

Universidade Federal de São Carlos  
Centro de Ciências Exatas e de Tecnologia  
Departamento de Estatística

# **Modelos Lineares Generalizados**

## **Modelo Gama**

Crystiane Fernanda de Souza  
Douglas de Paula Nestlehner

Março, 2022

# Sumário

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Resultados</b>	<b>3</b>
2.1	Análise Descritiva . . . . .	3
2.1.1	Variável Resposta . . . . .	3
2.1.2	Variáveis Preditoras . . . . .	4
2.2	Modelo Linear Generalizado . . . . .	7
2.2.1	Ajuste do Modelo . . . . .	7
2.2.2	ANODEV . . . . .	10
2.2.3	Análise de Diagnóstico . . . . .	13
<b>3</b>	<b>Conclusão</b>	<b>18</b>

# Capítulo 1

## Introdução

Nesse trabalho iremos realizar uma análise completa da base de dados **cyclones** utilizando técnicas de Modelos Linearmente Generalizados (MLG). Essa base está presente no pacote **GLMsData** disponível no software estatístico RStudio.

Os dados fornecem o número de ciclones tropicais severos e não severos de 1970 a 2005 na região australiana. Os ciclones severos são definidos com uma pressão central mínima inferior a 970 hPa.

Ademais, o Índice Oceânico Niño (ONI) é definido pela média móvel trimestral da anomalia de temperatura da superfície do mar para a região do Niño 3.4, por no mínimo, cinco meses consecutivos. Quando a anomalia é maior que  $0.5^{\circ}\text{C}$ , o ONI estará associado a El Niño, que refere-se às situações na qual o Oceano Pacífico Equatorial está mais quente. Quando a anomalia for inferior a  $-0.5^{\circ}\text{C}$ , o ONI estará associado a La Niña que refere-se às situações na qual o Oceano Pacífico Equatorial está mais frio.

Nesse contexto, tomamos:

- $Y$ : Número de ciclones graves registrados no ano corrente.
- $x_1$ : Índice Oceânico Niño, ou ONI, calculado ao longo dos meses de Janeiro à Março.
- $x_2$ : Índice Oceânico Niño, ou ONI, calculado ao longo dos meses de Abril à Junho.
- $x_3$ : Índice Oceânico Niño, ou ONI, calculado ao longo dos meses de Julho a Setembro.
- $x_4$ : Índice Oceânico Niño, ou ONI, calculado ao longo dos meses de Outubro à Dezembro.

Nesse sentido, no Capítulo 2 serão apresentados os resultados, sendo a análise descritiva da base de dados, a estratégia de análise usando as técnicas de MLG, a seleção de variáveis, o ajuste e interpretação do modelo, a análise de diagnóstico, e por fim, o modelo ajustado. No Capítulo 3 estão as conclusões obtidas e no Apêndice A, temos os códigos utilizados para a realização do trabalho através do software estatístico RStudio.

# Capítulo 2

## Resultados

Neste capítulo, apresentamos os resultados obtidos.

### 2.1 Análise Descritiva

A base de dados utilizada possui 37 observações e 4 covariáveis. Para a análise descritiva, foi realizada a divisão entre a variável resposta e as variáveis preditoras.

#### 2.1.1 Variável Resposta

Na Tabela 2.1 estão apresentadas algumas observações referentes a variável resposta.

<hr/> Y <hr/>	
1	3
2	3
3	9
...	...
25	5
26	5
27	8

Tabela 2.1: Base de dados da variável resposta.

Para a variável resposta, foi feito de algumas medidas descritivas, apresentadas na Tabela 2.2.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
3.00	4.00	5.00	5.46	7.00	11.00

Tabela 2.2: Medidas descritivas da variável resposta.

Dessa forma, têm-se que entre os anos de 1970 a 2005, em média, ocorreram 5.46 ciclones severos. Ademais, o mínimo de ciclones ocorridos em um ano foi de 3, enquanto o máximo em um ano foi de 11.

Na Figura 2.1, temos a representação do Box-plot e um Gráfico de Barras da variável resposta, para observar seu comportamento.

