

## CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ  
ISSN on-line 2317-4536 / ISSN impresso 1413-9022 - v.55, p.1 - 23, 2023  
DOI: 10.12957/cadest.2023.83809

### POPULAÇÃO EM SITUAÇÃO DE RUA: POR QUAIS MOTIVOS UMA PESSOA EM SITUAÇÃO DE RUA RESPONDERIA QUE NÃO DESEJA SAIR DESTA SITUAÇÃO?

Guilherme Caldas de Castro

UERJ

[guilherme.caldas@ime.uerj.br](mailto:guilherme.caldas@ime.uerj.br)

Letícia Castilho Homem Souza

UERJ

[leticiaacastilho.uerj@gmail.com](mailto:leticiaacastilho.uerj@gmail.com)

Yan de Aquino Pereira

UERJ

[Yanaquino02@hotmail.com](mailto:Yanaquino02@hotmail.com)

#### Resumo

*População em situação de rua sempre foi um assunto delicado de se tratar e ao mesmo tempo muito importante para saúde pública e vivência em sociedade. Dentre essa população, podemos identificar pessoas que estão nessa situação por falta de oportunidade, dinheiro ou até por questões pessoais como falta de contato com familiares por conta de brigas. Por consequência do aumento constante dessa população, a sociedade foi se organizando e criando abrigos para oferecer alguma dignidade à essa população. Em meio à tantos fatores que podem levar uma pessoa para essa situação, um muito importante é o uso de drogas tanto lícitas quanto ilícitas. Para entender os aspectos e características dessas pessoas, começaram a ser feitos censos como o realizado pelo Instituto Pereira Passos (IPP), que será utilizado nesse trabalho. A partir desses censos e consequentemente coleta de dados, é possível fazer estudos estatísticos para entender os comportamentos e padrões dessa parcela da sociedade. Sendo assim, esse trabalho tem como objetivo identificar através de um modelo de Regressão Logística Binária, quais são as características que mais contribuem para uma pessoa em situação de rua responder que não tem intenção de sair dessa situação.*

**Palavras-chave:** Regressão Logística Binária, População em Situação de Rua, Matriz de Confusão, Estatística.

## 1. Introdução

Cidadãos em situação de rua é um assunto muito importante de ser tratado por ter grande impacto na vivência em sociedade e um estudo a respeito desse tema pode gerar políticas públicas que realmente vejam o bem-estar e melhoria na condição de vida das pessoas que se encontram nessa situação.

Os motivos que levam uma pessoa a viver nessa situação são diversos e podem ser ocasionados tanto por questões financeiras, profissionais, pessoais ou até pela falta de oportunidade. O estudo desse fenômeno é importante para que o país possa reduzir o crescimento de pessoas nessa situação e consequentemente reduzir os problemas que isso acarreta.

Para a definição do objeto de estudo, foi adotado o conceito legal, definido no Decreto Federal nº 7.503, de 23/12/2009, art.1º, parágrafo único (BRASIL, 2023):

“(..) considera-se população em situação de rua o grupo populacional heterogêneo que possui em comum a pobreza extrema, os vínculos familiares interrompidos ou fragilizados e a inexistência de moradia convencional regular, e que utiliza os logradouros públicos e as áreas degradadas como espaço de moradia e de sustento, de forma temporária ou permanente, bem como as unidades de acolhimento para pernoite temporário ou como moradia provisória”.

Como uma maneira de tentar controlar a situação, passou-se a construir abrigos para acolher esses cidadãos e tentar dar uma vida mais digna a eles. Além disso, também passaram a existir trabalhos voluntários com o objetivo de reinserir essas pessoas nas atividades da sociedade.

O objetivo deste estudo é identificar quais características explicam o motivo de algumas pessoas em situação de rua responderem que não querem sair dessa situação. Para isto, foi realizado o modelo de Regressão Logística Binária com os dados do censo feito em 2020 pelo IPP - Instituto Pereira Passos (PREFEITURA DA CIDADE DO RIO DE JANEIRO, 2023). Com isso, será possível avaliar a relação que o uso de drogas lícitas, ilícitas, o motivo que o fez estar na rua, entre outros fatores que possam ter a ver com o fato de o cidadão não querer sair dessa situação.

## 2. Metodologia

### 2.1. Modelo Linear Generalizado (MLG)

Dentre inúmeros métodos estatísticos, os modelos de regressão são uns dos mais utilizados para análise das relações entre as variáveis. A equação dos modelos é dividida

entre variáveis explicativas e variável resposta. O modelo final será para estimar a variável resposta a partir do quanto cada variável explicativa tem de relação com a variável resposta.

Proposto por Nelder e Wedderburn (1972) o Modelo Linear Generalizado é uma generalização da regressão linear em que permite que as variáveis de resposta não necessariamente tenham uma distribuição normal e a estimativa dos parâmetros seja feita pela máxima verossimilhança.

Dependendo da base de dados que se tem disponível, o modelo de regressão que será utilizado deverá ser alterado. Quando os dados são normalmente distribuídos e as variáveis são numéricas, o modelo que é utilizado é o modelo de regressão linear.

No caso deste estudo, como a variável resposta é categórica, ou seja, não é quantitativa e sim qualitativa, o modelo de regressão linear não é o adequado. Como nossa variável resposta possui duas categorias (Quero sair da rua e não quero sair da rua), o modelo utilizado foi o de Regressão Logística Binária.

## **2.2. Regressão Logística Binária**

A Regressão Logística Binária (HOSMER & LEMESHOW, 2013) é uma técnica estatística com o objetivo de determinar um modelo que deverá ser capaz de prever uma resposta que será atribuída à variável dependente, que nesse caso é binária, a partir de variáveis independentes que podem ser contínuas ou categóricas.

Essa técnica é amplamente utilizada em diversos campos, como, em ciências médicas e sociais. Em medicina, por exemplo, essa regressão pode ser utilizada para classificar se um determinado paciente possui uma doença a partir de seus resultados de outros fatores. Neste estudo, foi utilizada essa técnica para uma visão social a respeito de pessoas em situação de rua. O objetivo é encontrar um modelo de regressão capaz de classificar se uma pessoa responderia que não deseja sair da rua a partir de sua resposta para outras perguntas, como por exemplo, se faz uso de alguma droga, sua idade, sexo, raça, entre outros aspectos.

Ao contrário da regressão linear que possui uma variável resposta contínua, na regressão logística a variável dependente é binária, ou seja, é classificada como 0 ou 1, sim ou não. Portanto, ao gerar um gráfico dos modelos, na regressão linear obtém-se uma reta e na regressão logística, um comportamento probabilístico no formato da letra S.

### 2.2.1. Especificando o Modelo Logístico

Nesse método de análise, não se pode mais assumir a distribuição normal visto que a variável resposta será codificada como 0 ou 1, sendo possuir ou não a característica de interesse. Logo, como a variável resposta passa a ser binária, deve-se supor a distribuição Bernoulli.

