



Universidade de Brasília  
Departamento de Estatística

Modelagem do número de reincidência criminal feminina no regime  
carcerário do Distrito Federal via regressão binomial negativa

Raíssa Alvim Teixeira

Projeto apresentado para o Departamento  
de Estatística da Universidade de Brasília  
como parte dos requisitos necessários para  
obtenção do grau de Bacharel em Es-  
tatística.

Brasília  
2023

**Raíssa Alvim Teixeira**

**Modelagem do número de reincidência criminal feminina no regime  
carcerário do Distrito Federal via regressão binomial negativa**

Orientador(a): Eduardo Yoshio Nakano  
Coordenador(a): Leandro Tavares Correia

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília  
2023**

---

Dedico este estudo à minha querida mãe Márcia Alvim Pereira Teixeira (in memoriam), ao meu querido pai Gil Teixeira Filho, às minhas queridas irmãs Letícia Alvim Teixeira e Thaís Alvim Teixeira, que nunca mediram esforços para o meu crescimento acadêmico e pessoal e sempre me apoiaram em todas as minhas decisões da vida.

---

# Agradecimentos

Expresso minha imensa gratidão a todos os profissionais, funcionários e estudantes da Universidade de Brasília que de alguma forma me deram apoio durante minha trajetória na Instituição e acreditaram em mim.

---

# Resumo

O estudo teve como objetivo verificar possíveis fatores que possam influenciar a reincidência de jovens infratoras no Distrito Federal a fim de entender a volta delas às unidades prisionais em um dado período de tempo. Utilizando o modelo de regressão binomial negativo, foi verificado que características como idade e composição do domicílio onde reside a jovem foram fatores significativos na reincidência criminal entre jovens mulheres no Distrito Federal.

Palavras-chave: Regressão Poisson, regressão binomial negativa, jovens infratoras, reincidência criminal.

# Sumário

<b>1 Introdução</b> . . . . .	8
<b>2 Metodologia</b> . . . . .	10
2.1 Regressão Poisson . . . . .	10
2.2 Regressão Binomial Negativa . . . . .	11
2.3 Função de verossimilhança . . . . .	12
2.4 Teste Wald . . . . .	13
2.5 Análise de resíduo . . . . .	14
2.5.1 Resíduo Bruto . . . . .	14
2.5.2 Resíduo de Pearson . . . . .	14
2.5.3 Resíduo de <i>deviance</i> . . . . .	15
<b>3 Resultados</b> . . . . .	16
3.1 Análise Exploratória . . . . .	16
3.2 Correlação de Pearson Para as Variáveis <i>Reincidência</i> e <i>Idade</i> . . . . .	22
3.3 Ajuste no modelo Poisson . . . . .	23
3.4 Ajuste no modelo Binomial Negativo. . . . .	26
3.5 Discussão dos resultados. . . . .	28
<b>4 Conclusão</b> . . . . .	30
<b>5 Referências</b> . . . . .	31

# 1 Introdução

Historicamente, em muitas sociedades, o crime feminino foi relatado em menor incidência em comparação ao masculino. No entanto, segundo Assis e Constantino (2001), o número de crimes cometidos por mulheres e adolescentes do sexo feminino vem crescendo ao longo do tempo.

Em se tratando da análise da criminalidade feminina, existem diversos contextos onde mudanças nas normas sociais e a percepção à resposta em relação ao crime feminino se alteram ao longo do tempo.

Alguns fatores como mudanças nas aplicações das leis, mudança no tratamento das mulheres infratoras, desigualdades de salário e oportunidades, rompimento de laços afetivos, falta da figura paterna, rompimento de estigmas associados às mulheres, afastamento da escola e o uso de drogas, são alguns dos motivos que podem desencadear no aumento dessa tendência associada a mulheres que cometem um crime.

As peculiaridades e o crescimento desse fenômeno têm contribuído para que, em diversos países, se amplie o interesse científico por esse tema (Constantino, 2001; Paiva, 2007).

De acordo com o Panorama Nacional de Medidas Socioeducativas de Internação (BRASIL, 2012), 54% dos adolescentes que cumpriam medida de internação em toda região Centro-Oeste no ano de 2012 eram reincidentes no ato criminal. As medidas sócio-educativas aplicadas nos adolescentes que se encontravam em regime de internação deveriam servir para educar esses adolescentes a fim de retornarem à família, escola e sociedade, repensando e mudando seus atos. Entretanto, observa-se uma ausência de acolhimento e assistência adequada para que esses jovens mudem seus comportamentos (BRASIL, 2012).

Neste contexto, o objetivo deste trabalho é identificar quais são os perfis de jovens com maior risco de reincidência em torno do Distrito Federal, a fim de entender a volta delas às unidades prisionais em um dado período de tempo. Essas informações poderão guiar um plano mais adequado de ressocialização dessas jovens na sociedade, uma vez que esse conhecimento se torna crucial para o desenvolvimento de políticas públicas mais eficazes.

Na busca por uma compreensão mais profunda dos fatores associados à reincidência criminal entre mulheres infratoras, como também na formulação de estratégias e

políticas mais eficazes para a ressocialização dessas mulheres na sociedade, foram escolhidos modelos que permitissem quantificar a relação entre as variáveis independentes obtidas no estudo e a reincidência criminal, identificando quais fatores têm maior influência no aumento do risco da reincidência entre essas mulheres.

A escolha das regressões binomial negativa e poisson puderam analisar a relação entre múltiplas variáveis independentes e a reincidência, uma vez que, a reincidência criminal pode não ser um evento comum e, portanto, variar em diferentes contextos.

A análise estatística foi realizada utilizando o programa R Core Team 2023. Foram utilizados o modelo de regressão de Poisson e o modelo de regressão Binomial Negativa para quantificar a relação entre as variáveis independentes e a reincidência das jovens.

Espera-se que este estudo forneça *insights* sobre os fatores que contribuem para a reincidência criminal em mulheres infratoras. Os resultados podem ser utilizados para orientar no desenvolvimento de estratégias de intervenção de políticas públicas mais direcionadas às questões sociais associadas a essas mulheres.

## 2 Metodologia

A aplicação de modelos de regressão, como os modelos de Poisson e Binomial Negativa, é essencial para compreender e prever fenômenos relacionados a uma ocorrência de eventos. Sendo assim, é fundamental delinear o processo metodológico detalhado para a aplicação e interpretação desses modelos, garantindo uma análise robusta e precisa das relações entre as variáveis estudadas.