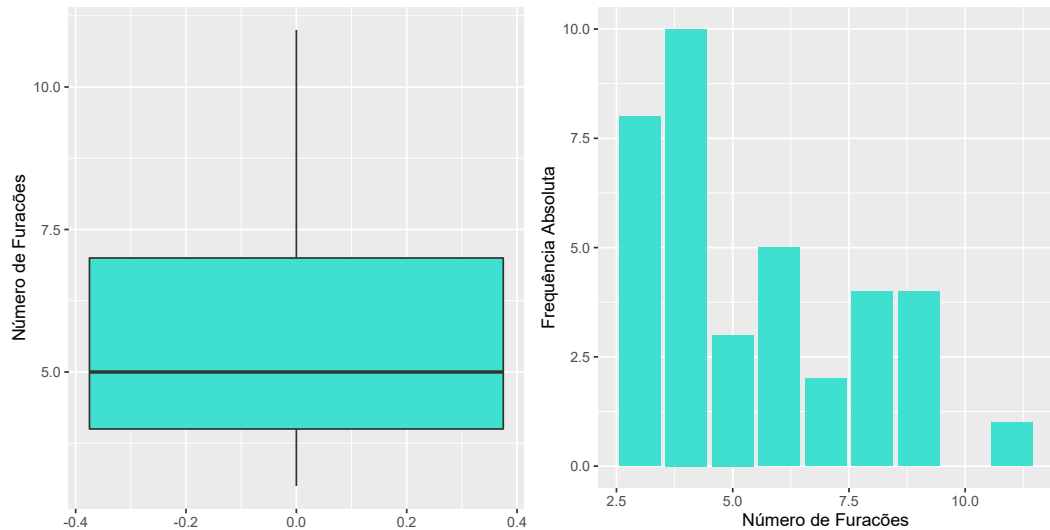


Figura 2.1: Gráficos da variável resposta Y.

Nesse sentido, ao analisarmos a Figura 2.1, podemos concluir algumas observações feitas através das medidas descritivas. Ademais, no gráfico de barras, pode-se observar que a maior frequência de ciclones tropicais severos ocorridos, foi de 4 ciclones/ano, enquanto a menor frequência foi de 11 ciclones/ano.

### 2.1.2 Variáveis Preditoras

Na Tabela 2.3 estão apresentadas algumas observações referentes as covariáveis.

	$x_1$	$x_2$	$x_3$	$x_4$
1	1.0	0.6	0.4	0.8
2	0.3	0.0	-0.8	-0.9
3	-1.3	-0.8	-0.8	-1.0
...	...	...	...	...
35	0.9	-0.1	0.5	0.6
36	0.3	0.3	0.8	0.8
37	0.5	0.4	0.3	-0.4

Tabela 2.3: Base de dados das covariáveis.

Inicialmente, foi realizado o cálculo de algumas medidas descritivas para as covariáveis, sendo essas: mínimo, 1º quartil, mediana, média, 3º quartil e máximo. Na Tabela 2.4 têm-se os resultados de cada medida.

Covariável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
$x_1$	-1.70	-0.50	0.20	0.05946	0.50	2.00
$x_2$	-0.90	-0.50	0.10	0.05135	0.50	1.20
$x_3$	-1.30	-0.40	0.10	0.06757	0.50	2.00
$x_4$	-2.00	-0.90	-0.10	0.06216	0.80	2.50

Tabela 2.4: Medidas descritivas das covariáveis.

Dessa forma, é possível notar que o menor e maior Índice Oceânico Niño (ONI), foi calculado nos meses de Outubro à Dezembro ( $x_4$ ). Um outro ponto, é que os valores mínimos do Índice em todas as covariáveis estão associados a La Niña, enquanto os valores máximos estão todos associados ao El Niño. Ademais, a mediana dos Índices são iguais calculados nos meses de Abril à Junho ( $x_2$ ) e nos meses de Julho à Setembro ( $x_3$ ).

Na Figura 2.2 estão representados o Box-plot e o Histograma para cada covariável que representam o Índice Oceânico Niño (ONI) de trimestres, com o intuito de conhecer o comportamento das observações de cada covariável.

