



3º Atividade Avaliativa de Modelos Lineares Generalizados

PROFESSORA: THEMIS ABENSUR LEÃO

ALUNA: MEILYN LEIENE MACHADO BARBOSA



Análise do Número de Homicídios no Brasil em 2020

Modelo Binomial Negativo

Contexto

- ▶ Segundo o IBGE, no ano de 2020 o Brasil registrou **49 868** homicídios;
- ▶ **45 915** vítimas eram homens;
- ▶ **3 833** vítimas eram mulheres;
- ▶ Segundo dados da PNAD contínua de 2017, 43,4% da população residente era **branca**, 8,6% **preta** e 47,1% **parda**, não houve grandes alterações nas proporções de 2017 para o ano de 2022 (pesquisa mais recente).

Apresentação dos Dados

Tabela 1 – Número de Homicídios no Brasil por Raça e Faixa Etária no ano de 2020

Cor/Raça	Faixa Etária				Total
	0-14	15-29	30-59	60+	
Branca	118	4 398	4 984	790	10 290
Preta	38	2 304	1 568	117	33 786
Parda	409	18 590	13 711	1 076	4 027
Total	565	25 292	20 263	1 983	48 103*

(*) : diverge do total apresentado anteriormente, pois no total de registros de homicídio estão inclusos casos em que não há informação de idade da vítima.

Análise Descritiva

Fig. 1 - N° de Homicídios por Raça

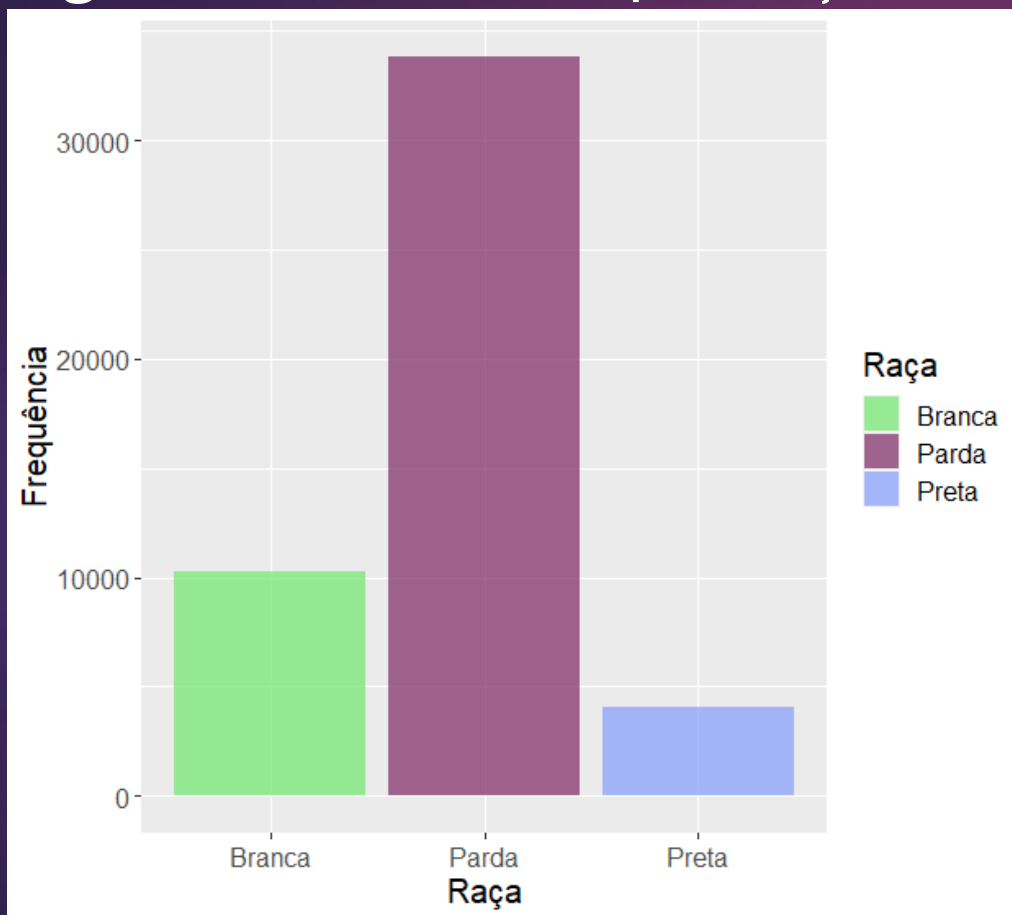
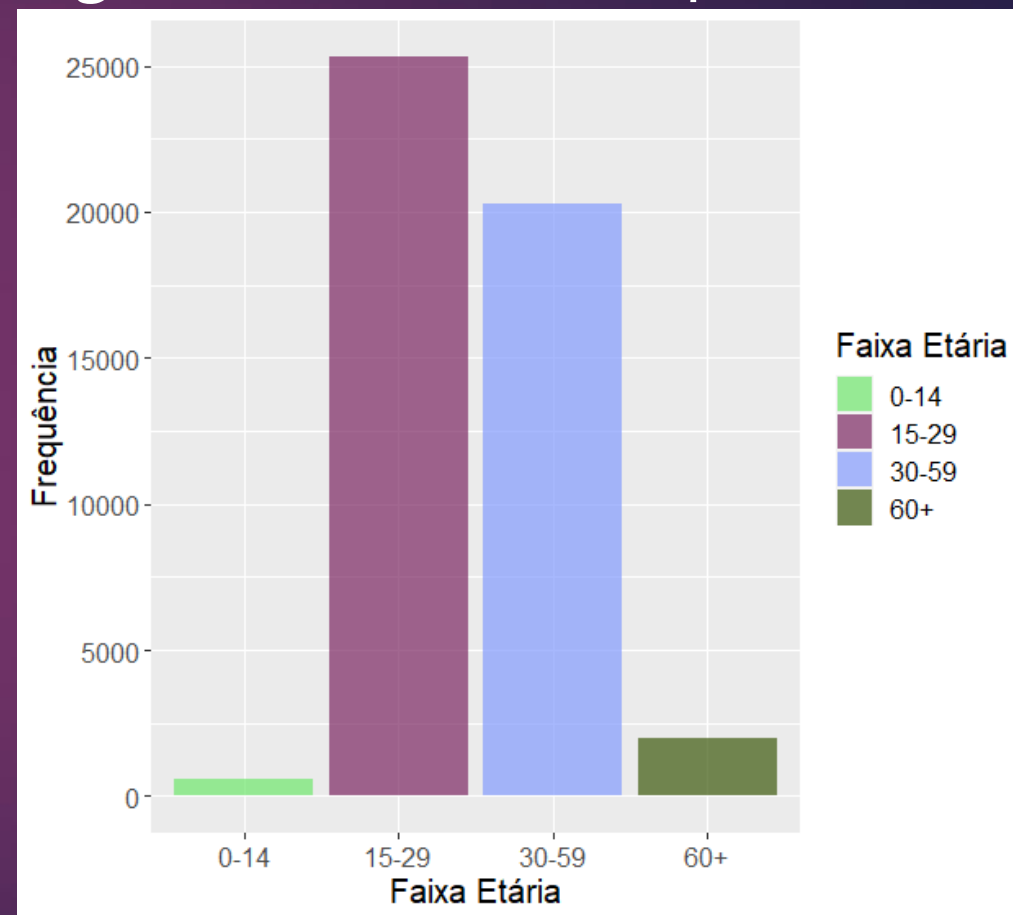


