

Trabalho 3 - Modelos Lineares Generalizados

Vítor Pereira

1 Modelando o banco de dados

Modelaremos o banco de dados de um experimento para avaliar o desempenho de cinco tipos de turbinas de alta velocidade, levando em consideração 10 motores dos 5 tipos avaliados, analisando o tempo (em unidades de milhões de ciclos) até a perda da velocidade.

1.1 Utilizando a Distribuição Gamma

Começaremos com a Distribuição Gamma, que é utilizada para modelar valores de dados positivos que são assimétricos à direita e maiores que 0.

1.1.1 Primeiro Ajuste

Então começaremos a análise da Distribuição Gamma, considerando todos os tipos variáveis dummies e analisaremos sua significância:

Tabela 1: Primeiro Ajuste - Gamma

	Estimativa	Desvio padrão	Estatística t	P.valor
(Intercept)	2.370	0.144	16.422	<0.001*
tipo2	-0.570	0.204	-2.791	0.008*
tipo3	-0.214	0.204	-1.047	0.301
tipo4	-0.087	0.204	-0.428	0.67
tipo5	0.319	0.204	1.562	0.125

Notamos, que os tipos não são completamente significativos, assim realizaremos junções buscando que as variáveis dummies sejam significativas.

1.1.2 Segundo Ajuste

Iremos aglutinar os grupos 3 e 4 em um só, visto que foram os grupos que obtiveram maior p-valor na tabela anterior, assim temos:

Tabela 2: Segundo Ajuste - Gamma

	Estimativa	Desvio padrão	Estatística t	P.valor
(Intercept)	2.370	0.145	16.318	<0.001*
tipo2	-0.570	0.205	-2.773	0.008*
tipo4	-0.149	0.178	-0.835	0.408
tipo5	0.319	0.205	1.552	0.128

Percebe-se que ainda não obtivemos significância em todos os tipos.

1.1.3 Terceiro Ajuste

Agora iremos juntar os tipos 3 e 4 com o tipo 1, logo obtêm-se:

Tabela 3: Terceiro Ajuste - Gamma

	Estimativa	Desvio padrão	Estatística t	P.valor
(Intercept)	2.273	0.083	27.31	<0.001*
tipo2	-0.473	0.166	-2.84	0.007*
tipo5	0.415	0.166	2.49	0.016*

Desse modo conseguimos significância em todas as variáveis e ficamos com 3 grupos, sendo 1 aglomerados, os grupos são: Tipo 1, 3 e 4, Tipo 2 e Tipo 5.

1.2 Utilizando a Distribuição Normal Inversa

Agora utilizaremos a Distribuição Normal Inversa (NI), que também é utilizada para modelar valores de dados positivos e maiores que 0.

1.2.1 Primeiro ajuste

Considerando todos os tipos variáveis dummies, a significância fica:

Tabela 4: Primeiro Ajuste - NI - Canônica

	Estimativa	Desvio padrão	Estatística t	P.valor
(Intercept)	0.009	0.003	3.155	0.003*
tipo2	0.019	0.007	2.624	0.012*
tipo3	0.005	0.005	0.988	0.328
tipo4	0.002	0.004	0.397	0.693
tipo5	-0.004	0.003	-1.264	0.213

Tabela 5: Primeiro Ajuste - NI - log

	Estimativa	Desvio padrão	Estatística t	P.valor
(Intercept)	2.370	0.158	14.951	<0.001*
tipo2	-0.570	0.198	-2.872	0.006*
tipo3	-0.214	0.213	-1.003	0.321
tipo4	-0.087	0.219	-0.398	0.692
tipo5	0.319	0.244	1.305	0.199

Notamos, que os tipos não são completamente significativos, assim realizaremos agregações em ambos modelos, buscando que as variáveis dummies sejam significativas.

1.2.2 Segundo Ajuste

Iremos unir os grupos 3 e 4 em um só, visto que foram os grupos que obtiveram maior p-valor na tabela anterior, assim temos:

Tabela 6: Segundo Ajuste - NI - Canônica

	Estimativa	Desvio padrão	Estatística t	P.valor
(Intercept)	0.009	0.003	3.121	0.003*
tipo2	0.019	0.007	2.596	0.013*
tipo4	0.003	0.004	0.809	0.423
tipo5	-0.004	0.003	-1.250	0.218

Tabela 7: Segundo Ajuste - NI - log

	Estimativa	Desvio padrão	Estatística t	P.valor
(Intercept)	2.370	0.160	14.793	<0.001*
tipo2	-0.570	0.200	-2.841	0.007*
tipo4	-0.149	0.192	-0.775	0.442
tipo5	0.319	0.247	1.291	0.203

Nota-se que ainda não obtivemos significância em todos os tipos, em nenhuma das distribuições NI.

