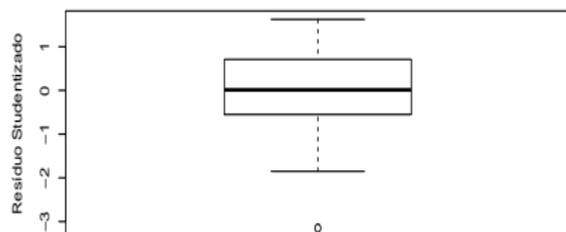
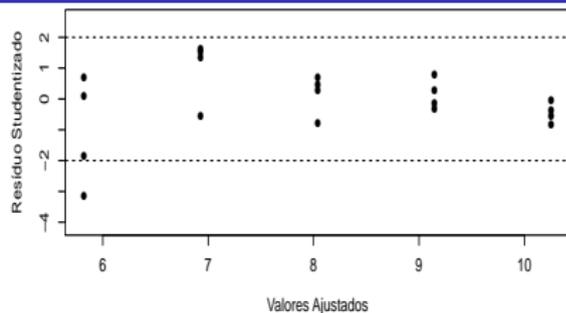
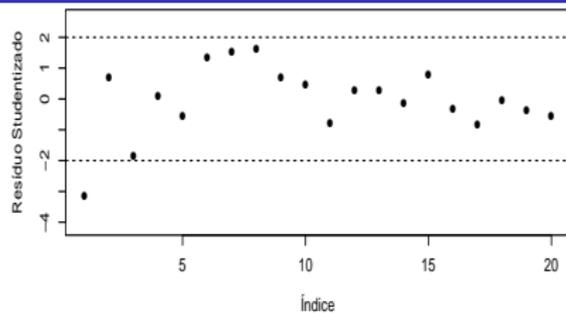
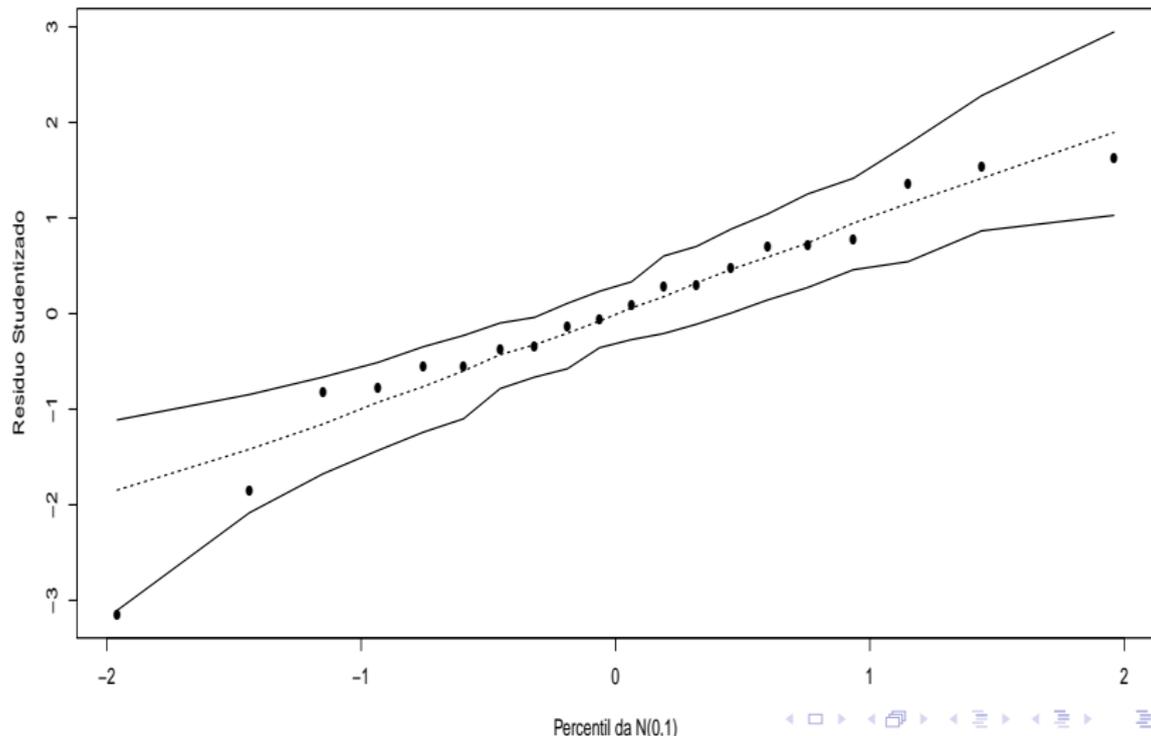


Gráfico de envelopes para os resíduos



Comentários

- Aparentemente os resíduos apresentam uma distribuição próxima da normal padrão.
- Presença de heterocedasticidade nos resíduos.
- Ausência de correlação entre os resíduos.
- O ponto que aparece destacado é devido ao fato de que o modelo linear não capta bem a relação entre o nível de fósforo e produção de milho.
- $R^2 = 0,56$ e $\bar{R}^2 = 0,54$.