A relação que a função logística irá propor é entre a variável resposta (codificada em 0 ou 1) e as variáveis preditivas  $x_i = (x_1 + x_2 + \dots + x_p)$ . Essa função logística é definida pela equação (1):

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1)$$

A partir dessa equação, para que a regressão logística seja realizada será necessário fazer a linearização desta equação aplicando o logaritmo, resultando na seguinte equação *logit* (2):

$$G(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2)$$

Como o objetivo do modelo é encontrar a probabilidade da ocorrência de um evento, a equação logística (3) será transformada em uma equação de probabilidade (HAIR *et al*, 200

$$\text{Prob(Evento)} = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (3)$$

Na regressão linear, o método para estimar os coeficientes  $\beta_0, \beta_1, \dots, \beta_n$  é a partir do método dos mínimos quadrados que tem como objetivo minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados. Já na regressão logística, o método utilizado é o da máxima verossimilhança que tem como objetivo maximizar a probabilidade dos dados observados (HOSMER & LEMESHOW, 2013). A equação para esse cálculo é (4):

$$L(\beta) = f(y_1, y_2, \dots, y_n) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (4)$$

A partir dessa equação, são feitos alguns tratamentos matemáticos como aplicar o logaritmo e depois derivar a função em relação à  $\beta_0$  e  $\beta_j$  e igualar as derivadas à zero. Os resultados gerados por esses tratamentos são chamados de estimativa da máxima verossimilhança e são denotados por (5):

$$\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^T \quad (5)$$

Em um modelo de regressão, quanto mais variáveis independentes presentes no modelo, maiores são as chances de criar um modelo mais dependente de dados e menos intuitivo.

O método escolhido para avaliar quais seriam as melhores variáveis para construir o modelo foi o método AIC (Critério de Informação de Akaike). O Critério de Informação de Akaike é uma métrica que da qualidade de um modelo estatístico, fazendo comparações entre possíveis modelos e selecionando aquele que possui o menor valor de AIC (LUCAMBIO, 2020).

### 2.2.2 Capacidade Preditiva do Modelo

Um dos métodos para avaliar a capacidade de previsão do modelo é a partir da Matriz de Confusão (tabela 1) que é uma tabela de classificação preditiva (NARKHED, 2023). Ou seja, com a criação do modelo estatístico, é possível realizar previsões e estimar a taxa de acertos e erros. As medidas de ajustes mais conhecidas são chamadas de sensibilidade e especificidade.

Tabela 1. Matriz de confusão

		Valor Predito	
		Sim	Não
Valor Real	Sim	Verdadeiro Positivo (A)	Falso Negativo (B)
	Não	Falso Positivo (C)	Verdadeiro Negativo (D)

Fonte: Autores (2023)

A partir do modelo, as previsões são feitas a partir de uma premissa. A premissa utilizada no projeto é considerar como 1 quando a probabilidade for superior à 0.5 e 0 quando for menor ou igual à 0.5. Essa probabilidade é uma estimativa da capacidade de

previsão do modelo. Será classificado como 1 a proporção de verdadeiros positivos e como 0 a proporção de verdadeiros negativos (Figura 1).

Figura 1. Classificação das previsões do modelo

$$\hat{y}_i \begin{cases} 1, \text{ se } P(y_i = 1) > 0.5 \\ 0, \text{ se } P(y_i = 0) \leq 0.5 \end{cases}$$

Fonte: Autores (2023)

Definindo o parâmetro da previsão, será possível obter os valores de A, B, C e D da matriz de confusão e conseqüentemente será possível calcular a sensibilidade e especificidade.

A sensibilidade mede a capacidade do teste em identificar corretamente a proporção de verdadeiros positivos e seu cálculo é definido por (6):

$$S = P(\hat{y}_i = 1 | y_i = 1) = \frac{d}{c+d} \quad (6)$$

A especificidade mede a capacidade do teste em identificar corretamente a proporção de verdadeiros negativos e seu cálculo é definido por (7):

$$E = P(\hat{y}_i = 0 | y_i = 0) = \frac{a}{a+b} \quad (7)$$

Essas medidas são eficientes, pois determinam o quão bom está a capacidade preditiva do modelo. Porém quanto mais uma medida cresce, mais a outra diminui. Ou seja, se o valor da sensibilidade está muito alto, o valor da especificidade será muito baixo e assim vice e versa.

Por último, a taxa global ou acurácia é uma medida que avalia quanto o modelo foi capaz de prever corretamente todas as previsões possíveis. Seu cálculo é feito da seguinte maneira (8):

$$TG = \frac{a+d}{a+b+c+d} \quad (8)$$

Para os cálculos realizados foi utilizado o Pacote Estatístico R (R CORE TEAM, 2023).

### 3. Materiais

O Censo de População em Situação de Rua de 2020, da Cidade do Rio de Janeiro, foi realizado entre os dias 26 a 29 de outubro, sob a coordenação do IPP (Instituto Municipal de Urbanismo Pereira Passos) e da Secretaria Municipal de Assistência Social e Direitos Humanos (SMASDH) com a parceria da Secretaria Municipal de Saúde. Para realizar a coleta e tratamento dos dados, optou-se pela contratação de uma empresa especializada, a Qualitest - Inteligência em Pesquisa. Contratação feita através de licitação pública na modalidade de técnica e preço.

Após definida a equipe que trabalharia na coleta de dados, foi o momento de definir a metodologia a ser utilizada.

“O Censo foi realizado por meio da aplicação de questionários diferentes para os três tipos básicos de situação: pessoa na rua, em cenas de uso de drogas e acolhidas. Para crianças até doze anos usou-se um questionário curto e para os indivíduos que não puderam ser diretamente entrevistados foram coletadas algumas informações por observação” (PREFEITURA DA CIDADE DO RIO DE JANEIRO, 2023).

Foi realizado um mapeamento do público-alvo com base em informações disponíveis na SMASDH, permitindo delimitar os trechos da cidade do Rio de Janeiro onde havia maior concentração dessas pessoas. Dividiu-se a cidade em quatro distritos, já conhecidos: Zona Oeste, Zona Norte, Zona Sul e Centro. Cada um desses distritos foi percorrido em um único dia, a fim de evitar a dupla contagem, que era uma preocupação da coordenação do censo, já que as pessoas em situação de rua se movimentam pela cidade durante o dia todo.

Foram 68 horas de trabalho de campo, realizadas por 137 profissionais contratados pela empresa de pesquisa. Além desses profissionais, 400 pessoas da Prefeitura acompanharam as equipes durante esses quatro dias de trabalho.

Após toda a coleta e tratamento de dados, o Censo resultou num total de 7.272 pessoas em situação de rua na cidade do Rio de Janeiro, no ano de 2020. Dentre essas pessoas, 5.469 encontravam-se nas ruas (sendo 1.109 desses, em cenas de uso de drogas) e 1.803 pessoas acolhidas em abrigos.

Os formulários utilizados resultaram em mais de 100 variáveis com informações de cada uma das pessoas entrevistadas ou observadas durante o trabalho de campo.

Para esta análise, foram selecionadas 23 variáveis específicas que fazem mais sentido para o estudo, que são:

- Rua Acolhimento, indica se a pessoa estava na rua ou no abrigo no momento da coleta de dados;
- Estado, informa o estado natal da pessoa;
- Bairro, onde a pessoa se encontrava no local da coleta;
- Faixa Etária;
- Classificação Idade, agrupamento das idades em criança, adolescente, adulto ou idoso;
- Sexo;
- Gênero;
- Cor/Raça;
- Contato\_Família, se a pessoa mantinha contato com a família;
- Motivo\_dormir\_ rua, por qual motivo a pessoa dorme na rua;
- Tempo\_ rua\_RJ, tempo em que a pessoa se encontra na rua;
- Atividade\_remunerada, se a pessoa realiza alguma atividade remunerada;
- Atividade\_realizada, qual a atividade remunerada realizada;
- Recebe\_outras\_fontes\_de\_renda (se recebe outras fontes, como: bolsa família, auxílio emergencial, BPC, aposentadoria/pensão, auxílio-doença, aluguel social e outros);
- Sabe\_ler\_escrever
- Escolaridade\_corrigida, último ano concluído com aprovação;
- Tabaco;
- Álcool;
- Maconha\_Haxixe;
- Crack\_similares;
- Cocaína;
- Inalantes cola\_solvente\_tiner;
- Necessidade\_sair\_situação, o que mais precisa para sair da situação de rua.