1.2.3 Terceiro Ajuste

Agora iremos juntar os tipos 3 e 4 com o tipo 1, logo obtêm-se:

Tabela 8: Terceiro Ajuste - NI - Canônica

	Estimativa	Desvio padrão	Estatística t	P.valor
(Intercept)	0.011	0.002	5.74	<0.001*
tipo2	0.017	0.007	2.47	0.017*
tipo5	-0.006	0.003	-2.37	0.022*

Tabela 9: Terceiro Ajuste - NI - log

	Estimativa	Desvio padrão	Estatística t	P.valor
(Intercept)	2.273	0.087	26.09	<0.001*
tipo2	-0.473	0.148	-3.21	0.002*
tipo5	0.415	0.205	2.02	0.049*

Desse modo conseguimos significância em todas as variáveis e ficamos com 3 grupos, sendo 1 aglomerados, os grupos são: Tipo 1, 3 e 4, Tipo 2 e Tipo 5, tanto na Normal Inversa com ligação canônica, quanto na com ligação log.

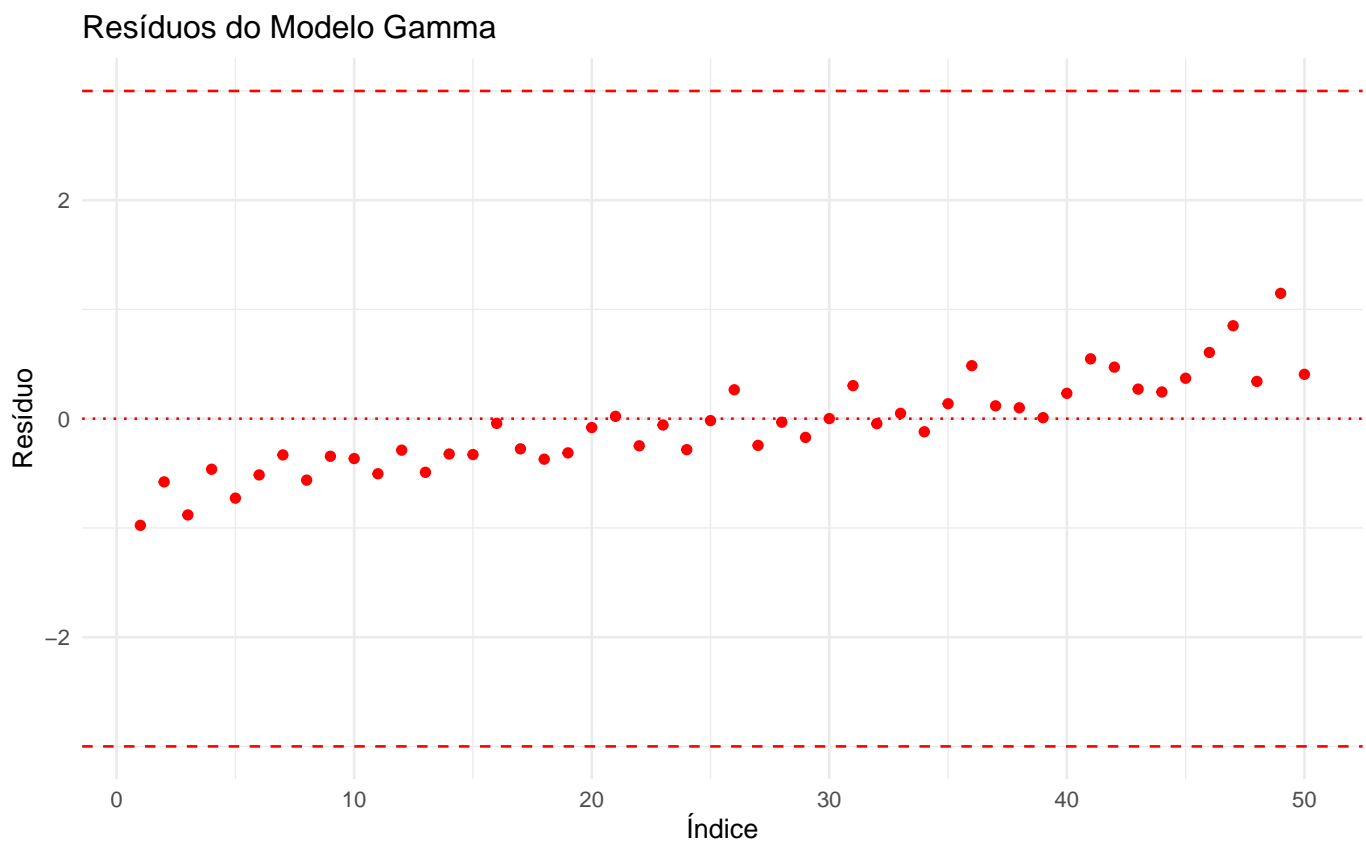
2 Análise de Influência

Nesta seção será realizada uma busca de observações atípicas no banco de dados, que assim possam estar influenciado a análise, também influenciado pelas junções de tipos realizados anteriormente, assim utilizaremos 5 análises para a verificação de pontos de influência: Análise de Resíduos Deviance, Envelope Simulado, Distância de Cook, Alavancagem e DFFits.