O modelo de Poisson é comumente usado para análise de dados de contagem, como o número de eventos que ocorrem em um intervalo de tempo fixo ou em uma área específica. É útil quando se trata de eventos discretos e não relacionados, como contagem de acidentes, número de ligações recebidas em um call center, entre outros.

Já a regressão Binomial Negativa é mais frequentemente utilizada quando se está lidando com proporções ou taxas de sucesso em experimentos com um número fixo de tentativas. Embora possa parecer semelhante à regressão de Poisson, é especialmente útil quando há superdispersão nos dados binomiais, ou seja, uma variabilidade maior do que o esperado.

### 2.1 Regressão Poisson

A regressão de Poisson também é conhecida como Modelo Log-Linear de Poisson, faz parte da família de Modelos Lineares Generalizados (MLG) e é adequada para a modelagem de variáveis que envolvam dados de contagem ou taxas.

A distribuição de Poisson pode ser escrita como:

$$P(Y = y) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, 2, \dots \quad (2.1.1)$$

em que  $Y$  é a variável aleatória que representa o número de ocorrências e  $\mu$  é o parâmetro que representa o valor esperado de  $Y$ .

No modelo de regressão linear normal, o valor médio da resposta na presença de  $k$  variáveis explicativas,  $x' = (1, x_1, x_2, \dots, x_k)$  é dado por:  $E(y|x) = x'\beta$ , em que  $\beta = (\beta_0, \beta_1, \dots, \beta_k)$  é o vetor dos parâmetros desconhecidos.

No entanto, essa mesma representação não é possível no modelo de regressão de Poisson, visto que sua média  $\mu$  é positiva. Sabendo que a função logarítmica é a função de

ligação natural para o modelo Poisson, o modelo log-linear é considerado:

$$\log(\mu) = \beta_0 + \sum_{j=1}^k \beta_j x_j \quad (2.1.2)$$

que resulta em:

$$\mu = \mu(x) = e^{x'\beta} \quad (2.1.3)$$

em que  $\beta' = (\beta_0, \beta_1, \dots, \beta_k)$  é vetor de parâmetros associado ao vetor de covariáveis  $x' = (1, x_1, x_2, \dots, x_k)$ .

A função de ligação é fundamental em modelos lineares generalizados, como a regressão de Poisson. Ela descreve a relação entre a média da variável de resposta (nesse caso, o número de ocorrências) e as variáveis explicativas (ou preditoras).

No caso da regressão de Poisson, a função de ligação é a função logarítmica. Isso significa que, para modelar como as variáveis explicativas afetam a média da variável de resposta (representada por  $\mu$ ), usamos o logaritmo natural.

O modelo log-linear é uma forma de expressar essa relação. A equação  $\log(\mu) = \beta_0 + \sum_{j=1}^k \beta_j x_j$  mostra como os valores das variáveis explicativas  $x_1, x_2, \dots, x_k$  estão relacionados ao logaritmo da média da variável de resposta.

Quando aplicamos a função exponencial a essa equação, obtemos a expressão para a média  $\mu$ , que é  $\mu = e^{x'\beta}$ . Aqui,  $e$  é a base do logaritmo natural,  $x'$  é o vetor de covariáveis e  $\beta$  é o vetor de parâmetros.

Essa relação log-linear e a função de ligação logarítmica permitem interpretar o efeito das variáveis explicativas na média das ocorrências. Quando os coeficientes  $\beta$  são positivos, o aumento em uma variável explicativa está associado a um aumento proporcional na média do número de ocorrências ( $\mu$ ), e vice-versa para coeficientes negativos.

Essa função de ligação é crucial para entender como as variáveis independentes influenciam a variável de contagem ( $Y$ ) na regressão de Poisson.

## 2.2 Regressão Binomial Negativa

A distribuição binomial negativa é uma generalização da distribuição Geométrica. As formas especiais desta distribuição surgiram, em 1679, com Pascal e Fermat. A distribuição Binomial Negativa pode ser utilizada para a modelagem de variáveis aleatórias que representam uma contagem, sendo uma boa alternativa em relação à distribuição de

Poisson. A distribuição Binomial Negativa pode ser escrita como:

$$P(Y = y) = \frac{\Gamma(y + k)}{\Gamma(k)y!} p^y (1 - p)^y, \quad y = 0, 1, 2, \dots$$

em que  $Y$  é a variável aleatória que representa o número de ocorrências e  $p$  ( $0 < p < 1$ ) e  $k$  ( $k > 0$ ) são os parâmetros da distribuição. Note que, quando  $k$  é um inteiro, a distribuição binomial negativa também é denominada como Pascal e representa o número de tentativas (independentes e identicamente distribuídas) fracassadas até a ocorrência do  $k$ -ésimo sucesso.

No modelo de regressão binomial negativa, o efeito das variáveis explicativas na variável  $Y$  é modelado através do parâmetro  $p$ . Visto que  $0 < p < 1$ , as covariáveis são associadas por meio da função de ligação logito, isto é:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{j=1}^k \beta_j x_j,$$

que resulta em:

$$p = \frac{e^{\beta_0 + \sum_{j=1}^k \beta_j x_j}}{1 + e^{\beta_0 + \sum_{j=1}^k \beta_j x_j}}.$$

em que  $\beta' = (\beta_0, \beta_1, \dots, \beta_k)$  é o vetor de parâmetros associado ao vetor de covariáveis  $x' = (1, x_1, \dots, x_k)$ .

### 2.3 Função de verossimilhança

Ao considerar uma amostra aleatória, a função de verossimilhança para  $n$  observações do modelo de regressão de Poisson é dada por:

$$\ell(\beta) = \prod_{i=1}^n \frac{\mu_i e^{-\mu_i}}{y_i!}. \quad (2.3.1)$$

O logaritmo da verossimilhança para esse modelo é:

$$\ell(\beta) = \log[L(\beta)] = \sum_{i=1}^n (y_i \log(\mu(x_i)) - \mu(x_i) - \log(y_i!)) \quad (2.3.2)$$

Segundo a ligação dada por (2.1.3):

$$\ell(\beta) = \sum_{i=1}^n (y_i x_i' \beta - e^{x_i' \beta} - \ln(y_i!)) \quad (2.3.3)$$

O Estimador de Máxima Verossimilhança (EMV) de  $\beta$  da equação é obtido resolvendo o seguinte sistema de equações:

$$\begin{cases} \frac{\partial \ell(\beta)}{\partial \beta_0} = 0 \\ \frac{\partial \ell(\beta)}{\partial \beta_1} = 0 \\ \dots \\ \frac{\partial \ell(\beta)}{\partial \beta_k} = 0 \end{cases}$$

cuja solução pode ser obtida por meio de métodos numéricos como, por exemplo, o de Newton-Raphson ou outros métodos computacionais.