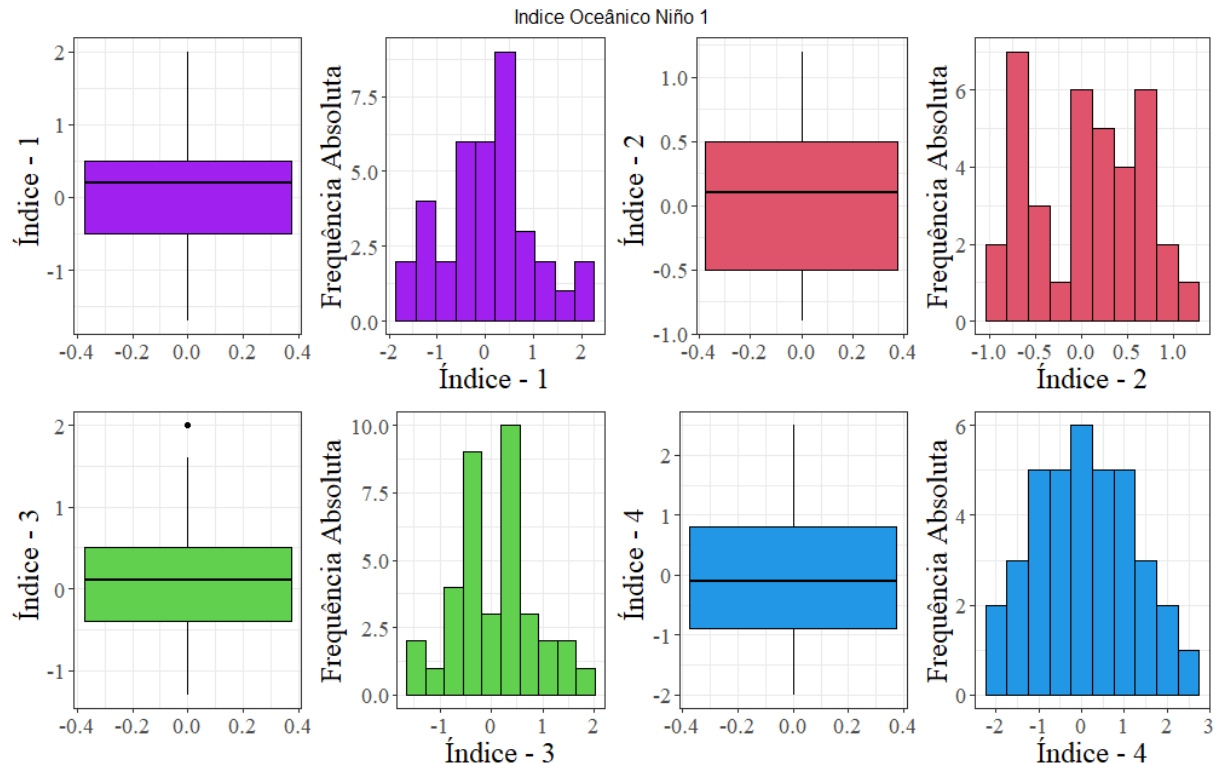


Figura 2.2: Box-plot e Histograma da covariáveis.

Ao analisar a Figura 2.2, é possível notar que a variabilidade entre as covariáveis  $x_1$  e  $x_3$  são semelhantes, enquanto a variabilidade da covariável  $x_2$  é superior a ambas. Além

disso, para a covariável  $x_4$ , pode-se considerar que a média do ONI calculado ao longo dos meses de Outubro a Dezembro, está próximo de 0.

Ademais, calculamos e plotamos a matriz de correlação entre as covariáveis, representados pela Tabela 2.5 e Figura 2.3.

	$x_1$	$x_2$	$x_3$	$x_4$
$x_1$	1.00	0.68	0.13	0.01
$x_2$	0.68	1.00	0.70	0.60
$x_3$	0.13	0.70	1.00	0.94
$x_4$	0.01	0.60	0.94	1.00

Tabela 2.5: Matriz de correlação das covariáveis.

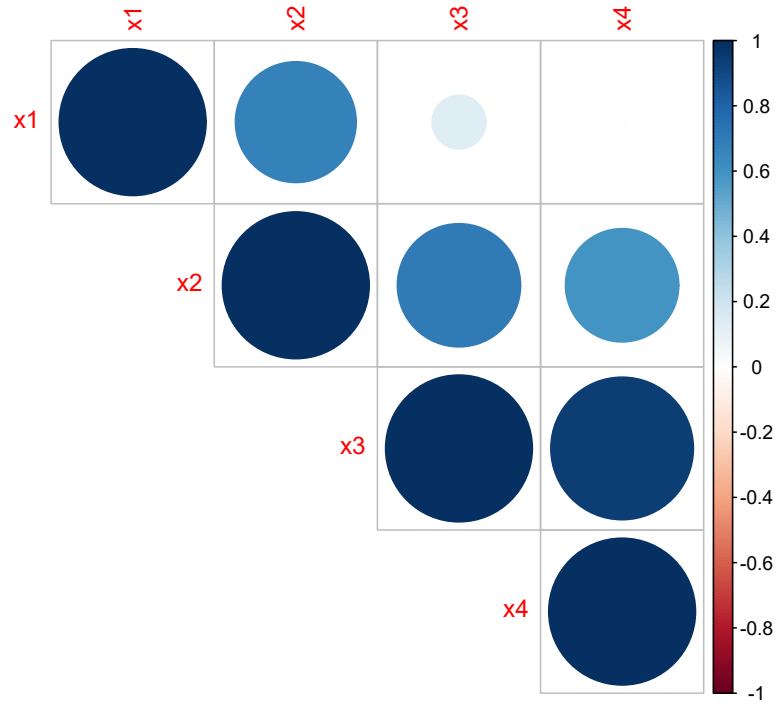


Figura 2.3: Matriz de correlação das covariáveis.

Analisando a Tabela 2.5 e a Figura 2.3, nota-se que as covariáveis  $x_3$  e  $x_4$  estão fortemente correlacionadas, podendo ocasionar problemas de multicolinearidade.

Porém como o número de covariáveis presente na nossa base de dados é baixo (apenas quatro) e também como cada covariável é sobre a média trimestral do ano, não é necessário e nem recomendável a exclusão de covariáveis altamente correlacionadas, uma vez que serão fundamentais para o ajuste do modelo.

## 2.2 Modelo Linear Generalizado

Nesta seção, será realizado a análise completa dos dados utilizando as técnicas de Modelos Linearmente Generalizados, em que será definido o componente aleatório, o componente sistemático e a função de ligação.