2.1 Ajuste com a Gamma

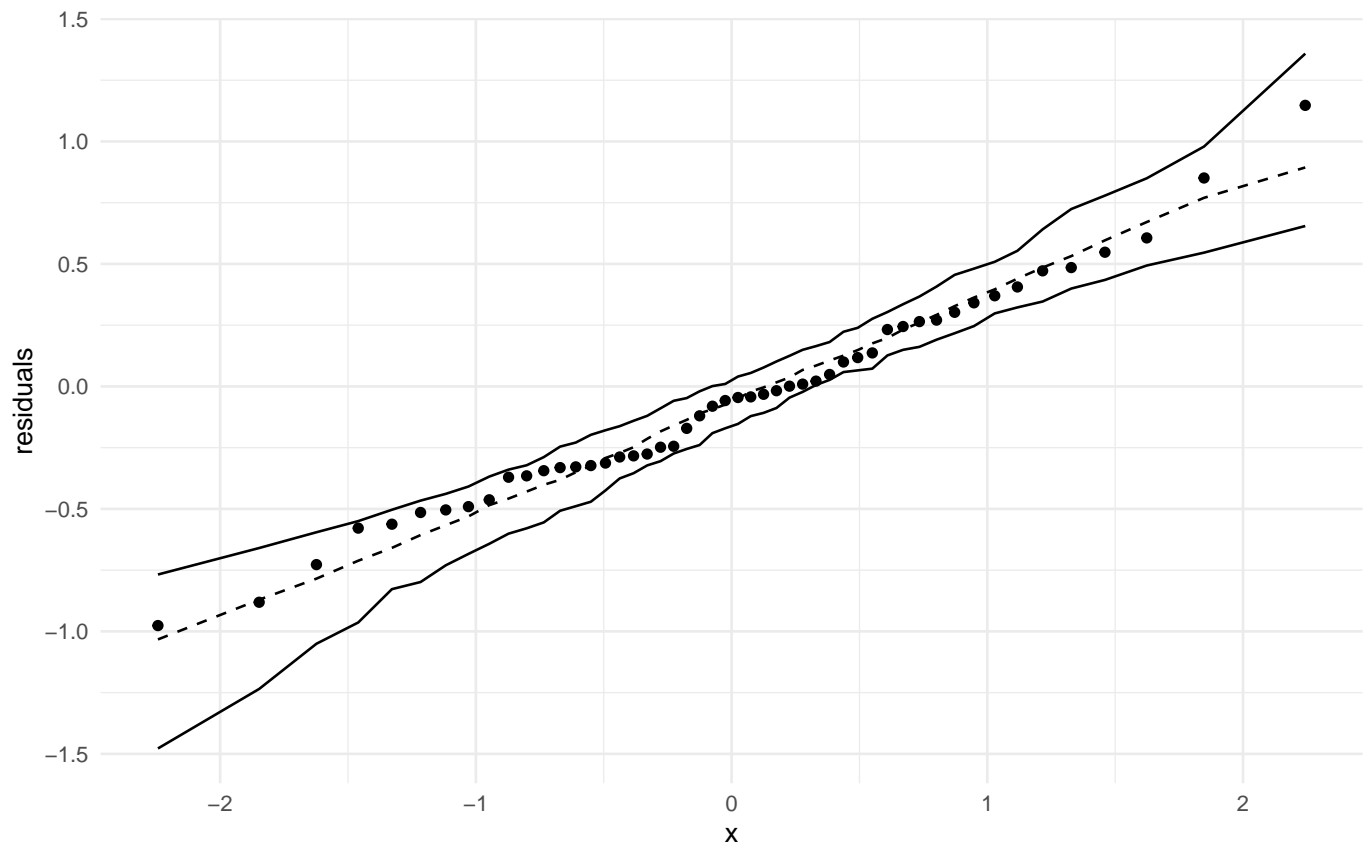
Começaremos a análise de influência com a distribuição Gamma.

