

Questão 3

Introdução

Um estudo da Faculdade de Medicina da USP sobre esforço cardiopulmonar em pacientes com insuficiência cardíaca tinha como objetivo comparar diferentes etiologias cardíacas quanto às respostas respiratórias e metabólicas. As etiologias cardíacas são: CH - chagásticos; ID - idiopáticos; IS = isquêmicos; e C - controle. O objetivo do nosso modelo é explicar a variação do consumo de oxigênio no limiar anaeróbico (ml/(kg.min)) em função da carga usada na esteira ergométrica.

Temos para a nossa análise dados de 118 pacientes que foram submetidos ao teste na esteira. Para tal análise será usada a metodologia dos modelos lineares homocedásticos, metodologias de verificação da qualidade do ajuste e comparação de modelos apropriado, veja Azevedo (2019). Todas as análises serão feitas com auxílio computacional do R.

Análise descritiva

Pelas figuras 1 e 2 observamos que a carga se relaciona com positivamente com o consumo e que uma reta parece modelar bem a relação. Olhando as etiologias individualmente vemos que individualmente uma reta para cada caso, quatro retas, parece também fazer sentido. Porém na figura 2 sugere que talvez a relação entre o consumo e carga sejam as mesmas para todos os casos, ou seja apenas uma reta, seja uma suposição válida.

Análise Inferencial

Primeiramente ajustaremos um modelo completo de regressão linear homocedástica, que considera a carga, a etiologia e a interação entre elas. O modelo será ajustado segundo o método dos mínimos quadrados ordinários e a significância de cada parâmetro será testada a partir de estatísticas, veja Azevedo(2019).

Temos que Y_i é o consumo da i -ésima observação, o modelo é

$$Y_i = \beta_{0j} + \beta_1 x_i + \beta_{2j} x_i + \varepsilon_i$$

para $i = (1, 2, \dots, 118)$ e $j = (1, 2, 3, 4)$. Onde β_{0j} é o efeito devido a etiologia no consumo, com casela de referência em β_{01} , $\beta_1 x_i$ é o efeito da carga no consumo e $\beta_{2j} x_i$ é o efeito de interação da carga com etiologia no consumo, com casela de referência em β_{21} . Com a parte aleatória sendo ε_i .

Esse modelo é válido sobre as seguintes suposições: (i) $\varepsilon_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$; (ii) as observações são independentes; (iii) e a variância é constante.

O modelo foi segundo o método dos mínimos quadrados ordinários. O resultado pode ser visto na tabela 1, os valores para o efeito da etiologia não foram significativos, o que confirma a nossa suspeita que talvez uma reta para todos os dados fosse mais apropriada que quatro retas. Essa análise do ajuste também está de acordo com o resultado da análise de variância, tabela 2, que também acusa a não significância dos efeitos da etiologia. As retas preditas podem ser vistas na figura 3 e realmente são bem similares.

A análise de resíduos se encontra na figura 4, reparamos nela que a suposições de homoscedasticidade e de normalidade dos dados não são válidas. No gráfico de resíduos por valor ajustado vemos uma variância crescer, no histograma e no gráfico quantil-quantil vemos uma assimetria dos resíduos. A análise de pontos influentes nos dá alguns candidatos, a observação 32 e 40. A retirada desses pontos altera bastante a estimativa e altera nossa inferência e conclusões. A retirada do ponto 32 deixa o efeito etiologia significativo para um nível e a retirada do ponto 40 deixa o efeito da etiologia menos significativo.

