

Um estudo sobre o número de mortes provindo do terrorismo nas províncias do Afeganistão

Ângelo Ghigiarelli, Clézio Lopes, Mayara Formenton

Universidade Federal de São Carlos - UFSCar
Centro de Ciências Exatas e Tecnológicas - CCET
Departamento de Estatística - DEs

22 de novembro de 2019



- 1 Introdução
- 2 Objetivo
- 3 Conjunto de dados
 - Variável Resposta
 - Variáveis preditoras
- 4 Metodologia
- 5 Fundamentação teórica
- 6 Resultados
 - Ajuste de Modelos
 - Seleção de Variáveis
 - Análise de Diagnóstico
- 7 Conclusão
- 8 Referências



O terrorismo pode ser explicado como um processo que engloba tanto as características psicológicas quanto o seu contexto político e social. John Horgan e Max Taylor (2006) defendem que este processo é constituído pelas seguintes variáveis:

- Os fatores pessoais;
- Contexto social, político e organizacional.



É possível analisar as causas associadas ao terrorismo em dois aspectos: *push* e *pull factors*.

Push factors: causas relacionadas às estruturas do ambiente que abrem espaço para o surgimento do terrorismo.

Exemplo: pobreza, desemprego ou discriminação.

Pull factors: são fatores desencadeantes - os que levam à radicalização.

Exemplo: ideologia do grupo, sensação de pertencimento ou perspectiva de fama.



De acordo com pesquisas secundárias, 75% de todos os ataques terroristas registrados no mundo se concentraram em dez países: Iraque, Afeganistão, Índia, Paquistão, Filipinas, Somália, Turquia, Nigéria, Iêmen e Síria.

Segundo especialistas, esse total pode estar subestimado, pois a *Global Terrorism Database* exclui sistematicamente os incidentes que não são noticiados pela imprensa.



Objetivo

Explicar o número de baixas (mortes) relacionadas ao terrorismo nas províncias do Afeganistão através de variáveis demográficas da região.



Os dados são referentes a 34 províncias do Afeganistão coletados de 1994 a 2008.

As variáveis preditoras são compostas por:

- **Províncias existentes no Afeganistão**
- Total de incidentes de atos terroristas
- **Total de baixas (mortes) nos atos terroristas**
- Média de hectares de cultivo de ópio
- População em cada província
- Área total de cada província em 1000s km^2
- % de áreas montanhosas por províncias



- % da população alfabetizada
- % da população com acesso a água potável
- % da população que ingere menos do que o mínimo de calorias
- % de estradas sem restrições de peso
- Taxa de mortalidade de crianças até os 5 anos
- Se a maioria da população da província é Pashtun:
$$\begin{cases} 1, & \text{Se a maioria da população da província é Pashtun} \\ 0, & \text{caso contrário} \end{cases}$$



- Média de tropas militares estrangeiras.

$$T_{i1} = \begin{cases} X_{i12}, & \text{representa a média 1225} \\ X_{i13}, & \text{representa a média 4256.2} \\ X_{i14}, & \text{representa média 5193.7} \end{cases}$$

Observação: O nível de referência representa a média 1000.



Variável resposta

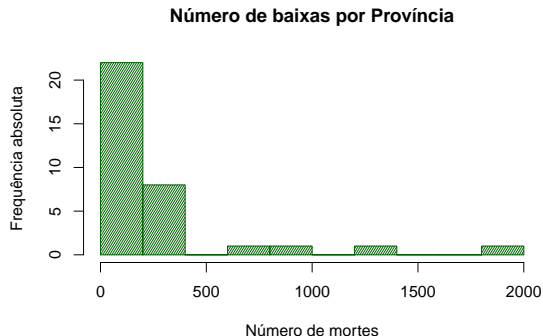


Tabela 1: Descritiva da variável resposta

Mín	0.0
1º Q.	26.5
Mediana	113.5
Média	247.4
3º Q.	265.0
Máx	1839.0
Var	153327.6
DP	391.6

Figura 1: Histograma da variável resposta (Nº de Mortes)