2.1.1 Resíduos deviances vs índices



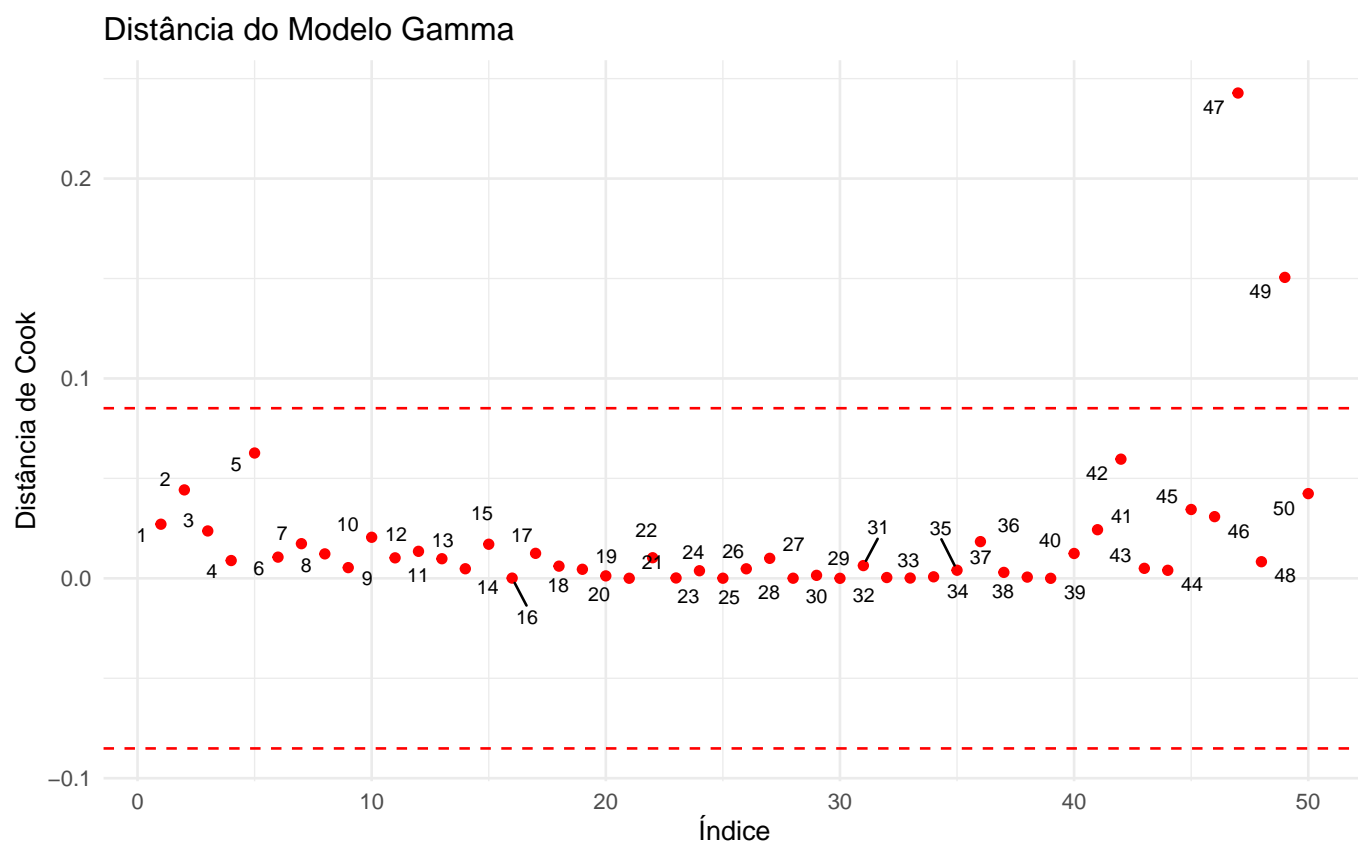
Não observa-se algum resíduo fora dos limites especificados, indicando que não exista pontos de influência.

2.1.2 Envelope Simulado



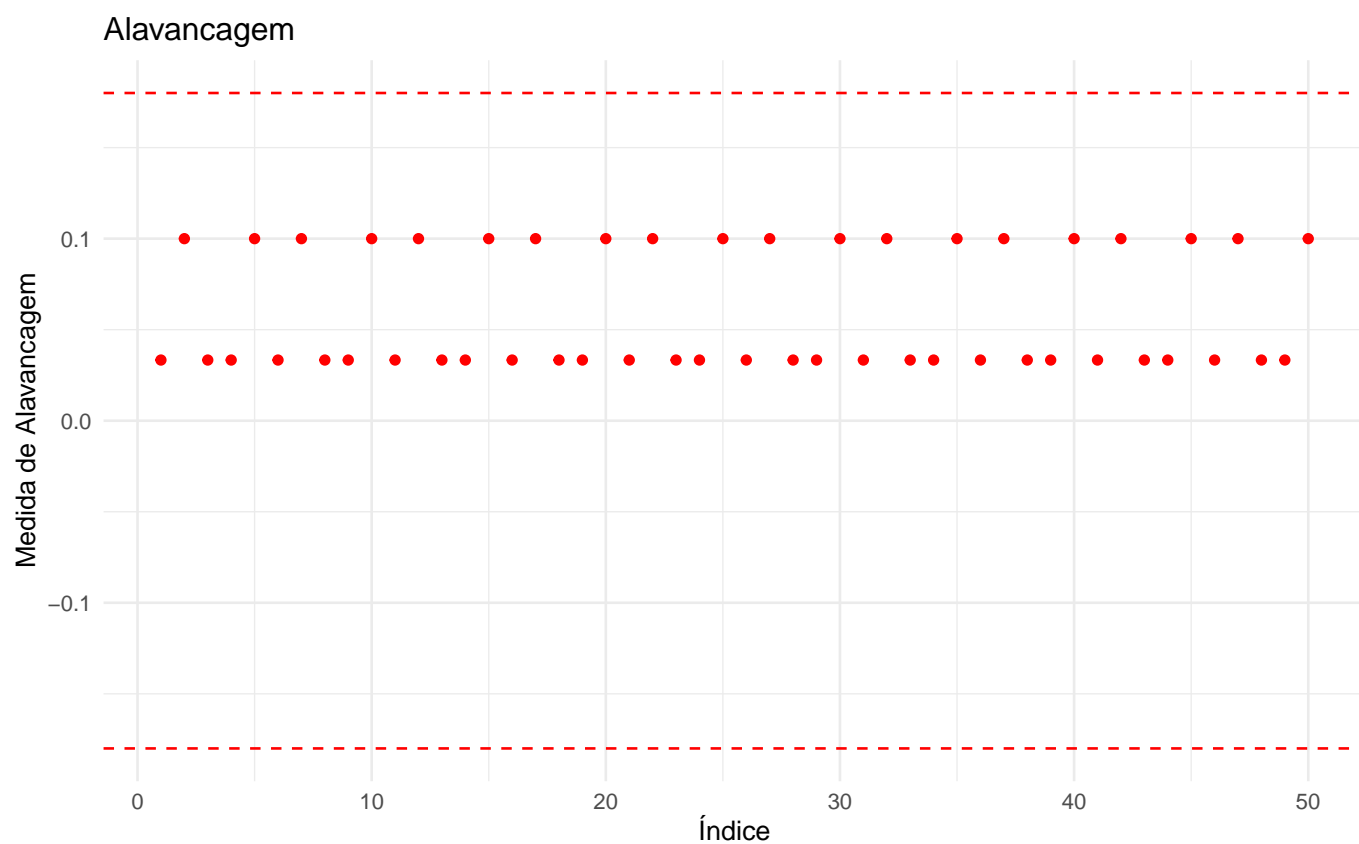
Todos os pontos estão dentro das bandas simuladas, indicando que a distribuição é adequada.