Fig. 2 - N° de Homicídios por Faixa Etária



Análise Descritiva

Fig. 3 - Boxplot Homicídios por Faixa Etária

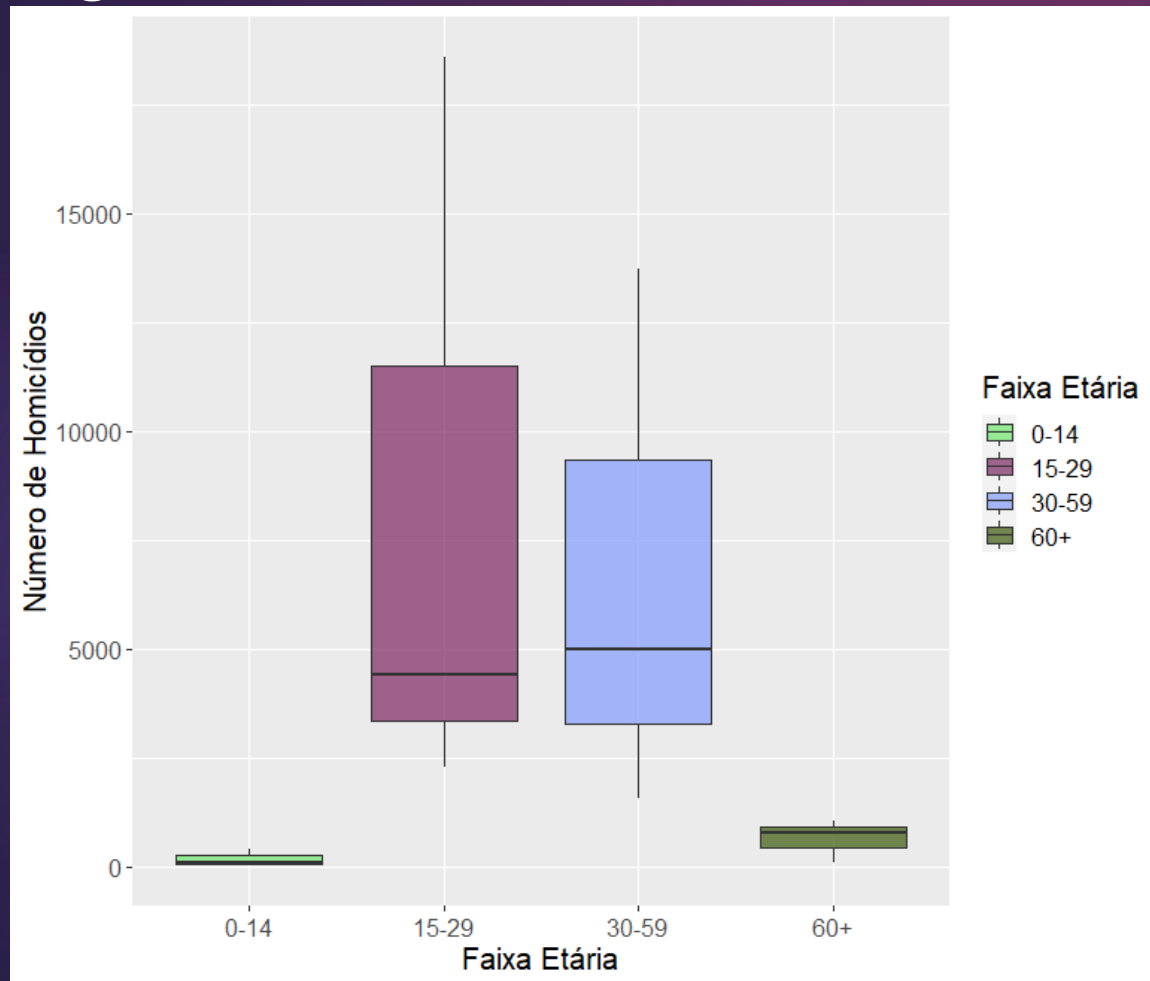
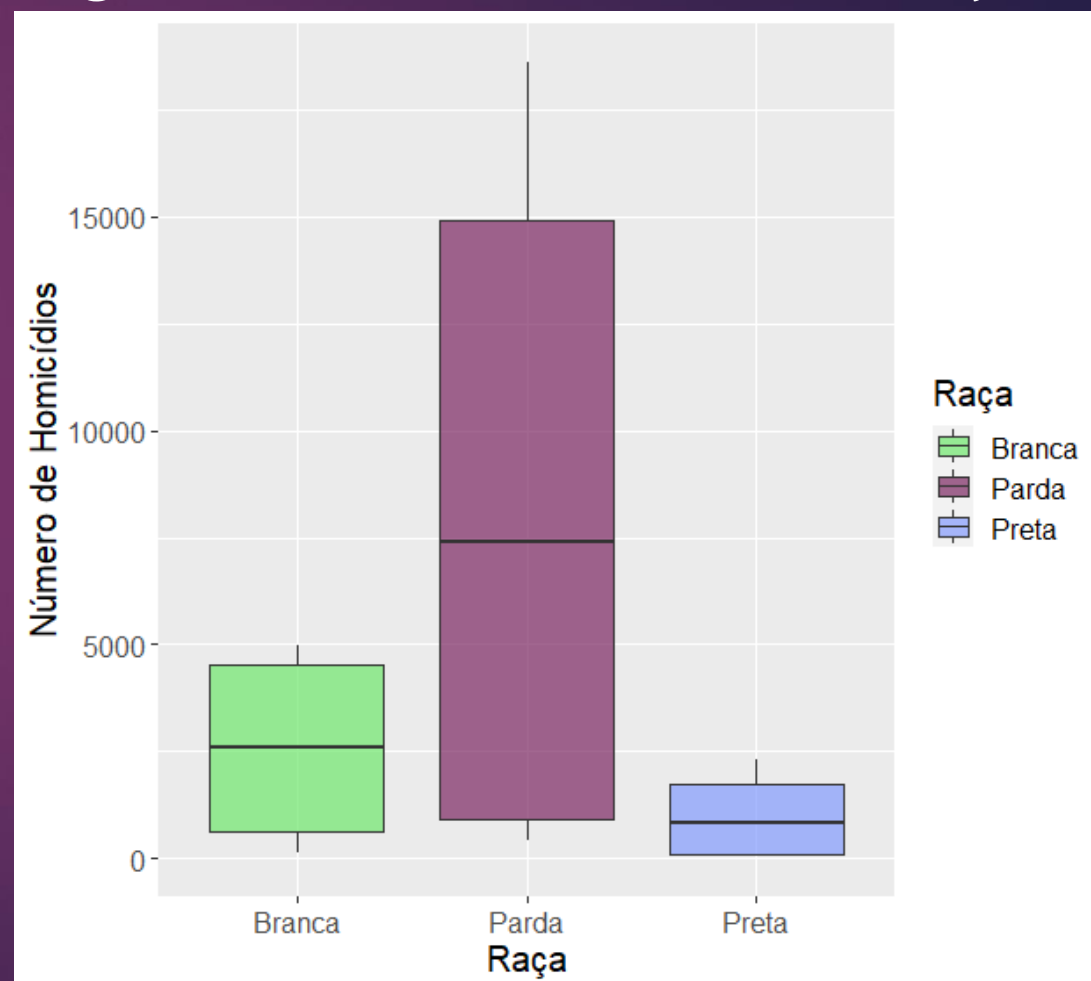