Em relação a base de dados **cyclones**, vimos na análise descritiva que a variável resposta trata-se sobre a contagem de ciclones severos ocorridos ao longo de 1970 à 2005, e ao se tratar de dados de contagem, a distribuição a ser considerada no ajuste do modelo linear generalizado é a distribuição **Poisson**.

Ao se tratar de contagem, não podemos estimar valores negativos, portanto a função de ligação mais indicada para o nosso problema é função de ligação canônica logarítmica.

### 2.2.1 Ajuste do Modelo

Nesse contexto, por meio das informações ditas anteriormente, pode-se definir:

- **Componente Aleatório:**  $Poisson(\lambda)$ .
- **Componente Sistemático:** Índice Oceânico Niño, ou ONI, calculado ao longo dos meses de Janeiro à Março ( $X_1$ ), Índice Oceânico Niño, ou ONI, calculado ao longo dos meses de Abril à Junho ( $X_2$ ), Índice Oceânico Niño, ou ONI, calculado ao longo dos meses de Julho à Setembro ( $X_3$ ) e Índice Oceânico Niño, ou ONI, calculado ao longo dos meses de Outubro à Dezembro ( $X_4$ ).
- **Função de Ligação:** Logarítmica.

Em seguida, iremos considerar três diferentes modelos considerando todas as covariáveis e algumas interações, para que possamos comparar e escolher qual o modelo mais adequado para seguir com a análise.

#### Modelo 1

Inicialmente, foi ajustado um modelo com todas as covariáveis apresentadas ao longo do trabalho. Nesse sentido, o modelo ficará da seguinte forma:

$$\log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

Na Tabela 2.6 temos as respectivas estimativas do primeiro modelo ajustado. Em que o valor de  $AIC$  é de 167.59 e o Desvio Residual é de 28.43.



Variável	Estimativa
Intercepto	1.684963
$X_1$	-0.009997
$X_2$	0.156619
$X_3$	0.153493
$X_4$	-0.237094

Tabela 2.6: Estimativas para o Modelo 1.

## Modelo 2

Também foi ajustado um modelo em que estavam presentes todas as covariáveis apresentadas e também as interações duas a duas. Logo, o modelo ficará da seguinte forma:

$$\log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_1 X_2 + \beta_6 X_1 X_3 + \beta_7 X_1 X_4 + \beta_8 X_2 X_3 + \beta_9 X_2 X_4 + \beta_{10} X_3 X_4$$

Na Tabela 2.7 temos as respectivas estimativas do segundo modelo ajustado.

Variável	Estimativa
Intercepto	1.761410
$X_1$	-0.135618
$X_2$	0.341716
$X_3$	0.279978
$X_4$	-0.371231
$X_1 X_2$	-0.175933
$X_1 X_3$	1.097612
$X_1 X_4$	-0.721443
$X_2 X_3$	-1.504682
$X_2 X_4$	0.913256
$X_3 X_4$	0.001543

Tabela 2.7: Estimativas para o Modelo 2.

Com um valor *AIC* de 175.17 e Desvio Residual de 24.02.

## Modelo 3

O terceiro modelo ajustado, além de considerar a interação duas a duas, foi considerado a interação entre as covariáveis  $X_1$ ,  $X_2$  e  $X_3$ . Nesse sentido, o modelo será de:

$$\log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_1 X_2 + \beta_6 X_1 X_3 + \beta_7 X_1 X_4 + \beta_8 X_2 X_3 \\ + \beta_9 X_2 X_4 + \beta_{10} X_3 X_4 + \beta_{11} X_1 X_2 X_3$$

As estimativas do terceiro ajuste estão representadas na Tabela 2.8.

Variável	Estimativa	Variável	Estimativa
Intercepto	1.74862	$X_1 X_3$	1.62675
$X_1$	0.14985	$X_1 X_4$	-1.33470
$X_2$	0.02669	$X_2 X_3$	-0.76325
$X_3$	0.48439	$X_2 X_4$	1.42202
$X_4$	-0.27281	$X_3 X_4$	-0.57581
$X_1 X_2$	-0.38737	$X_1 X_2 X_3$	-0.83908

Tabela 2.8: Estimativas para o Modelo 3.

Com um valor de  $AIC$  de 168.26 e Desvio Residual de 15.103.

#### Modelo 4

Por fim, para o quarto e último modelo, foi considerado todas as interações possíveis entre as covariáveis.

$$\log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_1 X_2 + \beta_6 X_1 X_3 + \beta_7 X_1 X_4 + \beta_8 X_2 X_3 \\ + \beta_9 X_2 X_4 + \beta_{10} X_3 X_4 + \beta_{11} X_1 X_2 X_3 + \beta_{12} X_1 X_2 X_4 + \beta_{13} X_1 X_3 X_4 + \beta_{14} X_2 X_3 X_4 \\ + \beta_{15} X_1 X_2 X_3 X_4$$

Na Tabela 2.9 têm-se as estimativas obtidas.