Como o objetivo do estudo é entender o perfil das pessoas que não desejam sair da rua ou do abrigo, a variável Necessidade\_sair\_situação foi escolhida como a variável dependente. Após isso, foram realizadas algumas limpezas e tratamentos na base, como:

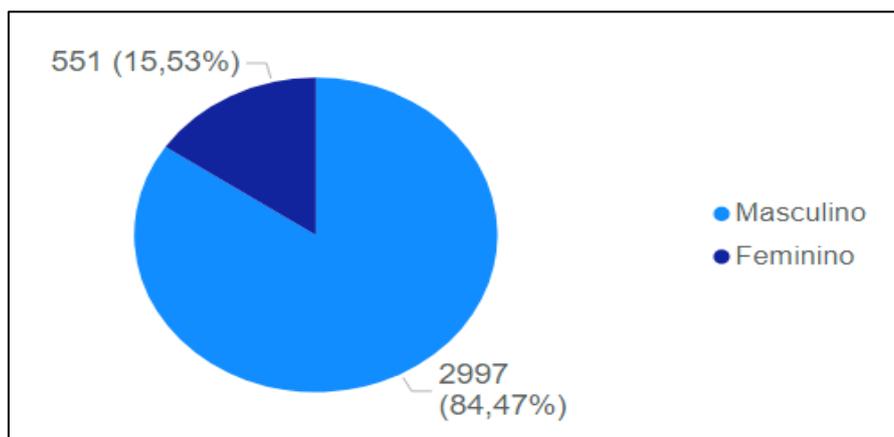
- Excluir todas as pessoas que responderam “NS/NR” e “Não se aplica” na variável Necessidade\_sair\_situação;
- Retirar a resposta NS/NR das variáveis: Estado, Gênero, Contato Família, Motivo dormir na rua, Tempo de rua, Atividade Remunerada, Atividade Realizada, Saber ler e escrever, Escolaridade Corrigida, Tabaco, Álcool, Maconha/Haxixe, Crack/Similares, Cocaína e Inalantes/Cola/Solvente/Tiner;
- Renomear todas as pessoas que responderam “Não quero sair do abrigo” para “Não quero sair da rua” e todas as outras respostas, que mostravam que a pessoa queria sair daquela situação, para “Quero sair da rua”. Esse tratamento foi realizado para que seja possível utilizar o modelo de Regressão Logística Binária no estudo.

Essas limpezas e tratamentos resultaram numa nova base de dados com 3.548 observações.

#### 4. Análise descritiva

Analisando as 3.548 observações da nossa base de dados, nota-se que 551 pessoas são do sexo feminino, enquanto 2.997, masculino, o que representa 15,53% e 84,47%, respectivamente (figura 2).

Figura 2. Distribuição de sexo entre as pessoas da amostra



Fonte: Autores (2023)

Quando perguntados sobre cor e raça, a classificação parda teve a maior frequência relativa, com um *share* de 41,83%, seguido da preta, com 34,95%, conforme mostrado na tabela 2.

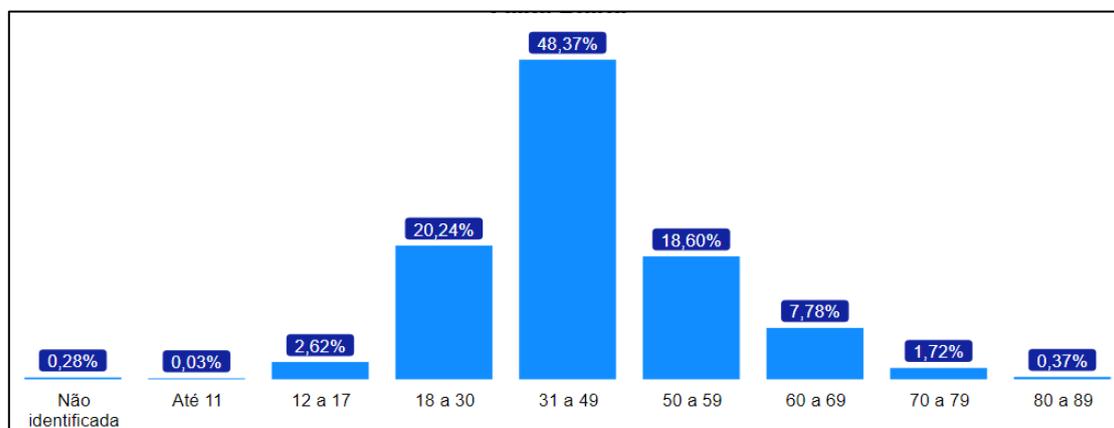
Tabela 2. Frequência de cor/raça na amostra

Cor/Raça	Frequência	Frequência Relativa (%)
Parda	1484	41,83
Preta	1240	34,95
Branca	712	20,07
Amarela	77	2,17
Não Identificada	19	0,54
Indígena	16	0,45
Total	3548	100

Fonte: Autores (2022)

Os indivíduos com idade entre 31 e 49 anos possuem maior representatividade na base, com 48,37%, seguido pela faixa etária de 18 a 30 anos, com 20,24%, e 50 a 59 anos, com 18,60%. Vale ressaltar que 2,65% da nossa base são menores de idade, conforme pode ser visto na figura 3.