Modelo linear 2: parábola

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \xi_i, i = 1, 2, \dots, 20$$

- x_i : quantidade de fósforo ministrada a i -ésima parcela.
- β_0 : valor esperado (média) da produção de milho quando a quantidade de fósforo aplicada é igual à 0.
- A interpretação isolada dos parâmetros β_1 e β_2 é complicada mas, podemos dizer que $\frac{-\beta_1}{2\beta_2}$ é o máximo (ou mínimo) do valor esperado da produção de milho.
- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.

Gráficos de resíduos

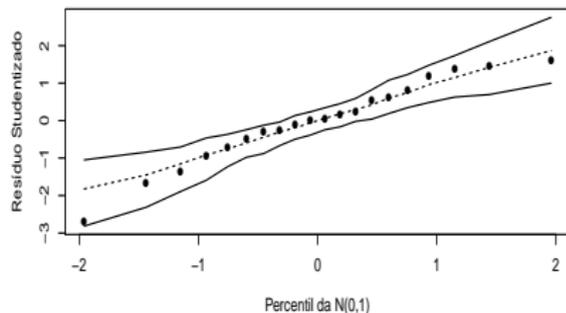
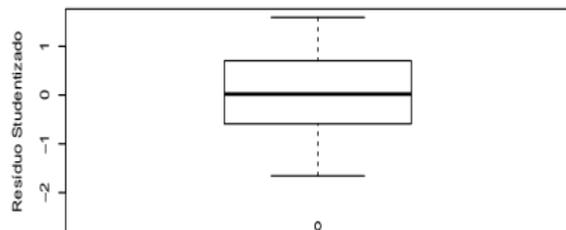
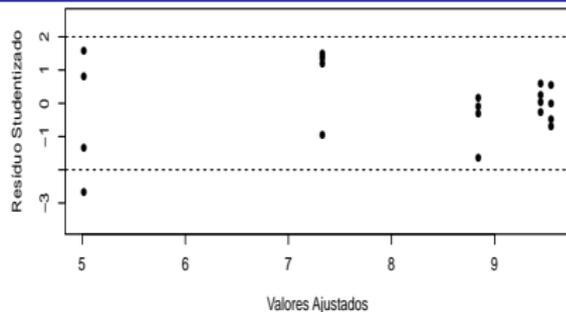
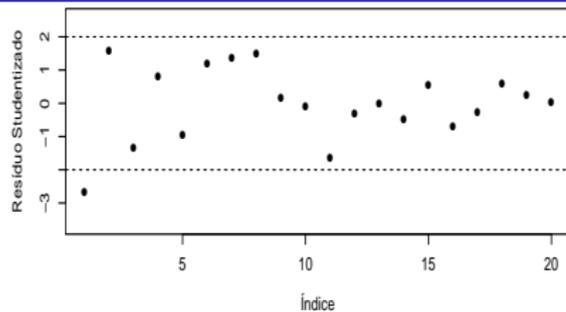
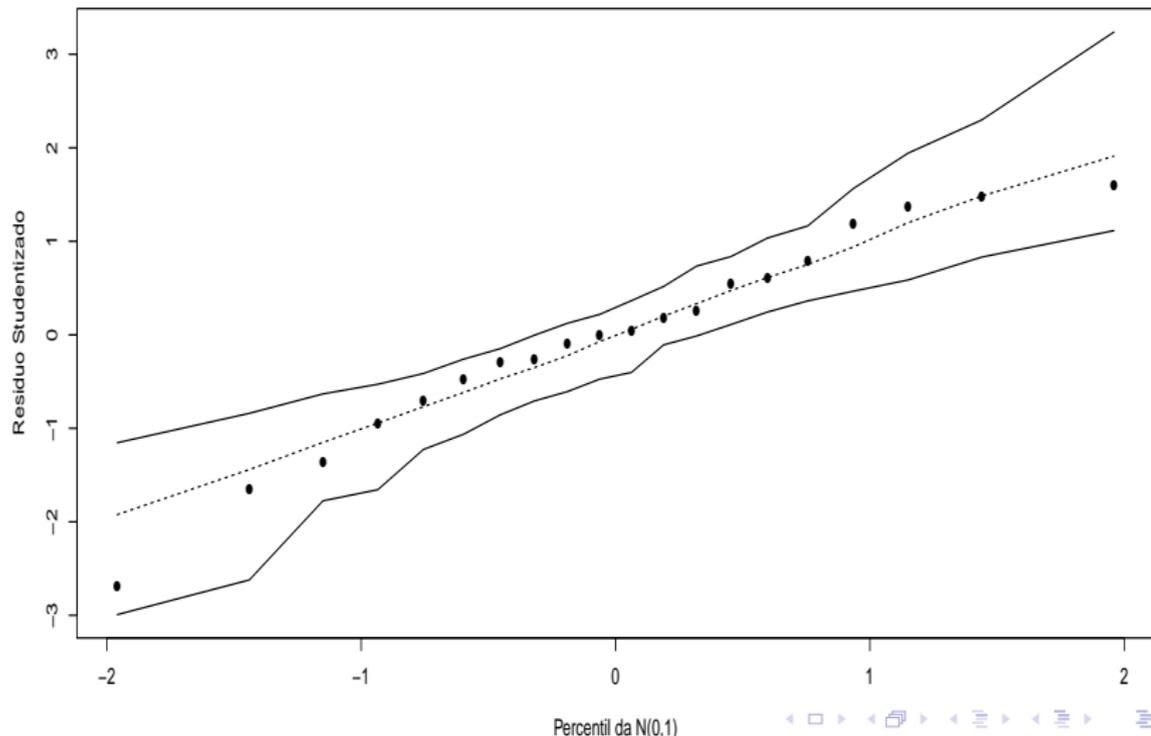