## 2.4 Teste Wald

O teste de Wald tem como objetivo avaliar a relação de significância de cada variável explicativa incluída no modelo. O teste de Wald testa a seguinte hipótese:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0, \text{ com } j = 0, 1, \dots, k \end{cases} \quad (2.4.1)$$

Sendo a estatística do teste, sob  $H_0$ :

$$W_j = \frac{\hat{\beta}_j}{\sqrt{\widehat{\text{var}}(\hat{\beta}_j)}} \sim \mathcal{N}(0, 1). \quad (2.4.2)$$

Em que:

- $\hat{\beta}_j$  é o estimador de máxima verossimilhança para  $\beta_j$ ;
- $\widehat{\text{var}}(\hat{\beta}_j)$  é o estimador de máxima verossimilhança para a variância de  $\hat{\beta}_j$ ;
- Sob  $H_0$ , a estatística de Wald segue aproximadamente uma distribuição normal padrão  $\mathcal{N}(0, 1)$ .

Essa expressão sugere que a estatística de teste  $W_j$ , relacionada à razão entre o estimador  $\hat{\beta}_j$  e o desvio padrão estimado de  $\hat{\beta}_j$ , segue aproximadamente uma distribuição normal com média zero e variância um. Isso é válido para amostras grandes de acordo com o teorema do limite central.

## 2.5 Análise de resíduo

A análise de resíduos é uma etapa crítica na validação de modelos estatísticos, pois permite verificar se as suposições do modelo estão sendo atendidas e se há estrutura ou padrões nos dados que não foram capturados pelo modelo. As sessões seguintes apresentam os resíduos bruto, de Pearson e de *deviance* aplicado ao modelo de regressão Poisson.

### 2.5.1 Resíduo Bruto

O resíduo bruto é a diferença entre a resposta observada e a estimativa do valor do modelo. A fórmula para o resíduo bruto estimado é:

$$\hat{r}_i = y_i - \exp(X_i \hat{\beta}) \quad (2.5.1)$$

Em que:

- $y_i$  é a resposta observada do  $i$ -ésimo indivíduo, onde  $i = 1, 2, \dots, n$ ;
- $X_i$  é o vetor de covariáveis do indivíduo  $i$ ;
- $\hat{\beta}$  é o Estimador de Máxima Verossimilhança (EMV) dos coeficientes do modelo.

### 2.5.2 Resíduo de Pearson

O resíduo de Pearson corrige a diferença das variâncias do resíduo bruto. E dado por:

$$\hat{p}_i = \frac{\hat{r}_i}{\hat{\phi} \exp(X_i \hat{\beta})} \quad (2.5.2)$$

Em que:

- $\hat{\phi} = \frac{1}{n-u} \sum_{i=1}^n \frac{(y_i - \exp(X_i \hat{\beta}))^2}{\exp(X_i \hat{\beta})}$  é o parâmetro de dispersão para controlar a super dispersão;
- $\hat{r}_i$  é o resíduo bruto apresentado em (2.1.13).

O resíduo de Pearson pode ser calculado por  $\hat{y}_i$  em relação a  $\hat{\phi}_i$ . Se o modelo estiver correto, os resíduos de Pearson devem apresentar uma dispersão constante, indicando uma boa adequação do modelo aos dados. No entanto, a presença de um padrão evidente

no gráfico, como um formato de funil ou qualquer padrão não aleatório, pode indicar uma violação da suposição de homogeneidade de variância. Isso sugere que o modelo pode necessitar de melhorias ou ajustes.

### 2.5.3 Resíduo de *deviance*

Os resíduos *deviance* são úteis na identificação de padrões de covariáveis não ajustados. A *deviance* do modelo é uma estatística de qualidade do ajuste que está baseada na função de log-verossimilhança. Está definido como:

$$d_j = \text{sign}(y_i - e^{x_i \hat{\beta}}) \sqrt{D_i} \quad (2.5.3)$$

Em que:

- $y_i$  é o valor da resposta do  $i$ -ésimo padrão de covariáveis;
- $e^{x_i \hat{\beta}}$  é o valor ajustado do  $i$ -ésimo termo de covariáveis;
- $D_i$  é a deviance do  $i$ -ésimo termo em (2.1.8).

Os resíduos de deviance são uma métrica que avalia a discrepância entre a probabilidade observada e a probabilidade prevista pelo modelo. Se esses resíduos se comportarem aleatoriamente em torno de zero, não mostrando um padrão claro à medida que os valores ajustados mudam, isso indica que o modelo está se ajustando bem aos dados. Por outro lado, se houver um padrão ou tendência nos resíduos conforme os valores ajustados mudam, pode sugerir que o modelo não está capturando completamente a estrutura dos dados.

### 3 Resultados

Esta aplicação foi realizada por meio do banco de dados do estudo realizado por Santos (2013), que avaliou o perfil de adolescentes do sexo feminino que estavam em conflito com as leis no Distrito Federal. A amostra é composta por 284 adolescentes do sexo feminino residentes no Distrito Federal que cumpriram medida socioeducativa de internação na Unidade de Internação do Plano Piloto – UIPP, entre os anos de 2004 e 2011. Posteriormente, Ribeiro (2015). Avaliou esse mesmo banco de dados, modelando a probabilidade de reincidência por meio da regressão logística. Neste trabalho, a o número de reincidências foi modelada por meio dos modelos de regressão Poisson e Binomial negativa.

#### 3.1 Análise Exploratória

Com base nas estatísticas, a maioria das jovens analisadas no estudo não apresentaram reincidência criminal, mas há algumas em que a reincidência foi relativamente alta. A distribuição verificada foi assimétrica, com uma inclinação à direita e caudas pesadas. O coeficiente de curtose sugere que a distribuição possui picos mais acentuados do que a distribuição normal.