Fig. 4 - Boxplot Homicídios por Raça



Ajuste de MLG Poisson

Inicialmente, foi ajustado um MLG Poisson, apropriado para analisar dados de contagem com todas as variáveis e utilizando função de ligação logarítmica.

Tabela 2 – Estimativas para o MLG Poisson

Coeficiente	Estimativa	Erro padrão	Valor crítico	Valor-p
Intercepto	4,79	0,04	111,58	< 0,0001
Faixa Etária [15,29)	3,80	0,04	89,37	< 0,0001
Faixa Etária [30,59)	3,57	0,04	83,93	< 0,0001
Faixa Etária (60+)	1,25	0,04	26,33	< 0,0001
Raça (parda)	1,18	0,01	105,59	< 0,0001
Raça (Preta)	-0,93	0,02	-50,47	< 0,0001

AIC: 844,28 ; $\hat{\phi}$: 129,10

Desvio nulo: 78 978,01

Desvio Residual: 726,02

* obs: as estimativas são para o logaritmo dos parâmetros.

Modelo Binomial Negativo

Considere Y_1, \dots, Y_n variáveis aleatórias independentes em que $Y_i \sim BN(\mu_i, \phi)$

- ▶ $f(y_i, \mu_i, \phi) = \frac{\Gamma(\phi + y_i)}{\Gamma(\phi)\Gamma(y_i + 1)} \left(\frac{\mu_i}{\mu_i + \phi}\right)^{y_i} \left(\frac{\phi}{\mu_i + \phi}\right)^\phi$, em que μ_i e $\phi > 0$,
 $y_i = 0, 1, 2, \dots$
- ▶ $E(y_i) = \mu_i$ e $Var(y_i) = \mu_i + \frac{\mu_i^2}{\phi}$;
- ▶ Assim como o modelo Poisson, este modelo pode ser usado para analisar dados de contagem;
- ▶ Devido ao fato de a variância estar especificada em termos da média, este modelo é adequado para estudar **superdispersão**.

Modelo Ajustado

Ajustou-se um MLG Binomial Negativo aos dados, considerando:

- **Número de homicídios** como variável resposta;
- **Raça e faixa etária** como variáveis explicativas;
- Foi escolhida a raça **branca** como casela de referência da variável **raça**;
- Tomamos **0-14** como casela de referência da variável **faixa etária**;
- Como função de ligação foi usada a **função logarítmica**.

Estimativas obtidas (modelo saturado)

Tabela 3 – Estimativas do modelo binomial negativo ajustado

Coeficiente	Estimativa	Erro padrão	Valor crítico	Valor-p
Intercepto	4,90	0,16	30,73	< 0,0001
Faixa Etária [15,29)	3,86	0,18	21,27	< 0,0001
Faixa Etária [30,59)	3,64	0,18	20,07	< 0,0001
Faixa Etária (60+)	1,38	0,18	7,49	< 0,0001
Raça (parda)	0,97	0,15	6,44	< 0,0001
Raça (Preta)	-1,22	0,16	-7,87	< 0,0001

AIC: 179,75 ; $\hat{\phi}$: 22,72

Desvio nulo: 676,59

Desvio Residual: 11,74

* obs: as estimativas são para o logaritmo dos parâmetros.