2.1.3 Distância de Cook



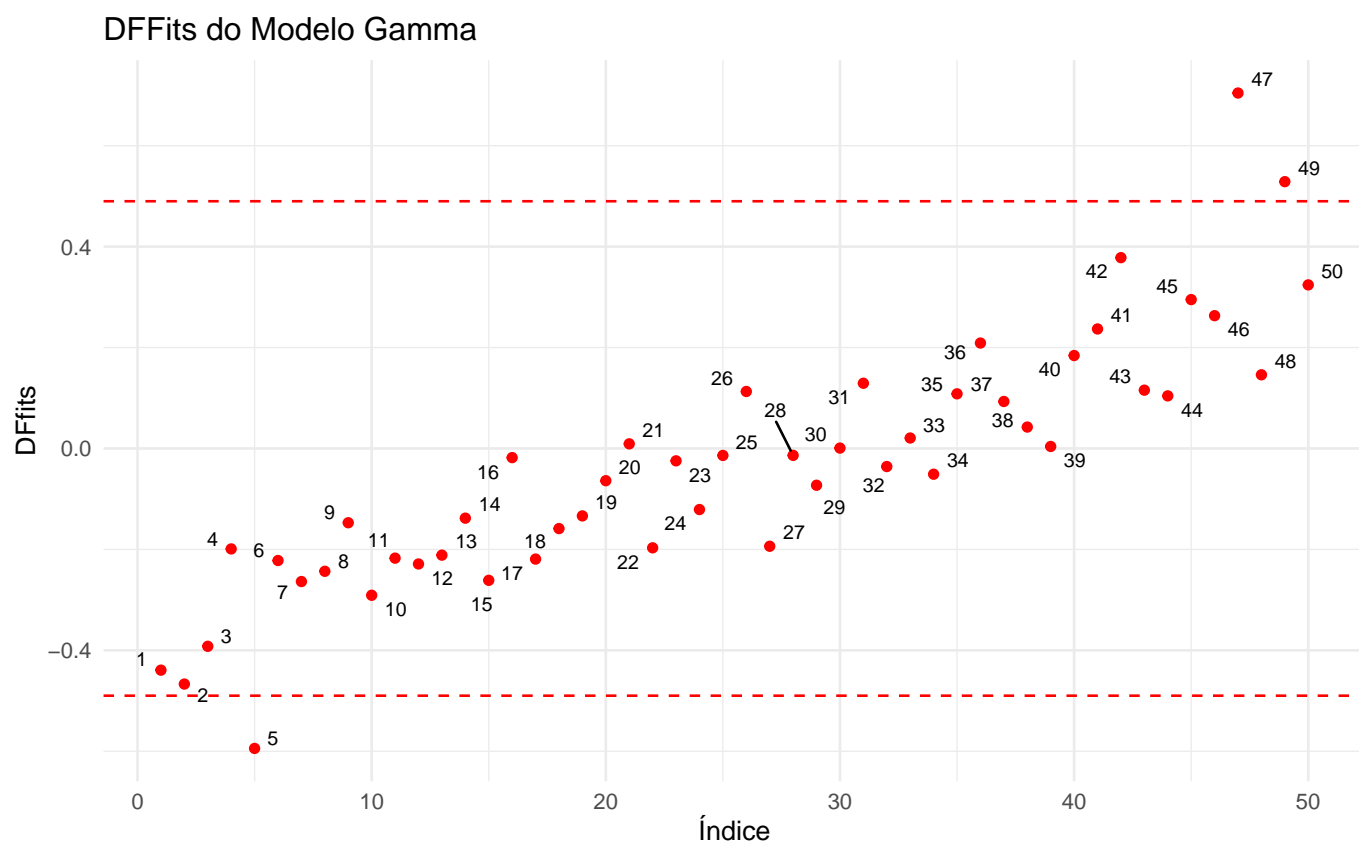
Nota-se que as observações 47 e 49 ficam fora dos limites estipulados, mas sem achatar o gráfico da distância de cook, indicam que são potenciais pontos de influência, assim iremos tomar a decisão sobre a sua remoção posteriormente.