Relação esperada

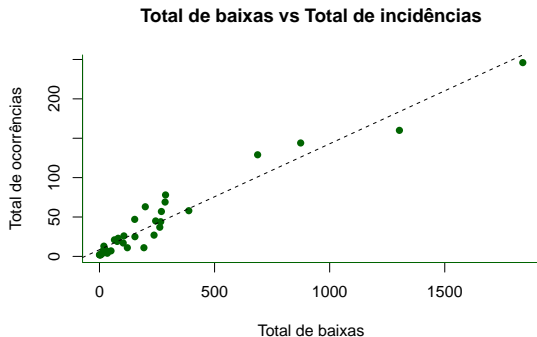


Tabela 2: Teste de correlação

Correlação	0.9686
L. Inferior	0.9376
L. superior	0.9844
p-valor	2.2e-16

Figura 2: Dispersão da variável resposta



Matriz de correlação

	Inc.	Mortes	Hect.	Pop.	Area	Mont.	Alfab.	Água	Cal.	Est.	M. Inf.	Pashtun	T. Est.
Inc.	1.00	-	-	-	-	-	-	-	-	-	-	-	-
Mortes	0.97	1.00	-	-	-	-	-	-	-	-	-	-	-
Hect.	0.39	0.36	1.00	-	-	-	-	-	-	-	-	-	-
Pop.	0.30	0.24	0.48	1.00	-	-	-	-	-	-	-	-	-
Área	0.34	0.33	0.48	0.51	1.00	-	-	-	-	-	-	-	-
Mont.	-0.40	-0.39	-0.20	-0.31	-0.29	1.00	-	-	-	-	-	-	-
Alfab.	0.06	0.15	-0.24	0.08	-0.26	0.14	1.00	-	-	-	-	-	-
Água	0.65	0.66	0.06	0.18	0.10	-0.68	0.25	1.00	-	-	-	-	-
cal.	0.09	0.08	0.26	0.07	0.38	-0.23	-0.46	-0.00	1.00	-	-	-	-
Est.	0.53	0.54	0.20	0.18	0.03	-0.65	0.17	0.64	-0.13	1.00	-	-	-
M. Inf.	-0.47	-0.45	-0.27	-0.31	-0.25	0.22	-0.40	-0.57	0.13	-0.38	1.00	-	-
Pashtun	0.61	0.50	0.25	0.10	-0.06	-0.28	-0.12	0.49	0.02	0.51	-0.38	1.00	-
T. Est.	0.20	0.13	0.23	-0.24	-0.19	0.15	-0.25	-0.01	0.23	0.15	-0.04	0.40	1.00