Análise de Diagnóstico

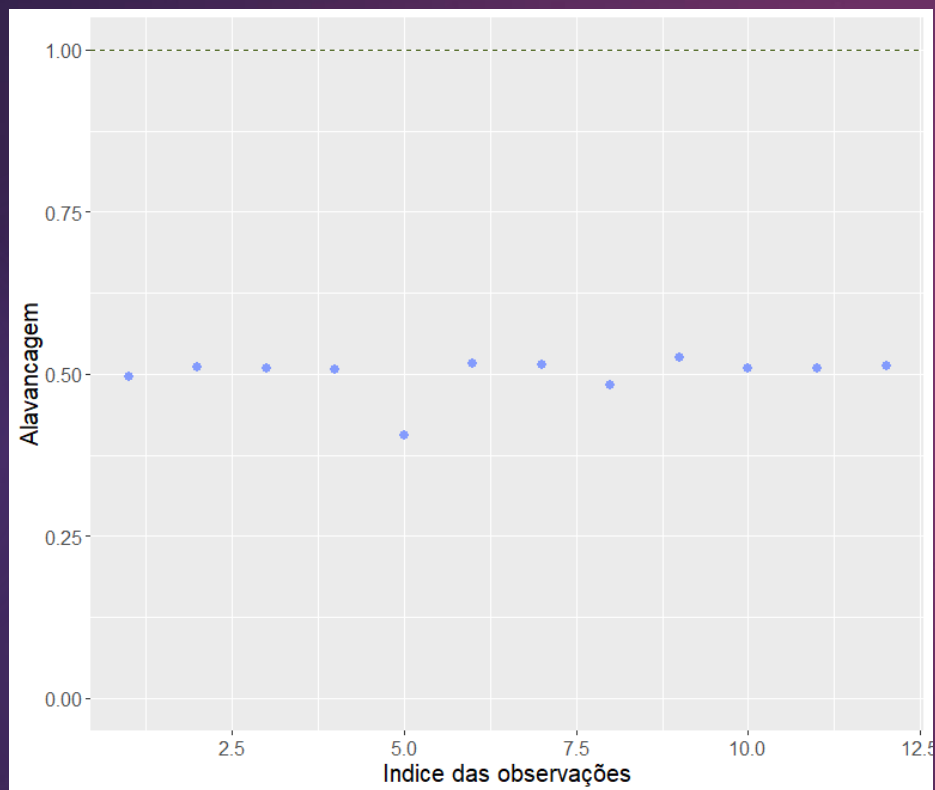


Fig. 5 - Detecção de pontos de Alavanca.

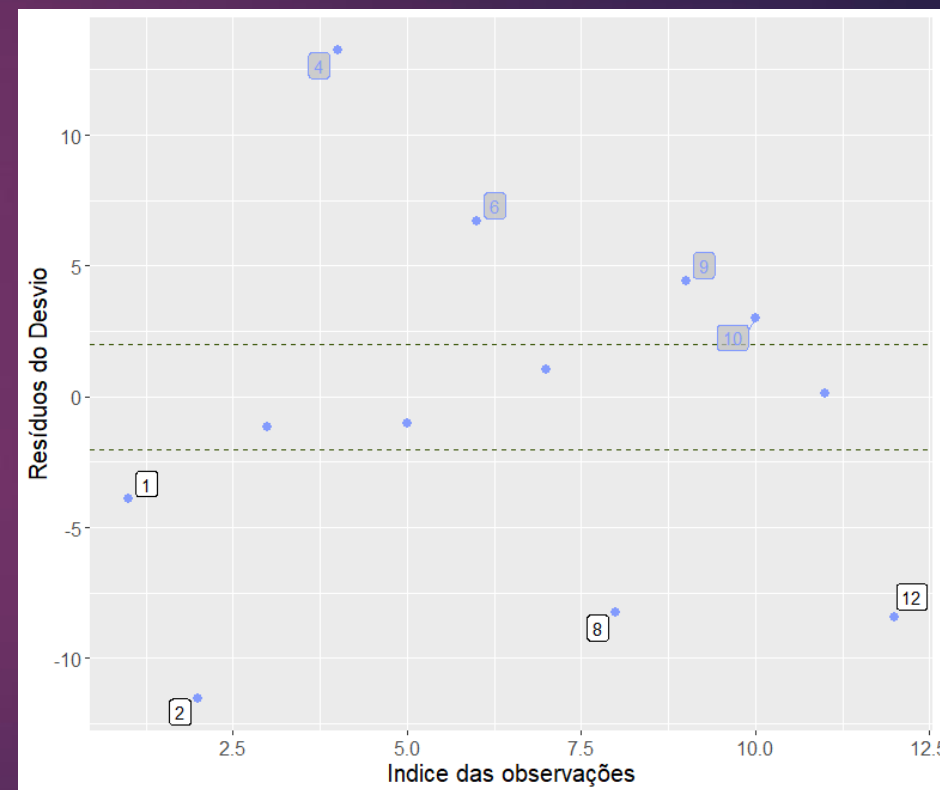


Fig. 6 - Resíduos vs. Índice das Observações

Análise de Diagnóstico

Tabela 4: Observações fora dos Limites da Fig. 6.

