Universidade de Brasília
Departamento de Estatística

Modelo Linear Generalizado de Poisson: Uma análise da ocorrência de homicídios e as desigualdades raciais no Brasil.

Bruno Soares Jorge

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
2021

Bruno Soares Jorge

Modelo Linear Generalizado de Poisson: Uma análise da ocorrência de homicídios e as desigualdades raciais no Brasil.

Orientador(a): Prof(a). Ana Maria Nogales Vasconcelos
Coorientador(a): Prof(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
2021

Agradecimentos

Agradeço a minha família, que me deu todo o suporte e amor necessários para viver, durante toda minha vida, e durante todo o curso. Em especial minha mãe e meu pai, que sempre me deram todo o apoio emocional e financeiro, e são meus exemplos.

Agradeço também meu irmão, todos os amigos, e a minha namorada, que fazem parte dessa caminhada. Acompanharam todos os momentos, bons e ruins, que fizeram parte da graduação.

Obrigado aos professores e funcionários do Departamento de Estatística da Universidade de Brasília. E a instituição como um todo, que permite o conhecimento adquirido se traduzir em benefícios para a sociedade.

Resumo

No Brasil, durante o ano de 2020, 77.23% das vítimas de homicídio eram negras. Quando a população considerada é a de jovens, de 15 a 29 anos, esse número sobe para 81.28%. Esse trabalho busca, por meio da implementação de um Modelo Linear Generalizado, evidências estatísticas que caracterizem a desigualdade racial e que se expressam na violência juvenil. A partir dos dados do Sistema de Informações de Mortalidade (SIM) do Ministério da Saúde, análises descritivas dos dados de mortalidade, por todas as causas foram feitas, para o ano de 2020. Seguindo a Classificação Internacional de Doenças (CID-10), as mortes por homicídio, de pessoas de todas as faixas etárias, e dos jovens, separadamente, também foram objeto da análise descritiva. Foram ajustados aos dados de contagem dos números de homicídios de jovens (15 a 29 anos), Modelos Lineares Generalizados (MLGs), a fim de definir aquele que melhor se comporta, e estimar os parâmetros que estipulam a relação entre Raça/Cor e esses números de assassinatos. Os dados são provenientes do SIM, do Ministério da Saúde. Na modelagem, foram consideradas características sociodemográficas das vítimas, em especial, a raça/cor. Na modelagem, foram consideradas características sociodemográficas das vítimas, em especial, a raça/cor.

Os dados de homicídio, tem o comportamento de dados de contagem, por se tratar do número de ocorrências. Assume valores inteiros não negativos. Alguns modelos foram implementados, e seus resultados comparados, com o objetivo de definir um modelo final, que melhor se ajusta aos dados. O modelo de Poisson teve o fenômeno de sobredispersão quando não consideradas as interações. O modelo Binomial Negativo foi a alternativa escolhida para contornar esse problema. Mas quando as interações entraram no modelo, a estimativa do parâmetro de dispersão foi de um número muito grande. Convergindo então, assintoticamente, para a distribuição de Poisson. O modelo toma como referência, indivíduos da Raça/Cor Branca, com menos de 8 anos de estudo, da Região Norte e Sexo Feminino.

Os resultados apontam para diferença significativa na contagem de mortos por homicídio, tendo um aumento para pessoas negras, quando comparado com a população

branca. Esse aumento foi multiplicativo de 11.02 vezes, se comparada a população branca. O incremento pode ser ainda maior, também como consequência racial, uma vez que interações entre Raça/Cor e outras variáveis também se mostraram significativas. O valor do parâmetro referente a interação entre Raça/Cor e Sexo, indica que o impacto da Raça/Cor na contagem de homicídios é maior para indivíduos do sexo masculino. E o sexo masculino, individualmente, também aponta para um impacto positivo, com o incremento multiplicativo de 15.36. Ou seja, jovens negros do sexo masculino têm um valor predito 216.45 vezes, ou 21645% maior do que o valor de referência. Existe uma concentração de homicídios de negros, com menos de 8 anos de estudo, da região Nordeste e do sexo masculino. Resultados na análise descritiva indicam esse comportamento, que foi melhor investigado na modelagem dos dados.

A partir da interpretação dos parâmetros estimados no modelo final, concluiu-se que a Raça/Cor tem interferência no número de homicídios de jovens de 15 a 29 anos, sendo maior o valor esperado de assassinatos de pessoas negras. Isto é, associação entre pertencer a Raça/Cor dos pretos e pardos e o número de mortes, se mostrou estatísticamente significante.

Palavras-chaves: Modelos Lineares Generalizados, Racismo, Desigualdade Racial, Negro, Homicídio, Contagem, Poisson, Binomial Negativa.

Lista de Tabelas

Lista de Figuras

Sumário

1 Introdução

O Brasil é um país em que aproximadamente 54.6% da população tema raça/cor autodeclarada como negros ou pardos (IBGE, 2018), e forma a maior comunidade negra fora da África. Essa representatividade, porém, não se mostra em todos os âmbitos sociais. Acesso a saneamento básico, educação, moradia, habitação e saúde são medidos por indicadores que invariavelmente refletem a desigualdade. O reconhecimento da diferença racial no país é recente (ARAÚJO, 2009) e na saúde especificamente, a quantidade de pesquisas que leva essa variável em consideração é muito reduzida e escassa, visto que as consequências da discriminação historicamente apontam para grandes diferenças.

No ano de 2020, 42391 pessoas foram assassinadas, segundo o Sistema de Informações de de Mortalidade (SIM), do Ministério da Saúde, divulgado pelo Departamento de Informática do Sistema Único de Saúde (DATASUS). Dessas, 77.23% tinham "Pardo"ou "Preto"como classificação de Raça/Cor.

Entende-se nesse estudo, raça/cor como uma variável social, que não se limita a sequenciamento genético, mas uma questão indentitária como consequência estrutural (FERREIRA; CAMARGO, 2011). Tal abordagem é possível por meio de dois métodos: autoclassificação, quando o entrevistado responde por si a qual raça se identifica, e heteroclassificação, que é feita por terceiros. A forma implementada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) é na prática, uma mistura dessas duas, sem ter como definir a quantidade de cada uma delas. O instituto dá preferência a definir a raça/cor do indivíduo por meio da autoclassificação, porém por vezes não é possível fazê-lo, seja por ausência no momento da visita, ou por incapacidade, por exemplo de crianças entrevistadas (OSORIO, 2013). É considerado então, que o método utilizado é a autoclassificação, diferentemente de quando é implementada a heteroatribuição realizada pelo entrevistador.

Todavia, pesquisas para aferir a possível diferença de classificação entre os dois métodos não mostram grandes diferenças. Hoje, a classificação do IBGE é definida pela cor da pele, tendo como possíveis respostas: "branca", "preta", "parda", "amarela"e como exceção a esse critério, foi incluída a categoria "indígena"em 1991. A luta pela igualdade racial no Brasil é representada pelo termo "negro", que engloba aqueles que são classificados como pretos ou pardos. Existem discussões a respeito da inclusão dessa categoria nas pesquisas, mas se entende que isso afetaria as demais, e resultados anteriores. Além disso, indivíduos já encontram representatividade nas alternativas "preto"e "pardo".

A abordagem em saúde sob a ótica étnico-racial no Brasil é recente. Exemplo

disso é que apenas em 1995/1996 foi incluído no SIM, o campo de raça/cor. Este tipo de limitação não se apresenta como a causa da falta de interesse acadêmico sobre o tema, mas como uma consequência da discriminação racial (CHOR; LIMA, 2005). Com esse cenário, é urgente a necessidade da produção de pesquisa na área da saúde com enfoque nas diferenças raciais. A comunidade científica também está contaminada com essas consequências, e projetos que visam diminuir ou trazer à luz essa disparidade podem contribuir para maior possibilidade futura de aproximação à equidade.

A branquitude é tratada como ideal por consequência da mentalidade eurocentrista. Com isso, existe um padrão de comportamento de negação da identidade negra, por traços fisicos menos característicos de quem a reproduziu naqueles que alcançam boas condições financeiras (OSORIO, 2013). Isso contraria a possibilidade do nível socioeconômico ser uma variável de confusão para mensurar diferenças entre populações de diferentes raças. A existência da relação entre nível socioeconômico e raça/cor é também entendida como consequencia da mesma discriminação que se busca investigar.

As consequências da desigualdade racial nos números de mortalidade, podem ser vistas em dados de mortes violentas. Segundo Waiselfisz (2012), existe uma tendência geral, desde 2002, de queda do número absoluto de homicídios na população branca, e o comportamento inverso para a população negra. Os números relativos também apontam para uma maior proporção de negros mortos por homicídios, em relação aos brancos, segundo Waiselfisz (2011). Por isso, esse estudo irá focar nas mortes por homicídio.

A população jovem, de 15 a 29 anos, também será analisada, e seus números de homicídios modelados, de forma separada. Waiselfisz (2011) mostra que, as taxas de homicídio, no ano de 2010, se concentram nessa faixa etária, e vão declinando a partir dessa idade. Além da faixa etária, Waiselfisz (2011) cita Mello, Minayo e UNICEF para afirmar que as mortes por homicídios são ocorrências marcadamente masculinas. Sobre a variação no território brasileiro: "A evolução dos homicídios considerando a cor das vítimas tem sido extremamente desigual entre as Unidades da Federação, obedecendo a fatores e determinantes locais."(WAISELFISZ, 2012) logo, as unidades da federação e macrorregiões também serão incluídas nas análises. Segundo Soares (2007), existe uma relação entre escolaridade e a probabilidade de ser vítima de homicídio. Assim, a escolaridade será incluída na análise.

Serão abordadas, primeiramente, todas as mortes registradas no SIM para o ano de 2020. Em seguida, apenas as mortes registradas como homicídios, segundo a Classificação Internacional de Doenças (CID-10), no ano de 2020. Os códigos referentes a mortes violentas são as categorias X85 a Y09, do CID-10. O SIM utiliza a mesma classi-

ficação racial do IBGE. A busca é por encontrar um modelo estatístico dentre os modelos lineares generalizados, que se adeque bem aos dados. Uma vez encontrada uma ou mais modelagens apropriadas, e discutido qual apresenta o melhor ajuste, será possível ter medidas que ajudem a interpretar a influência da variável raça/cor nos resultados observados. Outras variáveis também serão consideradas, e os estratos por diferentes níveis variáveis, para que possa definir quais têm essa influência.

Os dados de mortalidade tem as características daqueles que são considerados, na literatura, como dados de contagem, por se tratar de número de ocorrências. Ao fazer tal definição, restringe o número de famílias de distribuições que se ajustam bem, por não fazer sentido com o contexto tratar como alguma variável contínua. Para dados dessa natureza, são utilizadas as distribuições de Poisson, Binomial Negativa, Multinomial, Poisson Generalizada, entre outras. Os dados serão primeiramente ajustados a uma curva de regressão Poisson, e analisadas as funções de ligação mais comuns, com destaque para a ligação canônica log-linear. Será investigada a presença de sobredispersão e serão estudados as possíveis modelagens em caso de confirmação desse fenômeno.

O estudo visa abordar os resultados de mortalidade, e encontrar um modelo linear generalizado (MLG) que explique o comportamento desse indicador, dando ênfase na variável Raça/Cor, para avaliar as consequencias da desigualdade racial. Um modelo bem ajustado aos dados traz a possibilidade de interpretar a associação entre Raça/Cor e contagem de homicídios.

Foram analisadas as mortes por homicídio ocorridas no Brasil, no ano de 2020. Uma análise descritiva foi feita, a fim de levantar hipóteses sobre como cada possível variável explicativa interfere no número de homicídios. As mortes, por todas as causas, foram analisadas segundo cada causa, e faixa etária. Os registros de assassinatos foram analisados segundo Escolaridade, Sexo, Região e UF. A partir disso, alguns modelos foram implementados e, com os resultados da análise de diagnóstico, o que demonstrou ter melhor ajuste dos dados, foi definido como modelo final.

2 Revisão Metodológica

O interesse é estudar o número de ocorrências de mortes por homicídio, segundo os níveis de categorias das possíveis variáveis explicativas. Serão considerados os MLG's que se adequam bem aos dados de contagem, com tais características.

O autor Paula 2013 comenta, sobre quando a variável resposta de interesse é o número de ocorrências de um evento, segundo níveis de categorias:

> "Aqui, a suposição de distribuição de Poisson para o número de ocorrências do evento em cada configuração de níveis das categorias leva a resultados equivalentes à suposição de distribuição multinomial para as caselas da tabela de contingência formada. Assim, muitas tabelas de contingência que seriam originalmente analisadas através de um modelo log-linear multinomial podem ser analisadas, alternativamente, por um modelo log-linear de Poisson. A vantagem disso é o fato do modelo log-linear de Poisson ser ajustado mais facilmente [...], além da possibilidade de todos os procedimentos desenvolvidos para os MLGs serem diretamente estendidos para o modelo log-linear de Poisson. (PAULA, 2013)"

2.1 Modelo Linear Generalizado (MLG)

Segundo Paula (2013), um MLG é definido por: $Y_1, ..., Y_n$ variáveis aleatórias independentes, cada uma com função densidade ou função de probabilidades na forma dada abaixo

$$f(y_i; \theta_i, \phi) = \exp\left[\phi\left\{y_i\theta_i - b(\theta)\right\} + c(y_i, \phi)\right]. \tag{2.1.1}$$

Parte sistemática:

$$g(\mu_i) = \eta_i.$$

Em que
$\eta_i = \mathbf{x_i}^T\boldsymbol{\beta}$ é o preditor linear
$\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T, p < n$ é um vetor de parâmetros desconhecidos a serem estimados
$\mathbf{x_i} = (x_{i1}, ..., x_{ip})^T$ representa o valor de variáveis explicativas
e $g(.)$ é uma função monótona e diferenciável, denominada **função de ligação**

2.1.1 Função Desvio

Sem perda de generalidade, conforme Paula (2013), vamos supor que o logaritmo da função de verossimilhança seja definido por:

$$L(\mu; y) = \sum_{n=1}^{n} L(\mu_i; y_i),$$

em que $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = \mathbf{x_i^T}\boldsymbol{\beta}$. Para todo modelo saturado ($p = n$) a função $L(\mu; y)$ é estimada por

$$L(y; y) = \sum_{n=1}^{n} L(y_i; y_i).$$

Ou seja, a estimativa de máxima verossimilhança de de μ_i fica nesse caso dada por $\tilde{\mu}_i = y_i$. Quando $p < n$, denotamos a estimativa de $L(\mu; y)$ por $L(\hat{\mu}; y)$. Aqui, a estimativa de máxima verossimilhança de μ_i será dada por $\hat{\mu} = g^{-1}(\hat{\eta}_i)$, em que $\hat{\eta}_i = \mathbf{x_i}^T\hat{\beta}$. A qualidade do ajuste de um MLG é avaliada através da função desvio

$$D^*(y, \hat{\mu}) = \phi D(y, \hat{\mu}) = 2[L(y; y) - L(\hat{\mu}; y)], \tag{2.1.2}$$

que é uma distância entre o logaritmo da função de verossimilhança do modelo saturado (com n parâmetros) e do modelo sob investigação (com p parâmetros) avaliado na estimativa de máxima verossimilhança $\hat{\beta}$.