Variável	Estimativa	Variável	Estimativa
Intercepto	1.74707	$X_2X_3$	-1.72119
$X_1$	0.25343	$X_2X_4$	1.88912
$X_2$	-0.09497	$X_3X_4$	-0.79386
$X_3$	0.48473	$X_1X_2X_3$	-1.41487
$X_4$	-0.37115	$X_1X_2X_4$	0.51873
$X_1X_2$	-0.34793	$X_2X_3X_4$	0.24125
$X_1X_3$	1.93546	$X_1X_2X_3X_4$	0.21838
$X_1X_4$	-0.93485		

Tabela 2.9: Estimativas para o Modelo 4.

Com um valor de  $AIC$  de 173.05 e Desvio Residual de 11.893.

### Comparação

No intuito de verificar qual é o modelo mais adequado para a base de dados, foi observado o Critério de Akaike ( $AIC$ ) e o Desvio Residual de cada modelo apresentados na Tabela 2.10.

Modelo	AIC	Desvio Residual
Modelo 1	167.59	28.43
Modelo 2	175.17	24.02
Modelo 3	168.26	15.10
Modelo 4	173.05	11.89

Tabela 2.10: Comparação medidas de qualidade

Nesse sentido, achamos interessante analisar melhor o Modelo 1, uma vez que possui o menor Critério de Akaike ( $AIC$ ), dando indícios de que o modelo é mais adequado em relação aos outros.

### 2.2.2 ANODEV

A análise de deviance (ANODEV) é uma generalização para a análise de variância, para os modelos lineares generalizados. O interesse da ANODEV é testar a significância da inclusão de novos termos. Nesse sentido, será feito a ANODEV para o Modelo 1, com intuito de checar a qualidade do modelo ajustado. Na Tabela 2.11 está representada essa análise.

	<b>G.L.</b>	<b>Desvio</b>	<b>G.L. Residuais</b>	<b>Desvio Residual</b>	<b>Valor-p</b>
NULL			36	32.220	
$X_1$	1	1.09035	35	31.130	0.2964
$X_2$	1	0.36621	34	30.763	0.5451
$X_3$	1	1.01636	33	29.747	0.3134
$X_4$	1	1.31393	32	28.433	0.2517

Tabela 2.11: Tabela ANODEV.

Ao analisar a Tabela 2.11 é possível identificar que ao nível de significância de 5%, nenhuma covariável será significativa para explicar a variável resposta que é a contagem de ciclones tropicais severos.

Dessa forma, foi identificado que a melhor saída era retirar a covariável menos significativa do modelo e ajustá-lo novamente. Logo, ajustando o modelo sem a covariável  $X_2$  temos as novas estimativas representadas na Tabela 2.12.

<b>Variável</b>	<b>Estimativa</b>
Intercepto	1.68497
$X_1$	0.05199
$X_3$	0.22486
$X_4$	-0.23342

Tabela 2.12: Novos coeficientes para o Modelo 1 sem a covariável  $X_2$ .

Além das estimativas, o novo  $AIC$  será de 165.88 e o Desvio Residual de 28.72. Sendo assim, será feito novamente a ANODEV que está representada pela Tabela 2.13.

	<b>G.L.</b>	<b>Desvio</b>	<b>G.L. Residuais</b>	<b>Desvio Residual</b>	<b>Valor-p</b>
NULL			36	32.220	
$X_1$	1	1.0903	35	31.130	0.2964
$X_3$	1	1.1186	34	30.011	0.2902
$X_4$	1	1.2829	33	28.728	0.2574

Tabela 2.13: Tabela ANODEV.

Analisando novamente a Tabela 2.13, nota-se que ao nível de significância de 5% nenhuma das covariáveis é significativa. Dessa forma, nenhuma das covariáveis é significativa para explicar a variável resposta, logo, será retirado outra variável que possui menos significância no modelo, que é o caso da covariável  $X_1$ .

Retirando a variável preditora  $X_1$ , ajustamos um novo modelo com as estimativas na Tabela 2.14.

Variável	Estimativa
Intercepto	1.68695
$X_2$	0.30124
$X_4$	-0.28426

Tabela 2.14: Novos coeficientes para o Modelo 1 sem a covariável  $X_1$ .

E o novo  $AIC$  será de 164.3 e o Desvio Residual de 29.14. Nesse contexto, será feito a ANODEV mais uma vez representada pela Tabela 2.15

	G.L.	Desvio	G.L. Residuais	Desvio Residual	Valor-p
NULL			36	32.220	
$X_2$	1	0.86383	35	31.356	0.3527
$X_4$	1	2.21507	34	29.141	0.1367

Tabela 2.15: Tabela ANODEV.

Através da Tabela 2.15 é possível verificar que em um nível de significância de 5% ainda não há indícios de que as variáveis preditoras são significantes, ou seja, nenhuma das variáveis preditoras está explicando sobre a variável resposta. Logo, vamos retirar novamente a próxima covariável com menor significância que é a  $X_2$  e ajustar o modelo mais uma vez.

Portanto, após ajustar o modelo novamente, temos as novas estimativas na Tabela 2.16.