A Tabela 1 a seguir fornece uma visão abrangente das características centrais e de dispersão da variável de uma reincidência no conjunto de dados. Um número  $n$  de 284 observações de casos de reincidência criminal feminina, com uma média de pontuação de reincidência igual a 1,05 e um desvio padrão de 1,61.

Tabela 1: Estatísticas descritivas para a reincidência criminal entre mulheres entre 2004 e 2011

	$n$	Média	$Sd$	Mediana	Mín	Máx	Amp.	$Cs$	$Ck$	$Se$
Reincidência	284	1.05	1.61	0	0	8	8	1.90	3.43	0.10

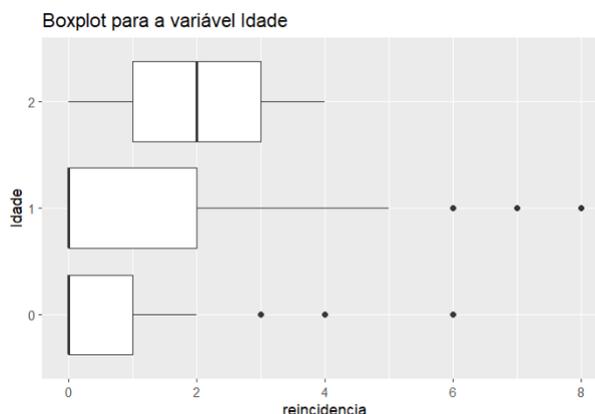
$n$ : Número de observações;  $Sd$ : Desvio-padrão;  $Amp.$ : Amplitude;  $Cs$ : Coeficiente de assimetria;  $Ck$ : Coeficiente de curtose;  $Se$ : Erro padrão da média

A mediana de 0 indica que metade das mulheres tem uma pontuação de reincidência menor ou igual a 0, e a outra metade tem uma pontuação maior que 0. O valor mínimo (Mín) de 0 e o valor máximo (Máx) de 8 indicam a faixa das pontuações de reincidência. A amplitude *Amp.* de 8 sugere a extensão total das pontuações de reincidência no conjunto de dados.

O coeficiente de assimetria *C<sub>s</sub>* de 1.90 indica uma assimetria positiva, sugerindo que há uma cauda mais longa à direita na distribuição e o erro padrão da média *Se* de 0.10 é uma estimativa da precisão da média amostral em relação à verdadeira média populacional.

A Figura 1 a seguir fornece uma visão rápida e intuitiva da distribuição da reincidência por idade. É possível observar que a faixa etária de 15 a 17 anos parece ter a maior barra, indicando uma contagem mais alta de reincidências nessas idades em comparação às outras.

Figura 1: Boxplot do número de reincidências segundo a idade



Cada estatística sumariza diferentes aspectos da distribuição da variável no conjunto de dados do estudo. Essas estatísticas resumem a distribuição dos dados de reincidência, fornecendo informações sobre a tendência central (média, mediana), a dispersão (mínimo, máximo, quartis) e a variabilidade dos valores dessa variável específica no conjunto de dados, como mostra a Tabela 2.

Tabela 2: Resumo estatístico da reincidência

Estatística	Mín	1º Qu.	Mediana	Média	3º Qu.	Máx.
Reincidência	0.000	0.000	0.000	1.046	1.250	8.000

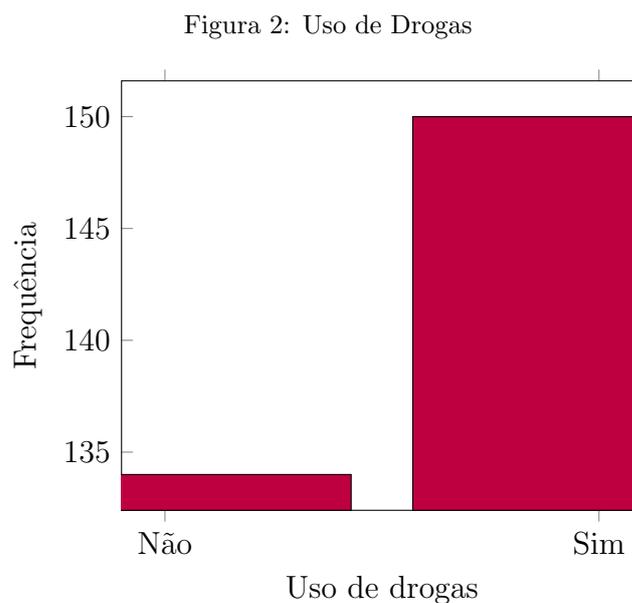
A Tabela 2 nos fornece estatísticas como o mínimo valor observado na variável *Reincidência*, o primeiro quartil representando o limite inferior para 25% das observações, a mediana representando o valor que divide o conjunto de dados em duas partes iguais, a média, o terceiro quartil representando o limite inferior para 75% das observações e o valor máximo observado para a variável *Reincidência*.

As observações de reincidência estão visualmente mais concentradas em valores baixos, com uma média de aproximadamente 1.046. O boxplot exibe visualmente a distribuição da reincidência segundo grupo etário. As principais características dessas observações incluem:

A mediana (linha dentro da caixa) em 0 está indicando que muitas observações se encontram sem reincidência. A caixa se estende do primeiro quartil (25% das observações) até o terceiro quartil (75% das observações), o que significa que a maioria dos dados está concentrada nessa faixa. Existem alguns pontos fora da caixa, indicando outliers.

A combinação dessas representações oferece uma visão abrangente da distribuição e padrões de reincidência em diferentes faixas etárias. A maioria das ocorrências é concentrada em idades mais jovens, com uma presença significativa de valores zero. A presença de outliers pode ser investigada mais a fundo para entender casos incomuns.