2.1.4 Alavancagem



Observamos basicamente duas retas para a medida de alavancagem, mas nenhum delas fora dos limites estipulados, então não indicando pontos de influência.

2.1.5 DFFits

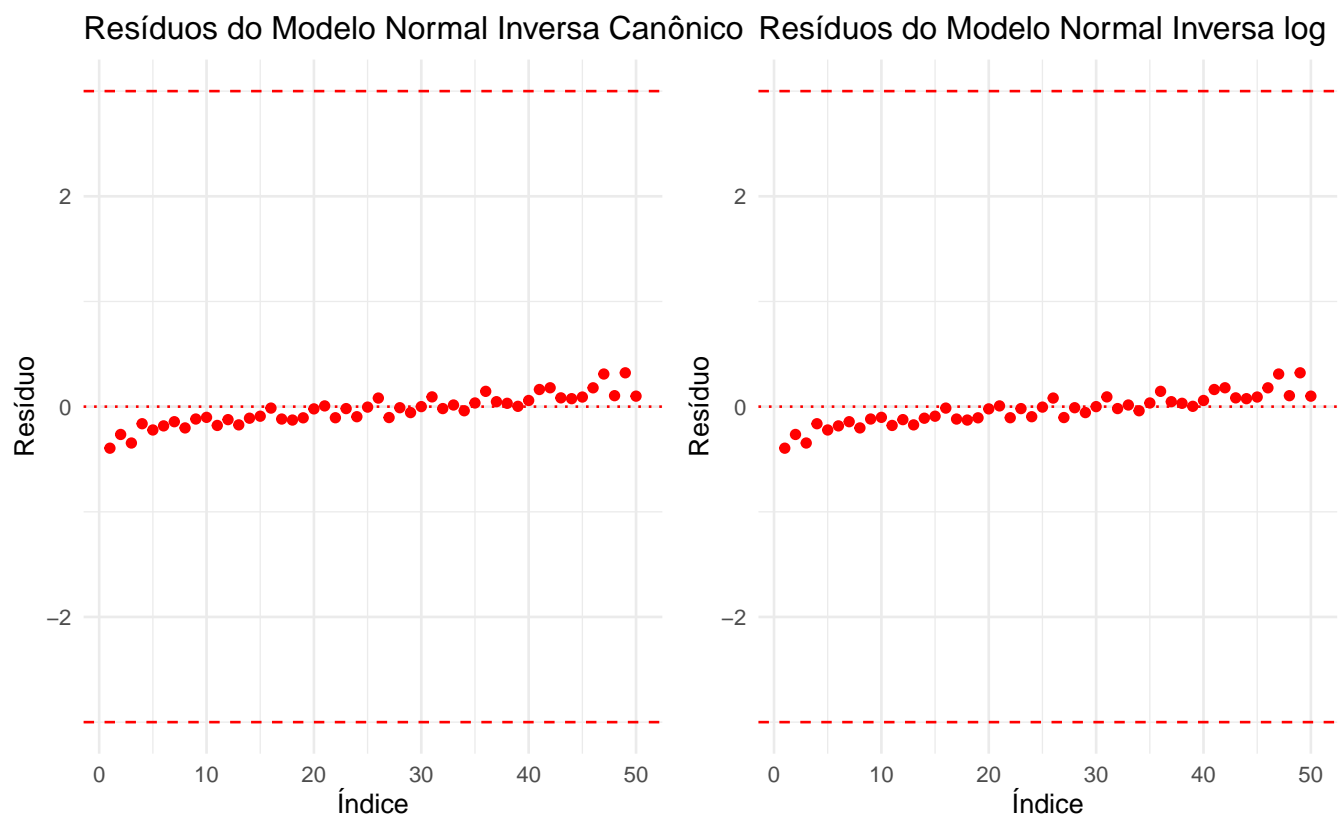


Observamos que os pontos 5, 47 e 49 ficam fora dos limites estipulados, assim são candidatos a pontos de influência.