Gráfico de envelopes para os resíduos



Comentários

- Aparentemente os resíduos apresentam uma distribuição próxima da normal padrão.
- Presença de heterocedasticidade nos resíduos.
- Ausência de correlação entre os resíduos.
- $R^2 = 0,67$ e $\bar{R}^2 = 0,63$.

Exemplo 4: consumo de combustível

Modelo 2

$$Y_i = \beta_0 + \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + \beta_3(x_{3i} - \bar{x}_3) + \beta_4(x_{4i} - \bar{x}_4) + \xi_i, \quad (1)$$

$$i = 1, \dots, 48, \bar{x}_j = \frac{1}{48} \sum_{i=1}^{48} x_{ij}, j = 1, 2, 3, 4$$

- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- β_0 : consumo esperado para estados com valor de cada covariável igual à sua respectiva média.
- β_j : incremento (positivo ou negativo) no consumo esperado para o aumento em uma unidade da variável $j, j = 1, 2, 3, 4$, mantendo-se as outras fixas.

Gráficos de resíduos

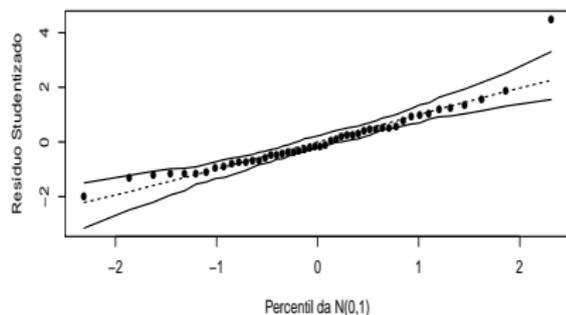
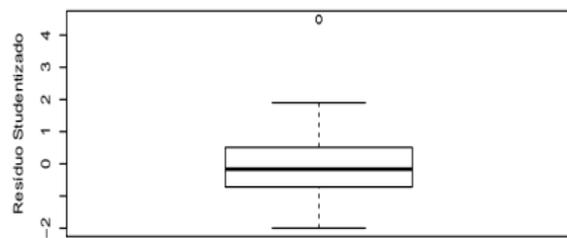