A Figura 2 a seguir referente à variável *Uso de Drogas*, fornece a distribuição da frequência de mulheres que participaram do estudo com base no uso de drogas:



O eixo X (horizontal) representa a variável *Uso de Drogas* com duas categorias

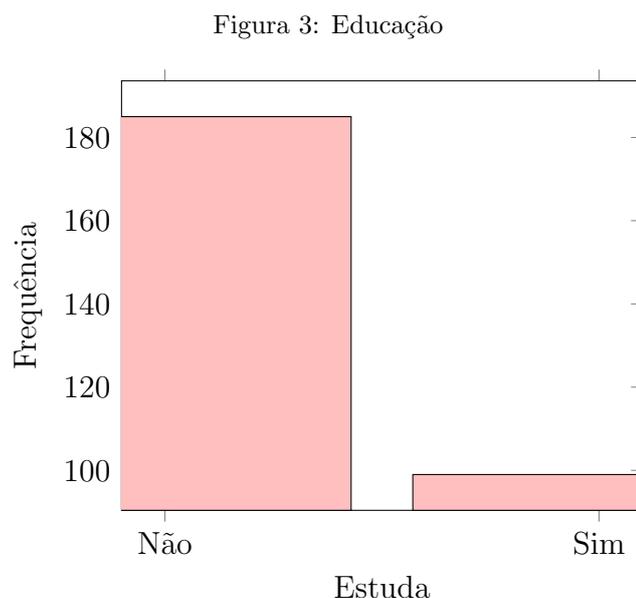
distintas: “Não” e “Sim”. O eixo Y (vertical) representa a frequência, ou seja, o número de mulheres que se enquadram em cada categoria. A altura de cada barra representa a frequência associada a cada categoria.

A Figura 2 indica que 134 mulheres que participaram do estudo não apresentaram histórico de uso de drogas e 150 apresentaram histórico do uso de drogas.

Entre as observações do conjunto de dados apresentados na Figura 3 em relação ao uso de drogas por mulheres que participaram do estudo, podemos perceber que há uma presença significativa de histórico com o uso de drogas em comparação àquelas mulheres sem histórico.

Além da análise do histórico de uso de drogas, a compreensão do perfil educacional das participantes torna-se essencial para compreender o perfil educacional das mulheres envolvidas no estudo, dividindo-as entre aquelas que estão envolvidas em atividades educacionais e aquelas que não estão. Essa representação visual oferece uma visão imediata da distribuição das frequências em relação ao status educacional das participantes, auxiliando na compreensão da influência dessa variável nos resultados observados.

A Figura 3 a seguir, referente à *Educação*, fornece as categorias da variável *Estuda*. As barras estão divididas em duas categorias: “Não” e “Sim”, indicando se as mulheres estudam ou não estudam. O Eixo Y (vertical) representa a frequência de casos observados para cada categoria.

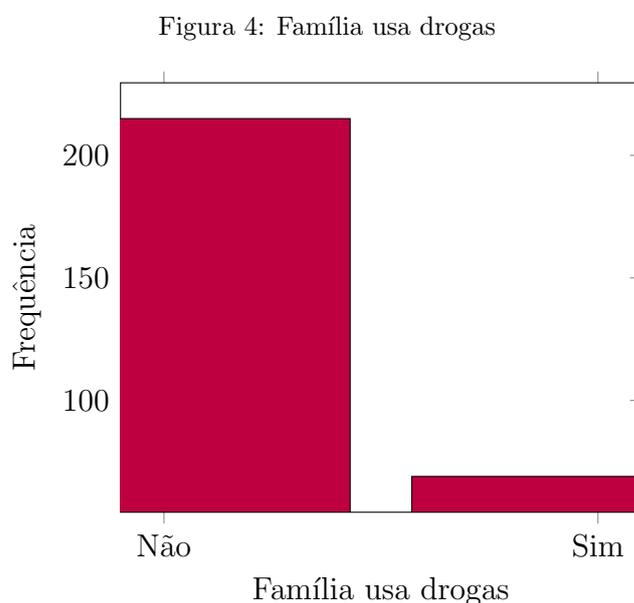


Cada barra representa uma categoria e a cor rosa indica a distribuição dos casos.

A primeira barra “Não” tem uma altura de 185, indicando que há 185 mulheres que não estudam e, portanto, a maioria. A segunda barra “Sim” tem uma altura de 99, indicando que há 99 mulheres que estudam e, portanto, a minoria.

Este gráfico fornece, portanto, uma visão da distribuição da variável *Estuda* entre mulheres que participaram do estudo, destacando assim, as diferenças na frequência entre as categorias.

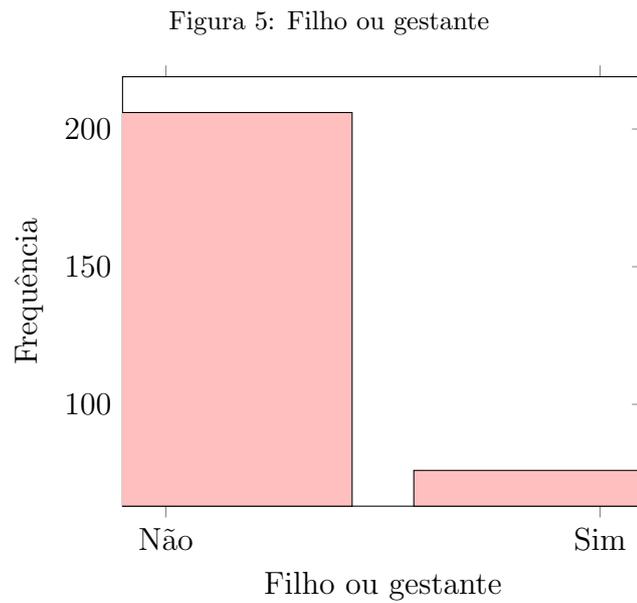
Oferecendo uma perspectiva adicional ao estudo, a Figura 4 a seguir, referente ao *Família usa drogas*, fornece a frequência de casos em que mulheres que participaram do estudo têm ou não têm histórico de uso de drogas por parte da família.



O eixo X (horizontal) representa as categorias da variável *Família usa drogas*. As barras estão divididas em duas categorias: “Não” e “Sim”, indicando se as famílias das mulheres têm ou não têm histórico de uso de drogas. O eixo Y (vertical) representa a frequência de casos observados para cada categoria. A altura de cada barra indica quantas mulheres estão na respectiva categoria.

A barra mais alta está na categoria “Não”, indicando que a maioria das mulheres tem famílias sem histórico de uso de drogas. A barra menor está na categoria “Sim”, indicando que uma menor proporção de mulheres tem famílias com histórico de uso de drogas.

Este gráfico fornece, portanto, uma visão da distribuição da variável *Família usa drogas* entre mulheres, destacando diferenças na frequência entre as categorias.