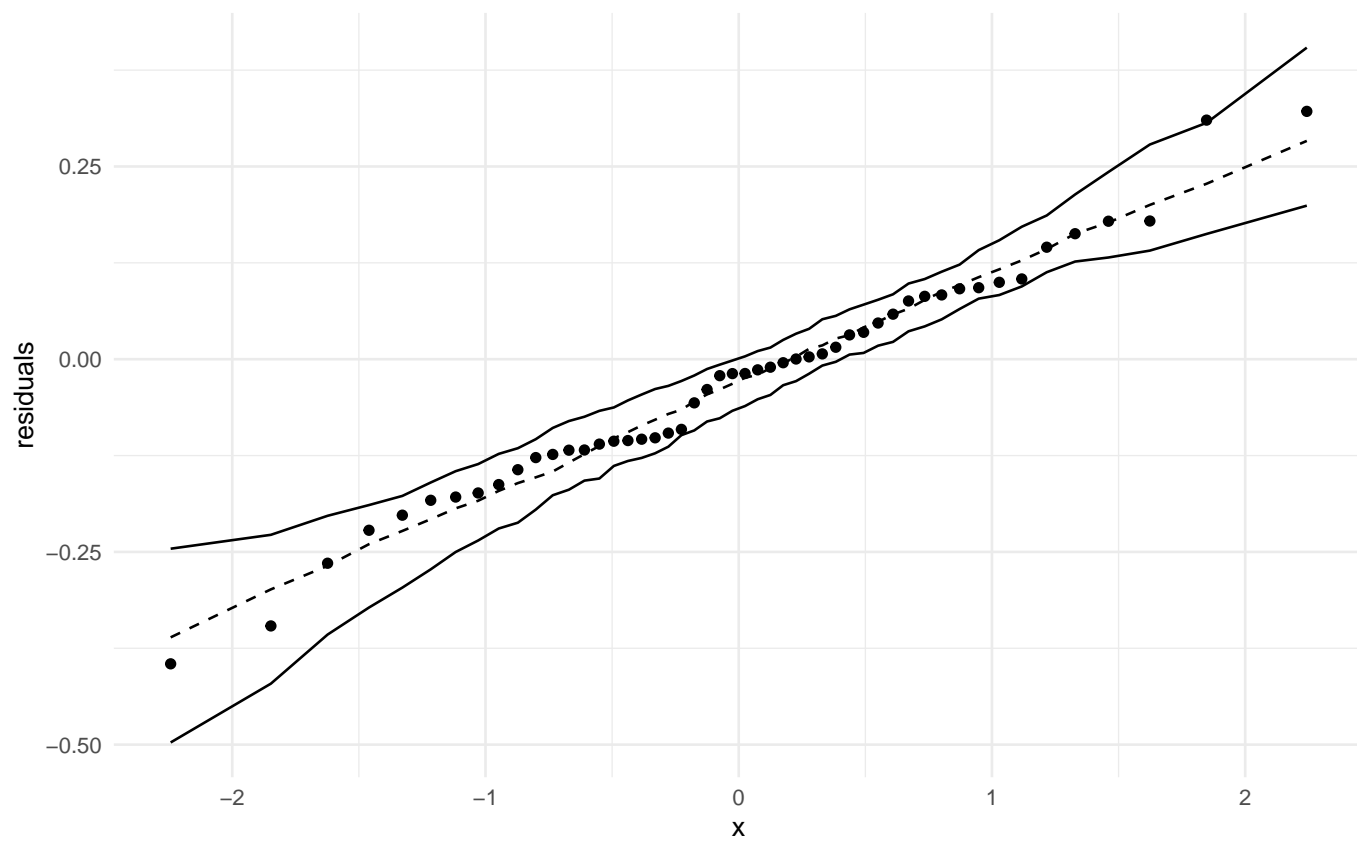
2.1.6 Conclusão

2.2 Ajuste com a Normal inversa com link $\frac{1}{\mu^2}$ e link log

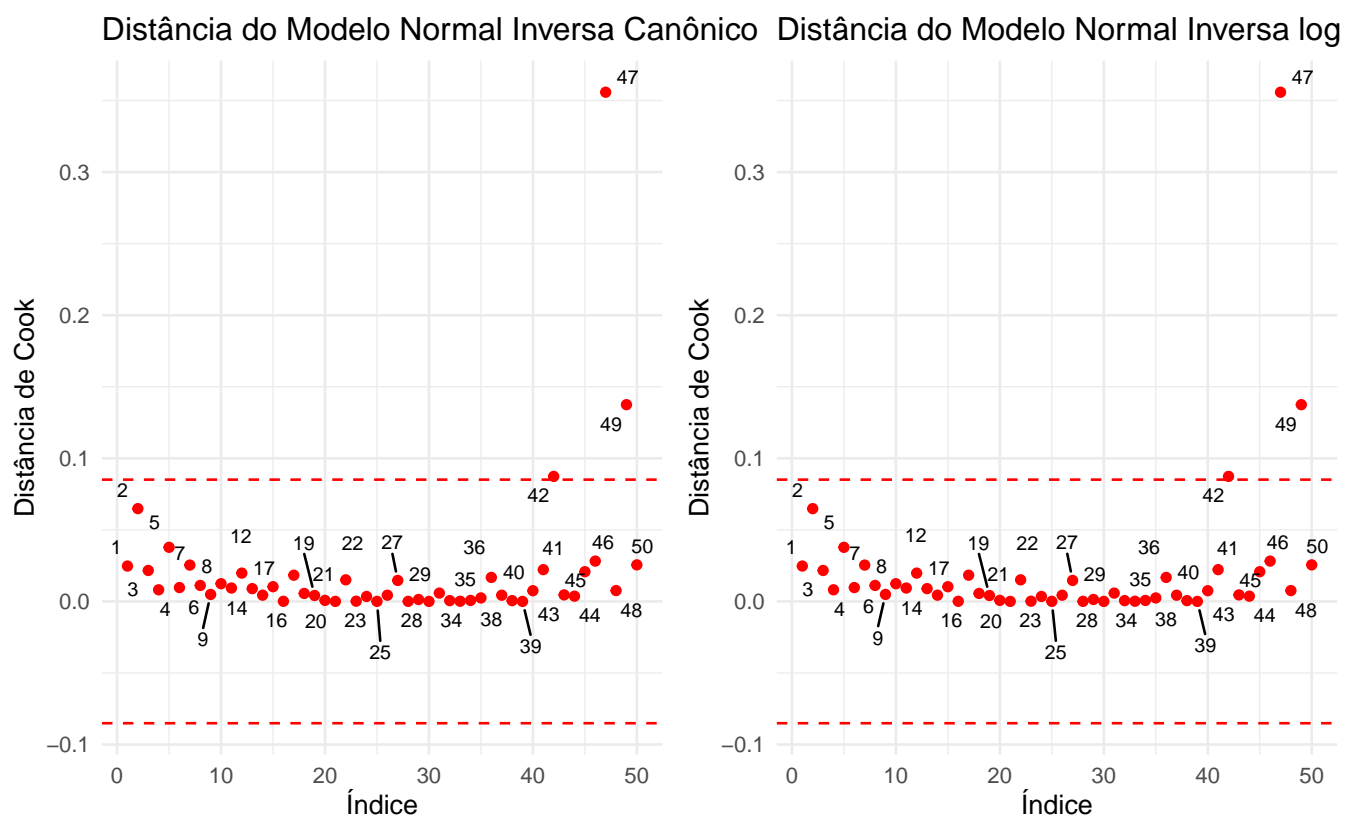