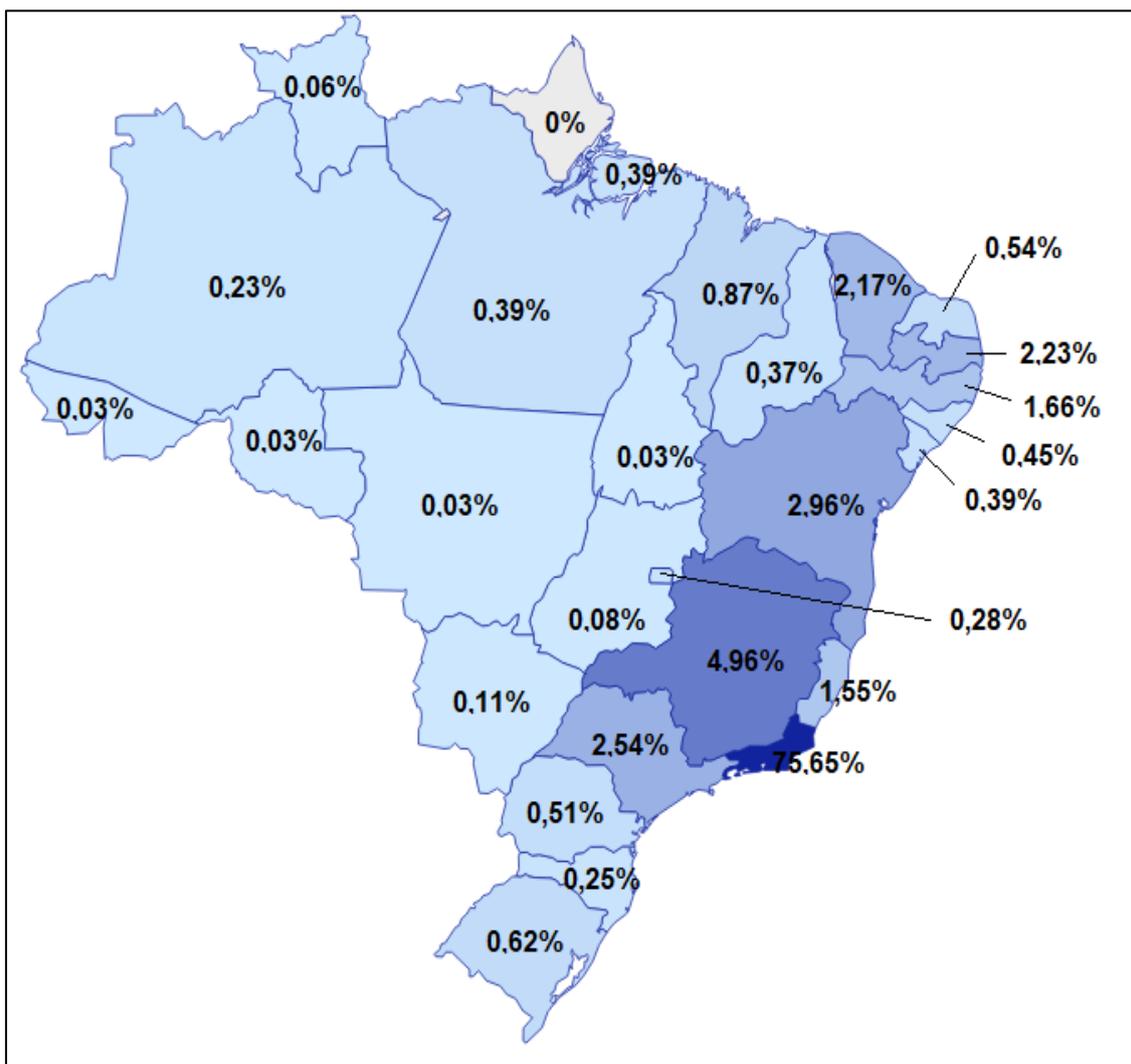
Figura 3. Distribuição das Faixas Etárias na amostra



Fonte: Autores (2023)

Já que o censo foi realizado na cidade do Rio de Janeiro, é esperado que a maioria das pessoas tenha como seu estado natal o Rio de Janeiro (RJ). Na amostra, retirada do censo, 75,65% das pessoas nasceram no estado do RJ, seguido por Minas Gerais com 4,96%, Bahia com 2,96% e São Paulo com 2,54%. Na figura 4 é apresentado o *share* de todos os estados no gráfico.

Figura 4. Mapa com a distribuição dos estados natais

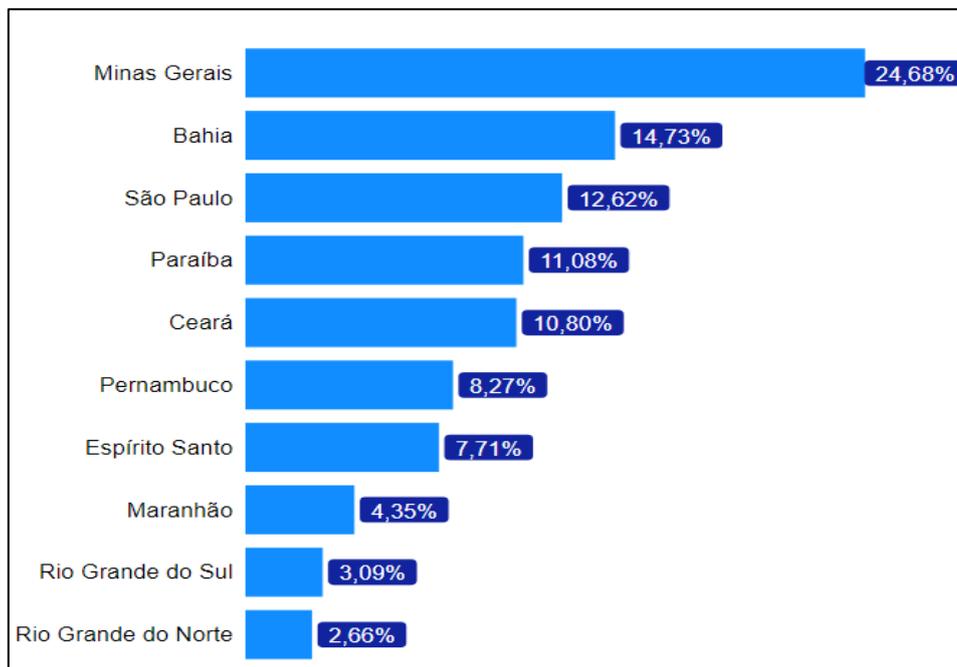


Fonte: Autores (2023)

Assim, um quantitativo de 23,34% da base estudada indicou não ter nascido no Estado do Rio de Janeiro. O *ranking* completo dos percentuais relativos aos estados natais é mostrado na figura 5.

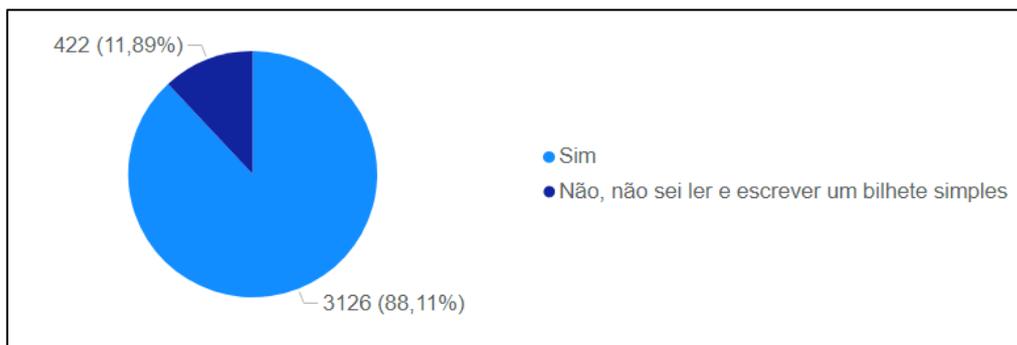
Quando perguntados com referência à alfabetização, 3.126 pessoas (88,11%) responderam que sabem ler e escrever, enquanto 422 (11,89%) responderam que não sabem ler e nem ao menos escrever um simples bilhete (figura 6).

Figura 5. Ranking dos estados natais



Fonte: Autores (2023)

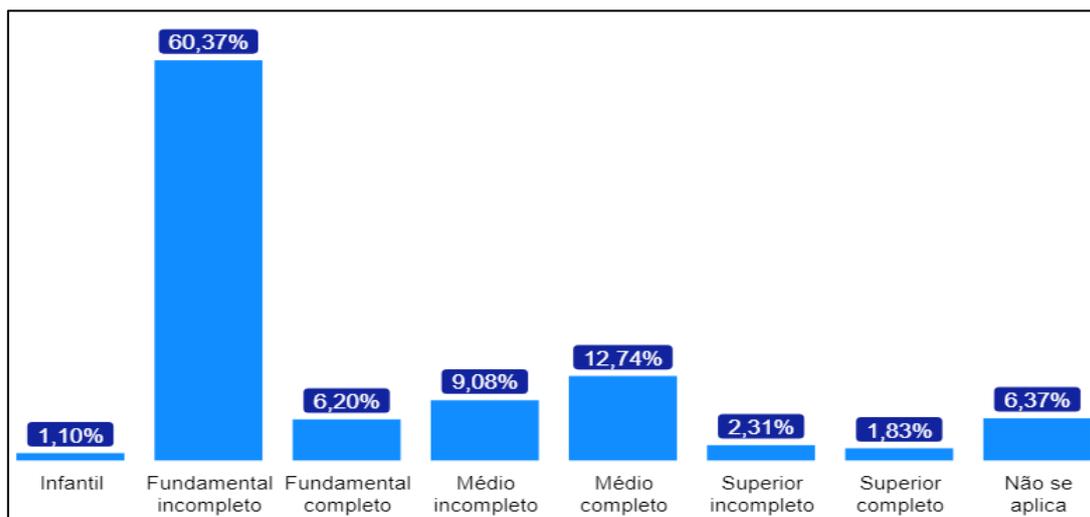
Figura 6. Percentual de pessoas que são alfabetizadas