Como vimos que o efeito da etiologia não é significativo vamos ajustar um modelo reduzido:

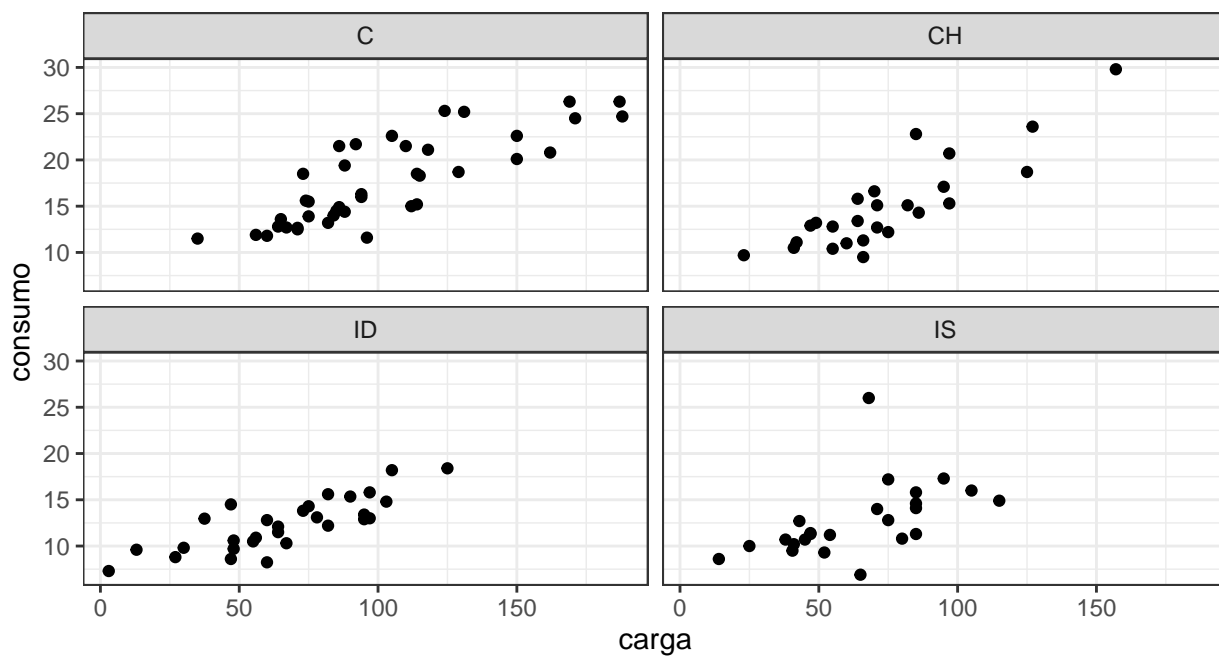


Figura 1: Gráfico de dispersão do consumo pela carga, separado por etiologias.