A Figura 5 referente à variável *Filho ou gestante* fornece a frequência de mulheres que participaram do estudo com base em sua condição de terem filho ou serem gestantes.

O eixo X (horizontal) representa as categorias da variável *Filho ou gestante*. As barras estão divididas em duas categorias: “Não” e “Sim”, indicando se a mulher gestante.

O eixo Y (vertical) representa a frequência de casos observados para cada categoria. A altura de cada barra indica quantas mulheres estão na respectiva categoria.

A barra mais alta está na categoria “Não”, indicando que a maioria das mulheres não possuem filho ou não são gestantes. A barra menor está na categoria “Sim”, indicando que uma menor proporção de mulheres possuem filho ou são gestantes.

Figura 6: Composição do domicílio



A Figura 6 mostra a *Composição do domicílio* e representa a frequência de mulheres com base em sua situação de moradia, especificamente se moram com genitores/outras familiares ou se moram sozinhas.

O eixo X (horizontal) representa as categorias da variável *Composição do domicílio*. As barras estão divididas em duas categorias: “Genitores/Outros Familiares” e “Sozinha”, indicando a situação de moradia das mulheres que participaram do estudo.

O eixo Y (vertical) representa a frequência de casos observados para cada categoria. A altura de cada barra indica quantas mulheres estão na respectiva categoria.

A barra mais alta está na categoria “Genitores/Outros Familiares”, indicando que a maioria das mulheres do estudo moram com seus genitores ou outros familiares. A barra menor está na categoria “Sozinha”, indicando que uma menor proporção de mulheres do estudo que moram sozinha.

### 3.2 Correlação de Pearson Para as Variáveis *Reincidência* e *Idade*

A análise de correlação de Pearson entre as variáveis *Reincidência* e *Idade* neste estudo possibilitou compreender a relação linear entre esses dois fatores no conjunto de dados e explorar a importância de entender a associação entre a idade e a reincidência criminal, examinando se a idade das mulheres está correlacionada com o número de rein-

cidências observado.

Tabela 3: Correlação e Teste de Significância entre *Reincidência* e *Idade*

	Correlação	Valor- <i>p</i>
Pearson	0.201	0.0006425

A Tabela 4 mostra, por meio da correlação de Pearson, que à medida que a idade das mulheres aumenta, a reincidência criminal também tende a aumentar. O valor-*p* associado à correlação de Pearson é menor que 0,05 (nível de significância comum). Isso sugere fortemente que a associação observada não é devida ao acaso, tornando a relação estatisticamente significativa. No entanto, é importante considerar que a significância estatística não implica necessariamente relevância prática, e a força da associação é relativamente fraca, mesmo que seja estatisticamente significativa.

Embora haja uma associação positiva e estatisticamente significativa entre as variáveis *Reincidência* e *Idade*, a correlação de 0,201 é relativamente baixa, sugerindo, que a relação linear entre idade e reincidência não é forte.

Isso indica que outros fatores podem influenciar a reincidência, e a relação observada não é suficientemente forte para prever a reincidência apenas com base na idade. Outras variáveis podem desempenhar um papel importante na compreensão da reincidência criminal feminina.

### 3.3 Ajuste no modelo Poisson

A tabela 5 a seguir apresenta a estimação dos coeficientes obtidos a partir do modelo de regressão de Poisson utilizado para analisar a relação entre as variáveis independentes e a reincidência. Esses coeficientes fornecem informações sobre a magnitude e a direção da influência que cada variável tem sobre a variável de resposta, considerando outras variáveis no modelo.

Tabela 4: Estimação dos Coeficientes do Modelo de Regressão de Poisson para Reincidência

Variável	Estimativa	Erro Padrão	Valor z	Valor p
(Intercepto)	0.01073	0.21695	0.049	0.960549
Idade1 (entre 15 e 17 anos)	0.52676	0.20038	2.629	0.008569*
Idade2 (18 ou mais)	1.13601	0.31475	3.609	0.000307*
Estuda (sim)	-0.38194	0.13847	-2.758	0.005811*
Família usa droga (sim)	-0.27183	0.14803	-1.836	0.066305
Filho ou gestante (sim)	0.03951	0.12970	0.305	0.760664
Comp. do domicílio (genitores/familiares)	-0.45643	0.12518	-3.646	0.000266*

Notas: Os valores de p significativos estão marcados com \*, com  $p < 0.05$  indicando significância.

A partir do levantamento de dados para a Tabela 5, o valor da estimativa do intercepto, de 0.01073 não é considerado estatisticamente significativo (valor-p alto).

A estimativa das variáveis representam o efeito que a presença ou mudança de uma unidade para cada variável tem no número de reincidências, mantendo todas as outras variáveis constantes. A variável *Idade* com coeficientes de valores positivos (0.52676 e 1.13601) e a variável *Composição do domicílio (genitores/familiares)* de valor negativo -0.45643, indicaram portanto, um aumento esperado no número de reincidências para mulheres com idades mais avançadas e que moram sozinhas.

A estimativa do coeficiente da variável *Estuda (sim)* foi negativo (-0.38194), indicando uma diminuição esperada no número de reincidências para aquelas mulheres que estudam.

Já as variáveis *Família usa droga* e *Filho ou gestante* não apresentaram influência estatisticamente significativa na contagem de reincidências do estudo, de acordo com o modelo de regressão de Poisson ( $p > 0.05$ ).

Além da estimativa dos coeficientes, a avaliação da adequação de um modelo inclui a análise de várias métricas, bem como a inspeção visual dos seus resíduos. Um exemplo disso é o Gráfico de Resíduos Deviance, como ilustrado na Figura 8. Esse gráfico fornece uma representação visual da discrepância entre os valores observados e os valores previstos pelo modelo, permitindo uma avaliação mais aprofundada da qualidade do ajuste do modelo Poisson.