Fonte: Autores (2023)

Mais do que saber ler e escrever, as pessoas foram perguntadas sobre o nível de escolaridade, e conclui-se que cerca de 76% da base amostral não concluíram os estudos, ressaltando que, aproximadamente, 61% nem sequer finalizaram o Ensino Fundamental (figura 7).

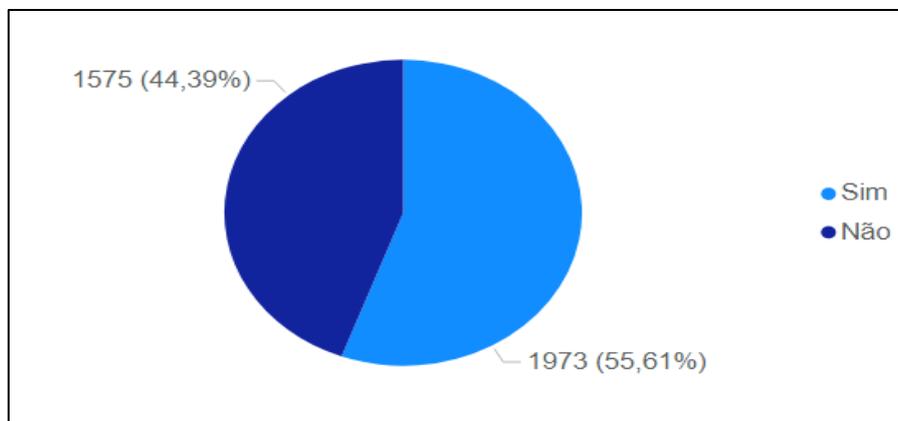
Figura 7. Distribuição do grau de escolaridade



Fonte: Autores (2023)

Seguindo para o aspecto social, cerca de 44% dessas pessoas não possuem contato com seus familiares conforme aponta a figura 8.

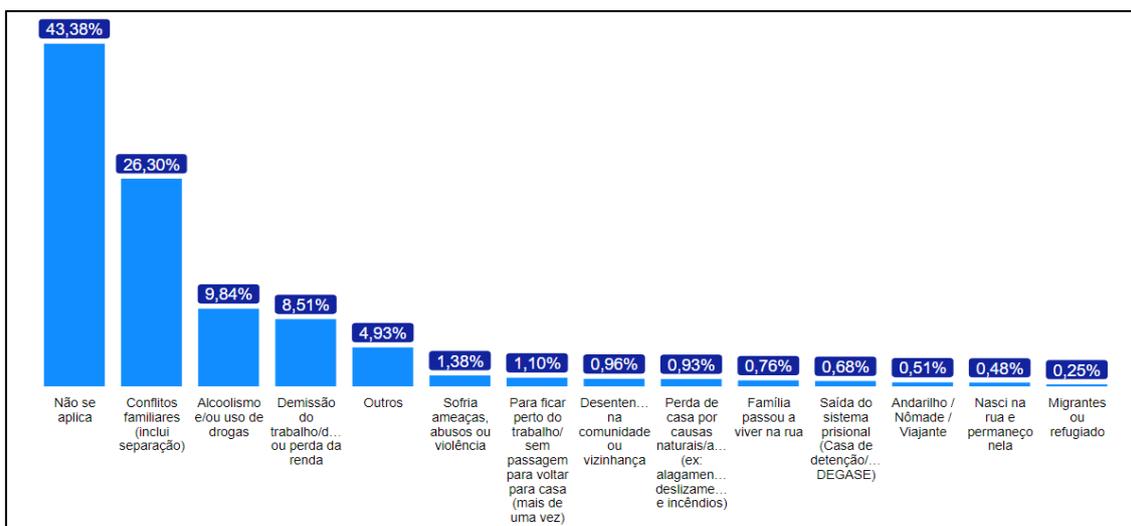
Figura 8. Quantitativo de pessoas que possuem contato com a família



Fonte: Autores (2023)

Quando perguntados sobre o porquê de dormirem na rua, o motivo mais identificado foi por conflitos familiares, com 26,30% das respostas, seguido de uso de álcool e drogas, com 9,84%, e logo em seguida, demissão do trabalho ou perda da renda, com 8,51% (figura 9).

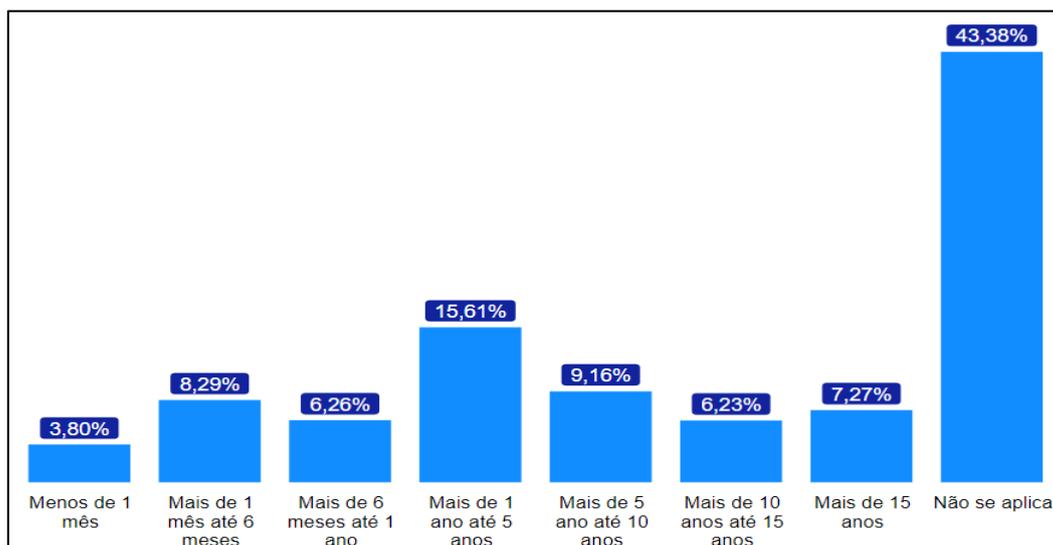
Figura 9. Motivos pelos quais as pessoas dormem na rua



Fonte: Autores (2023)

Analisando o período em que se encontram na rua, 22,66% responderam que estão há mais de 5 anos na rua. O tempo mais observado entre essas pessoas foi de mais de 1 ano até 5 anos, equivalente a 15,61% das respostas. Levando em consideração que para 43,38% essa pergunta não se aplicava, esse valor de 15,61% representa 27,58% das respostas válidas, enquanto o percentual de 22,66%, que equivale ao percentual de pessoas que estão há mais de 5 anos na rua, está para 40,02% (figura 10).