2.1.2 Função Escore e Informação de Fisher

Escore e Fisher para $\boldsymbol{\beta}$:

Considerando a partição $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ e denotando o logaritmo da função de verossimilhança por $L(\theta)$. De acordo com Paula (2013), para obter a função escore para o parâmetro $\boldsymbol{\beta}$, se calcula inicialmente as derivadas:

$$\begin{aligned}
\partial L(\boldsymbol{\theta})/\partial\beta_j &= \sum_{i=1}^{n} \phi \left\{ y_i \frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{d\eta_i}{\beta_j} - \frac{db(\theta_i)}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{d\eta_i}{d\beta_j} \right\} \\
&= \sum_{i=1}^{n} \phi\{y_i V_i^{-1}(d\mu_i/d\eta_i)x_{ij} - \mu_i V_i^{-1}(d\mu_i/d\eta_i)x_{ij}\} \\
&= \sum_{i=1}^{n} \phi \left\{ \sqrt{\frac{\omega_i}{V_i}}(y_i - \mu_i)x_{ij} \right\},
\end{aligned}$$

em que $\omega_i = (d\mu_i/d\eta_i)^2/V_i$. Logo, a função escore pode ser escrita na forma matricial

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \frac{\partial L(\theta)}{\partial \boldsymbol{\beta}} = \phi \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}), \tag{2.1.3}$$

em que $\mathbf{X}$ é uma matriz $n \times p$ de posto completo cujas linhas serão denotadas por $\mathbf{x_i}^T, i = 1, ..., n$, $\mathbf{W} = \text{diag}\{\omega_1, ..., \omega_n\}$ é a matriz de pesos, $\mathbf{V} = \text{diag}\{V_1, ..., V_n\}, \mathbf{y} = (y_1, ..., y_n)^T$ e $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)^T$.

Obtendo a matriz de informação de Fisher, a partir das derivadas

$$\begin{aligned}\partial^2 L(\boldsymbol{\theta})/\partial\beta_j\beta_\ell = \phi \sum_{i=1}^{n} (y_i - \mu_i) \frac{d^2\theta_i}{d\mu_i^2} \left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij} x_{i\ell} \\ + \phi \sum_{i=1}^{n} (y_i - \mu_i) \frac{d_i^\theta}{d\mu_i} \frac{d^2\mu_i}{d\eta^2} x_{ij} x_{i\ell} - \phi \sum_{i=1}^{n} \frac{d_i^\theta}{d\mu_i} \left(\frac{d\mu_i}{d\eta}\right)^2 x_{ij} x_{i\ell},\end{aligned}$$

cujos valores esperados são

$$\begin{aligned}E\{\partial^2 L(\boldsymbol{\theta})/\partial\beta_j\partial\beta_\ell\} &= -\phi \sum_{i=1}^{n} \frac{d\theta_i}{d\mu_i} \left(\frac{d\mu_i}{d\eta}\right)^2 x_{ij} x_{i\ell} \\ &= -\phi \sum_{i=1}^{n} \frac{(d\mu_i/d\eta_i)^2}{V_i} x_{ij} x_{i\ell} \\ &= -\phi \sum_{i=1}^{n} \omega_i x_{ij} x_{i\ell}.\end{aligned}$$

Logo, segue a informação de Fisher para $\boldsymbol{\beta}$ na forma matricial

$$\boldsymbol{K}_{\beta\beta}(\boldsymbol{\theta}) = E\left\{-\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right\} = \phi \mathbf{X}^T \mathbf{W} \mathbf{X}. \tag{2.1.4}$$

Escore e Fisher para ϕ:

A função escore para o parâmetro ϕ é

$$\begin{aligned}U_\phi(\boldsymbol{\theta}) &= \partial \frac{L(\boldsymbol{\theta})}{\partial\phi} \\ &= \sum_{i=1}^{n} \{y_i\theta_i - b(\theta_i)\} + \sum_{i=1}^{n} c'(y_i, \phi),\end{aligned}$$

em que $c'(y_i, \phi) = dc(y_i, \phi)/d\phi$. Para obter a matriz de informação de Fisher para ϕ temos que calcular $\partial^2 L(\boldsymbol{\theta})/\partial\phi^2$. Assim, a informação de Fisher para ϕ é

$$K_{\phi\phi}(\boldsymbol{\theta}) = -\sum_{i=1}^{n} E\{c''(Y_i, \phi)\}. \tag{2.1.5}$$

2.1.3 Estimação dos parâmetros

Estimação de $\boldsymbol{\beta}$

O processo iterativo de Newton-Raphson para a obtenção da estimativa de máxima verossimilhança de $\boldsymbol{\beta}$ é definido expandindo a função escore $\mathbf{U}_\beta$ em torno de um valor inicial $\beta_{(0)}$, tal que

$$\mathbf{U}_\beta \cong \mathbf{U}_\beta{}^{(0)} + \mathbf{U}_\beta{}'^{(0)}(\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)}),$$

em que $\mathbf{U}_\beta{}'$ denota a primeira derivada de $\mathbf{U}_\beta$ com respeito a $\boldsymbol{\beta}^{\boldsymbol{T}}$, sendo $\mathbf{U}_\beta{}'^{(0)}$ e $\mathbf{U}_\beta{}^{(0)}$, respectivamente, essas quantidades avaliadas em $\boldsymbol{\beta}^{(0)}$. Assim, repetindo o procedimento acima, têm-se o processo iterativo

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \{(-\mathbf{U}'_\beta)^{-1}\}^{(m)}\mathbf{U}_\beta^{(m)},$$

$m = 0, 1, ...$ Como a matriz $-\mathbf{U}'_\beta$ pode não ser positiva definida, a aplicação do método escore de Fisher substituindo a matriz $-\mathbf{U}'_\beta$ pelo correspondente valor esperado $\boldsymbol{K}_{\beta\beta}$ pode ser mais conveniente. Isso resulta no seguinte processo iterativo:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \{\boldsymbol{K}_{\beta\beta}^{-1}\}^{(m)}\mathbf{U}_\beta^{(m)},$$

$m = 0, ...$ Trabalhando o lado direito da expressão acima, se obtem um processo iterativo de mínimos quadrados reponderados

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^T\mathbf{W}^{(m)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{(m)}\mathbf{z}^{(m)}, \tag{2.1.6}$$

$m = 0, 1, ...,$ em que $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{1/2}\mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$. A quantidade $\mathbf{z}$ desempenha o papel de uma variável dependente modificada, enquanto $\mathbf{W}$ é uma matriz de pesos que muda a cada processo iterativo. É usual a utilização de $\boldsymbol{\eta}^{(0)} = \mathbf{g}(\mathbf{y})$ para inicialização do processo.

Estimação de ϕ

Igualando a função escore U_ϕ a zero, implica na solução

$$\sum_{i=1}^{n} c'(y_i, \hat{\phi}) = \frac{1}{2}D(\mathbf{y}; \hat{\boldsymbol{\mu}}) - \sum_{i=1}^{n}\{y_i\tilde{\theta}_i - b(\tilde{\theta}_i)\}, \tag{2.1.7}$$

sendo$d(\mathbf{y}; \hat{\boldsymbol{\mu}})$ denota o desvio do modelo sob investigação.

2.1.4 Distribuição Assintótica

Para demonstrar que $\hat{\boldsymbol{\beta}}$ e ϕ são assintoticamente normais e independentes, usaremos os resultados abaixo

$$E(\mathbf{U}_\theta) = \mathbf{0},$$
$$\text{Var}(\mathbf{U}_\theta) = \mathbf{K}_{\theta\theta},$$

com as funções escore de $\boldsymbol{\beta}$ e ϕ sendo, respectivamente, expressas nas formas $\mathbf{U}_\beta = \sum_{i=1}^{n} U_{i\beta}$, em que

$$U_{i\beta} = \phi\sqrt{\omega_i V_i^{-1}}(y_i - \mu_i)x_i \quad ,$$
$$U_\phi = \sum_{i=1}^{n} U_{i\phi},$$

com $U_{i\phi} = \{y_i\theta_i - b(\theta_i)\} + c'(y_i, \phi)$. Portanto, para n grande, temos que $\mathbf{U}_\theta \sim N_{p+1}(\mathbf{0}, \mathbf{K}_{\theta\theta})$. Em particular, assintoticamente $\mathbf{U}_\beta \sim N_p(\mathbf{0}, \mathbf{K}_{\beta\beta})$ e $U_\phi \sim N(0, K_{\phi\phi})$ e $\boldsymbol{U}_\beta$ e U_ϕ são independentes.

Expandindo $\boldsymbol{U}_{\hat{\theta}}$ em série de Taylor em torno de $\boldsymbol{\theta}$ obtemos

$$\mathbf{U}_{\hat{\theta}} \cong \mathbf{U}_\theta + \mathbf{U}'_\theta(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

em que $\mathbf{U}'_\theta = \partial\mathbf{U}_\theta/\partial\boldsymbol{\theta}^T$. Assim como $\hat{\boldsymbol{\theta}}$ é o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ temos que $\mathbf{U}_{\hat{\theta}} = \mathbf{0}$ e daí segue a relação

$$\hat{\boldsymbol{\theta}} \cong \boldsymbol{\theta} + (-\mathbf{U}'_\theta)^{-1}\boldsymbol{U}_{\boldsymbol{\theta}}.$$

Supondo que para n grande $-\boldsymbol{U}'_\theta \cong \mathbf{K}_{\theta\theta}$ (para ligação canônica $\mathbf{K}_{\beta\beta} = -\boldsymbol{U}'_\beta$), então obtemos

$$\hat{\boldsymbol{\theta}} \cong \boldsymbol{\theta} + \mathbf{K}_{\theta\theta}^{-1}\mathbf{U}_\theta,$$

ou seja, para n grande $\hat{\boldsymbol{\theta}} \sim N_{p+1}(\boldsymbol{\theta}, \mathbf{K}_{\theta\theta}^{-1})$. Como $\mathbf{K}_{\theta\theta} = \text{diag}\{\mathbf{K}_{\beta\beta}, \mathbf{K}_{\phi\phi}\}$ então assintoticamente segue que $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \mathbf{K}_{\beta\beta}^{-1})$ e $\hat{\phi} \sim N(0, K_{\phi\phi}^{-1})$ e $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$ são independentes.

2.1.5 Teste de Wald

Segundo Paula (2013), para testar $H_0 : \boldsymbol{\beta_1} = \boldsymbol{\beta_1^0}$ contra $H_1 : \boldsymbol{\beta_1} \neq \boldsymbol{\beta_1^0}$, a estatística Wald fica expressa na forma:

$$\xi_w = [\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{\beta}_{\mathbf{1}}^{\mathbf{0}}]^{\intercal} \hat{V}ar^{-1}(\hat{\boldsymbol{\beta}}_1)[\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \boldsymbol{\beta}_{\mathbf{1}}^{\mathbf{0}}],$$

em que $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ sai do vetor $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_{\mathbf{1}}^{\boldsymbol{T}}, \hat{\boldsymbol{\beta}}_{\mathbf{2}}^{\boldsymbol{T}})^T$. Usando resultados conhecidos de álgebra linear, mostramos que a variância assintótica de $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ é dada por

$$Var(\hat{\boldsymbol{\beta}}_{\mathbf{1}} = \phi^{-1}[\mathbf{X}_1^T \mathbf{W}^{1/2} \mathbf{M_2} \mathbf{W}^{1/2} \mathbf{X}_1]^{-1},$$

em que $\mathbf{X_1}$ sai da partição $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2})$, sendo portanto $n \times q$, $\mathbf{X_2}$ é $n \times (p-q)$, $\mathbf{M_2} = \mathbf{I_n} - \mathbf{H_2}$ e $\mathbf{H_2} = \mathbf{W^{1/2} X_2 (X_2^T W X_2)^{-1} X_2^T W^{1/2}}$ é a matriz de projeção ortogonal de vetores do $\mathbb{R}^n$ no subespaço gerado pelas colunas da matriz $\mathbf{W^{1/2} X_2}$. Em particular, no caso normal linear, temos as simplificações $\mathbf{H_2} = \mathbf{X_2 (X_2^T X_2)^{-1} X_2^T}$ e $Var(\hat{\beta}_{\mathbf{1}}) = \sigma^2 [\mathbf{X_1^T (I_n - H_2) X_1}]^{-1}$.

2.2 Análise de Diagnóstico

A análise de diagnóstico é a etapa da modelagem que verifica possível presença de violação das suposições feitas para o modelo proposto com destaque para o componente aleatório e parte sistemática. Também é verificada a presença de pontos discrepantes, pois estes poderiam apresentar um peso desproporcional na estimativa dos parâmetros do modelo.

Inicialmente, define-se o resíduo para a $i-$ésima observação como uma função $r_i = r(y_i, \hat{\mu}_i)$ que busca medir a distância entre o valor observado e o valor predito pelo modelo. A definição de resíduo mais utilizada é $r_i = y_i - \hat{\mu}_i$, o resíduo ordinário. O vetor de resíduos ordinários é definido por $\mathbf{r} = (r_1, ..., r_n)^T$.

A matriz de projeção foi proposta por Hoaglin e Welsch (1978), conforme citado por Paula (2013) é definida por $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}$, em que $\mathbf{X}$ denota a matriz modelo. O estudo da sua diagonal principal motivou a definição de pontos de alavanca.

2.2.1 Resíduos

Os resíduos mais utilizados em MLGs são definidos a partir dos componentes da função desvio. A versão padronizada é:

$$t_{D_i} = \frac{d^*(y_i, \hat{\mu}_i)}{\sqrt{1-\hat{h}_{ii}}} = \frac{\phi^{1/2} d(y_i, \hat{\mu}_i)}{\sqrt{1-\hat{h}_{ii}}}, \tag{2.2.1}$$

em que h_{ii} é o i-ésimo elemento da diagonal principal da matriz $\mathbf{H}$, e $d(y_i, \hat{\mu}_i) = \pm\sqrt{2}\{y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i))\}^{1/2}$. O sinal de $d(y_i, \hat{\mu}_i)$ é o mesmo de $y_i - \hat{\mu}_i$. De acordo com Paula (2013), ao citar McCullagh (1987) a distribuição de probabilidades de

$$\frac{d^*(Y_i, \mu_i) + \rho_{3i}/6}{\sqrt{1 + (14\rho_{3i}^2 - 9\rho_{4i})/36}}, \tag{2.2.2}$$

para os MLGs é aproximadamente N(0,1), em que ρ_{3i} e ρ_{4i} são os coeficientes de assimetria e curtose de $\partial L(\eta_i)/\partial\eta_i$, respectivamente, e $d^*(Y_i; \mu_i)$ é o i-ésimo componente do desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ avaliado no parâmetro verdadeiro.

2.3 Poisson

No caso de $Y \sim Poisson(\mu)$, a função de probabilidades fica dada por:

$$P(Y = y) = \frac{e^{-\mu}\mu^y}{y!} = \exp\{y \log \mu - \mu - \log y!\}. \tag{2.3.1}$$

Em que $\mu > 0$ e $y = 0, 1, ...$
então $\log \mu = \theta,\ b(\theta) = e^\theta,\ \phi = 1$ e $c(y, \phi) = -\log y!$.
Segue portanto que $V(\mu) = \mu$

2.3.1 Sobredispersão

Uma característica dos modelos de Poisson é o desvio padrão depender da média, isso inviabiliza a implementação de um modelo normal linear homocedástico para explicar o parâmetro λ. Existem formas de contornar essa limitação, como a transformação na resposta Y, de forma a garantir aproximada constância da variância.

Tal propriedade limita a possibilidade de adequação do modelo de Poisson. Assim, naturalmente existem vários exemplos de dados que apresentam o fenômeno de **sobredispersão** quando é aplicado este modelo para dados de resposta. Isso ocorre quando

a variância dessa resposta é maior do que a média, sendo que para ser adequado, estes deveriam ser iguais. A heterogeneidade das unidades amostrais é uma causa comum desse fenômeno.