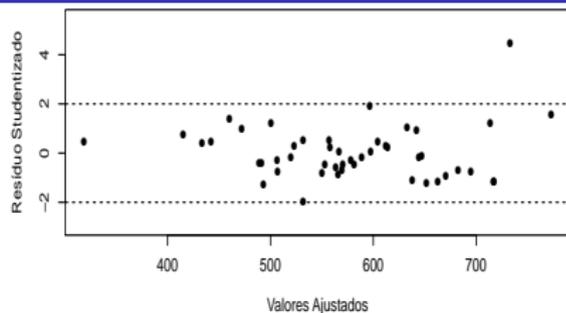
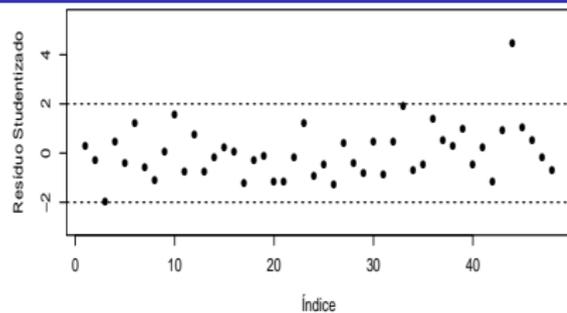
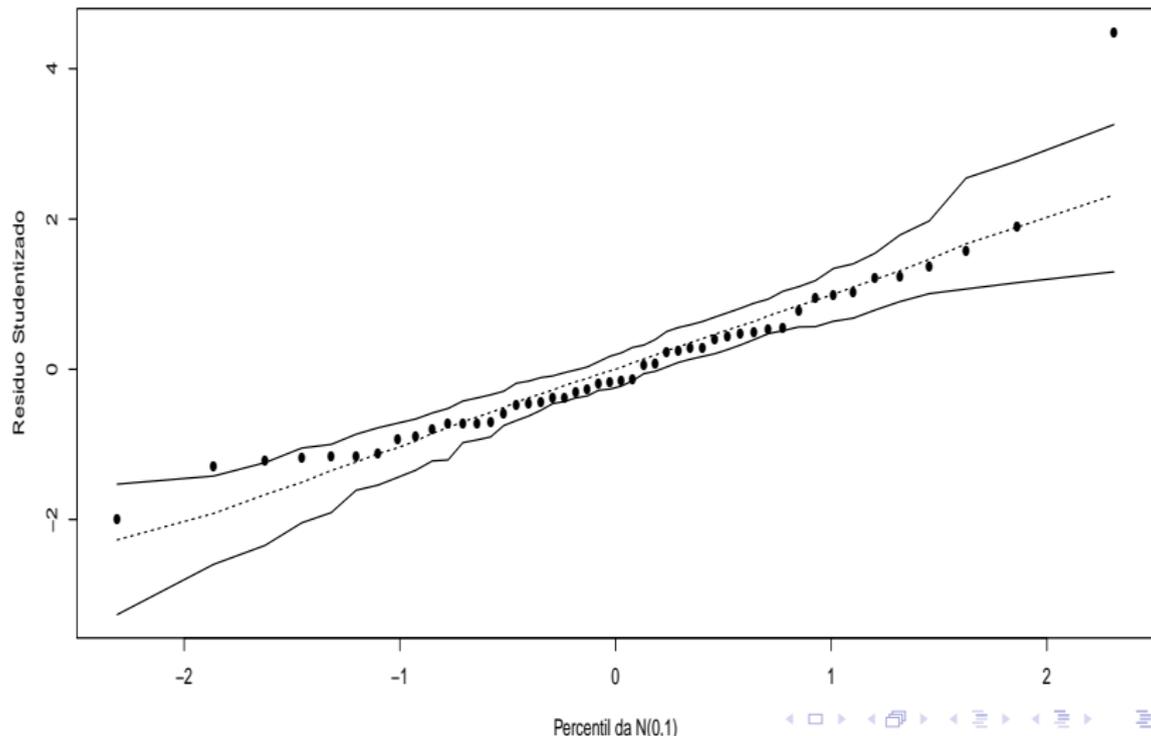


Gráfico de envelopes para os resíduos



Comentários

- Aparentemente os resíduos apresentam uma distribuição “levemente” assimétrica positiva.
- Presença de heterocedasticidade nos resíduos.
- Ausência de correlação entre os resíduos.
- $R^2 = 0,67$ e $\bar{R}^2 = 0,63$.

Exemplo 4: consumo de combustível

Modelo 2 (reduzido)

$$Y_i = \beta_0 + \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + \beta_3(x_{3i} - \bar{x}_3) + \xi_i, \quad (2)$$

$$i = 1, \dots, 48, \bar{x}_j = \frac{1}{48} \sum_{i=1}^{48} x_{ij}, j = 1, 2, 3, 4$$

- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- β_0 : consumo esperado para estados com valor de cada covariável igual à sua respectiva média.
- β_j : incremento (positivo ou negativo) no consumo esperado para o aumento em uma unidade da variável $j, j = 1, 2, 3, 4$, mantendo-se as outras fixas.