Índice	Nº Homicídios	F. Etária	Raça
1	118	0-14	Branca
2	4398	15-29	Branca
4	790	60+	Branca
6	2304	15-29	Preta
8	117	60+	Preta
9	409	0-14	Parda
10	18 590	15-29	Parda
12	1076	60+	Parda

Na Figura 7 nota-se que as observações 4 e 2 apresentam valores discrepantes das demais.

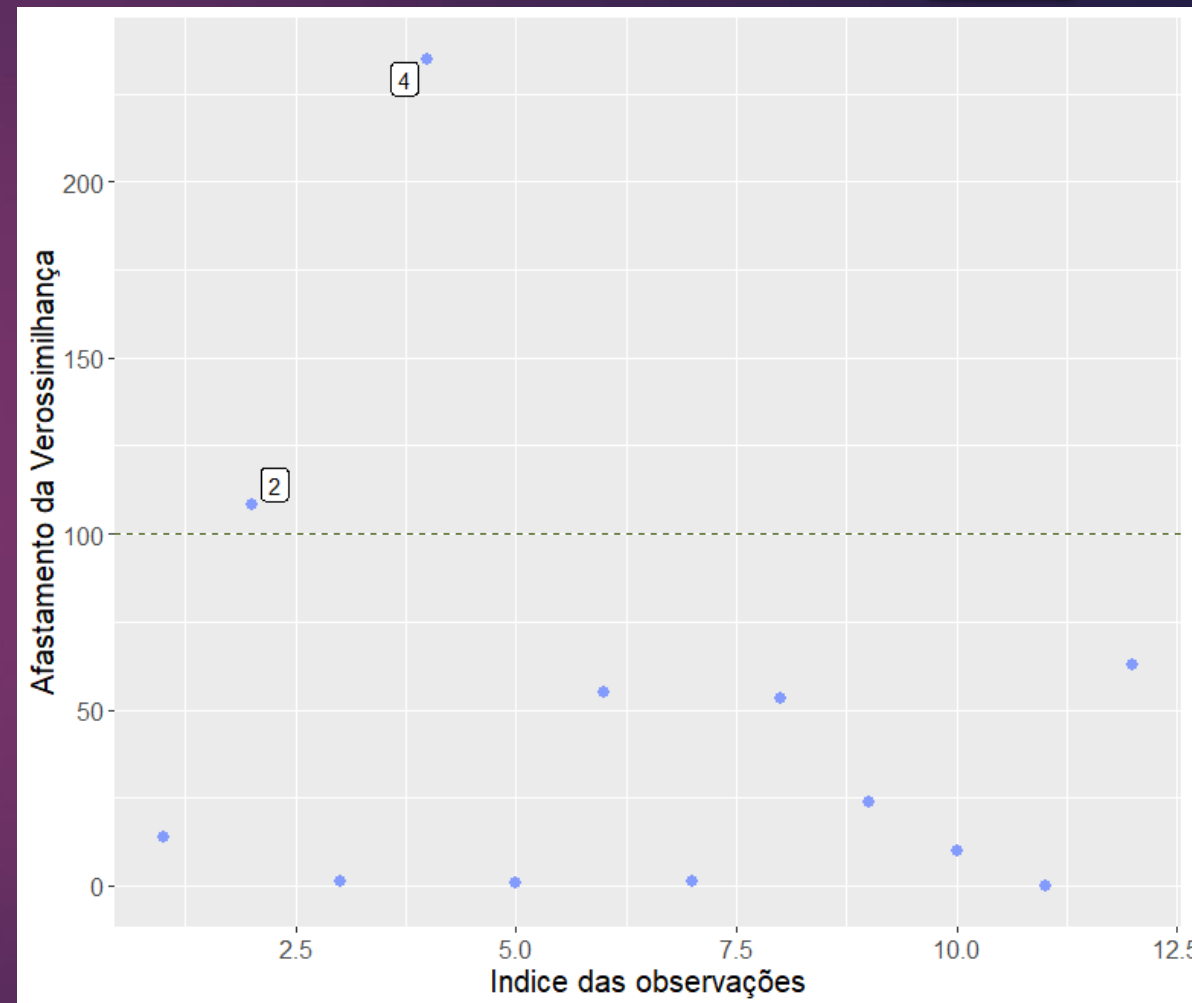


Fig. 7 – Detecção de Pontos Influentes

Análise de Diagnóstico

- Ausência de Correlação a partir da segunda observação.
- Densidade se assemelha à normalidade