A presença de padrões observada nos resíduos na Figura 8 como uma tendência crescente ou decrescente à medida que os valores ajustados aumentam, pode indicar pro-

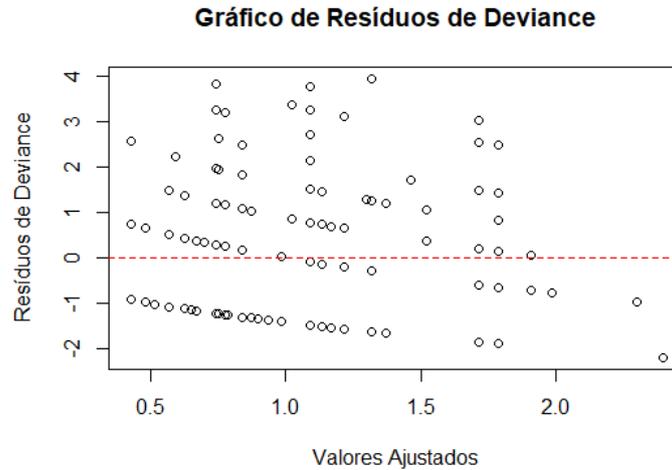


Figura 7: Gráfico de resíduos *deviance* para Poisson

blemas e sugerir que o modelo não está capturando completamente a estrutura dos dados.

A linha vermelha tracejada em zero serve como referência para avaliar a dispersão dos resíduos. Na Figura 8 é possível observar desvios sistemáticos dessa linha ou resíduos muito dispersos, o que indicam problemas no modelo.

Outra abordagem complementar para examinar a dispersão dos resíduos em relação à linha de regressão pode ser representada pelo gráfico de envelope simulado para o modelo.

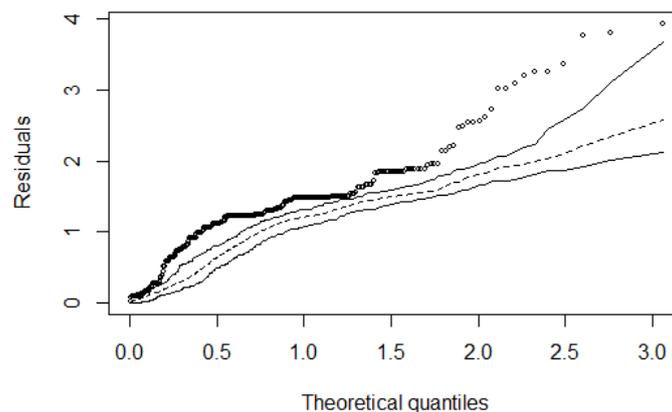


Figura 8: Gráfico de envelope simulado para Poisson

É possível observar da Figura 9 que os pontos se encontram fora das bandas de confiança, isto é, o modelo de Poisson apresenta discordância entre as previsões do modelo e os valores reais observados, indicando um mal ajuste do modelo.

### 3.4 Ajuste no modelo Binomial Negativo

A estimação dos coeficientes a partir do modelo de regressão Binomial Negativa utilizado para analisar a relação entre as variáveis independentes e a reincidência fornecem informações relevantes sobre a magnitude e a direção da influência que cada variável tem sobre a variável de resposta, considerando outras variáveis no modelo.

Tabela 5: Estimação dos Coeficientes do Modelo de Regressão Binomial Negativa para Reincidência

Variável	Estimativa	Erro Padrão	Valor-p
(Intercepto)	-0.03327	0.32169	0.9176
Idade1 (entre 15 e 17 anos)	0.59500	0.28606	0.0375*
Idade2 (18 ou mais)	1.32126	0.54195	0.0148*
Estuda (sim)	-0.37833	0.20838	0.0694
Familia usa droga (sim)	-0.27236	0.23026	0.2369
Filho ou gestante (sim)	0.05659	0.21457	0.7920
Comp. do domicílio (genitores/familiares)	-0.49873	0.20729	0.0161*

Notas: Os valores de  $p$  significativos estão marcados com \*, com  $p < 0.05$  indicando significância.

A partir do levantamento de dados para a Tabela 6, a estimativa das variáveis representam o efeito que a presença ou mudança de uma unidade para cada variável tem no número de reincidências, mantendo todas as outras variáveis constantes.

A variável *Idade* com coeficientes de valores positivos (0.59500 e 1.32126) e a variável *Composição do domicílio (genitores/familiares)* de valor negativo -0.49873, indicaram portanto, um aumento esperado no número de reincidências para mulheres com idades mais avançadas e que moram sozinhas.

Já as variáveis *Estuda*, *Família usa droga* e *Filho ou gestante* não apresentaram influência estatisticamente significativa na contagem de reincidências do estudo, de acordo com o modelo de regressão de Binomial Negativa ( $p > 0.05$ ).

Além da estimativa dos coeficientes e assim como na regressão de Poisson, na regressão Binomial Negativa, os resíduos *deviance* podem ser usados para verificar se o modelo está capturando adequadamente a variação nos dados, como ilustrado na Figura 10 a seguir:

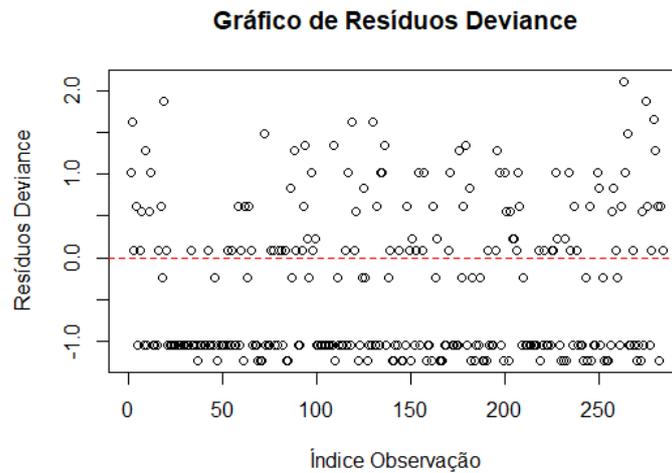


Figura 9: Gráfico de resíduos *deviance* para Binomial Negativa

A Figura 10 representa graficamente a dispersão dos resíduos de *deviance* ao redor do zero, proporcionando uma avaliação sobre a discrepância entre os valores observados e os previstos pelo modelo de regressão Binomial Negativa.

A ausência de uma tendência discernível, seja de crescimento ou decrescimento, indica uma distribuição aleatória dos resíduos, reforçando a ideia de que o modelo se ajusta adequadamente aos dados observados.

A presença da linha horizontal em zero atua como um ponto de referência, revelando que a maioria dos resíduos está contida dentro de limites superiores e inferiores próximos a zero.

A Figura 11 a seguir apresenta um gráfico de envelope simulado do modelo Binomial Negativo, como abordagem complementar para verificação do ajuste do modelo.