2.2.1 Resíduos deviances vs índices



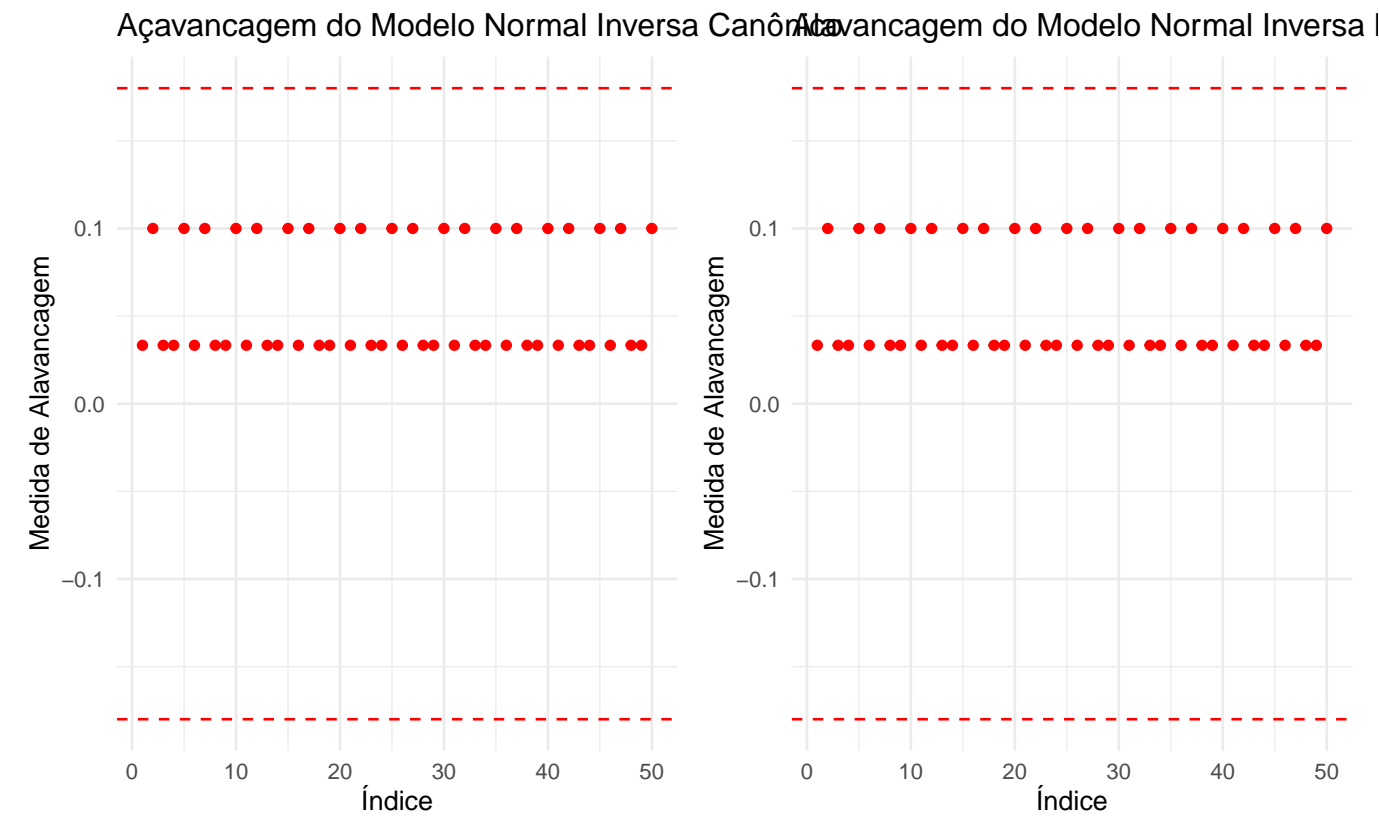
2.2.2 Envelope Simulado



2.2.3 Distância de Cook



2.2.4 Alavancagem



2.2.5 Conclusão