Como exemplo (PAULA, 2013), suponha que para um conjunto fixo $\mathbf{x} = (x_1, ..., x_p)^T$ de variáveis explicativas, $Y|z$ tem média z e variância z, e Z que é não observável varia nas unidades amostrais com $\mathbf{x}$ fixo, o que leva a $E(Z) = \mu$. Então,

$$\begin{aligned} E(Y) &= E[E(Y|Z)] = E[Z] = \mu, \\ Var(Y) &= E[Var(Y|Z)] + Var[E(Y|Z)] \\ &= \mu + Var(Z). \end{aligned} \tag{2.3.2}$$

Supondo que $Y|z$ tem distribuição de Poisson com média z e função de probabilidades $f(y|z)$ e que Z segue uma distribuição gama de média μ e parâmetro de dispersão ϕ com função densidade $g(z; \mu, \phi)$. Logo $E(Z) = \mu$ e $Var(Z) = \frac{\mu^2}{\phi}$, de onde segue que $E(Y) = \mu$ e $Var(Y) = \mu + \frac{\mu^2}{\phi}$. Temos

$$\begin{aligned} f(y|z) &= \frac{e^{-z} z^y}{y!}, \\ g(z; \mu, \phi) &= \frac{1}{\Gamma(\phi)} \left(\frac{z\phi}{\mu}\right)^{\phi} e^{-\frac{\phi z}{\mu}} \frac{1}{z}. \end{aligned}$$

A função de probabilidades de Y fica

$$\begin{aligned} Pr\{Y = y\} &= \int_0^{\infty} f(y|z) g(z; \mu, \phi)\, dz \\ &= \frac{1}{y!\phi} \left(\frac{\phi}{\mu}\right)^{\phi} \int_0^{\infty} e^{-z\left(1+\frac{\phi}{\mu}\right)} z^{\phi+y-1}\, dz. \end{aligned}$$

Fazendo a transformação de variável $t = z(1 + \frac{\phi}{\mu})$ temos que $\frac{dz}{dt} = (1 + \frac{\phi}{\mu})^{-1}$. Logo,

$$\begin{aligned} Pr\{Y = y\} &= \frac{1}{y!\Gamma(\phi)} \left(\frac{\phi}{\mu}\right)^{\phi} \left(1 + \frac{\phi}{\mu}\right)^{-(\phi+y)} \int_0^{\infty} e^{-t}\, t^{\phi+y-1}\, dt \\ &= \frac{\Gamma(\phi + y)\mu^y \phi^{\phi}}{\Gamma(\phi)\Gamma(y+1)(\mu+\phi)^{\phi+y}} \\ &= \frac{\Gamma(\phi + y)}{\Gamma(y+1)\Gamma(\phi)} \left(\frac{\mu}{\mu+\phi}\right)^{y} \left(\frac{\phi}{\mu+\phi}\right)^{\phi} \\ &= \frac{\Gamma(\phi + y)}{\Gamma(y+1)\Gamma(\phi)} (1-\pi)^{\phi}\, \pi^y, \;\; y = 0, 1, 2, ... \end{aligned} \tag{2.3.3}$$

com $\pi = \mu/(\mu + \phi)$. Portanto, Y segue uma distribuição **binomial negativa** de média μ e parâmetro de forma ϕ: $Y \sim BN(\mu, \phi)$. Este modelo, por sua vez, tem flexibilidade

para se adequar a dados com sobredispersão.

2.4 Binomial Negativa

Sendo $Y_1, ..., Y_n$ são variáveis aleatórias independentes tais que $Y_i \sim BinNeg(\mu_i, \phi)$. A função de probabilidades fica dada por:

$$f(y_i; \mu_i, \phi) = \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \left(\frac{\mu_i}{\mu_i + \phi}\right)^{y_i} \left(\frac{\phi}{\mu_i + \phi}\right)^{\phi}. \tag{2.4.1}$$

Temos que

$$E(Y_i) = \mu_i,$$
$$Var(Y_i) = \mu_i + \frac{\mu_i^2}{\phi}.$$

Assumimos a parte sistemática dada por $g(\mu_i) = \eta_i = \mathbf{x_i^T}$, em que $\mathbf{x_i} = (x_{i1}, ...x_{ip})^T$ contém os valores de variáveis explicativas, $\beta = (\beta_1, ..., \beta_p)^T$ um vetor de parâmetros desconhecidos e $g(.)$ é a função de ligação. Assim como nos modelos de Poisson, as ligações mais utilizadas são logarítmica ($g(\mu_i) = \log \mu_i$), raiz quadrada($g(\mu_i) = \sqrt{\mu_i}$) e identidade ($g(\mu_i) = \mu_i$)

3 Material

Esse estudo foi realizado considerando os registros de óbito por ocorrência. Segundo o manual de procedimentos sobre o SIM, é calculado o número de óbitos por ocorrência tendo como referência o número de atestados de óbito (SAÚDE, 2001). Portanto, é uma variável quantitativa discreta, em que cada unidade representa um óbito. Pertence à categoria dos Indicadores Demográficos, construídos a partir de eventos intrínsecos ao ciclo de vida da população.

3.1 Conjunto de dados

Microdados extraídos do Sistema de Informações sobre Mortalidade (SIM) do Ministério da Saúde. O download e pré-processamento dos microdados foram feitos com o uso do pacote "microdatasus" (SALDANHA; BASTOS; BARCELLOS, 2019).

O banco de dados possui registros de ocorrência de óbitos segundo 98 variáveis.

Foram consideradas, para a análise descritiva e para a modelagem as seguintes variáveis: Faixa Etária, a partir da Idade (Anos), Capítulo CID-10, Raça/Cor, Escolaridade, Região, Unidade de Federação e Sexo.

Tabela 1: Quadro com a Descrição das Variáveis Utilizadas para Análise Descritiva e Modelagem

Código da Variável	**Descrição**	**Tipo**	**Níveis**
CAUSABAS	Causa básica do óbito	Texto	Código da CID-10
CODMUNRES	Código do município de residência	Numérica	Código de 6 caracteres conforme padrão do IBGE
ESC	Escolaridade	Fator	1: Nenhuma 2: 1 a 3 anos 3: 4 a 7 anos 8 a 11 anos 12 e mais 8: de 9 a 11 anos 0;6;7;9;A: NA
IDADEanos	Idade em anos	Numérica	000;999: NA
RACACOR	Raça ou cor	Fator	1: Branca 2: Preta 3: Amarela 4: Parda 5: Indígena
SEXO	Sexo	Fator	1: Masculino 2: Feminino 0;9: NA

Os registros de morte também foram categorizados a partir das variáveis Idade e Escolaridade. A variável "ENSFUND"categoriza se o indivíduo tem, ou não, pelo menos 8 anos de escolaridade, que é a duração do ensino fundamental, segundo o Ministério da Educação e Cultura (MEC). A variável "FXETARIA"categoriza os indivíduos em 6 níveis de faixa etária.

Tabela 2: Quadro com a Descrição das Categorizações Feitas a Partir das Variáveis do Banco de Dados

Código da Nova Variável	**Descrição**	**Tipo**	**Níveis**
ENSFUND	Possui anos de escolaridade suficientes para conclusão do Ensino Fundamental (8 anos)	Fator	0: Não 1: Sim
FXETARIA	Faixa Etária	Fator	1: 0 a 4 anos 2: 5 a 14 anos 3: 15 a 29 anos 4: 30 a 49 anos 5: 50 a 69 anos 6: 70 anos ou mais

São cinco as classificações de Raça/Cor empregadas pelo IBGE: Branca, Preta, Amarela, Parda e Indígena. Dentro do contexto sociológico brasileiro, porém, as consequências do preconceito e da desigualdade racial se observam quando comparadas as raças branca e "negra", que considera pretos e pardos. Além disso, em 2020, brancos, pretos e pardos somam 96.40% do total de registros de morte no SIM, desconsiderando os que tiveram a Raça/Cor ignorada.

Por esse motivo, a análise descritiva foi feita também, considerando raça/cor como uma variável binária, que indica se o indivíduo é considerado pertencente da raça/cor

branca, ou "negra". E a modelagem dos dados foi feita apenas com essa variável binária, desconsiderando os Amarelos e Índigenas.

Tabela 3: Quadro com a Descrição da Categorização Feita por Raça/Cor

Código da Nova Variável	**Descrição**	**Tipo**	**Níveis**
RACACOR (Binária)	Raça ou cor: Branca ou Negra	Fator	0: Branca 1: Negra (Preta ou Parda)

Estão disponíveis os dados dos anos de 1996 a 2019, e os dados preliminares do ano de 2020. O período considerado no estudo foi o ano de 2020, que teve 2.65% dos registros de Raça/Cor respondidos como "Ignorado".

Na modelagem, a variável resposta é a contagem de homicídios, segundo os diferentes níveis das variáveis explicativas consideradas. Assume valores inteiros não negativos.

3.1.1 Dados de Homicídio

A análise do número de mortes, segundo capítulo CID-10, aponta para um maior contingente de negros que tem como causa de morte o capítulo XX, "causas externas de morbidade e mortalidade". Essa classificação inclui os códigos da CID-10 que se referem aos homicídios. Esses códigos são do X85 ao X99 e Y00 ao Y09, e suas respectivas descrições estão na tabela a seguir.

Código CID-10	Descrição
X85	Agressão Por Meio de Drogas, Medicamentos e Substâncias Biológicas
X86	Agressão Por Meio de Substâncias Corrosivas
X87	Agressão Por Pesticidas
X88	Agressão Por Meio de Gases e Vapores
X89	Agressão Por Meio de Outros Produtos Químicos e Substâncias Nocivas Especificados
X90	Agressão Por Meio de Produtos Químicos e Substâncias Nocivas Não Especificados
X91	Agressão Por Meio de Enforcamento, Estrangulamento e Sufocação
X92	Agressão Por Meio de Afogamento e Submersão
X93	Agressão Por Meio de Disparo de Arma de Fogo de Mão
X94	Agressão Por Meio de Disparo de Espingarda, Carabina ou Arma de Fogo de Maior Calibre
X95	Agressão Por Meio de Disparo de Outra Arma de Fogo ou de Arma Não Especificada
X96	Agressão Por Meio de Material Explosivo
X97	Agressão Por Meio de Fumaça, Fogo e Chamas
X98	Agressão Por Meio de Vapor de Água, Gases ou Objetos Quentes
X99	Agressão Por Meio de Objeto Cortante ou Penetrante
Y00	Agressão Por Meio de um Objeto Contundente
Y01	Agressão Por Meio de Projeção de um Lugar Elevado
Y02	Agressão Por Meio de Projeção ou Colocação da Vítima Diante de um Objeto em Movimento
Y03	Agressão Por Meio de Impacto de um Veículo a Motor
Y04	Agressão Por Meio de Força Corporal
Y05	Agressão Sexual Por Meio de Força Física
Y06	Negligência e Abandono
Y07	Outras Síndromes de Maus Tratos
Y08	Agressão Por Outros Meios Especificados
Y09	Agressão Por Meios Não Especificados

Tabela 4: Descrição dos Códigos referentes a agressão na CID-10

4 Resultados

4.1 Análise Descritiva de Mortalidade

A primeira análise realizada foi sobre os dados de mortalidade por todas as causas, para o ano de 2020. Foram feitos os recortes, dois a dois, de Raça/Cor por Faixa Etária e Causa básica do óbito. Os resultados para os anos de 1996 ate 2019 estão no Apêndice B.

4.1.1 Mortalidade por Faixa Etária e Raça/Cor

Os resultados de mortalidade para o ano de 2020, considerando as variáveis Faixa Etária e Raça/cor são:

Tabela 5: Mortalidade por faixa etária e raça/cor de 2020

Faixa Etária	Amarela	Branca	Ignorada	Indígena	Parda	Preta
0 a 4 anos	80	13331	2722	728	17981	1059
5 a 14 anos	13	2291	207	122	3433	365
15 a 29 anos	181	19772	1266	457	41633	5972
30 a 49 anos	557	61783	3998	738	88557	18451
50 a 69 anos	2241	216171	12193	1198	190495	46861
70 ou mais	6351	448932	20121	2020	259292	58723

Em 2020, a proporcão de mortes de indivíduos considerados de raça/cor branca ou negra (preta ou parda) sobre o total, desconsiderando os que ignoraram esse campo, é 96.44% das mortes.

Tabela 6: Mortalidade por faixa etária e raça/cor binária de 2020

Faixa Etária	Branca	Negra
0 a 4 anos	13331	19040
5 a 14 anos	2291	3798
15 a 29 anos	19772	47605
30 a 49 anos	61783	107008
50 a 69 anos	216171	237356
70 ou mais	448932	318015

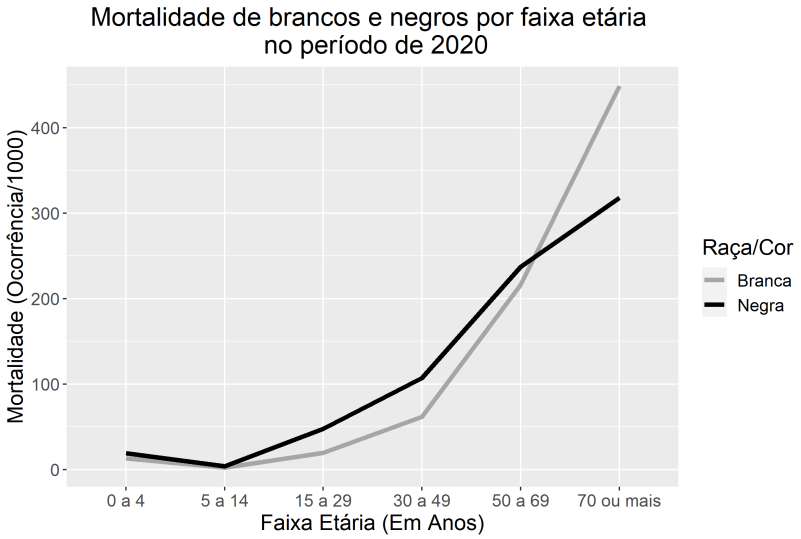


Figura 1: Mortalidade de Brancos e Negros segundo Faixa Etária no período de 2020

Considerando os dados do ano de 2020, a diferença entre a mortalidade observada de negros e brancos é notável. A mortalidade de negros se mantém mais elevada até a faixa etária de 50 a 69 anos, invertendo a ordem com a mortalidade de brancos apenas na faixa de 70 anos ou mais.

A análise do ano de 2020 teve enorme impacto causado pela pandemia do novo coronavírus. Destacando-se os idosos, pertencentes ao grupo de risco da doença.

4.1.2 Mortalidade por Capítulo CID - 10

A CID-10 existe para catalogar de forma padronizada as doenças e problemas de saúde e tem como referência a Nomenclatura Internacional de Doenças, da Organização Mundial de Saúde (OMS). Essas doenças e condições são divididas em capítulos, e estes foram os níveis da variável analisada.

São os capítulos CID:

Tabela 7: Descrição e Códigos dos Capítulos da CID-10

Capítulo CID-10	Descrição	Códigos da CID-10
I	Algumas doenças infecciosas e parasitárias	A00-B99
II	Neoplasmas (tumores)	C00-D48
III	Doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários	D50-D89
IV	Doenças endócrinas, nutricionais e metabólicas	E00-E90
V	Transtornos mentais e comportamentais	F00-F99
VI	Doenças do sistema nervoso	G00-G99
VII	Doenças do olho e anexos	H00-H59
VIII	Doenças do ouvido e da apófise mastóide	H60-H95
IX	Doenças do aparelho circulatório	I00-I99
X	Doenças do aparelho respiratório	J00-J99
XI	Doenças do aparelho digestivo	K00-K93
XII	Doenças do aparelho digestivo	K00-K93
XIII	Doenças do sistema osteomuscular e do tecido conjuntivo	M00-M99
XIV	Doenças do aparelho geniturinário	N00-N99
XV	Gravidez, parto e puerpério	O00-O99
XVI	Algumas afecções originadas no período perinatal	P00-P96
XVII	Malformações congênitas, deformidades e anomalias cromossômicas	Q00-Q99
XVIII	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte	R00-R99
XX	Causas externas de morbidade e de mortalidade	V01-Y98

Os dados relativos de mortalidade, segundo capítulos da CID-10 para o ano de 2020, são:

Tabela 8: Mortalidade relativa por Capítulo CID-10 e Raça/Cor no período de 2020

Capítulo CID-10	Branca	Preta	Amarela	Parda	Indígena	Ignorado
I	8.14%	1.55%	0.11%	6.63%	0.08%	0.53%
II	8.2%	1.12%	0.1%	4.85%	0.03%	0.36%
III	0.19%	0.04%	0%	0.17%	0%	0.01%
IV	2.77%	0.58%	0.04%	2.31%	0.02%	0.15%
V	0.48%	0.12%	0.01%	0.47%	0%	0.02%
VI	1.87%	0.16%	0.02%	0.77%	0%	0.07%
VII	0%	0%	0%	0%	0%	0%
VIII	0%	0%	0%	0%	0%	0%
IX	11.51%	2.11%	0.14%	8.46%	0.05%	0.54%
X	5.21%	0.74%	0.07%	3.39%	0.03%	0.25
XI	2.09%	0.35%	0.02%	1.67%	0.01%	0.11%
XII	0.22%	0.04%	0%	0.16%	0%	0.01%
XIII	0.21%	0.03%	0%	0.13%	0%	0.01%
XIV	1.6%	0.26%	0.02%	0.94%	0.01%	0.07%
XV	0.04%	0.02%	0%	0.07%	0%	0%
XVI	0.41%	0.03%	0%	0.64%	0.02%	0.11%
XVII	0.28%	0.02%	0%	0.26%	0.01%	0.04%
XVIII	2.59%	0.67%	0.03%	2.77%	0.03% XVIII	0.18%
XX	3.29%	0.64%	0.04%	5.11%	0.04%	0.17%

Brancos e negros (pretos e pardos) representam, juntos, 96,4% do total de mortes, desconsiderando os que tiveram o registro de Raça/Cor ignorado.