Fig. 9 – Autocorrelação dos Resíduos

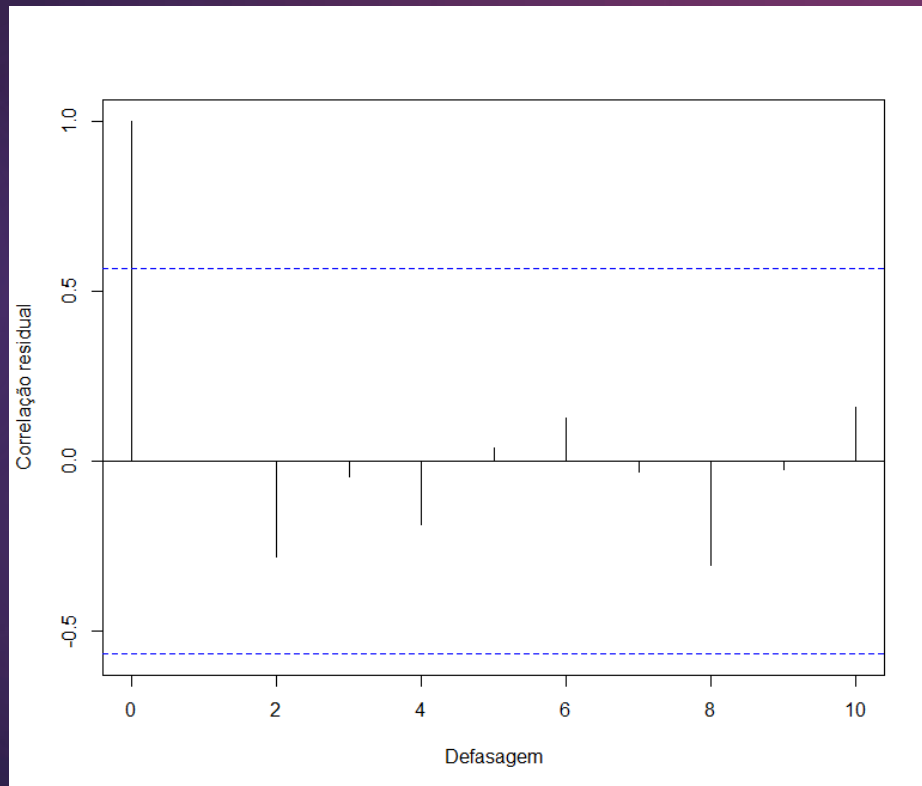
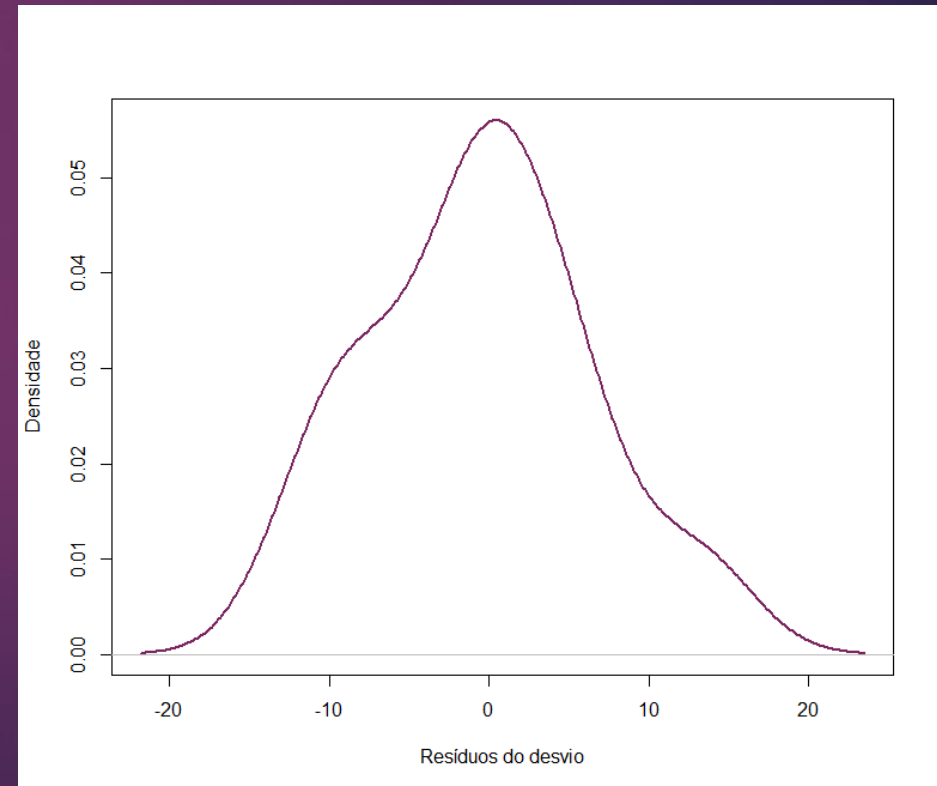
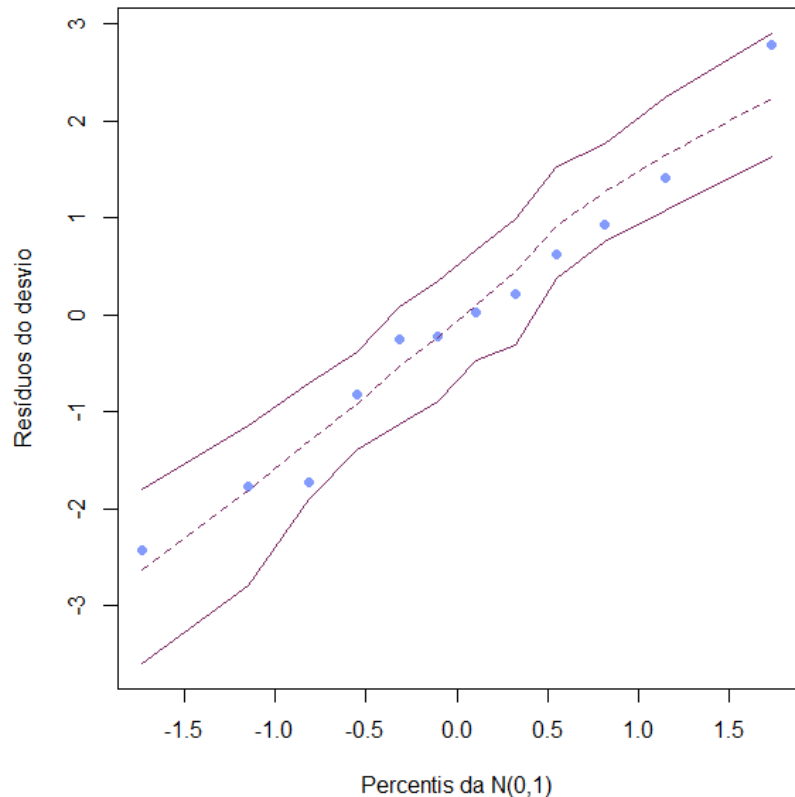