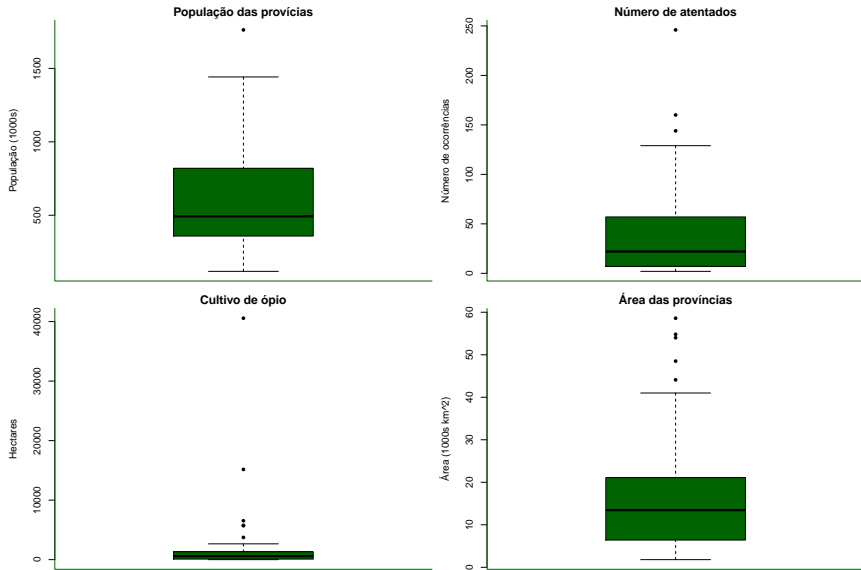


Figura 3: Análise descritiva



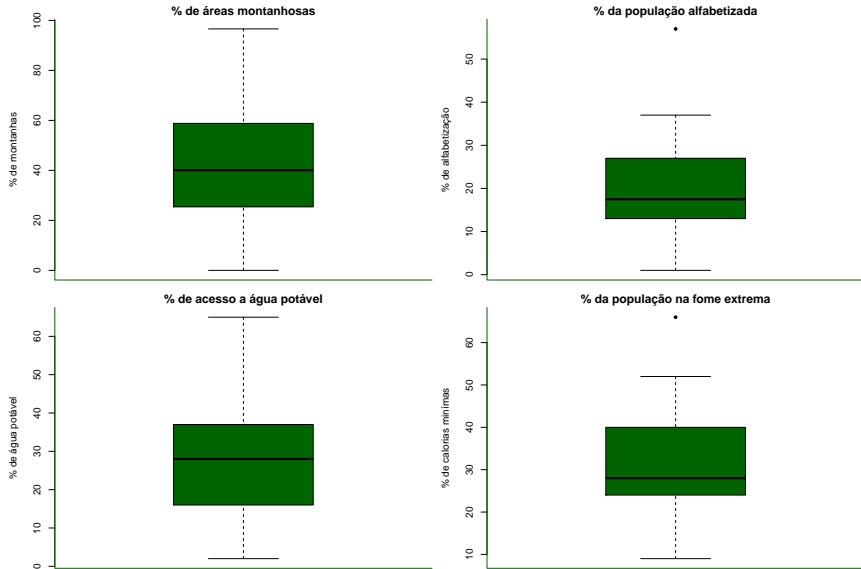


Figura 4: Análise descritiva



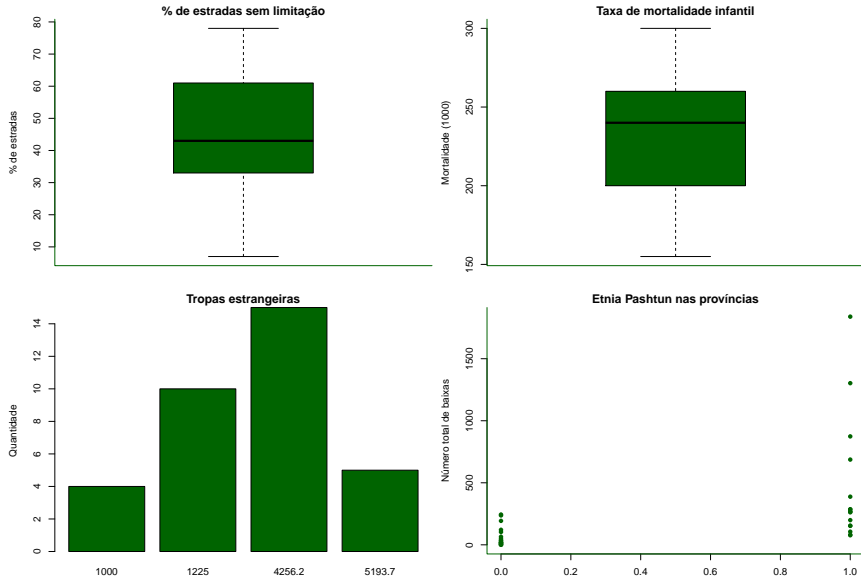


Figura 5: Análise descritiva



Será realizado um modelo linear generalizado para dados de contagem, uma vez que a variável resposta contabiliza o número de baixas em atentados terroristas.

Posteriormente, será diagnosticado se o modelo apresentado é o correto. E por fim, serão verificadas as interpretações dos resultados.



Ajuste de Modelos

Inicialmente foram ajustados três modelos e seus resultados foram os seguintes:

Tabela 3: Modelos Ajustados

Modelo	Distribuição	F. Lig.	AIC
1	B. Negativa	$\log(\mu_i)$	412.01
2	Poisson	$\sqrt{\mu_i}$	863.94
3	Poisson	$\log(\mu_i)$	1168.8



Seleção de variáveis

Utilizando-se o método de seleção de variáveis *stepwise-backward* obteve-se os seguintes resultados: **Modelo inicial, AIC = 412.01**

Tabela 4: Passos do *Stepwise*

Passo	Var. Remov.	Nome	AIC
1	X_{i12}, X_{i13} e X_{i14}	T. Est.	405.82
2	X_{i2}	Hect.	404.16
3	X_{i6}	Alfab.	402.78
4	X_{i1}	Inc.	401.84