Tabela 9: Mortalidade relativa de brancos e negros por Capítulo CID-10 no período de 2020

Capítulo CID-10	Brancos	Negros
I	8.14%	8.18%
II	8.2%	5.97%
III	0.19%	0.21%
IV	2.77%	2.89%
V	0.48%	0.59%
VI	1.87%	0.93%
VII	0%	0%
VIII	0%	0%
IX	11.51%	10.57%
X	5.21%	4.13%
XI	2.09%	2.02%
XII	0.22%	0.20%
XIII	0.21%	0.16%
XIV	1.6%	1.20%
XV	0.04%	0.09%
XVI	0.41%	0.67%
XVII	0.28%	0.28%
XVIII	2.59%	3.44%
XX	3.29%	5.75%

Os Capítulos VII, "Doenças do olho e anexos", e VIII, "Doenças do ouvido e da apófise mastóide", são os dois com menos ocorrências, para ambos os grupos considerados. Juntos somam apenas 169 mortes.

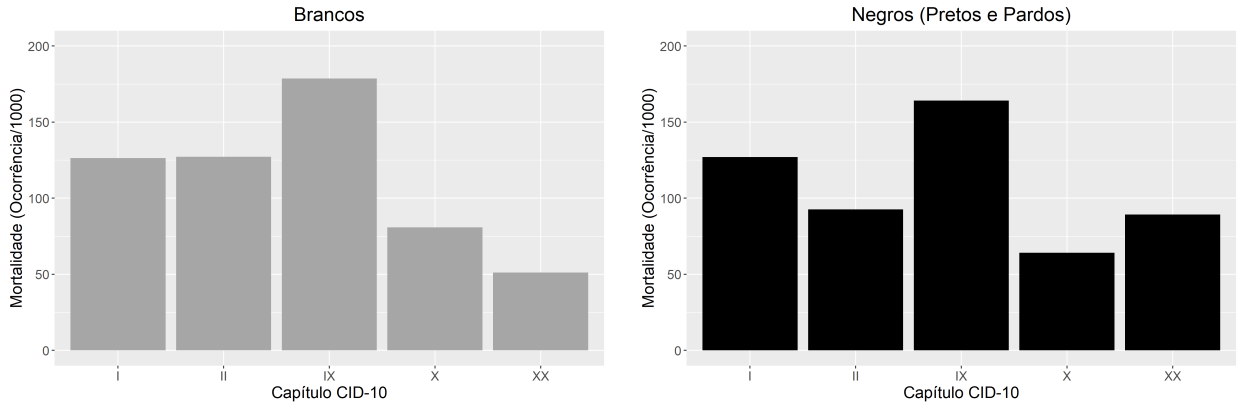
Tabela 10: 5 maiores mortalidades de brancos e negros por Capítulo CID-10 no período de 2020

Capítulo CID-10	Brancos
IX	178711
II	127260
I	126410
X	80844
XX	51155

Capítulo CID-10	Negros
IX	164090
I	127030
II	92664
XX	89342
X	64073

Destaca-se o elevado número de mortes tendo o capítulo I, "Algumas doenças infecciosas e parasitárias", como causa. É a segunda causa mais comum entre negros, e a terceira entre pessoas brancas. O ano de 2020 foi marcado pela pandemia do novo

coronavírus, que explicaria a diferença em relação ao período anterior, por se tratar de uma doença infecciosa.



(a) Raça/Cor = Branca

(b) Raça/Cor = Preta ou Parda

Figura 2: Mortalidade da população de brancos e negros por Capítulo CID-10 no período de 2020

Mortes registradas como causa do Capítulo CID-10 XX, "Causas externas de morbidade e mortalidade", é mais recorrente em 2020 entre os negros, com 89342 ocorrências, do que entre os brancos, com 51155. Esse capítulo inclui causas de morte como: acidentes, agressões, lesões autoprovocadas intencionalmente, entre outras.

4.2 Análise Descritiva de Homicídios

A seguinte análise será feita sobre os dados filtrados sobre os códigos da CID-10 referentes a homicídios, conforme a Tabela 4.

4.2.1 Faixa Etária e Raça/Cor

Para os dados de faixa etária dos respondentes, construídos a partir da idade, 1.39% tem dados faltantes, NA (590 pessoas).

Tabela 11: Homicídios por Faixa Etária e Raça/Cor no Período de 2020

Faixa Etária	Amarela	Branca	Indígena	Parda	Preta
0 a 4 anos	0	25	0	46	6
5 a 14 anos	0	68	5	262	24
15 a 29 anos	42	3637	65	16089	1720
30 a 49 anos	26	3423	58	10419	1092
50 a 69 anos	6	1180	10	2182	216
70 ou mais	4	212	5	280	31

Negros (pardos e pretos) e brancos somam 99.41% do total de homicídios do ano de 2020, desconsiderando aqueles que tiveram o campo Raça/Cor ignorado.

Tabela 12: Homicídios de Brancos e Negros por Faixa Etária no Período de 2020

Faixa Etária	Brancos	Negros
0 a 4 anos	25	52
5 a 14 anos	68	286
15 a 29 anos	3637	17809
30 a 49 anos	3423	11511
50 a 69 anos	1180	2398
70 ou mais	212	311

O número de homicídios de pessoas pretas e pardas é maior que o de brancos em todas as faixas etárias consideradas. Destaque para o intervalo que comporta o maior número de indivíduos: 15 a 29 anos. Nessa faixa etária, que representa 52.4% dos homicídios, tem 4.9 vezes mais homicídios de negros, em comparação aos brancos.

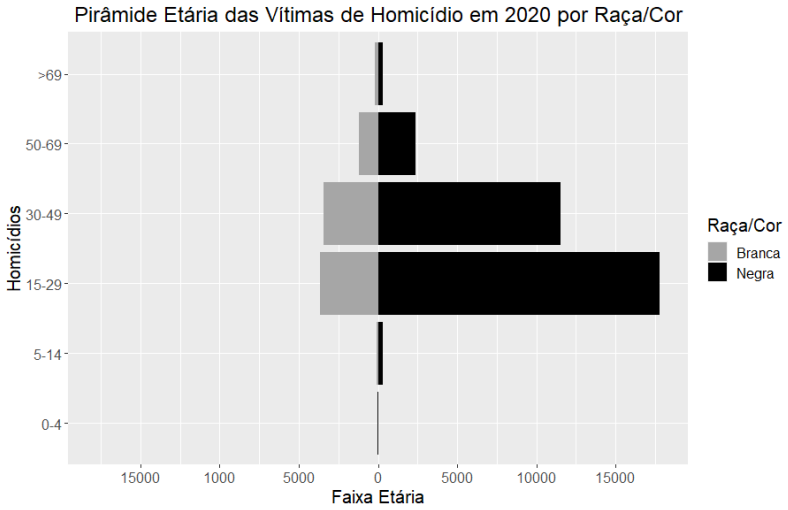


Figura 3: Pirâmide Etária de Brancos e Negros Vítimas de Homicídio no ano de 2020

O gráfico destaca visualmente a disparidade, sobretudo nas faixas etárias de 15 a 29 anos e de 30 a 49 anos. Nas idades mais extremas, o número pequeno de homicídios leva a crer que essa não é uma causa de morte muito relevante para crianças e idosos.

4.2.2 Escolaridade e Raça/Cor

Os dados de homicídio por tempo de escolaridade e raça/cor tiveram 20.11% de não respondentes, um total de 8525 NA's:

Tabela 13: Homicídios por anos de escolaridade e raça/cor em 2020

Anos de Escolaridade	Amarela	Branca	Indígena	Parda	Preta
Nenhuma	5	156	22	1176	121
1 a 3 anos	10	875	22	4804	507
4 a 7 anos	22	2649	48	10972	1184
8 a 11 anos	24	2727	30	6708	704
12 anos e mais	5	349	3	435	33

Mais uma vez, em todos os tempos de escolaridade considerados, os homicídios são mais recorrentes entre pretos e pardos do que entre brancos. O número de indígenas e amarelos também é irrelevante.

Tabela 14: Homicídios por anos de escolaridade e raça/cor em 2020

Anos de Escolaridade	Brancos	Negros
Nenhuma	156	1297
1 a 3 anos	875	5311
4 a 7 anos	2649	12156
8 a 11 anos	2727	7412
12 anos e mais	349	468

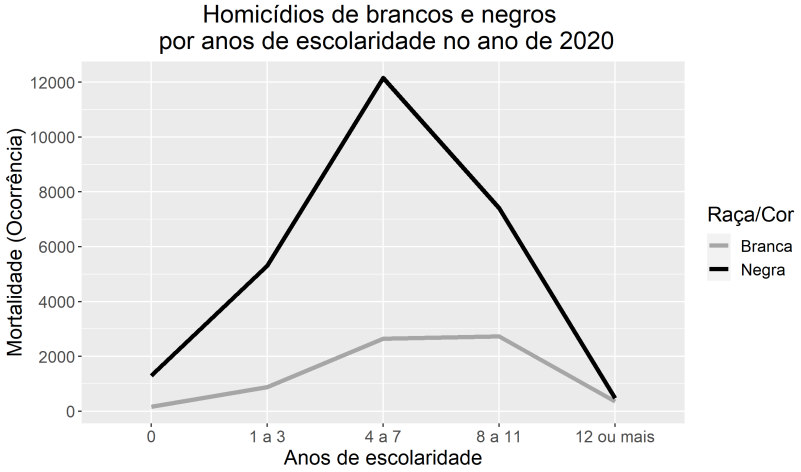


Figura 4: Homicídios de Brancos e Negros segundo Anos de Escolaridade no ano de 2020

É possível traçar um paralelo entre os resultados observados em escolaridade e faixa etária. No tempo de escolaridade nem tão alto, nem zero, observa-se um aumento na disparidade entre as raças observadas.

Os resultados também foram analisados, segundo escolaridade, considerando apenas a população de 15 a 29 anos.

Tabela 15: Homicídios de jovens (15 a 29 anos) brancos e negros por anos de escolaridade e raça/cor em 2020

Anos de Escolaridade	Brancos	Negros
Nenhuma	22	279
1 a 3 anos	302	2620
4 a 7 anos	1315	7562
8 a 11 anos	1225	4301
12 anos ou mais	79	156

Se observa grande concentração de registros de homicídios de negros, na faixa de

4 a 7 anos de escolaridade. São 7562 registros, do total de 12156, se consideradas todas as faixas etárias (62.21%).

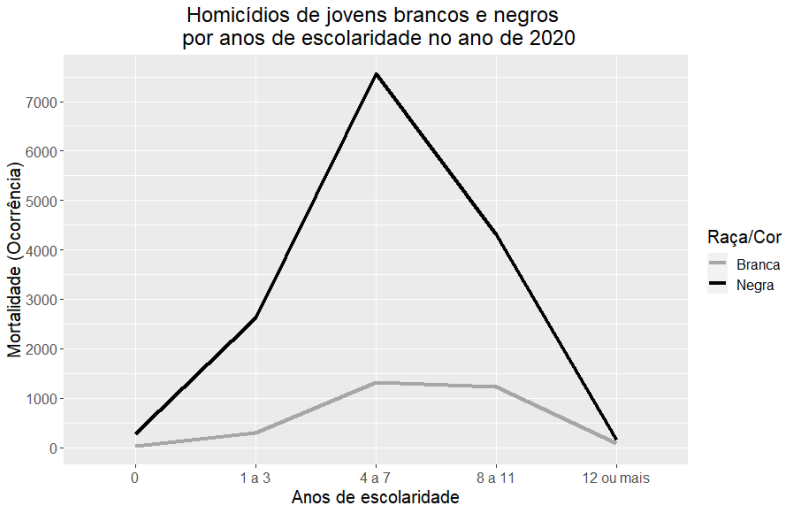


Figura 5: Homicídios de Jovens (15 a 29 Anos) Brancos e Negros segundo Anos de Escolaridade no ano de 2020

O padrão do gráfico se repete, se comparado o observado na Figura 4, que considera todas as faixa etárias.

4.2.3 Ensino Fundamental e Raça/Cor

A partir dos resultados segundo escolaridade e raça/cor, foi criada uma variável binária ("ENSFUND"), que identifica se o respondente tem pelo menos 8 anos de escolaridade. Esse número foi definido levando em consideração a duração do ensino fundamental, segundo o Ministério da Educação e Cultura (MEC).

- **Não**: Não tem anos de escolaridade o suficiente para ter completado o ensino fundamental (Anos de Escolaridade < 8).
- **Sim**: Tem anos de escolaridade o suficiente para ter completado o ensino fundamental (Anos de Escolaridade ≥ 8).

Assim, cruzando essa variável com raça/cor, o resultado está na tabela Tabela 16.

Tabela 16: Homicídios por conclusão do ensino fundamental e raça/cor em 2020

Ensino Fundamental	Amarela	Branca	Indígena	Parda	Preta
Não	37	3680	92	16952	1812
Sim	29	3076	33	7143	737

E os resultados para brancos e negros:

Tabela 17: Homicídios de brancos e negros em 2020, por conclusão do ensino fundamental

Ensino Fundamental	Brancos	Negros
Não	3680	18764
Sim	3076	7880

A maioria (55.41% do total, desconsiderando os NA's) dos assassinatos registrados em 2020 foram de negros, que não tem anos de escolaridade o suficiente para ter completado o ensino fundamental.

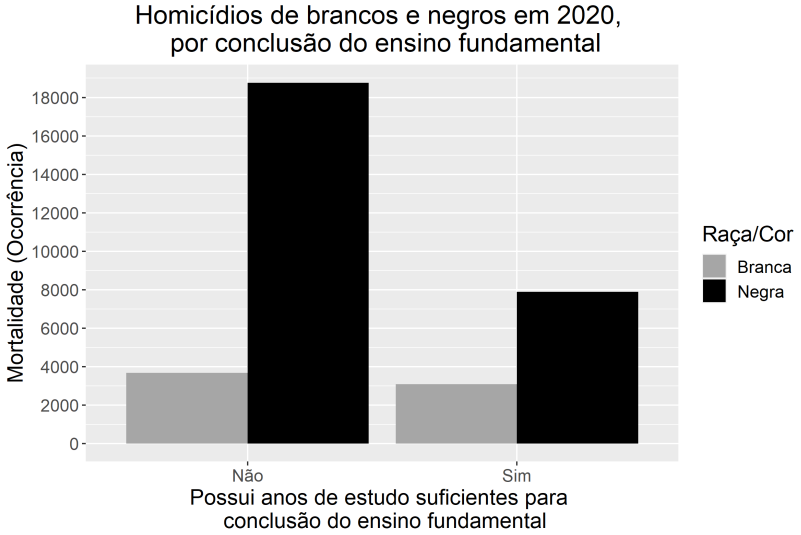


Figura 6: Homicídios de brancos e negros em 2020, por conclusão do ensino fundamental

A análise dessa variável também foi realizada considerando somente a população jovem, de 15 a 29 anos, considerando apenas Brancos e Negros:

Tabela 18: Homicídios de jovens (15 a 29 anos) brancos e negros em 2020, por conclusão do ensino fundamental

Ensino Fundamental	Brancos	Negros
Não	1639	10461
Sim	1304	4457

A proporção de negros, que não tiveram anos de escolaridade o suficiente para conclusão do ensino fundamental é de 58.57% do total. O comportamento se assemelha ao recorte por todas as faixas etárias.