Gráficos de resíduos

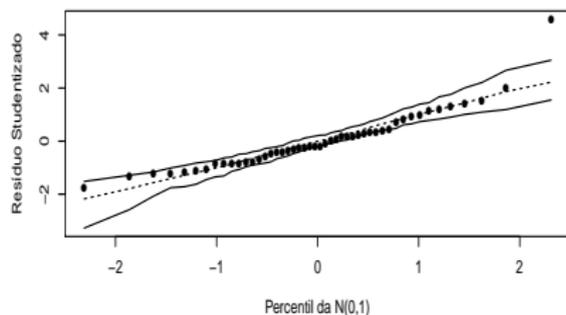
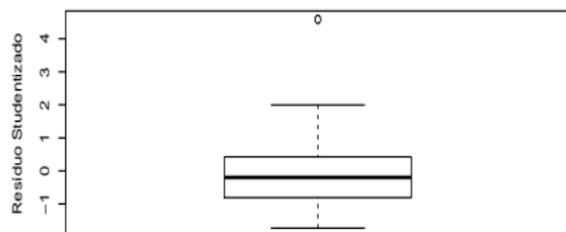
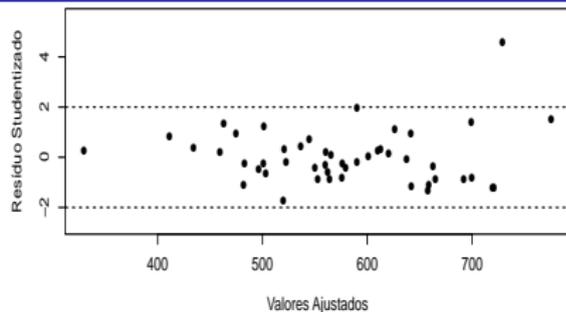
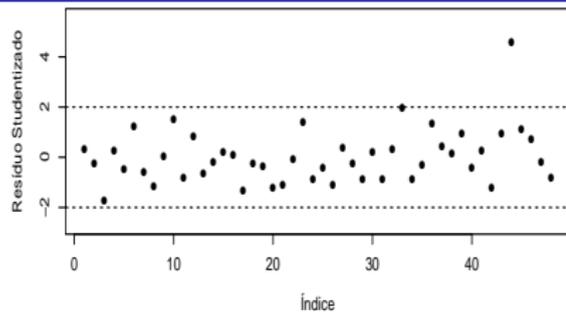
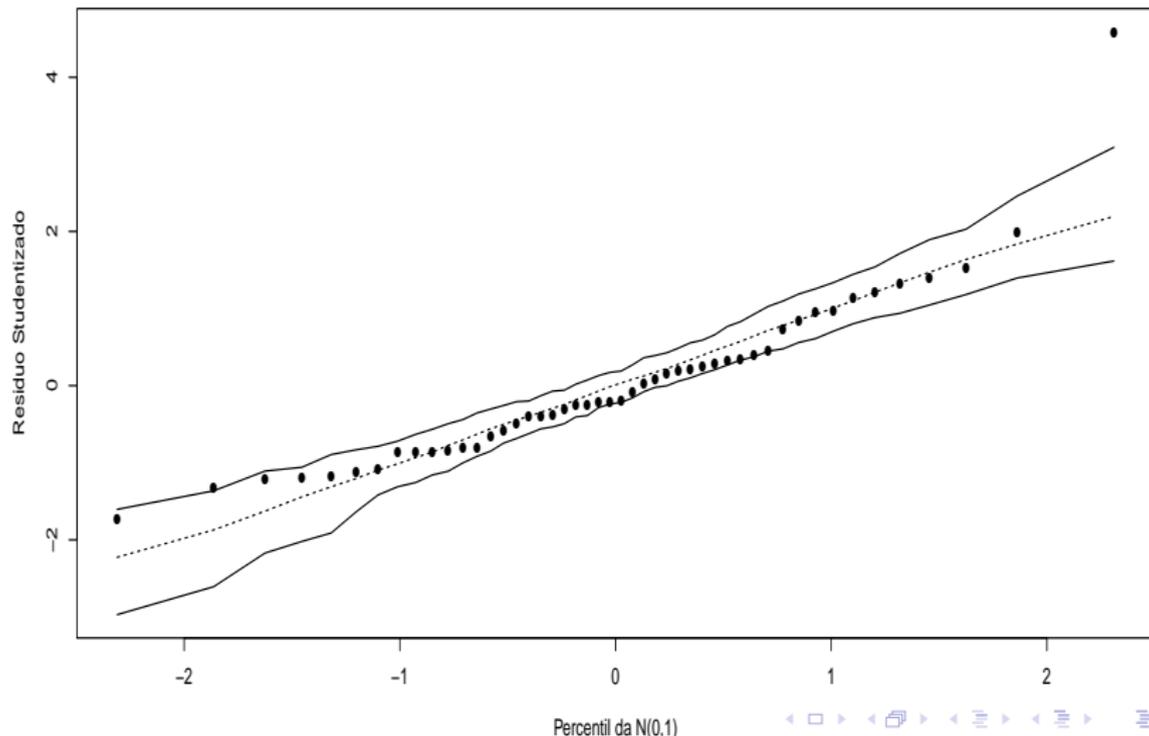