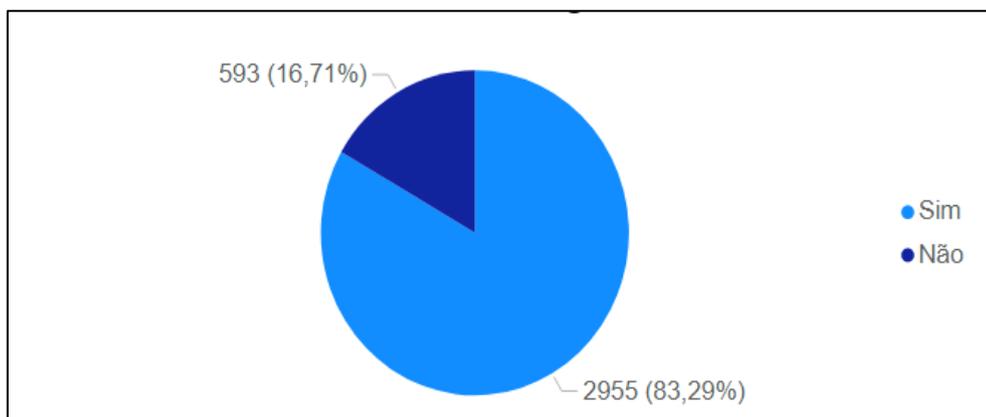
Figura 10. Tempo que as pessoas permanecem na rua



Fonte: Autores (2023)

Cerca de 83% da base amostral faz uso de pelo menos um tipo de droga, seja: tabaco, álcool, maconha/haxixe, crack/similares, cocaína e inalantes, como cola, solvente e tiner (figura 11).

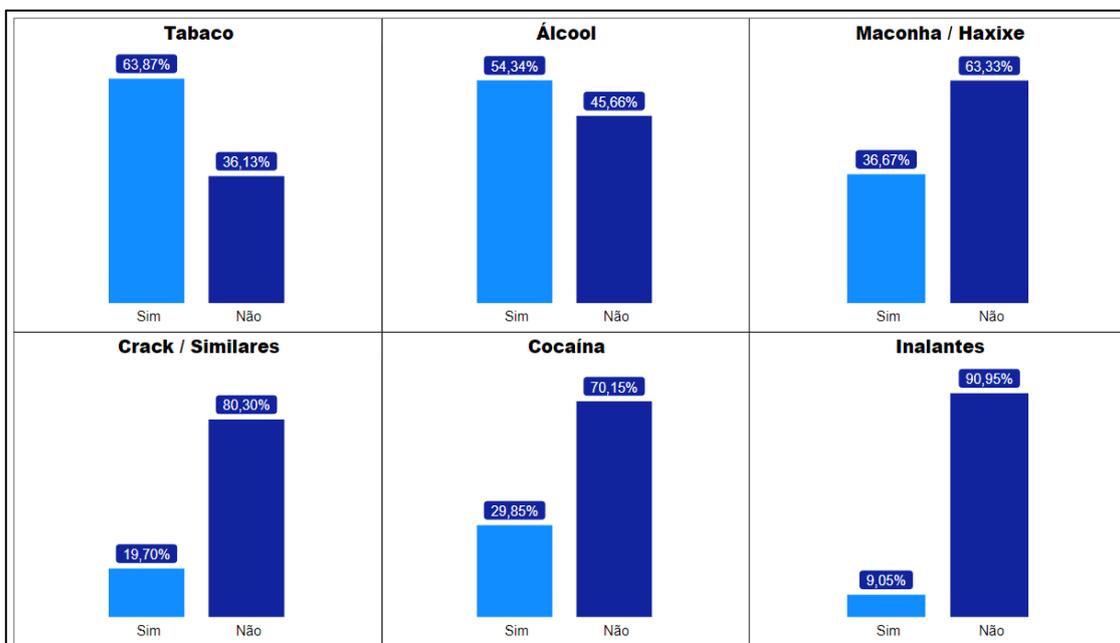
Figura 11. Percentual de pessoas que usam drogas



Fonte: Autores (2023)

Quando analisadas as drogas separadamente, é visto que as mais utilizadas são tabaco e álcool, com 63,87% e 54,34%, respectivamente. As menos usadas são inalantes e crack/similares, com 9,05% e 19,70%, respectivamente (figura 12).

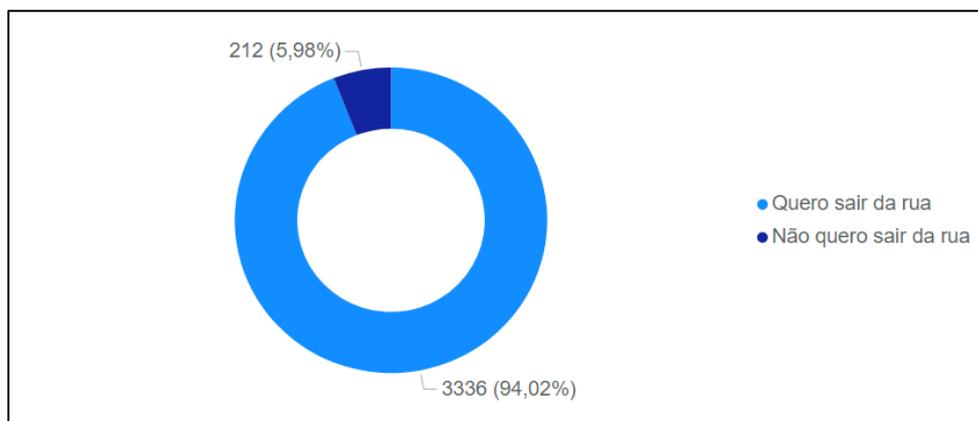
Figura 12. Tipo de droga utilizada



Fonte: Autores (2023)

Quando perguntados sobre o desejo de sair dessa situação de rua, cerca de 6% da nossa base amostral respondeu que não deseja sair da rua, o que nos faz questionar: quais os motivos levariam uma pessoa a não querer sair dessa situação? (figura 13).

Figura 13. Quantitativo de pessoas que não desejam sair das ruas



Fonte: Autores (2023)

## 5. Resultado do Modelo de Regressão

Aqui são apresentados todos os resultados encontrados a partir do modelo de regressão logística. Como dito anteriormente, esse estudo tem o intuito de mostrar os fatores mais importantes para a decisão de uma pessoa em situação de rua preferir continuar nessa situação a tentar reverter.

As conclusões serão tomadas observando os resultados dos modelos de regressão em relação ao intervalo de confiança de 95%, os p-valores do teste de Wald e os valores observados da estatística de Wald (MIGUEL, 2015).

### 5.1 Modelo Uso de Drogas

Primeiramente, foi criado um modelo de Regressão Logística Binária para o uso de drogas. Foi levado em consideração as variáveis independentes: tabaco, álcool, maconha/haxixe, crack/similares, cocaína, inalantes/cola/solventes/tíner.

A partir desse modelo, é observado que a variável que tem maior significância (abaixo de 0.05) com o fato de uma pessoa não querer sair da rua é o Crack/Similares (tabela 3).