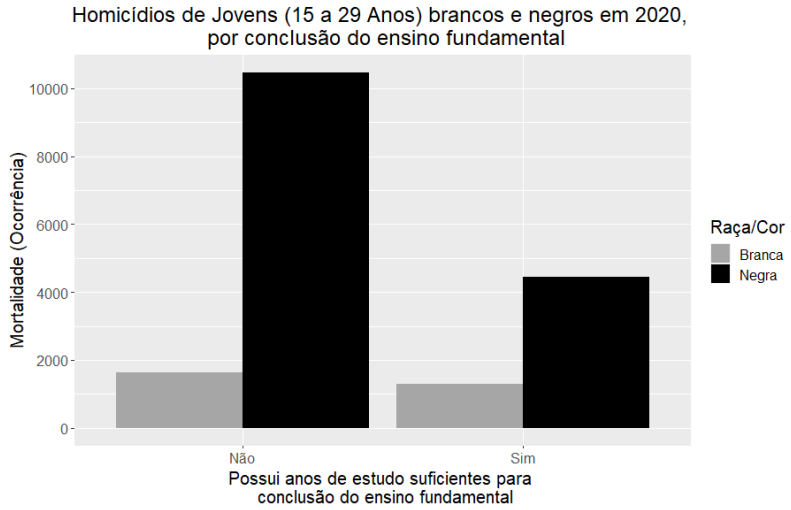


Figura 7: Homicídios de jovens (15 a 29 anos) brancos e negros em 2020, por conclusão do ensino fundamental

4.2.4 Idade e Escolaridade

Análise da idade das vítimas de homicídio, segundo anos de escolaridade:

Tabela 19: Estatísticas de Homicídio por Idade e Tempo de Escolaridade.

Anos de Escolaridade	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Nenhuma	3	31	42	43.54	54	92
1 a 3 anos	7	23	30	32.99	40	97
4 a 7 anos	7	21	26	29.39	35	90
8 a 11 anos	12	22	28	30.57	36	104
12 anos e mais	17	29	36	38.06	45	85

Os maiores valores de média, mediana, primeiro e terceiro quartis, são do grupo com nenhuma escolaridade. Apesar da idade mais elevada, não tiveram nenhum ano de estudo.

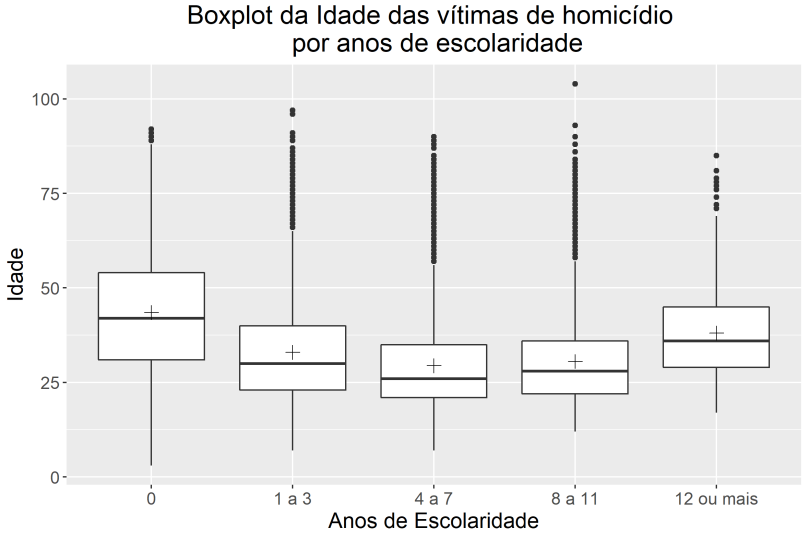


Figura 8: Boxplot da Idade das vítimas de homicídio em 2020 por anos de escolaridade

O boxplot da Figura 8 aponta para comportamentos não muito diferentes. Além do grupo sem anos de escolaridade, apenas os que tiveram 12 ou mais anos, aparentam ter as medidas descritivas um pouco acima das demais.

Assim como na análise da escolaridade e raça/cor, é feita a separação das ocorrências de mortalidade entre indivíduos que tiveram, ou não, pelo menos 8 anos de escolaridade.

Tabela 20: Estatísticas de Homicídio por Idade e Tempo de Escolaridade ≥ 8.

Anos de Escolaridade ≥ 8	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Não	3	21	28	31.31	38	97
Sim	12	22	29	31.13	37	104

Com exceção do mínimo, todas as outras medidaas descritivas apresentadas na Tabela 20 são muito próximas para os dois grupos comparados. O fato de ter completado, ou não, oito anos de escolaridade, parece não ter associação com a idade das vítimas de homicídio de 2020.

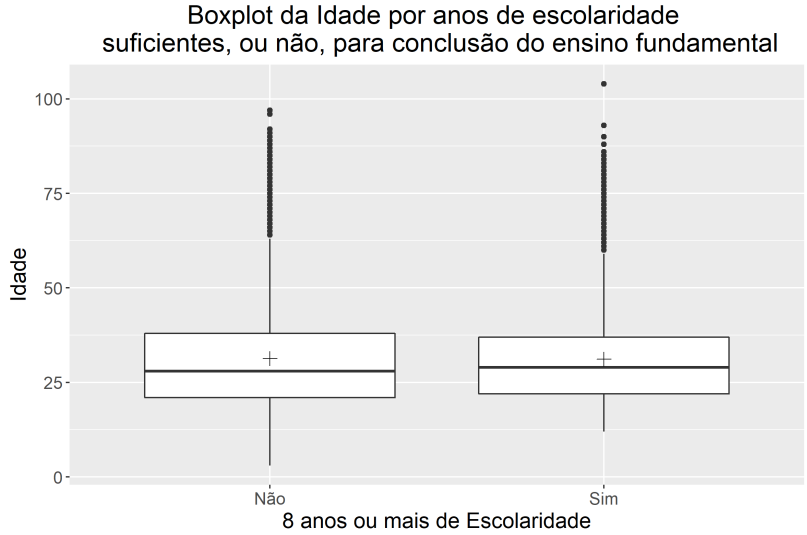


Figura 9: Boxplot da Idade das vítimas de homicídio em 2020 por anos de escolaridade suficientes para conclusão do ensino fundamental

4.2.5 Região e Raça/Cor

Historicamente, as macrorregiões brasileiras apresentam grandes diferenças sociais. Assim, o recorte dos homicídios foi feito levando em conta esse critério. Para essa variável, nenhuma resposta faltante foi observada.

Tabela 21: Homicídios por região e raça/cor no período de 2020

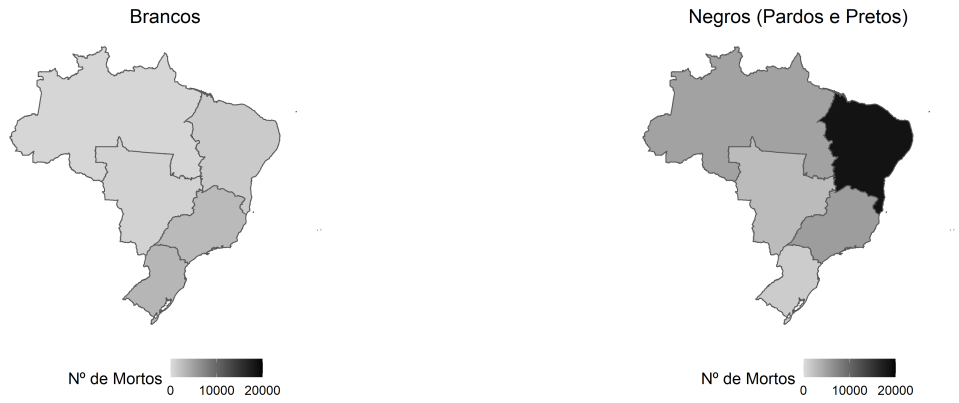
Região	Amarela	Branca	Indígena	Parda	Preta
Norte	19	460	78	4559	274
Nordeste	39	1494	27	17343	1284
Sudeste	8	2750	9	4279	998
Sul	9	3121	12	952	325
Centro-Oeste	10	776	36	2490	234

O baixo número de homicídios de amarelos e indígenas também aparece no recorte por região. É destaque, também, a enorme concentração na população parda do Nordeste, com mais de 17 mil assassinatos.

Tabela 22: Homicídios de brancos e negros por região no período de 2020

Região	Brancos	Negros
Norte	460	4833
Nordeste	1494	18627
Sudeste	2750	5277
Sul	3121	1277
Centro-Oeste	776	2724

A região Sul é a única em que o total de assassinatos de brancos é maior que de negros. Nas demais, homicídios de negros são muito mais comuns.



(a) Raça/Cor = Branca (b) Raça/Cor = Preta ou Parda

Figura 10: Homicídios da população de brancos e negros por região no período de 2020

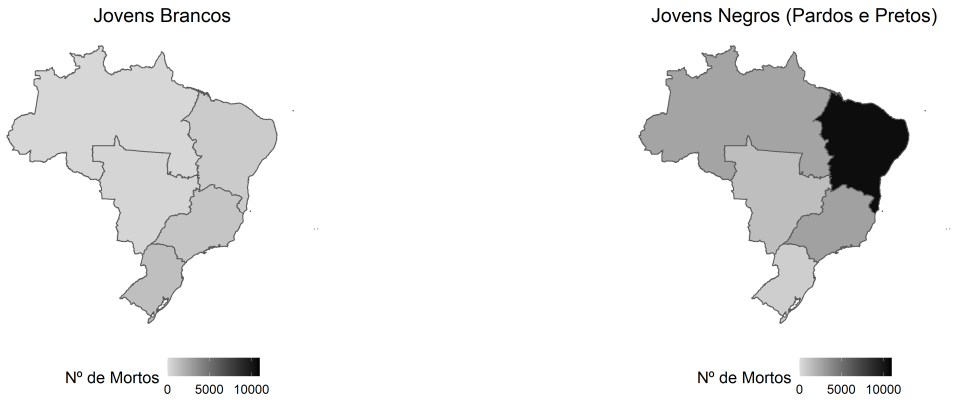
Os mapas da Figura 10 ilustram bem a concentração de homicídios no grupo formado por pessoas negras. O Nordeste também se destaca, na Figura 10 (b).

A análise por macrorregiões brasileiras, também foi feita considerando apenas a população jovem, de 15 a 29 anos.

Tabela 23: Homicídios de jovens (15 a 29 anos) brancos e negros por região no período de 2020

Região	Brancos	Negros
Norte	205	2559
Nordeste	763	10554
Sudeste	1027	2705
Sul	1310	627
Centro-Oeste	332	1364

A região Sul também é a única em que o total de assissinatos de jovens brancos é maior do que de negros. No Nordeste, 56.66 das vítimas de homicídio em 2020, eram jovens.



(a) Raça/Cor = Branca

(b) Raça/Cor = Preta ou Parda

Figura 11: Homicídios da população jovem (15 a 29 anos) de brancos e negros por região no período de 2020

4.2.6 UF e Raça/Cor

O recorte por unidade da federação pode ir mais a fundo nos resultados obtidos por região.

Tabela 24: Homicídios por Unidade da Federação e raça/cor no ano de 2020

UF	Amarela	Branca	Indígena	Parda	Preta
Rondônia	0	87	2	307	28
Acre	0	34	1	239	1
Amazonas	11	87	42	1130	17
Roraima	0	4	20	42	2
Pará	5	164	13	2247	156
Amapá	2	15	0	279	10
Tocantins	1	69	0	315	60
Maranhão	9	183	7	1534	204
Piauí	0	70	1	497	46
Ceará	12	203	0	3578	30
Rio Grande do Norte	0	110	7	1165	13
Paraíba	1	77	1	941	17
Pernambuco	10	490	4	2954	123
Alagoas	2	5	2	1133	7
Sergipe	1	47	0	868	46
Bahia	4	309	5	4673	798
Minas Gerais	2	567	3	1229	271
Espírito Santo	3	124	1	760	107
Rio de Janeiro	0	556	4	1096	372
São Paulo	3	1.503	1	1194	248
Paraná	1	1.162	5	515	83
Santa Catarina	0	486	3	141	28
Rio Grande do Sul	8	1473	4	296	214
Mato Grosso do Sul	1	124	31	275	23
Mato Grosso	1	176	4	586	68
Goiás	4	380	1	1398	103
Distrito Federal	4	96	0	231	40

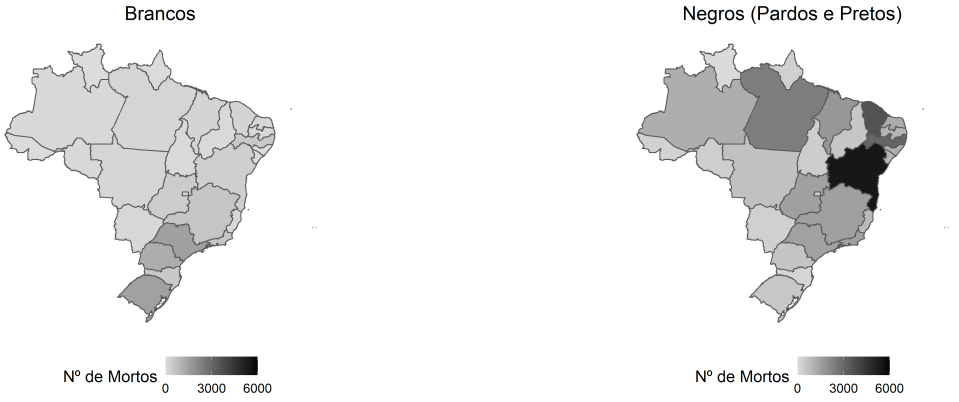
Mais uma vez, os números para amarelos e indígenas são muito reduzidos. Bahia, Ceará e Pernambuco são os estados com maior número de asssassinatos de pardos.

Tabela 25: Homicídios por Unidade da Federação e raça/cor no ano de 2020

UF	Brancos	Negros
Rondônia	87	335
Acre	34	240
Amazonas	87	1147
Roraima	4	44
Pará	164	2403
Amapá	15	289
Tocantins	69	375
Maranhão	183	1738
Piauí	70	543
Ceará	203	3608
Rio Grande do Norte	110	1178
Paraíba	77	958
Pernambuco	490	3077
Alagoas	5	1140
Sergipe	47	914
Bahia	309	5471
Minas Gerais	567	1500
Espírito Santo	124	867
Rio de Janeiro	556	1468
São Paulo	1503	1442
Paraná	1162	598
Santa Catarina	486	169
Rio Grande do Sul	1473	510
Mato Grosso do Sul	124	298
Mato Grosso	176	654
Goiás	380	1501
Distrito Federal	96	271

Apenas nos estados de São Paulo, Paraná, Rio Grande do Sul e Santa Catarina, os assassinatos de brancos são mais registrados que de negros. Destaque para a região sul, que tem esse comportamento em todos os três estados que a compõe.

O fato dos estados de Roraima, Alagoas e Amapá terem registro muito baixo de homicídios de pessoas brancas, possivelmente dificultará a utilização da variável UF na modelagem dos dados.



(a) Raça/Cor = Branca

(b) Raça/Cor = Preta ou Parda

Figura 12: Homicídios da população de brancos e negros por UF no período de 2020

4.2.7 Sexo e Raça/Cor

Foi feita a análise dos homicídios por Raça/Cor e Sexo. Apenas 0.13% das respostas foram NA's (57 respondentes).

Tabela 26: Homicídios por sexo e raça/cor no ano de 2020

Sexo	Amarela	Branca	Indígena	Parda	Preta
Masculino	77	7613	125	27544	2889
Feminino	7	988	34	2076	225

O número de amarelos e indígenas do sexo masculino não é tão baixo como em outras variáveis. Porém ainda representam, somados, uma parcela não muito significante da população.

Tabela 27: Homicídios por sexo de brancos e negros no ano de 2020

Sexo	Brancos	Negros
Masculino	7613	30433
Feminino	988	2301

A ocorrência de homicídios é maior entre negros, e também é maior entre os homens, em relação às mulheres. As duas variáveis analisadas parecem influenciar a quantidade de assassinatos. Destaque para a quantidade de homens negros, 30433 representando 71.79% do total de homicídios.