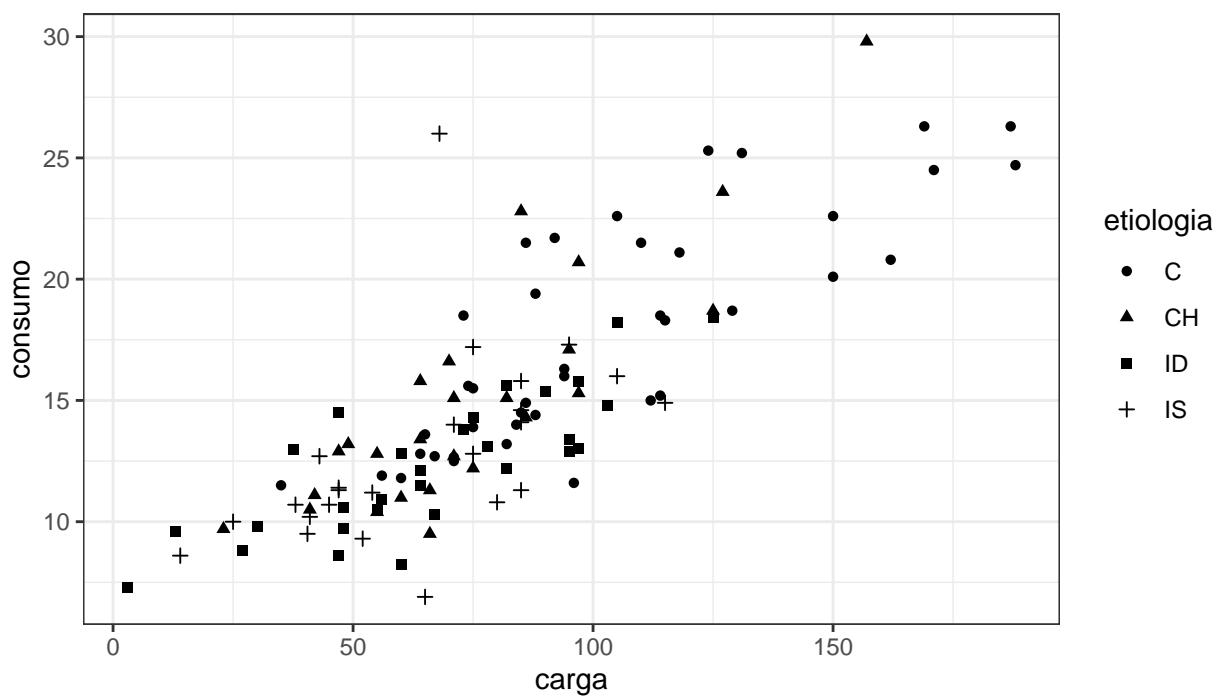


Figura 2: Gráfico de dispersão geral.

Tabela 1: Estimativa para os parâmetros do modelo completo.

Termo	Estimativa	Erro Padrão	IC (95%)	Estatística t	p-valor
β_{01}	7,169	1,234	[4,751 ; 9,587]	5,811	< 0.001
β_1	0,102	0,011	[0,08 ; 0,124]	9,053	< 0.001
β_{02}	-2,749	1,902	[-6,477 ; 0,979]	-1,445	0,151
β_{03}	-0,029	1,753	[-3,464 ; 3,406]	-0,017	0,987
β_{04}	0,480	1,928	[-3,299 ; 4,259]	0,249	0,804
β_{22}	0,039	0,021	[-0,002 ; 0,081]	1,856	0,066
β_{23}	-0,023	0,021	[-0,064 ; 0,017]	-1,128	0,262
β_{24}	-0,022	0,024	[-0,069 ; 0,026]	-0,891	0,375

Tabela 2: Análise de convariância para o modelo completo

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrados Médios	Estatística	p-valor
β_1	1	1736,8158	1736,8158	249,2441	0,0000
β_0	3	53,8421	17,9474	2,5756	0,0575
β_2	3	53,1662	17,7221	2,5432	0,0599
Resíduos	110	766,5165	6,9683	NA	NA

$$Y_i = \beta_1 + \beta_1 x_i + \varepsilon_i$$

onde Y_i é o consumo da i -ésima observação, para $i = (1, 2, \dots, 118)$ e $j = (1, 2, 3, 4)$. E β_0 é o consumo quando não há carga, $\beta_1 x_i$ é o feito da carga no consumo e parte aleatória é ε_i . Esse modelo é válido sobre as seguintes suposições: (i) $\varepsilon_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$; (ii) as observações são independentes; (iii) e a variância é constante. O resultado do seu ajuste pode ser visto na tabela 3, todos os parâmetros são significativos e vemos que os seus valores não se alteraram muito do modelo completo. Na tabela 4 vemos o resultado da análise de variância que nos diz que há regressão. Observando a reta ajustada e o intervalo de predição, figura 6, o modelo parece se ajustar bem aos dados.

Observando a análise de resíduos na figura 7, vemos os mesmos problemas do modelo completo heterocedasticidade e a ausência de normalidade. Na figura 8 ainda temos candidatos a pontos influentes ou de alavanca, mas a retirada dos mesmos não afeta mais que 5% as nossas estimativas e não muda a conclusão.