Gráfico de envelopes para os resíduos



Comentários

- Aparentemente os resíduos apresentam uma distribuição “levemente” assimétrica positiva.
- Presença de heterocedasticidade nos resíduos.
- Ausência de correlação entre os resíduos.
- $R^2 = 0,67$ e $\bar{R}^2 = 0,63$.

Exemplo 5: Modelo (casela de referência)

$$Y_{ij} = \mu + \alpha_i + \xi_{ij},$$

$i = 1, 2, \dots, 5$ (grupos); $j = 1, \dots, 5$ (unidades experimentais)

- $\xi_{ij} \stackrel{ind.}{\sim} N(0, \sigma^2)$. Parte sistemática: $\mu_i = \mu + \alpha_i$, é a média populacional relacionada ao i -ésimo fator, $\alpha_1 = 0$ (restrição de identificabilidade) .
- μ : é a média populacional do grupo de referência, $\mu_1 = \mu$.
- $\alpha_i = \mu_i - \mu_1, i = 2, \dots, 5$, é o incremento (positivo ou negativo) entre a média do grupo i e a média do grupo de referência.
- Grupos : grupo 1(E50), grupo 2(E70), grupo 3(EAW), grupo 4(M1M), grupo 5(MAW).

Gráficos de resíduos

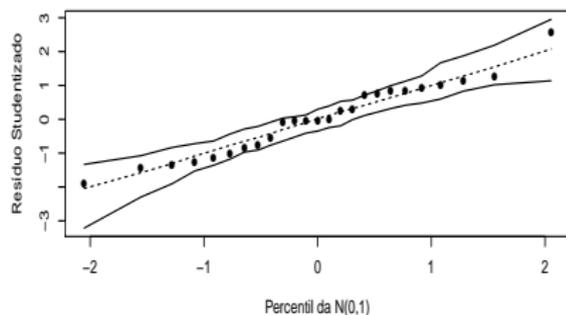
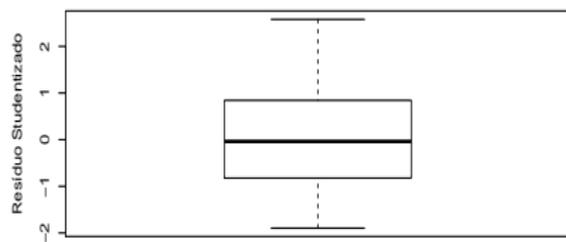
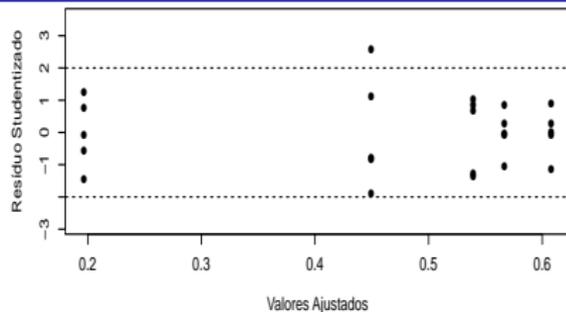
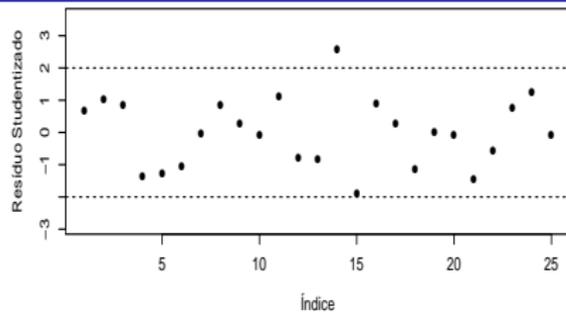
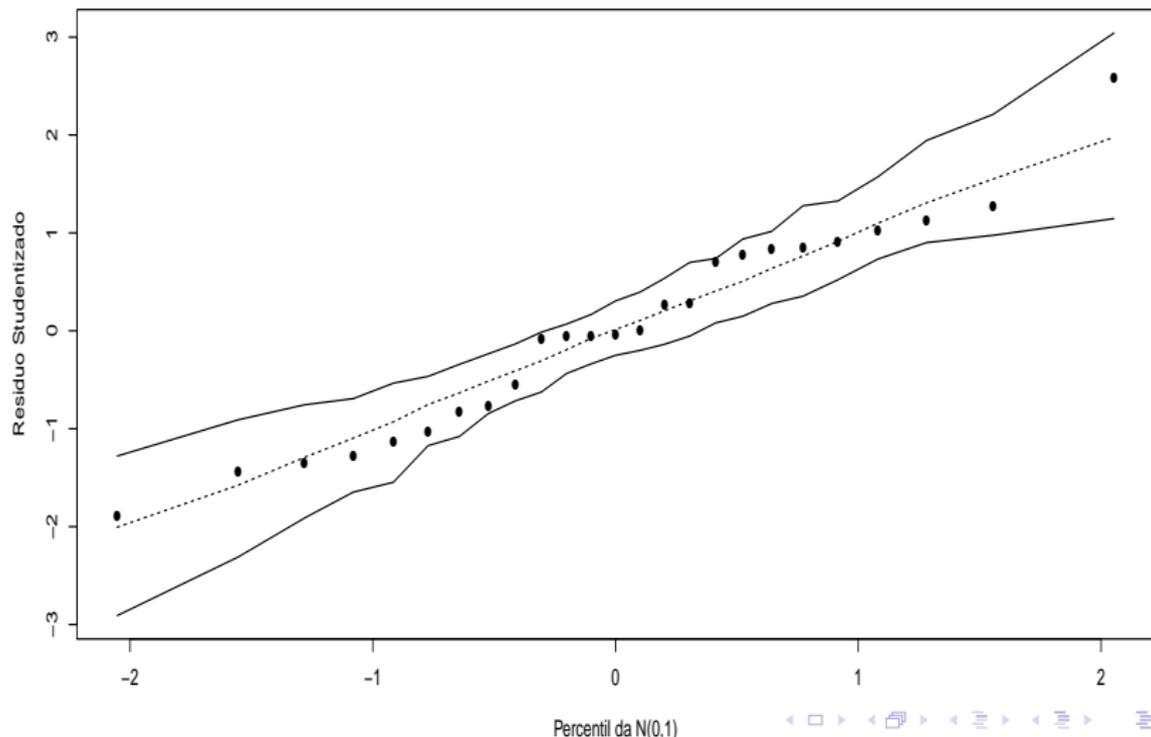