Variável	Estimativa
Intercepto	1.69808
$X_4$	-0.08698

Tabela 2.16: Novos coeficientes para o Modelo 1 sem a covariável  $X_2$ .

E o  $AIC$  passará a ser 163.5 e o Desvio Residual será de 30.34. Na Tabela 2.17 têm-se a representação da ANODEV.

	G.L.	Desvio	G.L. Residuais	Desvio Residual	Valor-p
NULL			36	32.220	
$X_4$	1	1.8765	35	30.343	0.1707

Tabela 2.17: Tabela ANODEV.

Analisando a Tabela 2.17 nota-se que ao nível de significância de 5%, têm-se que o modelo composto apenas pela covariável  $X_4$  não é significativo para explicar a contagem de ciclones severos na região australiana.

Nesse sentido, para que o modelo seja significativo, a covariável  $X_4$  também será retirada e será ajustado o modelo nulo. Na Tabela 2.18 temos a nova estimativa.

Variável	Estimativa
Intercepto	1.69735

Tabela 2.18: Modelo nulo ajustado.

Para a base de dados, o ideal seria encontrar um modelo que estivesse entre o modelo nulo e o modelo saturado, e que fosse adequado aos dados. Contudo, como não encontramos o modelo ideal, deveria ser utilizado o Modelo Nulo que é o modelo mais simples, em que a variável resposta vai variar entre si, apenas por causa do componente aleatório que é a distribuição Poisson.

Todavia, consideramos que não é correto que as covariáveis não expliquem nada sobre a variável resposta, uma vez que o Índice Oceânico Ninõ é definido pela média móvel trimestral da anomalia de temperatura da superfície do mar e a partir desse Índice é possível prever o número de ciclones tropicais graves na região australiana. Dessa forma, mesmo que através da análise de deviance (ANODEV) as covariáveis foram retiradas, seguiremos com o Modelo 1 em que estão presentes todas as variáveis preditoras.

### 2.2.3 Análise de Diagnóstico

Por fim, será realizado a análise de diagnóstico para verificar se há indícios de problemas no ajuste do Modelo 1 que inclui as variáveis preditoras  $X_1$ ,  $X_2$ ,  $X_3$  e  $X_4$ .

#### Envelope

Para iniciar a análise de diagnóstico, será feito o gráfico do Envelope, representado na Figura 2.4 para verificar se o modelo é adequado.

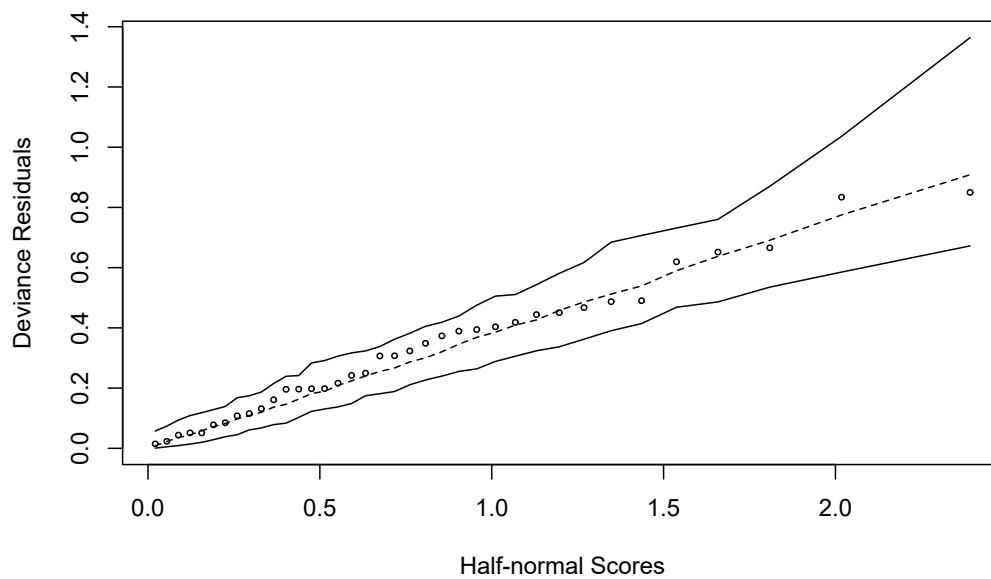


Figura 2.4: Gráfico do envelope.

Através da Figura 2.4 é possível notar que todos os pontos estão dentro do limite, ou seja, todos os pontos estão dentro do envelope. Sendo assim, há indícios de que o modelo foi bem ajustado.

### Função de Ligação

Para verificar se a função de ligação é adequada para os dados, foi feito o gráfico dos valores preditos *versus* resíduos de Pearson, representado pela Figura 2.5.