Conclusão

Pelo resultado dos ajustes dos modelos, o que afeta o consumo de oxigênio aparenta ser apenas a carga e não a etiologia. Porém vimos que os modelos ajustados não são apropriados e quebraram várias suposições, por isso é provável que as inferências não estejam corretas. Deveriam ser feitos mais análises considerando-se um modelo mais apropriado, que comporte a heterocedasticidade e a não normalidade dos dados. Por exemplo, adequar a uma regressão com variáveis positivas, com o modelo gamma ou normal inversa. Também para o modelo completo detectamos pontos influentes, por isso deve-se conversar com o pesquisador a respeito dessas observações e de como tratá-las, para propor um modelo que considere melhor esses pontos.

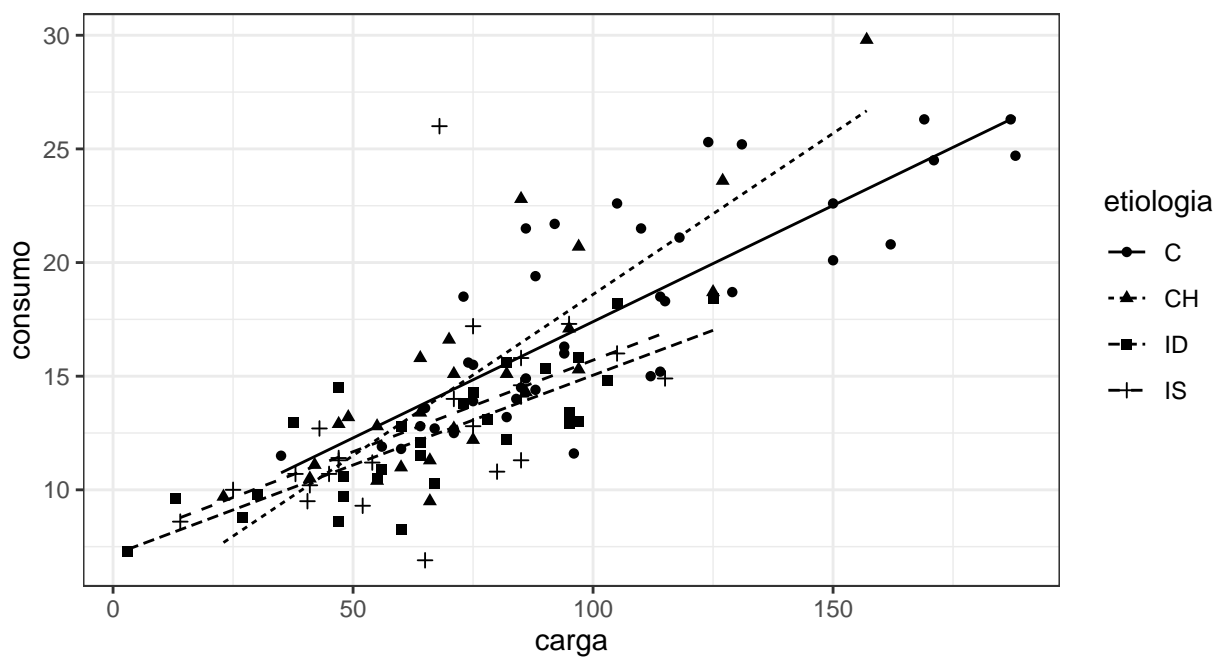


Figura 3: Valores preditos do modelo completo e os dados originais.

Tabela 3: Estimativa para os parâmetros do modelo reduzido.