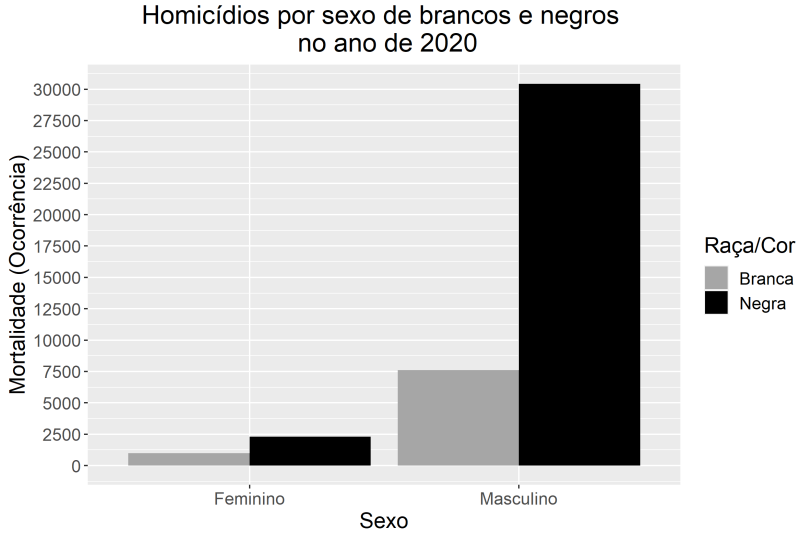


Figura 13: Homicídios da população de brancos e negros por sexo no período de 2020

A análise por Sexo, considerando apenas os jovens de 15 a 29 anos:

Tabela 28: Homicídios por sexo de jovens (15 a 29 anos) brancos e negros no ano de 2020

Sexo	Brancos	Negros
Masculino	3301	16766
Feminino	336	1043

Assim como quando consideradas todas as faixas etárias, existe uma concentração de homicídios de jovens negros do sexo masculino.

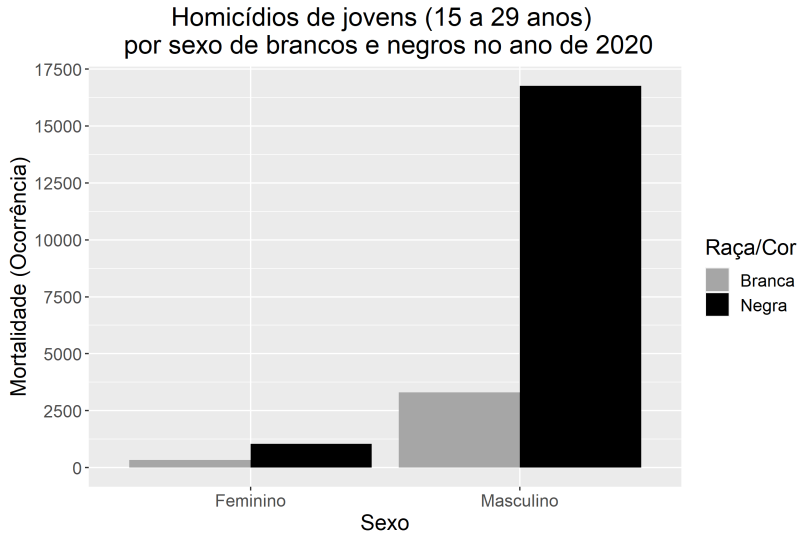


Figura 14: Homicídios da população jovem (15 a 29 anos) de brancos e negros por sexo no período de 2020

4.3 Modelagem

O modelo de Poisson foi o primeiro a ser considerado, e uma de suas suposições é a variância da resposta ser igual a sua média. Quando a variância é maior que a média, o fenômeno de sobredispersão pode ser observado. Uma alternativa é o modelo com resposta Binomial Negativa, que tem a flexibilidade necessária para se adequar a dados com essa característica.

"Raça/Cor"é uma das variáveis explicativas, a fim de verificar a intensidade das consequências da desigualdade racial. Todas as variáveis presentes na análise descritiva, e as interações entre elas, também entraram no modelo e suas significâncias foram avaliadas por meio de testes de hipóteses. Conforme a Tabela 12 indica, mais da metade dos registros de homicídios são de jovens, na faixa etária de 15 a 29 anos. Portanto, a modelagem

considerando apenas essa população foi realizada. O nível de significância utilizado foi de 0.05.

4.3.1 Modelo Poisson

O primeiro modelo implementado, supõe que a variável resposta de contagem de homicídios segue uma distribuição de Poisson. Inicialmente, serão consideradas como candidatas a variáveis explicativas: Raça/Cor (β_i), Faixa Etária (γ_j), ENSFUND (δ_k), Região (λ_l) e Sexo (τ_m). Os indivíduos de todas as idades foram considerados.

Sendo Y_{ijklmr} o r-ésimo indivíduo da i-ésima raça/cor, j-ésima Faixa Etária. k-ésima definição de anos suficientes para conclusão do ensino médio, l-ésima região e m-ésimo sexo, a descrição do modelo é:

- $Y_{ijklmr} \sim Poisson(\mu_{ijklmr})$;
- i = 1,2. (Raça/Cor: 1 - Branca, 2 - Negra);
- j = 1, 2, 3, 4, 5, 6. (Faixa Etária: 1 - 0 a 4 anos, 2 - 5 a 14 anos, 3 - 15 a 29 anos, 4 - 30 a 49 anos, 5 - 50 a 69 anos, 6 - Mais de 69 anos).
- k = 1, 2. (1 - Não possui 8 anos de estudo, 2 - Possui pelo menos 8 anos de estudo).
- l = 1, 2, 3, 4, 5. (1 - Norte, 2 - Nordeste, 3 - Sudeste, 4 - Sul, 5 - Centro-Oeste).
- m = 1,2. (1 - Feminino, 2 - Masculino).
- $g(\mu_{ijklmr}) = \alpha + \beta_i + \gamma_j + \delta_k + \lambda_l + \tau_m$. A definir a função de ligação $g(.)$.
- $\beta_1 = \gamma_1 = \delta_1 = \lambda_1 = \tau_1 = 0$

Tabela 29: 1º Modelo: Comparação das Medidas AIC e BIC para as diferentes funções de ligação.

Função de Ligação	AIC	BIC	Função Desvio
log	10691.94	10734.35	9693.86
raiz quadrada	65245.93	65288.35	64247.86
identidade	63590.92	63633.34	62592.85
inversa	9282.02	9324.44	8283.94

Para as funções de ligação raiz quadrada, identidade e inversa, o algoritmo não convergiu. Desta forma, a função de ligação escolhida foi $g(.) = log(.)$.

Então, foi considerado como "Modelo 1": Variável resposta Poisson, função de ligação logarítmica e variáveis explicativas: Raça/Cor, Faixa Etária, ENSFUND, Região e Sexo.

Tabela 30: Resultados dos Parâmetros, Intervalos de Confiança e Testes de Hipóteses para o Modelo 1.

Parâmetro	Estimativa	EP	IC(95%)	Valor Z	p-valor
α (Intercepto)	-3.69	0.58	(-4.82 ; -2.56)	-6.38	< 0.0001
β_2 (Raça/Cor - Negra)	1.37	0.01	(1.34 ; 1.40)	100.54	< 0.0001
γ_2 (Faixa Etária - 5 a 14 Anos)	2.84	0.58	(1.70 ; 3.98)	4.90	< 0.0001
γ_3 (Faixa Etária - 15 a 29 Anos)	6.93	0.58	(5.80 ; 8.07)	12.01	< 0.0001
γ_4 (Faixa Etária - 30 a 49 Anos)	6.54	0.58	(5.41 ; 7.67)	11.32	< 0.0001
γ_5 (Faixa Etária - 50 a 69 Anos)	5.08	0.58	(3.95 ; 6.22)	8.80	< 0.0001
γ_6 (Faixa Etária - $>$ 69 Anos)	3.15	0.58	(2.02 ; 4.29)	5.44	< 0.0001
δ_2 (Anos de estudo ≥ 8)	-0.71	0.01	(-0.74 ; -0.69)	-61.15	< 0.0001
λ_2 (Região - Nordeste)	1.28	0.02	(1.25 ; 1.31)	76.73	< 0.0001
λ_3 (Região - Sudeste)	0.25	0.02	(0.21 ; 0.29)	12.87	< 0.0001
λ_4 (Região - Sul)	-0.28	0.02	(-0.32 ; -0.23)	-12.34	< 0.0001
λ_5 (Região - Centro Oeste)	-0.45	0.02	(-0.49 ; -0.40)	-18.88	< 0.0001
τ_2 (Sexo - Masculino)	2.48	0.02	(2.44 ; 2.52)	120.94	< 0.0001

Todos os coeficientes foram significativos para entrar no modelo. Porém, o desvio $D(\mathbf{y}, \tilde{\boldsymbol{\mu}}) = 9693.86$ $(p - valor \approx 0)$, que considera como adequada a aproximação pela distribuição $\chi^2_{(180)}$ indica que o modelo não se ajustou bem aos dados.

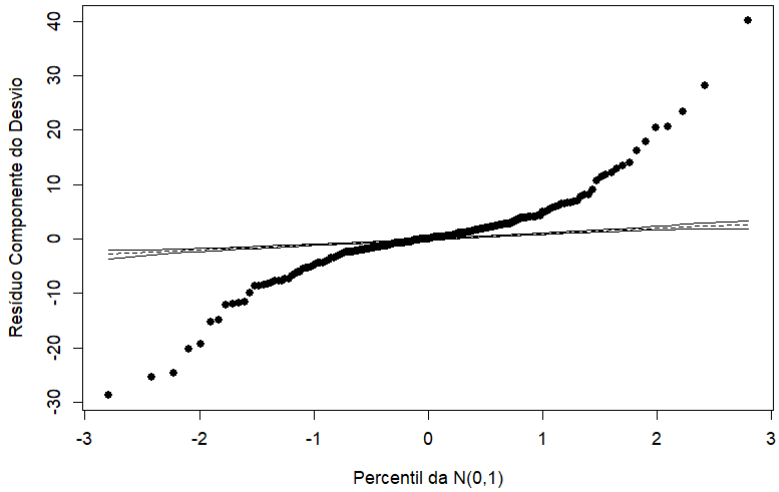


Figura 15: Gráficos de Envelope - Modelo 1

O gráfico de envelope do modelo 1 reforça que o ajuste não é adequado. O fenômeno de sobredispersão será investigado.

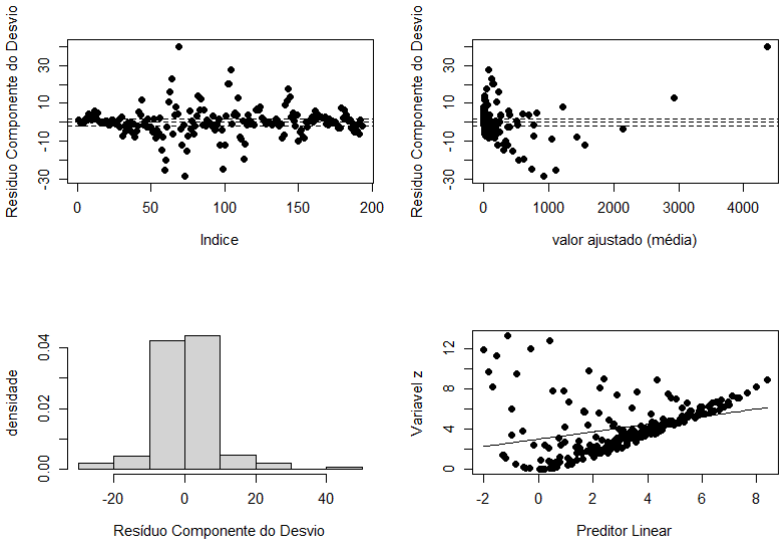


Figura 16: Gráficos de Diagnóstico - Modelo 1

Na Figura 16, os gráficos de resíduo componente do desvio, pelo índice e pelo valor ajustado, tem muitos pontos extrapolando os limites -2 e 2 da normal. O que pode indicar que alguma variabilidade não está sendo explicada no modelo.

[1]Poucas observações.

Tabela 31: Média e Variância dos Diferentes Níveis das Variáveis Explicativas do Modelo 1.

Variável-Nível	Média	Variância
Raça/Cor - Branca	70.33	14791.99
Raça/Cor - Negra	274.41	624458.1
Faixa Etária - 0 a 4 anos	1	0[1]
Faixa Etária - 5 a 14 anos	8.36	280.43
Faixa Etária - 15 a 29 anos	446.52	1110341
Faixa Etária - 30 a 49 anos	300.65	343107.7
Faixa Etária - 50 a 69 anos	70.2	16253.96
Faixa Etária - 69 anos ou mais	10.7	287.94
Anos de Estudo - Menos de 8 anos	217.95	548732.5
Anos de Estudo - 8 anos ou mais	121.69	78086.31
Região - Norte	117.46	85220.94
Região - Nordeste	434.21	1419714
Região - Sudeste	147.6	68087.22
Região - Sul	91.05	21.162
Região - Centro-Oeste	76.97	21946.89
Sexo - Feminino	27.02	2233.62
Sexo - Masculino	314.32	610012.5

Com exceção da faixa etária de 0 a 4 anos, que têm poucos registros de homicídios no ano de 2020, todas as demais respostas das variáveis têm uma média de homicídios muito menor do que a variância. Isso viola um dos pressupostos da distribuição Poisson, que têm os valores da média e variância iguais.

O próximo modelo considera apenas a população dos jovens de 15 a 29 anos. Variável resposta Poisson, considerando as variáveis explicativas: Raça/Cor (β_i), ENS-FUND (δ_k), Região (λ_l) e Sexo (τ_m).

Sendo Y_{iklmr} o r-ésimo indivíduo da i-ésima raça/cor, k-ésima definição de anos suficientes para conclusão do ensino médio, l-ésima região e m-ésimo sexo, a descrição do modelo é:

- $Y_{iklmr} \sim Poisson(\mu_{ijklmr})$;
- i = 1,2. (Raça/Cor: 1 - Branca, 2 - Negra);
- k = 1, 2. (1 - Não possui 8 anos de estudo, 2 - Possui pelo menos 8 anos de estudo).
- l = 1, 2, 3, 4, 5. (1 - Norte, 2 - Nordeste, 3 - Sudeste, 4 - Sul, 5 - Centro-Oeste).

- m = 1,2. (1 - Feminino, 2 - Masculino).
- $g(\mu_{ijklmr}) = \alpha + \beta_i + \gamma_j + \delta_k + \lambda_l + \tau_m$. A definir a função de ligação $g(.)$.
- $\beta_1 = \delta_1 = \lambda_1 = \tau_1 = 0$

Dessa forma, o segundo modelo tende a ser mais parcimonioso, uma vez que o número de parâmetros a serem estimados é menor. O foco na população jovem, pode também trazer uma análise mais eficiente, pois é a faixa etária que mais sofre com os assassinatos.

Tabela 32: 2º Modelo: Comparação das Medidas AIC e BIC para as diferentes funções de ligação.

Função de Ligação	AIC	BIC	Função Desvio
log	4193.75	4207.26	3912.99
raiz quadrada	13999.21	14012.73	13718.46
identidade	20851.69	20865.20	20570.94
inversa	2711.84	2725.35	2431.09

As funções de ligação raiz quadrada e identidade não convergiram. Entre as duas demais, a inversa apresenta valores menores para o AIC, BIC e desvio. Foi então a empregada para o modelo 2.

Modelo 2: Variável resposta Poisson, função de ligação inversa, e variáveis explicativas Raça/Cor, ENSFUND, Região e Sexo.

Tabela 33: Resultados dos Parâmetros, Intervalos de Confiança e Testes de Hipóteses para o Modelo 2.