Modelo *Stepwise*

Desta forma, um modelo inicial para trabalhar é dado por:

$$g(\mu_i) = \hat{\beta}_0 + \hat{\beta}_1 X_{i3} + \hat{\beta}_2 X_{i4} + \hat{\beta}_3 X_{i5} + \hat{\beta}_4 X_{i7} + \hat{\beta}_5 X_{i8} + \hat{\beta}_6 X_{i9} + \hat{\beta}_7 X_{i10} + \hat{\beta}_8 X_{i11},$$

sendo que:

X_{i3} = População em cada província;

X_{i4} = Área total de cada província em 1000s km^2 ;

X_{i5} = % de áreas montanhosas por província;

X_{i7} = % da população com acesso a água potável;

X_{i8} = % da população que ingere menos do que o mínimo de caloria;

X_{i9} = % de estradas sem restrições de peso;

X_{i10} = Taxa de mortalidade de crianças até os 5 anos;

X_{i11} = Se a maioria da população da província é Pashtun.



Análise de Diagnóstico

O modelo encontrado anteriormente necessita de algumas suposições para que seus resultados sejam confiáveis. É necessário atender os seguintes casos:

- $Y_i|X_i$ é independente de $Y_j|X_j$.
- ϕ não varia em função das variáveis preditoras.
- $Y_j|X_j$ tem distribuição que pertence a família exponencial linear e a correta distribuição foi escolhida para Y_j .
- A função de ligação escolhida é a correta.
- Todas as variáveis foram inseridas corretamente no modelo.