Termo	Estimativa	Erro Padrão	IC (95%)	Estatística t	p-valor
β_0	6,140	0,625	[4,915 ; 7,365]	9,823	< 0.001
β_1	0,109	0,007	[0,095 ; 0,123]	15,187	< 0.001

Referências

- Azevedo, C. L. N (2019). Notas de aula sobre Análise de regressão, http://www.ime.unicamp.br/~cnaber/Material_ME613_1S_2019.htm
- Paula, G. A. (2013). Modelos de regressão com apoio computacional, versão pré-eliminar, https://www.ime.usp.br/~giapaula/texto_2013.pdf

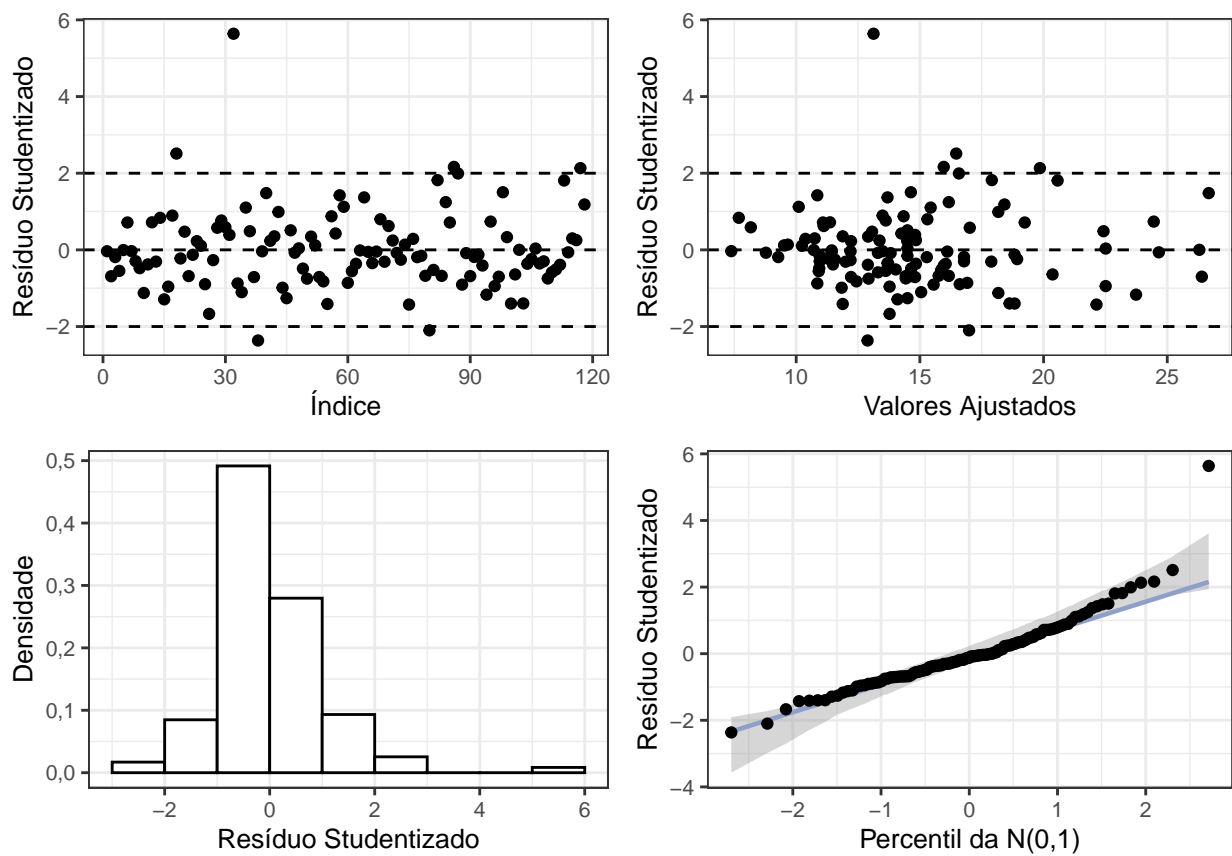


Figura 4: Diagnóstico do modelo completo

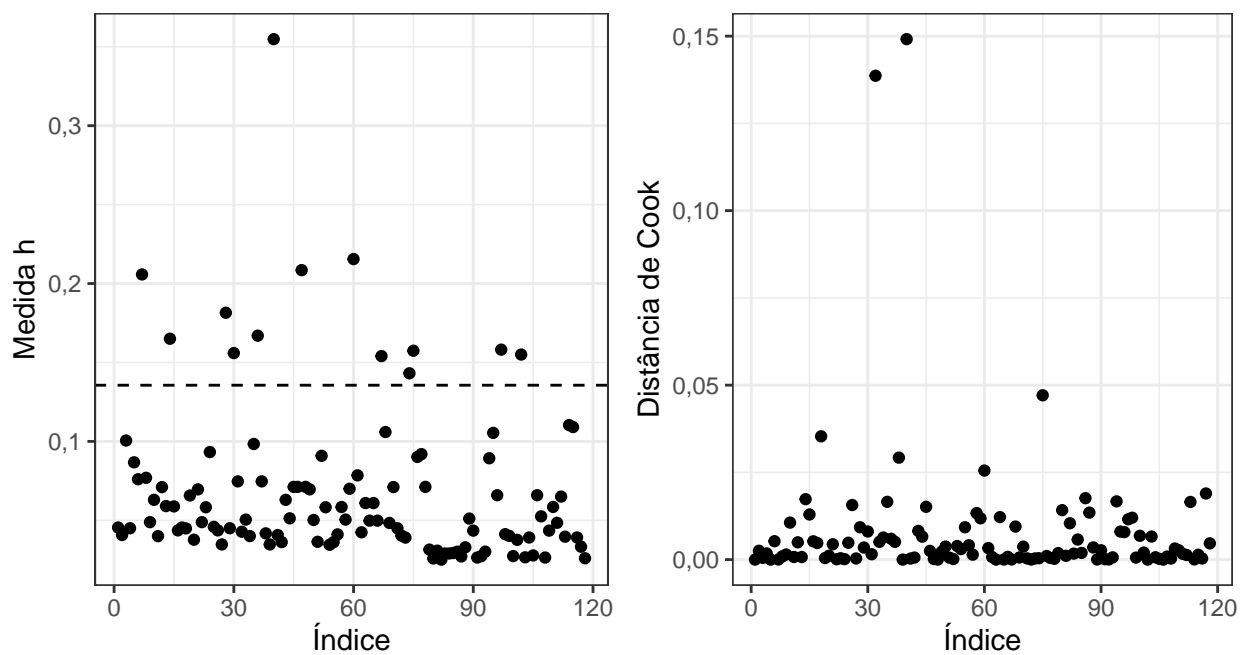