Tabela 3. Modelo de Regressão Logística com variáveis relacionadas a drogas

<b>Variáveis:</b>	<b>Estimativa</b>	<b>Erro padrão</b>	<b>z value</b>	<b>p-value</b>	
Coefficiente	2,76E+03	1,58E+02	17.418	<2e-16	***
TabacoSim	3,06E+02	1,96E+02	1.561	0.1186	
AlcoolSim	1,61E+02	1,88E+02	0.859	0.3905	
Maconha_HaxixeSim	-1,49E-02	2,07E+02	0.000	0.9999	
Crack_SimilaresSim	-5,64E+02	2,33E+02	-2.423	0.0154	*
CocainaSim	-3,96E+02	2,10E+02	-1.883	0.0596	.
Inalantes_Cola_Solvente_TinerSim	-8,97E+01	3,01E+02	-0.298	0.7658	

Fonte: Autores (2023)

## 5.2 Modelo Perfil

Nesse modelo, o intuito é avaliar separadamente as variáveis relacionadas ao perfil de cada indivíduo. Assim, pode-se avaliar quais características influenciam mais na decisão de não querer sair da rua. As variáveis consideradas foram: idade, sexo, gênero e cor/raça.

Não foram encontradas características com p-valores inferiores à 0,05 no modelo, indicando que, no intervalo de confiança de 95% não há variável estatisticamente significativa para tal conclusão, como pode ser visto na tabela 4.

Tabela 4. Modelo de Regressão Logística com variáveis relacionadas ao perfil

<b>Variáveis:</b>	<b>Estimativa</b>	<b>Erro padrão</b>	<b>z value</b>	<b>p-value</b>	
Coefficiente	2,89E+05	1,19E+05	2.431	0.0150	*
Classificacao_idadeAdulto	0.39537	0.49436	0.800	0.4239	
Classificacao_idadeCriança	1,40E+06	2,40E+08	0.006	0.9953	
Classificacao_idadeldoso	-0.50210	0.52626	-0.954	0.3400	
Classificacao_idadeSem informação	-1,72E+05	1,00E+05	-1.714	0.0865	.
SexoMasculino	-0.08982	0.94441	-0.095	0.9242	
GeneroHomem Trans	-1,60E+05	1,24E+05	-1.293	0.1960	
GeneroMulher Cis	-0.57517	0.96773	-0.594	0.5523	
GeneroMulher Trans	1,36E+06	4,97E+07	0.027	0.9781	
GeneroOutro	-1,08E+05	0.55134	-1.959	0.0501	.
GeneroTravesti	1,36E+06	7,99E+07	0.017	0.9865	
Cor_racaBranca	-0.26322	0.62549	-0.421	0.6739	
Cor_racaIndígena	-1,24E+05	1,25E+05	-0.997	0.3188	
Cor_racaNão identificada	-0.48706	1,21E+05	-0.402	0.6877	
Cor_racaParda	-0.22736	0.61509	-0.370	0.7117	
Cor_racaPreta	-0.04305	0.62024	-0.069	0.9447	

Fonte: Autores (2023)

### 5.3 Modelo Grau de Educação

Como um último modelo individual, o objetivo é entender se o grau de educação de uma pessoa em situação de rua, ou seja, se ele avançou muito nos estudos ou não, tem significância na decisão de querer sair da rua. As variáveis utilizadas para a criação desse modelo foram: saber ler e escrever e escolaridade (tabela 5).

Tabela 5. Modelo de Regressão Logística com variáveis relacionadas à educação

Variáveis:	Estimativa	Erro padrão	z value	p-valor	
Coefficiente	2,93E+05	0.51105	5.724	1.04e-08	***
Sabe_ler_escreverSim	0.02564	0.36007	0.071	0.943	
Escolaridade_corrigidaFundamental incompleto	-0.18037	0.37989	-0.475	0.635	
Escolaridade_corrigidaInfantil	0.36230	1,11E+05	0.327	0.744	
Escolaridade_corrigidaMédio completo	-0.11761	0.44024	-0.267	0.789	
Escolaridade_corrigidaMédio incompleto	-0.14248	0.46891	-0.304	0.761	
Escolaridade_corrigidaNão se aplica	-0.48453	0.58048	-0.835	0.404	
Escolaridade_corrigidaSuperior completo	-0.77624	0.59519	-1.304	0.192	
Escolaridade_corrigidaSuperior incompleto	-0.11778	0.69602	-0.169	0.866	

Fonte: Autores (2023)

Conclui-se que nenhuma dessas características são importantes para a decisão final. Logo, a pessoa ter concluído o ensino superior ou ensino médio não aumenta estatisticamente a probabilidade de ela responder que não quer sair da rua visto que o p-valor de quem tem o ensino superior completo é menor do que quem não tem.

### 5.4 Modelo Final

Como explicado anteriormente, foi utilizado o Critério de Informação de Akaike (AIC) para definir o melhor modelo com as variáveis disponíveis para estudo. Esse método selecionou as seguintes variáveis que obtiveram o melhor AIC: faixa etária, sexo, motivo para dormir na rua, crack/similares e cocaína, com um AIC de 1071.4 (tabela 6).

Com esse resultado, foram observadas algumas características interessantes. As respostas que mais influenciam na decisão de alguém não querer sair da situação de rua com um nível de significância de 95% são, estar na faixa etária de 70 a 79 anos, ser do sexo masculino, o motivo de estar na rua ser andarilho, nômade e/ou viajante, ser um migrante ou refugiado, para ficar perto do trabalho, ter saído do sistema prisional, fazer uso de crack/similares e fazer uso de cocaína.

Tabela 6. Modelo de Regressão Logística com variáveis definidas pelo AIC

Variáveis:	Estimativa	Erro padrão	z value	p-valor	
Coefficiente	1,91E+04	0.5740	3.331	0.000865	***
Faixa_etaria18 a 30	0.6702	0.5240	1.279	0.200844	
Faixa_etaria31 a 49	0.8154	0.5061	1.611	0.107138	
Faixa_etaria50 a 59	0.2929	0.5272	0.556	0.578504	
Faixa_etaria60 a 69	-0.5386	0.5468	-0.985	0.324706	
Faixa_etaria70 a 79	-1,43E+04	0.6081	-2.344	0.019102	*
Faixa_etaria80 a 89	-1,00E+04	1,19E+04	-0.839	0.401337	
Faixa_etariaAté 11	1,41E+05	2,40E+07	0.006	0.995319	
Faixa_etariaNão identificada	-1,43E+04	1,07E+04	-1.346	0.178453	
SexoMasculino	0.5753	0.2138	2.691	0.007118	**
Motivo_dormir_uaAndarilho / Nômade / Viajante	-2,20E+04	0.8090	-2.718	0.006563	**
Motivo_dormir_uaConflitos familiares (inclui separação)	0.5726	0.3009	1.903	0.057046	.
Motivo_dormir_uaDemissão do trabalho/desemprego ou perda da renda	0.6832	0.4426	1.544	0.122662	
Motivo_dormir_uaDesentendimentos na comunidade ou vizinhança	0.8464	1,06E+04	0.801	0.423241	
Motivo_dormir_uaFamília passou a viver na rua	-0.4261	0.8046	-0.530	0.596422	
Motivo_dormir_uaMigrantes ou refugiado	-2,00E+04	0.9251	-2.166	0.030276	*
Motivo_dormir_uaNão se aplica	0.5827	0.2848	2.046	0.040781	*
Motivo_dormir_uaNasci na rua e permaneço nela	-0.7669	1,12E+04	-0.684	0.493902	
Motivo_dormir_uaOutros	-0.2871	0.3916	-0.733	0.463561	
Motivo_dormir_uaPara ficar perto do trabalho/ sem passagem para voltar para casa (mais de uma vez)	-1,27E+04	0.5832	-2.185	0.028879	*
Motivo_dormir_uaPerda de casa por causas naturais/acidentes (ex: alagamentos, deslizamentos e incêndios)	1,43E+05	4,94E+06	0.029	0.976817	
Motivo_dormir_uaSaída do sistema prisional (Casa de detenção/cadeia/penitenciária/ DEGASE)	-1,58E+04	0.6458	-2.451	0.014248	*
Motivo_dormir_uaSofria ameaças, abusos ou violência	-0.2214	0.6974	-0.318	0.750863	
Crack_SimilaresSim	-0.6683	0.2260	-2.957	0.003108	**
CocainaSim	-0.6262	0.2073	-3.020	0.002524	**