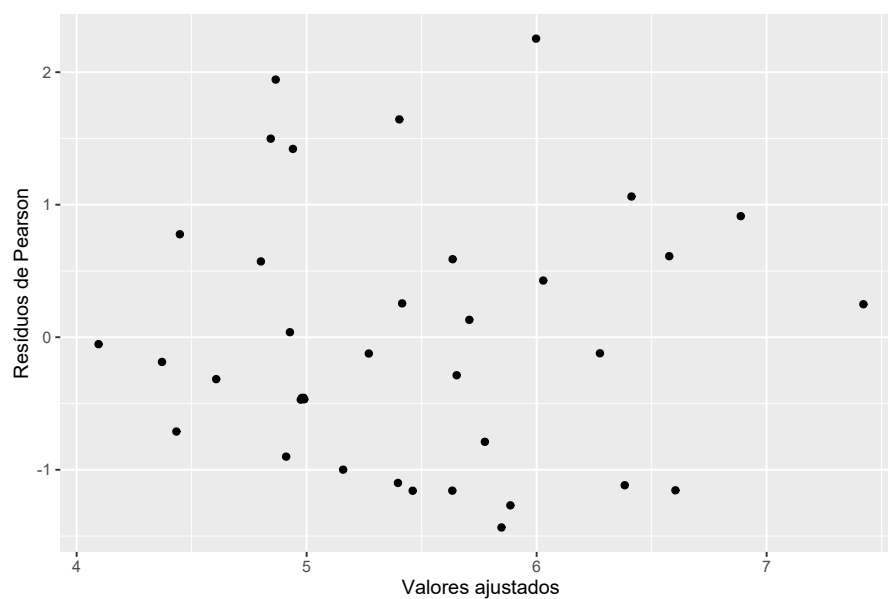


Figura 2.5: Gráfico de valores preditos versus resíduos de Pearson.

Nesse contexto, através da Figura 2.5 nota-se um padrão aleatório em torno do zero, dando indícios de que a função de ligação canônica logarítmica é adequada. Ademais, já indícios que a escala das variáveis preditoras utilizadas no modelo estão corretas.

### Componente do Desvio

Um outro modo de verificar se o modelo foi bem ajustado, é através do gráfico de valores preditos *versus* componente de desvio, representado na Figura 2.6.

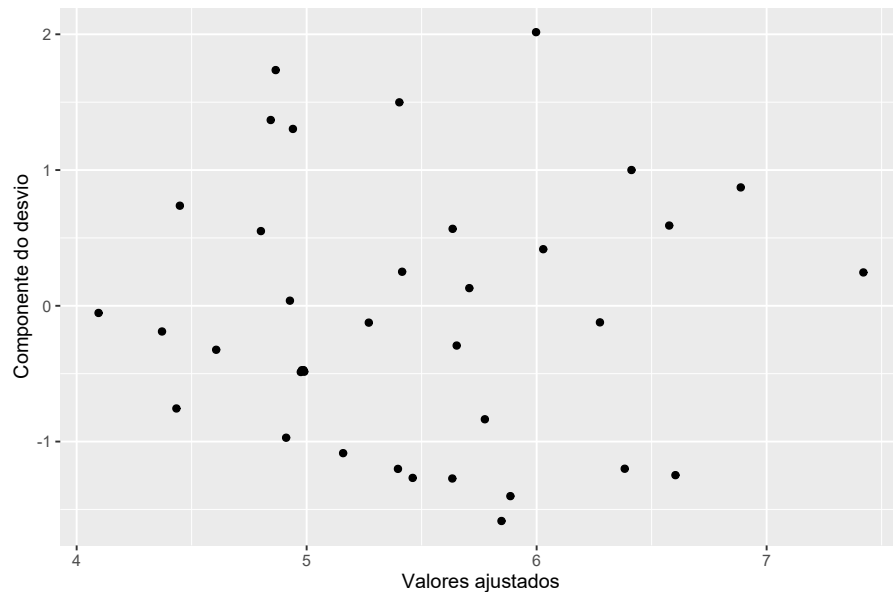


Figura 2.6: Gráfico de valores preditos versus componente de desvio.

Na Figura 2.6 nota-se também um padrão aleatório em torno do zero, dando indícios novamente de que o modelo foi bem ajustado.

### Independência

Um pressuposto que deve ser satisfeito também é o da independência, já que os dados possuem ordem de coleta de 1970 à 2005. Nesse sentido, foi realizado o gráfico da ordem de coleta *versus* resíduos de Pearson, representado na Figura 2.7.