Figura 5: Análise de influência e alavanca para o modelo completo

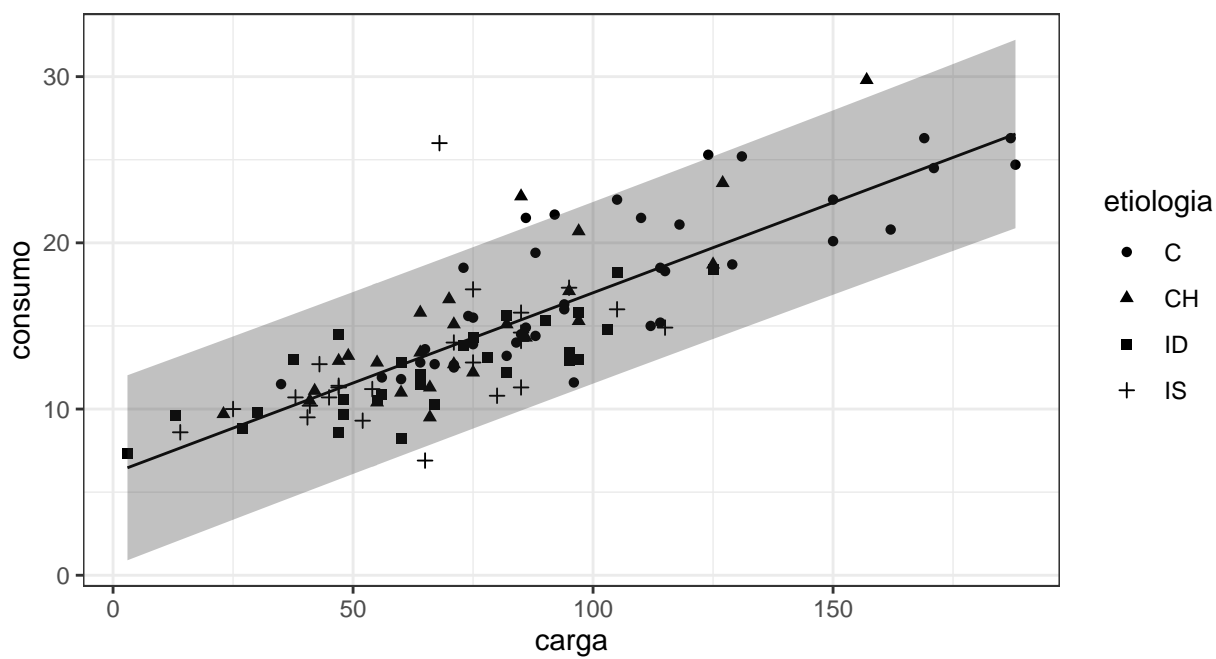


Figura 6: Valores preditos e intervalo de predição para o modelo reduzido

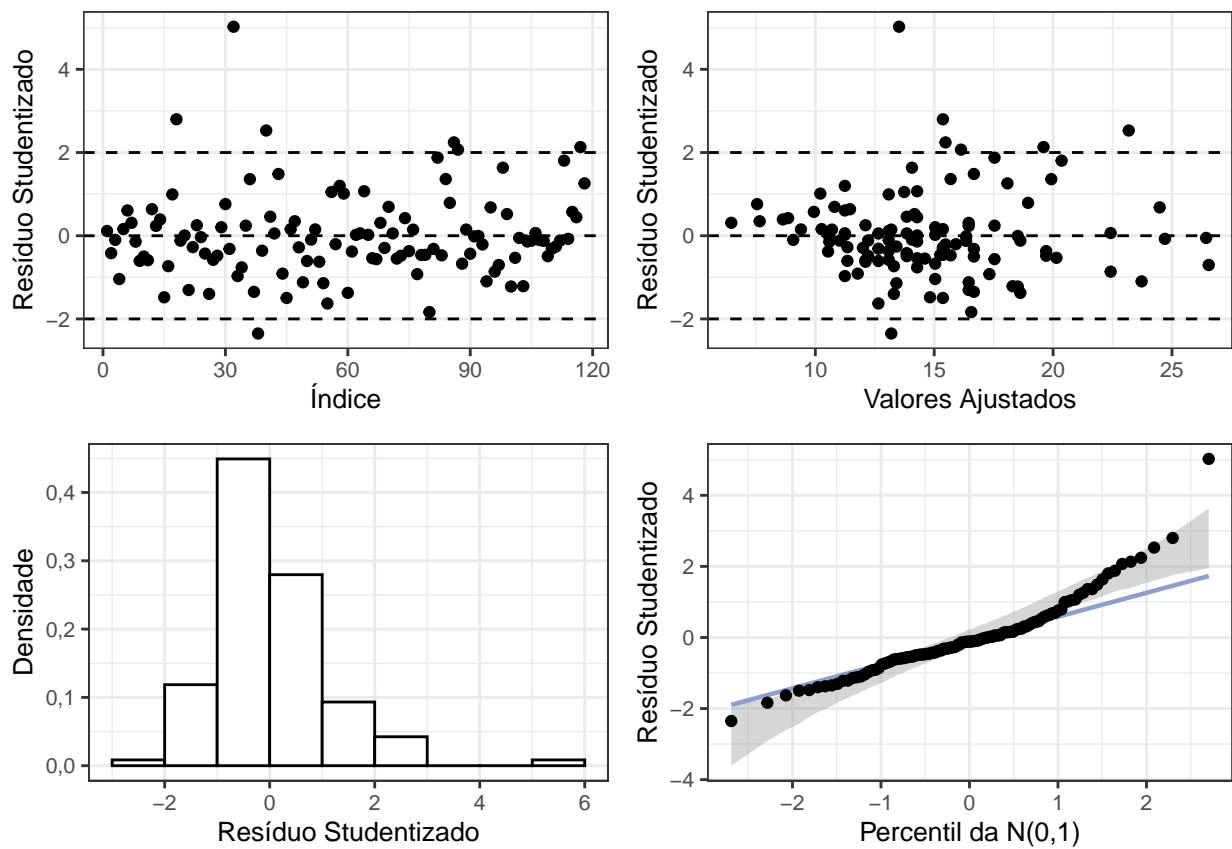


Figura 7: Diagnóstico do modelo reduzido.

Tabela 4: Análise de convariância para o modelo reduzido.