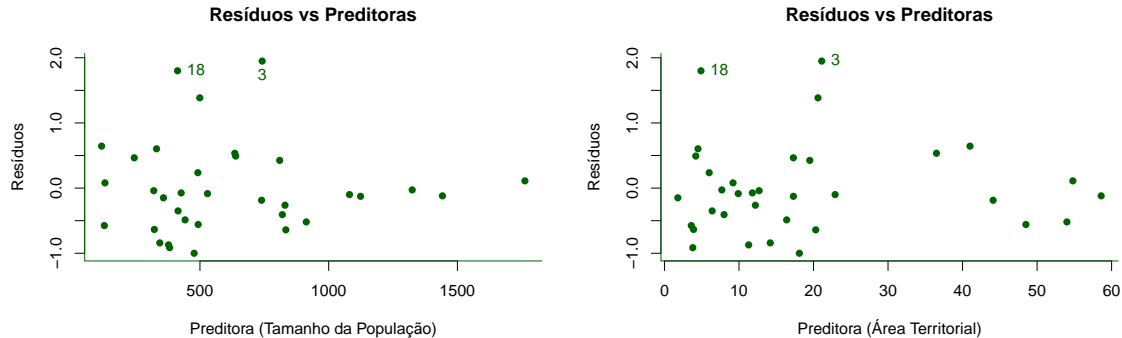


Figura 6: Análise de Diagnóstico - Resíduos vs. Predictoras



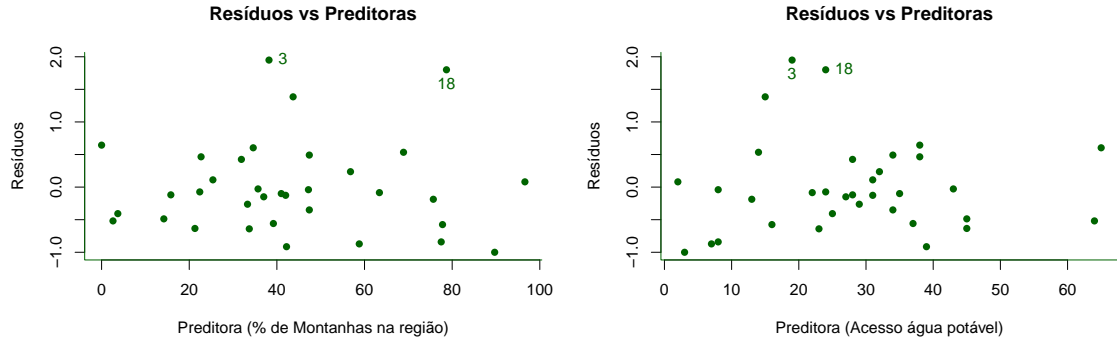


Figura 7: Análise de Diagnóstico - Resíduos vs Predictoras



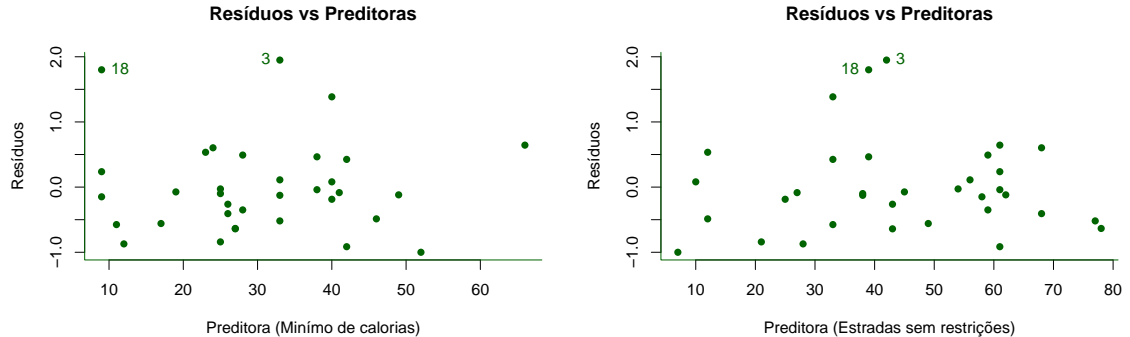


Figura 8: Análise de Diagnóstico - Resíduos vs Predictoras



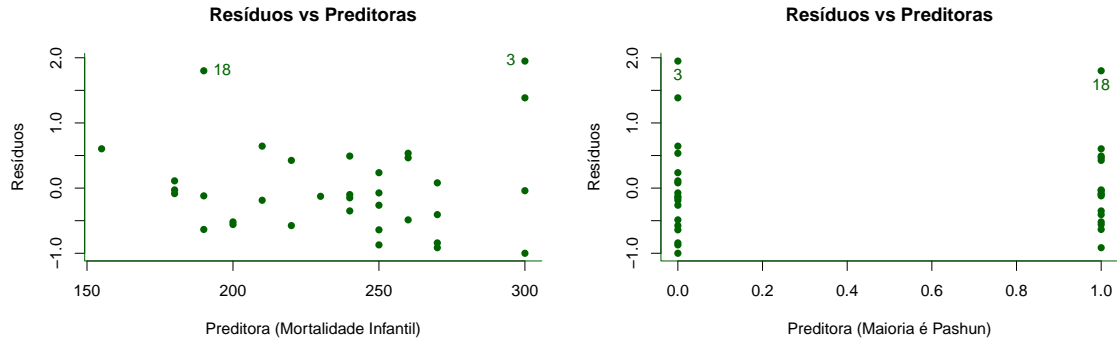


Figura 9: Análise de Diagnóstico - Resíduos vs Predictoras



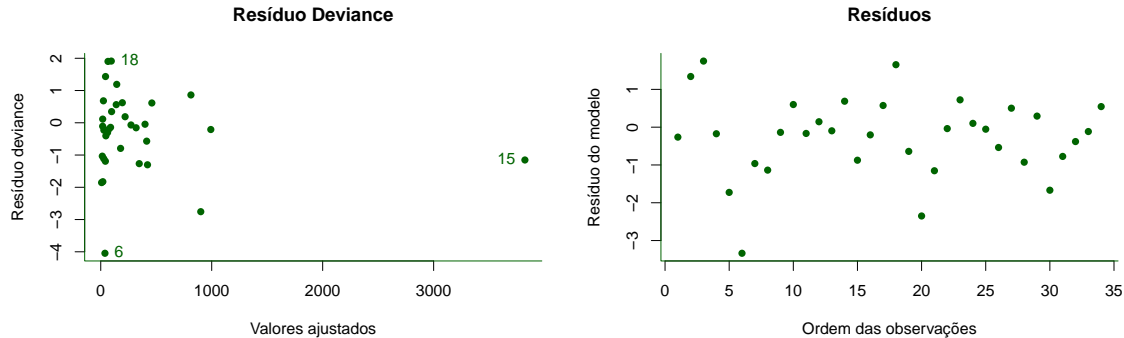


Figura 10: Análise de Diagnóstico - Tipos de Resíduos



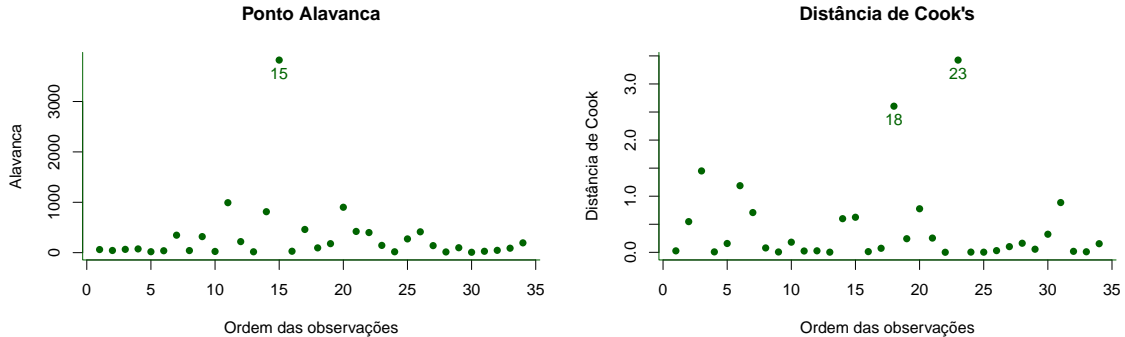


Figura 11: Análise de Diagnóstico - Alavanca e Distância de Cook's



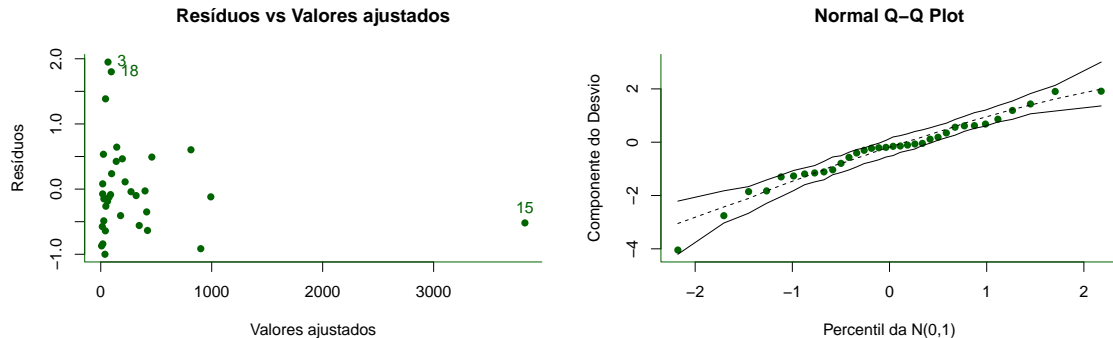


Figura 12: Análise de Diagnóstico - Envelope Simulado



Conforme **Figura 10** e **Figura 11** temos um destaque para as observações 18 e 23, pois ambas possuem alto valores de resíduos (Deviance e Pearson) e Distância de Cook's, sendo assim pode-ser que estas observações estejam interferindo negativamente no ajuste.