Parâmetro	Estimativa	EP	IC(95%)	Valor Z	p-valor
α (Intercepto)	0.0176	0.0005	(0.0166 ; 0.0185)	36.8727	< 0.0001
β_2 (Raça/Cor - Negra)	-0.0028	0.0001	(-0.0029 ; -0.0026)	-36.2327	< 0.0001
δ_2 (Anos de Estudo ≥ 8)	0.0004	< 0.0001	(0.0004 ; 0.0004)	32.6380	< 0.0001
λ_2 (Região - Nordeste)	-0.0006	< 0.0001	(-0.0007 ; -0.0006)	-32.4168	< 0.0001
λ_3 (Região - Sudeste)	0.0001	< 0.0001	(0.0000 ; 0.0001)	2.0174	< 0.0001
λ_4 (Região - Sul)	0.0017	0.0001	(0.0015 ; 0.0019)	18.1087	< 0.0001
λ_5 (Região - Centro-Oeste)	0.0010	0.0001	(0.0009 ; 0.0011)	16.1444	< 0.0001
τ_2 (Sexo - Masculino)	-0.0140	0.0005	(-0.0149 ; -0.0131)	-29.5378	< 0.0001

Observa-se na Tabela 33 que as estimativas dos parâmetros ficaram muito próximas de zero, apesar de significativos. Por esse motivo, a função de ligação foi definida como a logarítmica, que apresentou os melhores resultados de AIC e BIC, depois da inversa.

Modelo 2.1: Variável resposta Poisson, função de ligação logarítmica, e variáveis explicativas Raça/Cor, ENSFUND, Região e Sexo.

Tabela 34: Resultados dos Parâmetros, Intervalos de Confiança e Testes de Hipóteses para o Modelo 2.1.

Parâmetro	Estimativa	EP	IC(95%)	Valor Z	p-valor
α (Intercepto)	2.85	0.04	(2.77 ; 2.93)	72.41	< 0.0001
β_2 (Raça/Cor - Negra)	1.62	0.02	(1.58 ; 1.66)	80.47	< 0.0001
δ_2 (Anos de estudo ≥ 8)	-0.74	0.02	(-0.77 ; -0.71)	-46.36	< 0.0001
λ_2) (Região - Nordeste)	1.36	0.02	(1.31 ; 1.40)	59.98	< 0.0001
λ_3 (Região - Sudeste)	0.15	0.03	(0.09 ; 0.20)	5.36	< 0.0001
λ_4 (Região - Sul)	-0.46	0.03	(-0.52 ; -0.39)	-14.14	< 0.0001
λ_5 (Região - Centro Oeste)	-0.53	0.03	(-0.60 ; -0.47)	-16.05	< 0.0001
τ_2 (Sexo - Masculino)	2.70	0.03	(2.64 ; 2.76)	87.66	< 0.0001

Todos os coeficientes das variáveis explicativas consideradas, e o intercepto, deram significativos, a um nível de 0.05. Os intervalos de confiança não incluem o zero. O valor do desvio $D(\mathbf{y}, \tilde{\boldsymbol{\mu}}) = 3912.99$ ($p - valor \approx 0$) indica que o modelo não se ajusta bem aos dados, considerando adequada a aproximação pela $\chi^2_{(32)}$.

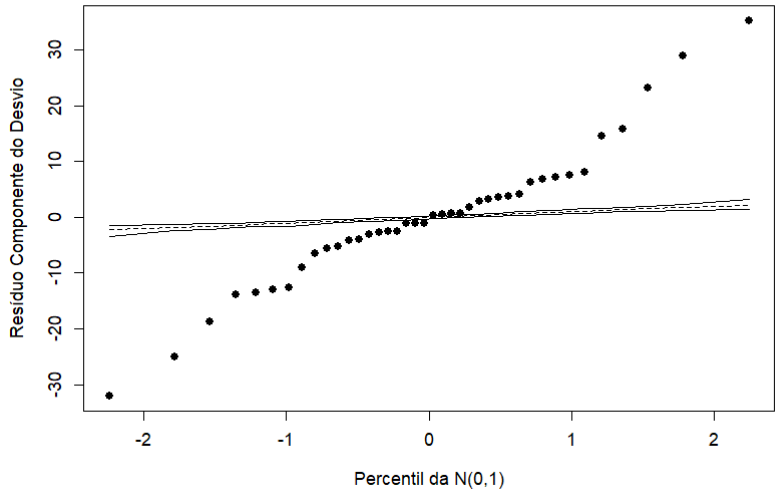


Figura 17: Gráficos de Envelope - Modelo 2.1

O gráfico de envelope mostra poucos pontos entre as bandas de confiança. O que é mais uma evidência da falta de qualidade do ajuste.

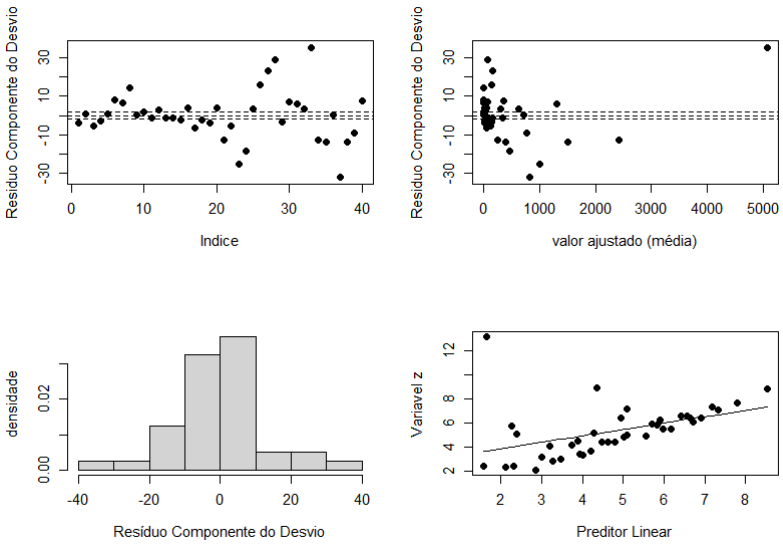


Figura 18: Gráficos de Diagnóstico - Modelo 2.1

A exemplo do modelo 1, que considerou as vítimas de assassinato de todas as faixas etárias, no modelo 2.1, os gráficos do resíduo componente do desvio pelo índice, e pelo valor ajustado, indicam haver alguma variabilidade não presente no modelo. Essa conclusão se dá pela grande quantidade de pontos fora das bandas de confiança.

Novamente, o fenômeno de sobredispersão parece estar presente, ao comparar as médias e variâncias dos níveis de cada variável explicativa considerada.

Tabela 35: Média e Variância dos Diferentes Níveis das Variáveis Explicativas do Modelo 2.1.

Variável-Nível	Média	Variância
Raça/Cor - Branca	147.15	29018.77
Raça/Cor - Negra	745.9	2061417.46
Anos de Estudo - Menos de 8 anos	605	2012430.53
Anos de Estudo - 8 anos ou mais	288.05	213818.47
Região - Norte	307.5	263994
Região - Nordeste	1193.63	4835368.84
Região - Sudeste	356.38	170977.41
Região - Sul	194.5	38024.86
Região - Centro-Oeste	180.63	55155.13
Sexo - Feminino	56.25	5216.30
Sexo - Masculino	836.8	1953242.69

Os resultados mostram que os valores das variâncias amostrais seguem muito distantes dos valores das médias amostrais. Isso atrapalha a implementação do modelo,

considerando a variável resposta Poisson.

4.3.2 Modelo Binomial Negativo

Os resultados apontam para a necessidade de contornar o problema de superdispersão. A distribuição binomial negativa é mais flexível e possibilita bom ajuste sobre dados de contagem, mesmo com suas variâncias muito acima das médias.

O terceiro modelo, então, teve como distribuição da variável resposta, a Binomial Negativa. Foram consideradas candidatas, as mesmas variáveis explicativas do modelo de Poisson. A população é a de jovens, no Brasil, de 15 a 29 anos, no ano de 2020.

Sendo Y_{iklmr} o r-ésimo indivíduo da i-ésima raça/cor, k-ésima definição de anos suficientes para conclusão do ensino médio, l-ésima região e m-ésimo sexo, a descrição do modelo é:

- $Y_{iklmr} \sim BinNeg(\mu_{iklmr}, \phi)$;
- i = 1,2. (Raça/Cor: 1 - Branca, 2 - Negra);
- k = 1, 2. (1 - Não possui 8 anos de estudo, 2 - Possui pelo menos 8 anos de estudo).
- l = 1, 2, 3, 4, 5. (1 - Norte, 2 - Nordeste, 3 - Sudeste, 4 - Sul, 5 - Centro-Oeste).
- m = 1,2. (1 - Feminino, 2 - Masculino).
- $g(\mu_{iklmr}) = \alpha + \beta_i + \delta_k + \lambda_l + \tau_m$. A definir a função de ligação $g(.)$.
- $\beta_1 = \delta_1 = \lambda_1 = \tau_1 = 0$

A definição da função de ligação $g(.)$ será feita a partir da comparação dos AIC e BIC das três funções mais utilizadas para a Binomial Negativa, segundo (PAULA, 2013): logarítmica, identidade e raiz quadrada.

Tabela 36: 3º Modelo: Comparação das Medidas AIC e BIC para as diferentes funções de ligação.

Função de Ligação	AIC	BIC	Função Desvio
log	482.18	497.38	42.11
raiz quadrada	-	-	-
identidade	-	-	-

O algoritmo do modelo não convergiu para as funções de ligação raiz quadrada e identidade. Assim, $g(.) = log(.)$ será a função de ligação do modelo 3.

Modelo 3: Variável resposta Binomial Negativa, função de ligação logarítmica e variáveis explicativas Raça/Cor, ENSFUND, Região e Sexo.

Tabela 37: Resultados dos Parâmetros, Intervalos de Confiança e Testes de Hipóteses para o Modelo 3.

Parâmetro	Estimativa	EP	IC(95%)	Valor Z	p-valor
α (Intercepto)	2.72	0.29	(2.16 ; 3.29)	9.51	< 0.0001
β_2 (Raça/Cor - Negra)	1.31	0.20	(0.92 ; 1.70)	6.59	< 0.0001
δ_2 (Anos de estudo ≥ 8)	-0.19	0.20	(-0.58 ; 0.20)	-0.97	0.33
λ_2 (Região - Nordeste)	1.26	0.31	(0.65 ; 1.88)	4.01	< 0.0001
λ_3 (Região - Sudeste)	0.62	0.32	(0.00 ; 1.24)	1.96	0.05
λ_4 (Região - Sul)	0.57	0.32	(-0.05 ; 1.19)	1.79	0.07
λ_5 (Região - Centro-Oeste)	-0.28	0.32	(-0.91 ; 0.34)	-0.89	0.38
τ_2 (Sexo - Masculino)	2.53	0.20	(2.14 ; 2.93)	12.70	< 0.0001
ϕ (Dispersão)	2.63	0.58	(1.49 ; 3.76)	-	-

O valor do desvio, considerando adequada a aproximação pela $\chi^2_{(32)}$, $D(\mathbf{y}, \tilde{\boldsymbol{\mu}}) = 42.11$ ($p - valor \approx 0.1088$) indica que o modelo se ajusta bem aos dados.

Os parâmetros δ_2 (Possuir pelo menos 8 anos de estudo) e λ_5 (Região Centro Oeste) tiveram os respectivos p-valores dos testes distantes do valor crítico definido (0.05). Um teste de Wald foi implementado para avaliar a possibilidade de retirada desses parâmetros do modelo:

Tabela 38: Teste de Wald para o Modelo 3.

H_0:	$\delta_2 = \lambda_5 = 0$
H_1:	$\delta_2 \neq 0$ ou $\lambda_5 \neq 0$
q_t:	1.73
p-valor:	0.4216

O p-valor de 0.4216, do teste de Wald, indica que não existe evidência estatística que os parâmetros testados são diferentes de zero. Então serão ambos retirados do modelo.

Para a variável explicativa Região, não houve diferença significativa entre o valor de referência λ_1 (Norte) e λ_5 (Centro Oeste). Assim, essas duas regiões serão agrupadas ($\lambda_1 = \lambda_5 = 0$).

Modelo 3.1: Variável resposta Binomial Negativa, função de ligação logarítmica e variáveis explicativas Raça/Cor, Região ($\lambda_5 = 0$) e Sexo.

Tabela 39: Resultados dos Parâmetros, Intervalos de Confiança e Testes de Hipóteses para o Modelo 3.1.

Parâmetro	Estimativa	EP	IC(95%)	Valor Z	p-valor
α (Intercepto)	2.46	0.22	(2.02 ; 2.90)	11.03	< 0.0001
β_2 (Raça/Cor - Negra)	1.36	0.20	(0.96 ; 1.76)	6.68	< 0.0001
λ_2 (Região - Nordeste)	1.42	0.28	(0.88 ; 1.97)	5.12	< 0.0001
λ_3 (Região - Sudeste)	0.75	0.28	(0.20 ; 1.30)	2.69	0.01
λ_4 (Região - Sul)	0.71	0.28	(0.16 ; 1.26)	2.55	0.01
τ_2 (Sexo - Masculino)	2.55	0.20	(2.15 ; 2.95)	12.53	< 0.0001
ϕ (Dispersão)	2.51	0.55	(1.44 ; 3.59)	-	-

Para todos os parâmetros estimados, os respectivos p-valores deram abaixo de 0.01. Como são todos significamente diferentes de zero, entrarão no modelo. Inclusive o λ_4 (Região Sul) que, no modelo 3, teve seu p-valor (0.07) acima, mas próximo, do valor crítico estabelecido. Ao retirar os demais parâmetros, ele se mostrou significativo.

Para o modelo 3.1, o resultado do desvio $D(\mathbf{y}, \tilde{\boldsymbol{\mu}}) = 42.03$ $(p{-}valor \approx 0.1621)$, $\chi^2_{(34)}$ aponta para um bom ajuste aos dados.

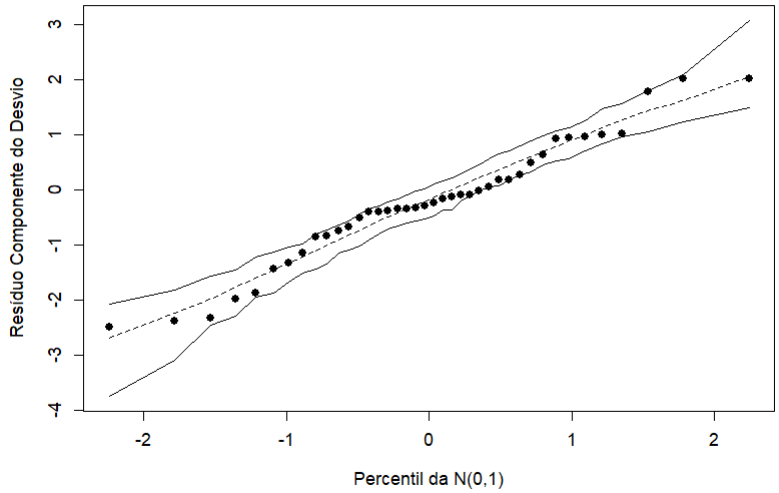


Figura 19: Gráficos de Envelope - Modelo 3.1

O fato de praticamente todos os pontos do gráfico de envelope, do modelo 3.1, estarem dentro das bandas de confiança, é mais uma evidência de bom ajuste.

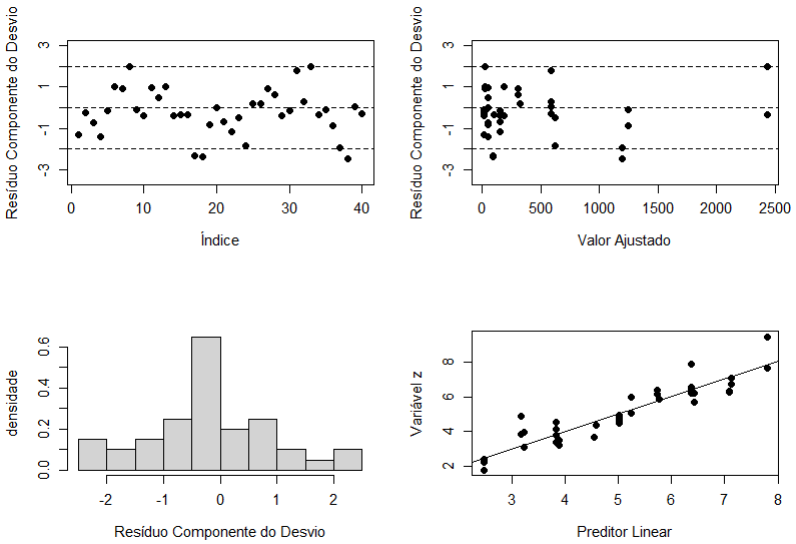


Figura 20: Gráficos de Diagnóstico - Modelo 3.1

A maioria dos resíduos componentes do desvio estão dentro dos limites -2 e 2 da normal. A variabilidade que não era explicada pelos modelos de Poisson, está contida no modelo Binomial Negativo. O problema de superdispersão foi corrigido.