Fonte: Autores (2023)

Portanto, pode-se levantar algumas hipóteses com esse resultado. Será que as pessoas que fazem uso de crack e cocaína estão tão dependentes da droga que já não têm mais perspectiva de sair dessa situação? Será que migrantes ou refugiados não tem as mesmas oportunidades que um cidadão comum e por isso não conseguem melhorar sua qualidade de vida? Às vezes a pessoa prefere não ter uma moradia para poder chegar a tempo ao trabalho e com isso acaba dormindo na rua. Algo bem comum na sociedade é o fato de uma pessoa que acabou de sair da prisão não conseguir um emprego de qualidade por julgamento da sociedade.

Essas análises são de extrema importância para conhecer as deficiências de uma sociedade e onde está faltando oportunidades. Pessoas em situação de rua é um caso de saúde pública e é do interesse (ou deveria ser) de toda a sociedade para solucionar esse problema. Estudos como esse apontam características muito importantes para se investigar mais a fundo e virarem uma proposta de lei. Por exemplo a situação do uso de crack e cocaína, será que a liberação dessas drogas traria uma melhoria para a sociedade

ou só deixariam mais pessoas dependentes e conseqüentemente mais pessoas nessa situação?

### 5.5 Matriz de Confusão do Modelo

A Matriz de Confusão é uma tabela com a função de relatar os casos de falsos positivos, falsos negativos, verdadeiros positivos e verdadeiros negativos (NARKHED, 2023). No caso do modelo do projeto, a matriz (tabela 7), foi gerada a partir de uma base de teste, usando o *SplitRation* de 0,7.

Tabela 7. Resultado da Matriz de confusão

		Classe Original	
		False	True
Classe Predita	Não quero sair da rua	1	63
	Quero sair da rua	3	998

Fonte: Autores (2023)

A partir da matriz, foram calculadas algumas métricas de avaliação para observar a qualidade do modelo (MARIANO & PAZ, 2020). As métricas foram sensibilidade, especificidade e a taxa global. A sensibilidade é o percentual que o modelo acertou em classificar quem realmente quer sair da rua, a especificidade calcula a taxa de acerto do modelo em classificar quem não quer sair da rua e a taxa global ou acurácia é uma métrica para avaliação geral do modelo, onde mostra o percentual de observações classificadas da maneira correta. Os cálculos são realizados das seguintes maneiras:

- Sensibilidade =  $998 / (998+3) = 99,70\%$
- Especificidade =  $1 / (1+63) = 1,56\%$
- Taxa Global =  $(1+998) / (1+63+998+3) = 93,80\%$

A partir da Sensibilidade e da Taxa Global conclui-se que o modelo está tendo uma alta taxa de acertos em suas previsões, porém, através da Especificidade, pode-se observar uma deficiência do modelo em prever esses casos.

## 6. Conclusão

Considerando a amostra de 3.548 observações retiradas do Censo de População em Situação de Rua, da cidade do Rio de Janeiro, realizado em 2020, pode-se retirar 4 *insights* aplicando o método estatístico de Regressão Logística Binária.

Relacionando o uso das drogas de forma separada das demais variáveis com o fato de as pessoas não desejarem sair da rua, nota-se que crack e similares possuem grande significância para essa tomada de decisão, com p-valor de 1,54%.

Ao relacionar o perfil (idade, sexo, gênero e cor/raça) e o grau de escolaridade dessas 3.548 pessoas presentes na amostra, também de formas separadas, conclui-se que não são fatores que podem influenciar na escolha de não sair da rua.

Aplicando o AIC no modelo composto por todas as variáveis selecionadas para o estudo, é visto que as variáveis faixa etária, sexo, motivo para dormir na rua, uso de crack/similares e uso de cocaína podem colaborar para que as pessoas não desejem sair dessa situação.

Detalhando essas variáveis, pessoas que: são do sexo masculino, possuem entre 70 e 79 anos, tem como motivo de estar na rua ser andarilho, nômade e/ou viajante, migrante ou refugiado, ficar mais próximo do trabalho, ter saído do sistema prisional, fazer uso das drogas crack/similares e cocaína, possuem maior tendência a decidirem por continuar na rua.

## Referências

BRASIL. Decreto Federal nº 7053 de 23/12/2009. Institui a Política Nacional para a População em Situação de Rua e seu Comitê Intersetorial de Acompanhamento e Monitoramento, e dá outras providências. Disponível em: <https://legis.senado.leg.br/norma/411537/publicacao/15750903> Acesso em: 06/04/2023.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise Multivariada de Dados**. Bookman editora. 2009.

HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**. New Jersey: John Wiley & Sons, 2013.

LUCAMBIO, F. **Seleção de modelos, Critério de Informação de Akaike**. 2020. Disponível em: <http://leg.ufpr.br/~lucambio/CE017/20202S/AIC.pdf>. Acesso em: 05/06/2023.

MARIANO, D.; PAZ, F. J. **Data Mining**. 1ª Edição. Porto Alegre: Sagah, 2020.

MIGUEL, M. G. R. Comparação do desempenho da estatística tipo-Wald com o de outras estatísticas de teste em modelos lineares generalizados. 2015. Dissertação (**Mestrado**) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015.

NARKHED, S. **Understanding confusion matrix**. 2018. Disponível em: <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>. Acesso em: 10/05/2023.

NELDER, J.; WEDDERBURN, R. Generalized Linear Models. Blackwell Publishing. **Journal of the Royal Statistical Society**, 1972.

PREFEITURA DA CIDADE DO RIO DE JANEIRO. Censo de População em Situação de Rua - 2020. Disponível em: <https://censurua-pcrj.hub.arcgis.com/>Acesso em: 06/04/2023.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019. Disponível em: <https://www.R-project.org>. Acesso: 06/06/2023.

## **HOMELESS POPULATION: FOR WHAT REASONS WOULD SOMEONE IN A HOMELESS SITUATION RESPOND THAT THEY DO NOT WANT TO EXIT THIS SITUATION?**

### **Abstract**

*The homeless population has always been a delicate matter to deal with and at the same time very important for public health and living in society. About that population, we can identify people that are in this situation because of the lack of opportunity, money or personal problems like no contact with relatives. As a result of the constant increase in this population, the society started organizing and creating shelters to offer some dignity to the homeless people. Among many factors that can lead a person to this situation, one very important is the use of drugs, including legal and illegal. To understand the aspects and characteristics of these people, censuses like the one carried out by the Pereira Passos Institute (IPP) began to be developed, which we will use in this work. From these census and consequently the collect of data, it's starts to be possible to make statistical studies to understand the behavior and patterns of this portion of society. The objective of this study is to identify through the Binary Logistic Regression Model, which are the aspects that contribute the most to someone respond that don't want to get out of the streets.*

**Keywords:** *Binary Logistic Regression, Homeless Population, Confusion Matrix, Statistics.*