Gráfico de envelopes para os resíduos



Comentários

- Aparentemente os resíduos não apresentam uma distribuição normal padrão.
- Presença de heterocedasticidade nos resíduos.
- Ausência de correlação entre os resíduos.
- $R^2 = 0,97$ e $\bar{R}^2 = 0,97$.

Exemplo 5: Modelo reduzido (casela de referência)

$$Y_{ij} = \mu + \alpha_i + \xi_{ij},$$

$i = 1, 2, 3, 4$ (grupos); $j = 1, \dots, 4$ (unidades experimentais)

- Parte sistemática: $\mu_i = \mu + \alpha_i$, é a média populacional relacionada ao i -ésimo fator, $\alpha_1 = 0$ (restrição de identificabilidade).
- μ : é a média populacional do grupo de referência, $\mu_1 = \mu$.
- $\alpha_i = \mu_i - \mu_1$, $i = 2, 3, 4$, é o incremento (positivo ou negativo) entre a média do grupo i e a média do grupo de referência.
- Grupos : grupo 1(E50/EAW), grupo 2(E70), grupo 3(M1M), grupo 4(MAW).

Gráficos de resíduos

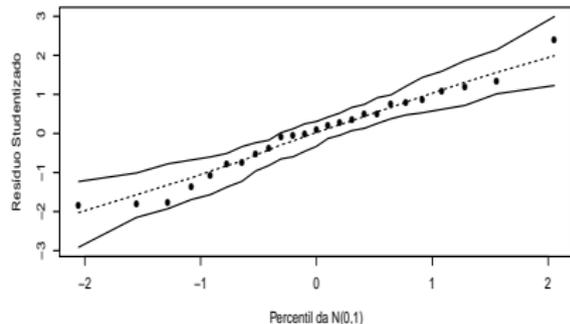
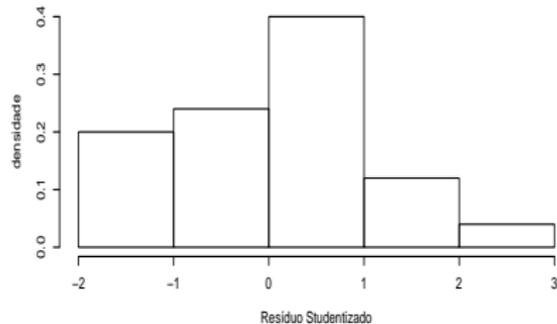
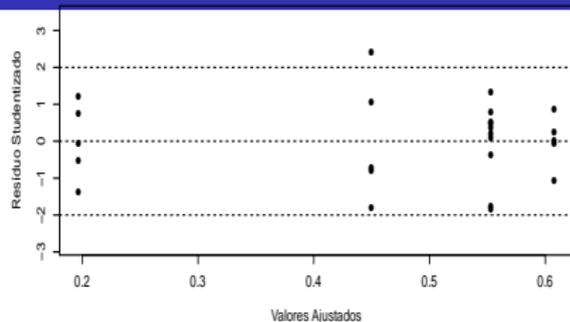
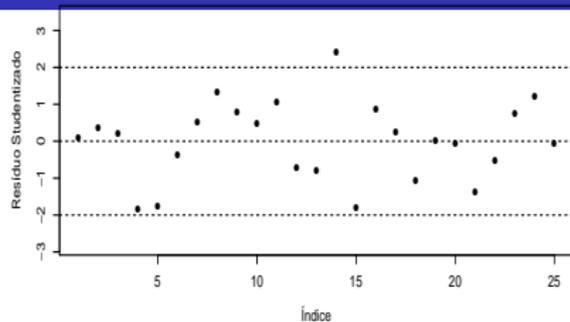
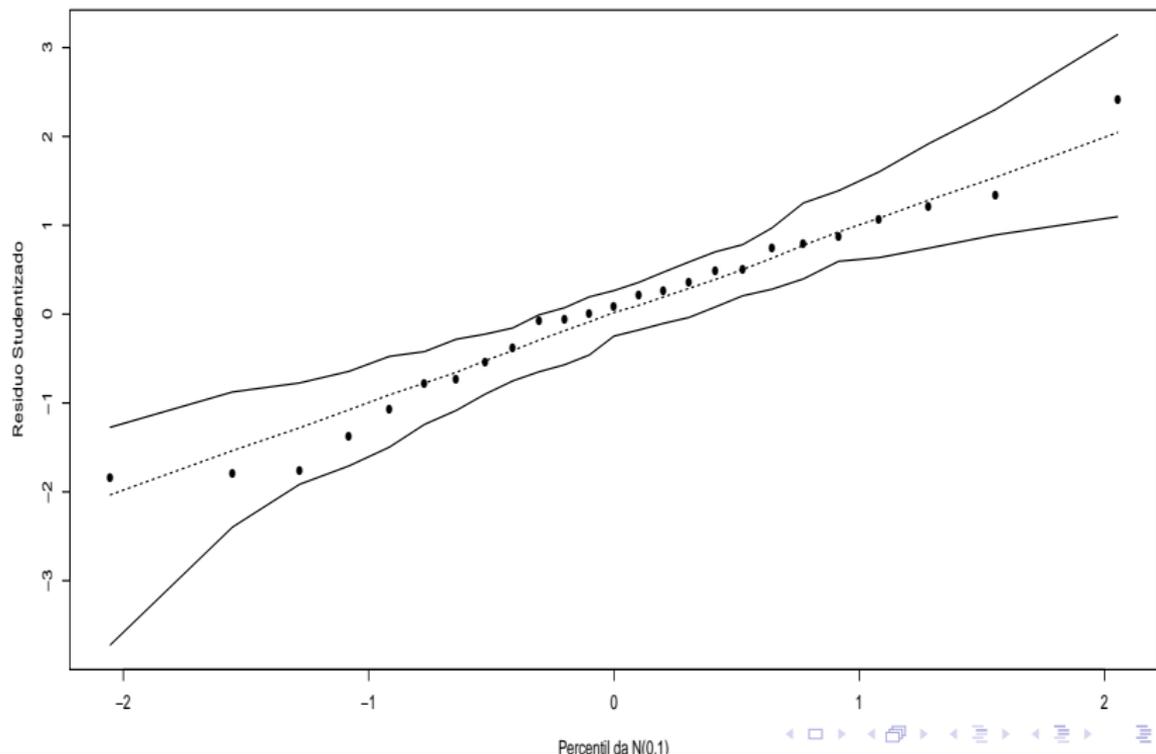


Gráfico de envelopes para os resíduos



Comentários

- Aparentemente os resíduos não apresentam uma distribuição normal padrão.
- Presença de heterocedasticidade nos resíduos.
- Ausência de correlação entre os resíduos.
- $R^2 = 0,97$ e $\bar{R}^2 = 0,97$.