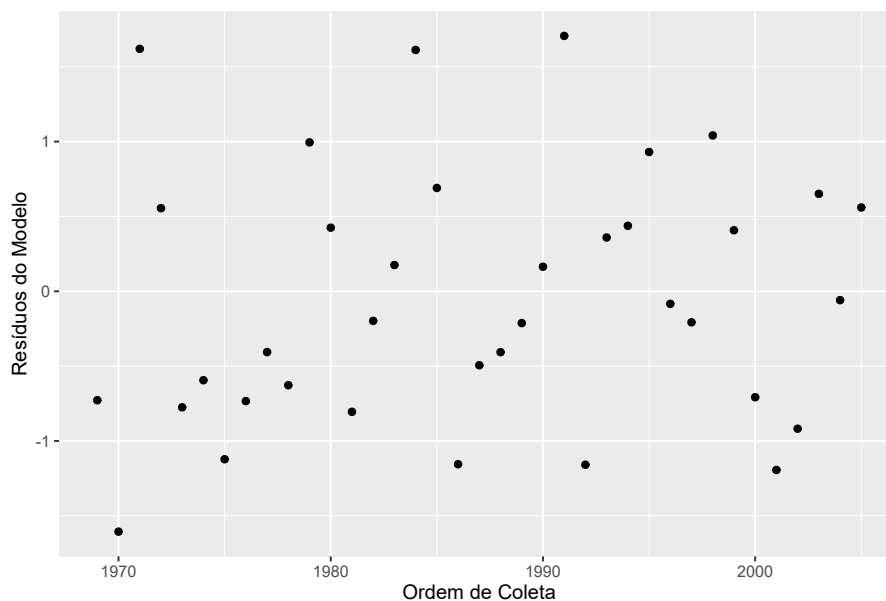


Figura 2.7: Gráfico de ordem de coleta versus resíduos de Pearson.

Dessa forma, através da Figura 2.7, nota-se que há dispersão dos pontos de forma aleatória, indicando que não há violação na suposição de independência por mais que haja ordem de coleta dos dados.

### Pontos de Alavanca

Por fim, foi verificado a existência de pontos de alavanca, uma vez que esses pontos podem influenciar nas estimativas dos parâmetros. Nesse sentido, na Figura 2.8 têm-se a representação desses pontos.

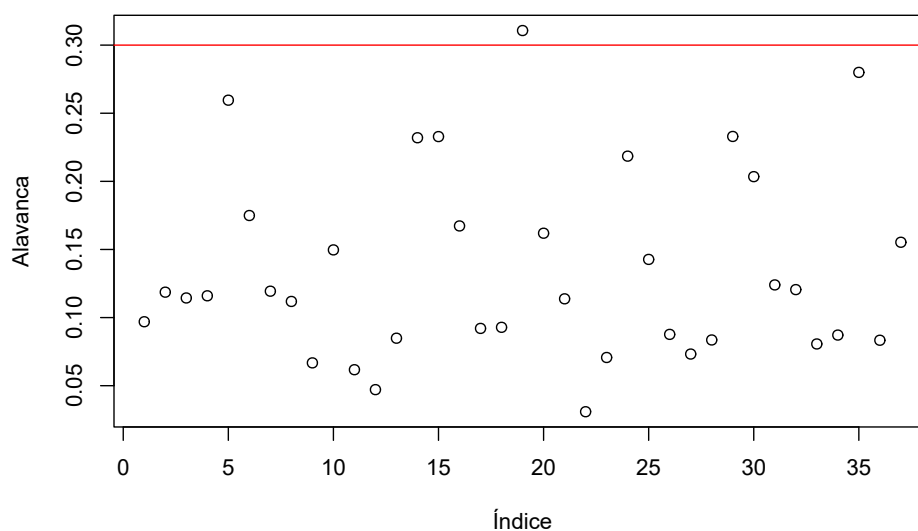


Figura 2.8: Gráfico com os pontos de alavanca.

Dessa forma, ao analisar a Figura 2.8 nota-se apenas uma observação superior em relação as outras. Contudo, como os pressupostos para a análise de resíduos foram satisfeitos, a observação não será retirada.

# Capítulo 3

## Conclusão

Nesse trabalho, o componente aleatório escolhido foi a distribuição Poisson, dado que a variável resposta trata-se da contagem de ciclones tropicais severos na região australiana. Para o componente sistemático, foi utilizado todas as covariáveis, sendo:  $X_1$ ,  $X_2$ ,  $X_3$  e  $X_4$ . Ademais, a função de ligação utilizada foi a função de ligação canônica logarítmica.

Após a definição do componente aleatório, componente sistemático e função de ligação, foi feito o ajuste de quatro modelos e utilizado o Critério de Akaike ( $AIC$ ) para escolher o modelo mais adequado. Entretanto, chegamos que o modelo ajustado não era significativo através da Análise de Deviance (ANODEV), o que na prática não faria sentido, uma vez que nenhuma covariável deveria ser removida, dado que são importantes para o ajuste do modelo.

Nesse contexto, resolvemos considerar o Modelo 1 da mesma forma e ao realizar a análise de diagnóstico, sendo visto que a análise foi satisfatória para todos os pressupostos. Portanto, mesmo que através da Análise de Deviance (ANODEV) o modelo ajustado não é significativo, consideramos que o Modelo 1 será um bom modelo para os dados, já que passou pelo crivo de análise de diagnóstico, confirmando que de há indícios de que o modelo foi bem ajustado e a função de ligação utilizada era correta.

Por fim, o Modelo 1 será dado pela seguinte equação:

$$\begin{aligned}\mu_i &= \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4) \\ &= \exp(1.684963 - 0.009997X_1 + 0.156619X_2 + 0.153493X_3 - 0.237094X_4),\end{aligned}$$

e que esse modelo é considerado adequado para a base de dados.