Para tentar identificar se de fato isso ocorre, foi ajustado um modelo sem estas observações e comparado as estimativas:



Tabela 5: Comparação das estimativas

Parâmetro	Com ambas	Sem 18	Sem 23	Sem ambas
$\hat{\beta}_0$	-4.3194	-5.9298	-5.4636	-6.1930
$\hat{\beta}_1$	0.0007	0.0009	0.0011	0.0011
$\hat{\beta}_2$	0.0221	0.0204	0.0201	0.0196
$\hat{\beta}_3$	0.0290	0.0290	0.0331	0.0306
$\hat{\beta}_4$	0.0535	0.0642	0.0601	0.0653
$\hat{\beta}_5$	0.0275	0.0385	0.0172	0.0332
$\hat{\beta}_6$	0.0416	0.0429	0.0364	0.0408
$\hat{\beta}_7$	0.0091	0.0127	0.0130	0.0137
$\hat{\beta}_8$	1.3050	1.1004	1.5368	1.2146



Sendo assim, ambas as observações 18, 23 não serão removidas da nossa base e o modelo final é dado por:

$$g(\mu_i) = \hat{\beta}_0 + \hat{\beta}_1 X_{i3} + \hat{\beta}_2 X_{i4} + \hat{\beta}_3 X_{i5} + \hat{\beta}_4 X_{i7} + \hat{\beta}_5 X_{i8} + \hat{\beta}_6 X_{i9} + \hat{\beta}_7 X_{i10} + \hat{\beta}_8 X_{i11}$$



As estimativas do modelo mencionado acima é dado por:

Tabela 6: Estimação dos Parâmetros

Variável	Estimativa	Std. Error	t value	p-valor
Intercepto	-4.3294	2.0366	-2.121	0.0339
X_{i3}	0.0007	0.0004	1.637	0.1016
X_{i4}	0.0221	0.0108	2.042	0.0412
X_{i5}	0.0290	0.0100	2.900	0.0037
X_{i7}	0.0534	0.0162	3.291	0.0009
X_{i8}	0.0275	0.0123	2.235	0.0253
X_{i9}	0.0416	0.0112	3.715	0.0002
X_{i10}	0.0091	0.0049	1.872	0.0612
X_{i11}	1.3050	0.3325	3.924	8.69e-05



Interpretação dos Parâmetros

$\exp(\hat{\beta}_1) = \exp(0.0007) = 1.0007$ Estima-se, em média, que o total de baixas relacionadas ao terrorismo nas províncias do Afeganistão sofre um acréscimo de 0.07% quando acrescenta-se uma unidade na variável população, mantida as demais variáveis constantes.

$\exp(\hat{\beta}_2) = \exp(0.0221) = 1.0223$ Estima-se, em média, que o total de baixas relacionadas ao terrorismo nas províncias do Afeganistão sofre um acréscimo de 2.23% quando acrescenta-se uma unidade na variável área total de cada província, mantida as demais variáveis constantes.

$\exp(\hat{\beta}_3) = \exp(0.0290) = 1.0295$ Estima-se, em média, que o total de baixas relacionadas ao terrorismo nas províncias do Afeganistão sofre um acréscimo de 2.95% quando acrescenta-se um ponto percentual na variável % de áreas montanhosas por província, mantida as demais variáveis constantes.



$\exp(\hat{\beta}_4) = \exp(0.0534) = 1.0549$ Estima-se, em média, que o total de baixas relacionadas ao terrorismo nas províncias do Afeganistão sofre um acréscimo de 5.49% quando acrescenta-se um ponto percentual na variável % da população com acesso a água potável, mantida as demais variáveis constantes.

$\exp(\hat{\beta}_5) = \exp(0.0275) = 1.0279$ Estima-se, em média, que o total de baixas relacionadas ao terrorismo nas províncias do Afeganistão sofre um acréscimo de 2.79% quando acrescenta-se um ponto percentual na variável % da população que ingere menos do que o mínimo de caloria, mantida as demais variáveis constantes.

$\exp(\hat{\beta}_6) = \exp(0.0416) = 1.0425$ Estima-se, em média, que o total de baixas relacionadas ao terrorismo nas províncias do Afeganistão sofre um acréscimo de 4.25% quando acrescenta-se um ponto percentual na variável % de estradas sem restrições de peso, mantida as demais variáveis constantes.



$\exp(\hat{\beta}_7) = \exp(0.0091) = 1.0092$ Estima-se, em média, que o total de baixas relacionadas ao terrorismo nas províncias do Afeganistão sofre um acréscimo de 0.92% quando acrescenta-se uma unidade na variável taxa de mortalidade de crianças até os 5 anos, mantida as demais variáveis constantes.




$\exp(\hat{\beta}_8) = \exp(1.3050) = 3.6876$ Estima-se, em média, que o total de baixas relacionadas ao terrorismo nas províncias do Afeganistão sofre um acréscimo de 268.76% quando a maioria da população da província não é Pashtun e passa a ser, mantida as demais variáveis constantes.



Desta forma o objetivo foi atendido, pois foi possível relacionar o número de baixas provindos de atentados terroristas no Afeganistão por meio de um modelo linear generalizado com resposta binomial negativa e função de ligação logarítmica.

Todas as suposições necessárias para a continuidade desse tipo de análise foram atendidas. Sendo assim, o modelo final é considerado como correto e ainda produziu parâmetros interpretáveis.



-  Teodoro, Ana Cláudia S. C., Ramalho, Diana Gonçalves, "O combate ao terrorismo: uma análise das principais causas e das estratégias e modelos contraterroristas" - 2017.
-  Piazza, James A. 2012. "O comércio de ópio e os padrões de terrorismo nas províncias do Afeganistão: uma análise empírica". Terrorismo e violência política (24 março): 213-234.
-  Pereira, Gustavo. Notas de Aula - Modelos Lineares Generalizados. 2019.