Fig. 10 – Densidade dos Resíduos



Análise de Diagnóstico



- No gráfico de envelope é possível observar que todas as observações estão contidas dentro das bandas de confiança;
- O teste de normalidade de Shapiro-Wilk resultou em um valor-p de 0,96. Considerando um nível de significância de 5% não é rejeitada a suposição de normalidade para os resíduos.

Fig. 7 - Gráfico de Envelope para os Resíduos

Tabela 5: Variação (%) dos parâmetros para os coeficientes do modelo ajustado

Observações	Intercepto	Faixa Etária [15,29)	Faixa Etária [30,59)	Faixa Etária (60+)	Raça (Parda)	Raça (Preta)
1	2,54	2,06	2,33	6,89	6,37	4,48
2	2,50	5,06	0,09	2,18	17,30	14,31
4	3,29	1,07	0,36	28,09	26,71	22,91
6	0,50	3,89	0,37	0,14	1,62	10,00
8	0,88	0,52	0,29	10,39	2,68	11,36
9	1,91	3,19	3,31	8,35	7,50	0,40
10	0,35	1,89	0,13	0,5	4,81	0,84
12	1,33	0,10	0,28	12,90	14,16	2,51

Interpretações dos Resultados

Tabela 6 – Estimativas para os parâmetros do modelo BN

Parâmetro	Estimativa
Intercepto	135,15
Faixa Etária [15,29)	47,50
Faixa Etária [30,59)	38,24
Faixa Etária (60+)	3,96
Raça (parda)	2,64
Raça (Preta)	0,29

- A taxa de homicídios de pessoas brancas e de 0-14 anos é de aproximadamente 135,15 mantidas todas as outras variáveis constantes;
- Há um aumento da taxa de homicídios quando as pessoas tem de 15 a 29 anos, comparado a pessoas de 0-14 anos.
- Há um aumento na taxa de homicídios quando as pessoas tem de 30 a 59 anos comparado a pessoas brancas de 0-14 anos.

Referências

Leão, A. T. (2023). Slides da disciplina de Modelos Lineares Generalizados. *Manaus: UFAM*.

Cordeiro, G. M., & Demétrio, C. G. (2008). Modelos lineares generalizados e extensões. *Piracicaba: USP*, 31.