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrados Médios	Estatística	p-valor
β_1	1	1736,8158	1736,8158	230,641	0
Resíduos	116	873,5247	7,5304	NA	NA

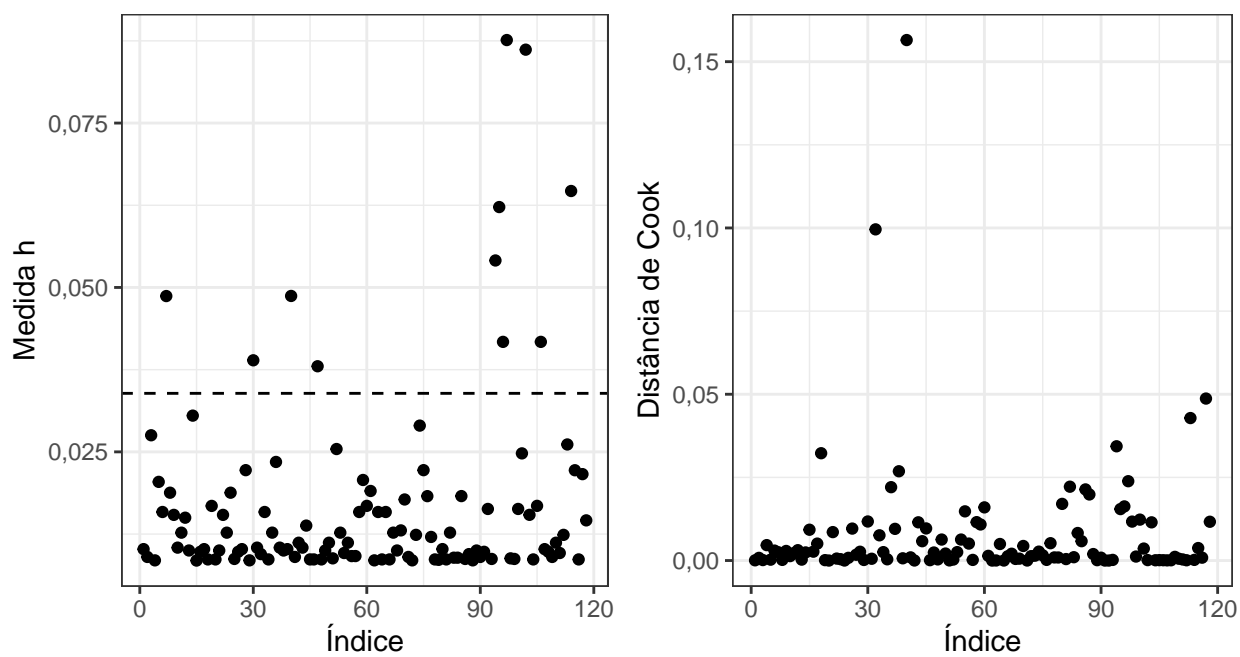


Figura 8: Análise de influência e alavanca para o modelo reduzido.