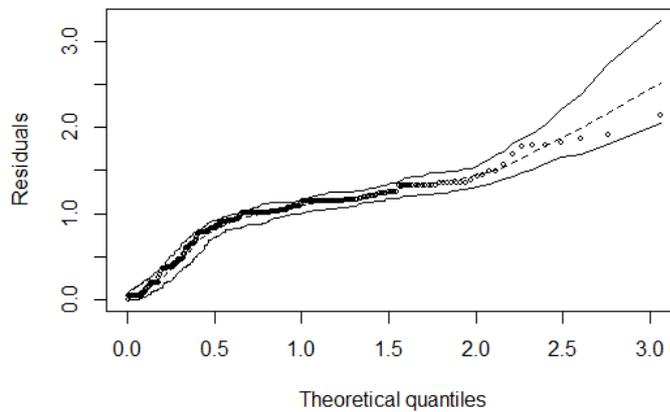


Figura 10: Gráfico de envelope simulado para Binomial Negativa

Pode ser observado da Figura 11 que todos os pontos se encontram dentro das bandas de confiança, que indicam um bom ajuste do modelo, isto é, uma boa concordância entre as previsões do modelo e os valores reais observados.

### 3.5 Discussão dos resultados

A avaliação do gráfico de envelope simulado revelou a inadequação do modelo de regressão de Poisson para esse conjunto de dados. Em contrapartida, este mesmo gráfico mostrou que o modelo de regressão Binomial Negativa apresentou um ajuste adequado aos dados observados. Isso reforça a confiança da capacidade do modelo Binomial Negativa em representar efetivamente a variabilidade e estrutura presentes nos dados estudados.

A diferença no ajuste entre os dois modelos pode indicar que os dados estudados possuem uma variabilidade maior do que o esperado pelo modelo de Poisson, justificando assim a escolha do modelo de regressão Binomial Negativa, que é mais flexível e capaz de lidar com essa superdispersão de forma mais eficaz.

A análise dos resultados obtidos através do modelo de regressão Binomial Negativa oferece *insights* significativos sobre a relação entre as variáveis independentes e a reincidência.

A significância estatística das variáveis *Idade* e *Composição do domicílio* ressalta sua influência na contagem de reincidências. Isso sugere que a idade e a composição do domicílio são fatores determinantes na ocorrência de reincidências.

Por outro lado, as variáveis *Educação*, *Família usa droga* e *Filho ou gestante*

demonstraram não possuir uma influência estatisticamente significativa na contagem de reincidências, de acordo com ambos os modelos de regressão. Esse resultado pode indicar que, dentro do contexto do estudo, essas variáveis não são determinantes ou possuem uma influência menor na ocorrência de reincidências.

É importante ressaltar que a não significância estatística de algumas variáveis não deve ser interpretada como falta de relevância prática. Outros fatores, como o tamanho da amostra e possíveis interações entre variáveis, podem influenciar os resultados.

## 4 Conclusão

Neste trabalho o modelo de regressão Binomial Negativa foi utilizado para descrever o comportamento das variáveis explicativas (idade, educação, uso de drogas, entre outras) e a taxa de reincidência criminal entre as mulheres estudadas. Ele forneceu *insights* sobre quais variáveis estavam associadas à reincidência e como essas variáveis afetam as taxas de reincidência.

Os resultados apresentados neste trabalho destacam a importância da idade e da composição do domicílio como fatores significativos na reincidência criminal entre mulheres no Distrito Federal. Observou-se que, à medida que a idade das mulheres aumenta, a taxa de reincidência tende a aumentar, o que sugere a necessidade de abordagens específicas para diferentes faixas etárias. Ademais, o fato de morar junto com os genitores/familiares diminui o número esperado de reincidência.

No entanto, outros fatores como educação, uso de drogas pela família, uso de drogas pela mulher e ter ou não filho (ou ser ou não gestante), não demonstraram uma relação estatisticamente significativa com o número de reincidências. Isso ressalta a complexidade do fenômeno da reincidência criminal feminina e a necessidade de considerar múltiplos fatores em abordagens de prevenção e intervenção.

É fundamental reconhecer que os resultados deste estudo são baseados em dados específicos do Distrito Federal. Portanto, a generalização para outras populações ou regiões deve ser feita com cautela.

Para avançar na compreensão da reincidência criminal feminina, é importante realizar pesquisas futuras que considerem uma gama mais ampla de variáveis e abordem questões socioeconômicas, psicológicas e culturais que podem influenciar o comportamento criminoso das mulheres. Além disso, a implementação de estratégias de prevenção e reabilitação direcionadas às necessidades específicas das mulheres é essencial para reduzir a reincidência e promover a reintegração na sociedade.

## 5 Referências

AGRESTI, Alan. An Introduction to Categorical Data Analysis. 2<sup>a</sup> ed. New York: John Wiley Sons, 2007.

ASSIS, S.G.; CONSTANTINO, P. Filhas do mundo: infração juvenil feminina no Rio de Janeiro. Rio de Janeiro: Fiocruz, 2001.

BARRETO, Alexandre Serra. Modelos de regressão: teoria e com o programa estatístico R. Ed. do autor, Brasília, 2011.

BRASIL. Panorama Nacional de Execução de Medidas Socioeducativas. Conselho Nacional de Justiça. Brasília, DF, 2012.

HOSMER, D. W.; LEMESHOW, S. Applied Logistic Regression, Second Edition. New York: Wiley, 2000.

NERY, 2015. Risco de concessão de crédito bancário para empresas: Uma aplicação dos modelos de regressão logística. Bacharelado em Estatística. Universidade de Brasília.

NETER, John et al. Applied Linear Statistical Models. McGraw-Hill Higher Education, 2005. Citado na página 22.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. Disponível em: <<http://www.R-project.org/>>. Acesso em: (data de acesso).

RIBEIRO, F.G. Uso da regressão logística na estimação da probabilidade de reincidência de jovens infratoras. Monografia de graduação. Bacharelado em Estatística. Universidade de Brasília, 2015. 34f.

SANTOS, 2013. Perfil de adolescentes do sexo feminino em conflito com a lei no Distrito Federal: violência, uso de drogas e reincidência. Monografia de graduação. Bacharelado em Terapia Ocupacional. Universidade de Brasília, 42f.

PEARSON, Karl. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, Series 6, v. 2, n. 11, p. 559-572, 1901. Disponível em: <https://doi.org/10>