A densidade dos resíduos é aproximadamente simétrica e a variável Z parece seguir um comportamento linear, o que era esperado para um modelo bem ajustado.

4.3.3 Modelo Binomial Negativo com Interações

A partir dos gráficos de perfis, percebe-se a possibilidade de presença de interações entre as variáveis explicativas. O modelo que as inclui define se são, ou não, significativas.

Modelo 4: Variável resposta Binomial Negativa, função de ligação logarítmica, variáveis explicativas Raça/Cor, Região, ENSFUND e Sexo. Interações dois a dois entre todos os níveis de Raça/Cor e Região, Raça/Cor e Sexo, Raça/Cor e ENSFUND, Região e ENSFUND, Região e Sexo, Sexo e ENSFUND.

- $Y_{iklmr} \sim BinNeg(\mu_{iklmr}, \phi)$;
- $log(\mu_{iklmr}) = \alpha + \beta_i + \lambda_l + \tau_m + \delta_k + (\beta\lambda)_{il} + (\beta\tau)_{im} + (\beta\delta)_{ik} + (\lambda\tau)_{lm} + (\lambda\delta)_{lk} + (\tau\delta)_{mk}$
- $\beta_1 = \delta_1 = \lambda_1 = \tau_1 = (\beta\delta)_{1k} = (\beta\delta)_{i1} = (\beta\lambda)_{1l} = (\beta\lambda)_{i1} = (\beta\tau)_{1m} = (\beta\tau)_{i1} = (\lambda\delta)_{1k} = (\lambda\delta)_{l1} = (\lambda\tau)_{1m} = (\lambda\tau)_{l1} = (\tau\delta)_{1k} = (\tau\delta)_{m1} = 0 \ , \ \forall \ i, j, k, l, m$
- i = 1,2. (Raça/Cor: 1 - Branca, 2 - Negra);

- l = 1, 2, 3, 4, 5. (1 - Norte, 2 - Nordeste, 3 - Sudeste, 4 - Sul, 5 - Centro-Oeste);
- m = 1,2. (1 - Feminino, 2 - Masculino);
- k = 1, 2. (1 - Não possui 8 anos de estudo, 2 - Possui pelo menos 8 anos de estudo).

Tabela 40: Resultados dos Parâmetros, Intervalos de Confiança e Testes de Hipóteses para o Modelo 4.

Parâmetro	Estimativa	EP	IC(95%)	Valor Z	p-valor
α (Intercepto)	1.92	0.14	(1.66 ; 2.19)	14.15	< 0.0001
β_2 (Raça/Cor - Negra)	2.44	0.11	(2.22 ; 2.66)	21.53	< 0.0001
λ_2 (Região - Nordeste)	1.39	0.12	(1.15 ; 1.64)	11.26	< 0.0001
λ_3 (Região - Sudeste)	1.47	0.13	(1.20 ; 1.73)	10.90	< 0.0001
λ_4 (Região - Sul)	1.78	0.15	(1.49 ; 2.08)	11.75	< 0.0001
λ_5 (Região - Centro-Oeste)	0.08	0.16	(-0.24 ; 0.41)	0.51	0.61
τ_2 (Sexo - Masculino)	2.68	0.12	(2.45 ; 2.91)	22.54	< 0.0001
δ_2 (Anos de estudo ≥ 8)	0.08	0.08	(-0.08 ; 0.25)	0.97	0.33
$(\beta\lambda)_{22}$ (Negra $\times$ Nordeste)	0.01	0.09	(-0.17 ; 0.18)	0.06	0.95
$(\beta\lambda)_{23}$ (Negra $\times$ Sudeste)	-1.52	0.09	(-1.69 ; -1.34)	-17.20	< 0.0001
$(\beta\lambda)_{24}$ (Negra $\times$ Sul)	-3.20	0.09	(-3.39 ; -3.02)	-33.97	< 0.0001
$(\beta\lambda)_{25}$ (Negra $\times$ Centro-Oeste)	-1.04	0.10	(-1.24 ; -0.84)	-10.10	< 0.0001
$(\beta\tau)_{22}$ (Negra $\times$ Masculino)	0.24	0.09	(0.07 ; 0.41)	2.78	0.01
$(\beta\delta)_{22}$ (Negra $\times$ Estudo ≥ 8)	-0.34	0.05	(-0.43 ; -0.25)	-7.21	< 0.0001
$(\lambda\tau)_{22}$ (Nordeste $\times$ Masculino)	0.09	0.10	(-0.10 ; 0.28)	0.96	0.34
$(\lambda\tau)_{32}$ (Sudeste $\times$ Masculino)	-0.16	0.11	(-0.38 ; 0.06)	-1.43	0.15
$(\lambda\tau)_{42}$ (Sul $\times$ Masculino)	-0.17	0.13	(-0.43 ; 0.10)	-1.25	0.21
$(\lambda\tau)_{52}$ (Centro-Oeste $\times$ Masculino)	0.08	0.14	(-0.19 ; 0.35)	0.61	0.54
$(\lambda\delta)_{22}$ (Nordeste $\times$ Estudo ≥ 8)	-0.44	0.05	(-0.54 ; -0.35)	-9.06	< 0.0001
$(\lambda\delta)_{32}$ (Sudeste $\times$ Estudo ≥ 8)	0.29	0.06	(0.18 ; 0.41)	5.03	< 0.0001
$(\lambda\delta)_{42}$ (Sul $\times$ Estudo ≥ 8)	0.26	0.07	(0.12 ; 0.40)	3.59	< 0.0001
$(\lambda\delta)_{52}$ (Centro-Oeste $\times$ Estudo ≥ 8)	0.53	0.07	(0.40 ; 0.66)	7.77	< 0.0001
$(\tau\delta)_{22}$ (Masculino $\times$ Estudo ≥ 8)	-0.48	0.06	(-0.60 ; -0.35)	-7.46	< 0.0001

Para o modelo 4, o resultado do desvio $D(\mathbf{y}, \tilde{\boldsymbol{\mu}}) = 19.17$ ($p-valor \approx 0.3189$), $\chi^2_{(17)}$ não rejeita a hipótese nula de bom ajuste do modelo.

Os parâmetros $\lambda_5, \delta_2, (\beta\lambda)_{22}, (\lambda\tau)_{22}, (\lambda\tau)_{32}, (\lambda\tau)_{42}$ e $(\lambda\tau)_{52}$ não são significativamente diferentes de zero, e tiveram seus respectivos p-valores distantes do valor de significância definido.

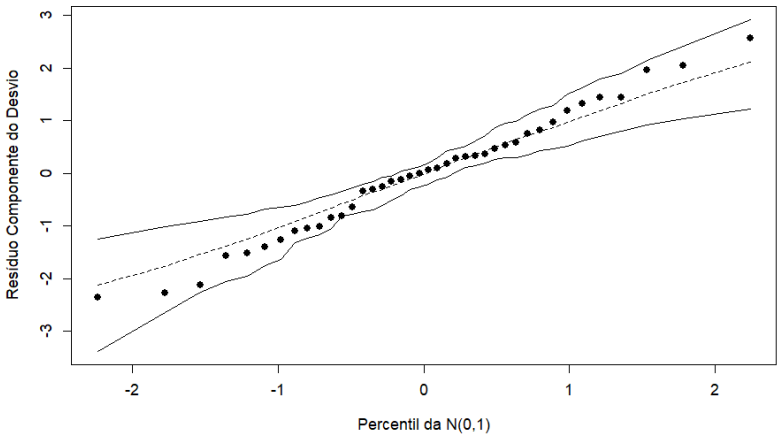


Figura 21: Gráficos de Envelope - Modelo 4

O gráfico de envelope do modelo 4 indica bom ajuste, pois mostra todos os pontos dentro das bandas de confiança.

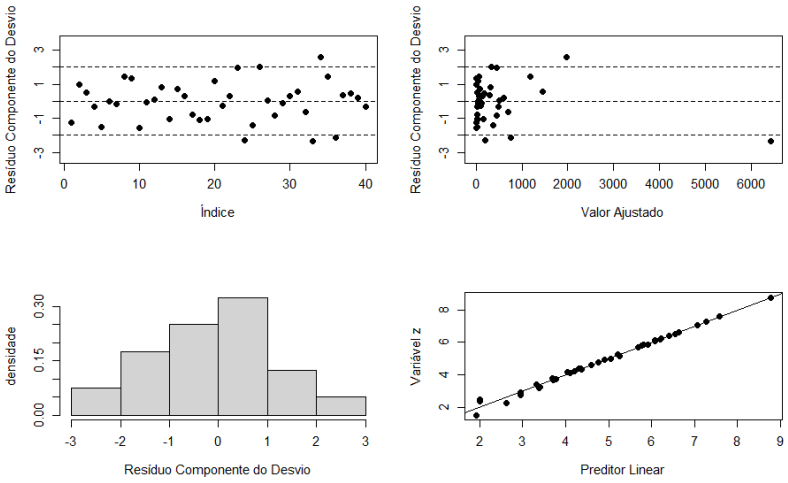


Figura 22: Gráficos de Diagnóstico - Modelo 4

Os gráficos de diagnóstico também apontam para um bom ajuste, com poucos pontos fora dos limites -2 e 2 da normal, para os resíduos componentes do desvio, e a relação entre o preditor linear e a variável z tem um comportamento muito aproximado ao linear.

Apesar de algumas evidências de um bom ajuste para o modelo 4, o parâmetro de dispersão ϕ foi estimado em 11468256232 e seu desvio padrão 408789110203. Segundo (JøRGENSEN, 1996), uma distribuição Binomial Negativa, quando $\phi \to \infty$, converge assintoticamente para uma distribuição Poisson.

4.3.4 Modelo Poisson com Interações

O valor muito alto do parâmetro de dispersão do modelo 4 leva a conclusão de que, com as interações entre as variáveis explicativas, os dados devem se ajustar melhor se considerada a variável resposta com distribuição de Poisson.

Modelo 4.1: Variável resposta Poisson, função de ligação logarítmica, variáveis explicativas Raça/Cor, Região, Sexo e ENSFUND. Interações dois a dois entre todos os níveis de Raça/Cor e Região, Raça/Cor e Sexo, Raça/Cor e ENSFUND, Região e ENSFUND, Região e Sexo, Sexo e ENSFUND.

Tabela 41: Resultados dos Parâmetros, Intervalos de Confiança e Testes de Hipóteses para o Modelo 4.1.

Parâmetro	Estimativa	EP	IC(95%)	Valor Z	p-valor
α (Intercepto)	1.92	0.14	(1.66 ; 2.19)	14.15	< 0.0001
β_2 (Raça/Cor - Negra)	2.44	0.11	(2.22 ; 2.66)	21.53	< 0.0001
λ_2 (Região - Nordeste)	1.39	0.12	(1.15 ; 1.64)	11.26	< 0.0001
λ_3 (Região - Sudeste)	1.47	0.13	(1.20 ; 1.73)	10.90	< 0.0001
λ_4 (Região - Sul)	1.78	0.15	(1.49 ; 2.08)	11.75	< 0.0001
λ_5 (Região - Centro-Oeste)	0.08	0.16	(-0.24 ; 0.41)	0.51	0.61
τ_2 (Sexo - Masculino)	2.68	0.12	(2.45 ; 2.91)	22.54	< 0.0001
δ_2 (Anos de estudo ≥ 8)	0.08	0.08	(-0.08 ; 0.25)	0.97	0.33
$(\beta\lambda)_{22}$ (Negra $\times$ Nordeste)	0.01	0.09	(-0.17 ; 0.18)	0.06	0.95
$(\beta\lambda)_{23}$ (Negra $\times$ Sudeste)	-1.52	0.09	(-1.69 ; -1.34)	-17.20	< 0.0001
$(\beta\lambda)_{24}$ (Negra $\times$ Sul)	-3.20	0.09	(-3.39 ; -3.02)	-33.97	< 0.0001
$(\beta\lambda)_{25}$ (Negra $\times$ Centro-Oeste)	-1.04	0.10	(-1.24 ; -0.84)	-10.10	< 0.0001
$(\beta\tau)_{22}$ (Negra $\times$ Masculino)	0.24	0.09	(0.07 ; 0.41)	2.78	0.01
$(\beta\delta)_{22}$ (Negra $\times$ Estudo$\geq$ 8)	-0.34	0.05	(-0.43 ; -0.25)	-7.21	< 0.0001
$(\lambda\tau)_{22}$ (Nordeste $\times$ Masculino)	0.09	0.10	(-0.10 ; 0.28)	0.96	0.34
$(\lambda\tau)_{32}$ (Sudeste $\times$ Masculino)	-0.16	0.11	(-0.38 ; 0.06)	-1.43	0.15
$(\lambda\tau)_{42}$ (Sul $\times$ Masculino)	-0.17	0.13	(-0.43 ; 0.10)	-1.25	0.21
$(\lambda\tau)_{52}$ (Centro-Oeste $\times$ Masculino)	0.08	0.14	(-0.19 ; 0.35)	0.61	0.54
$(\lambda\delta)_{22}$ (Nordeste $\times$ Estudo$\geq$ 8)	-0.44	0.05	(-0.54 ; -0.35)	-9.06	< 0.0001
$(\lambda\delta)_{32}$ (Sudeste $\times$ Estudo$\geq$ 8)	0.29	0.06	(0.18 ; 0.41)	5.03	< 0.0001
$(\lambda\delta)_{42}$ (Sul $\times$ Estudo$\geq$ 8)	0.26	0.07	(0.12 ; 0.40)	3.59	< 0.0001
$(\lambda\delta)_{52}$ (Centro-Oeste $\times$ Estudo$\geq$ 8)	0.53	0.07	(0.40 ; 0.66)	7.77	< 0.0001
$(\tau\delta)_{22}$ (Masculino $\times$ Estudo$\geq$ 8)	-0.48	0.06	(-0.60 ; -0.35)	-7.46	< 0.0001

O resultado do desvio do modelo 4.1 $D(\mathbf{y}, \tilde{\boldsymbol{\mu}}) = 19.17$ $(p - value \approx 0.3189), \chi^2_{(17)}$ não rejeita a hipótese nula de bom ajuste. O valor do AIC foi 329.92 e BIC 368.77.

Os valores estimados dos parâmetros foram os mesmos do modelo 4, consequencia da convergência assintótica para a distribuição de Poisson, quando a Binomial negativa tem um valor muito alto para seu parâmetro de dispersão. Os parâmetros λ_5, δ_2, $(\beta\lambda)_{22}$, $(\lambda\tau)_{22}$, $(\lambda\tau)_{32}$, $(\lambda\tau)_{42}$ e $(\lambda\tau)_{52}$ não tiveram seus respectivos p-valores significativamente diferentes de zero. Porém, os p-valores dos parâmetros $(\lambda\tau)_{32}$ e $(\lambda\tau)_{42}$ respectivamente, 0.15 e 0.21 não são tão distantes do valor crítico quanto os demais. Um teste de Wald foi implementado, para avaliar a possibilidade da retirada dos parâmetros e interações do modelo.

Tabela 42: Teste de Wald para o Modelo 4.1.

H_0:	$\lambda_5 = \delta_2 = (\beta\lambda)_{22} = (\lambda\tau)_{22} = (\lambda\tau)_{52} = 0$
H_1:	$\lambda_5 \neq 0$ ou $\delta_2 \neq 0$ ou $(\beta\lambda)_{22} \neq 0$ ou $(\lambda\tau)_{22} \neq 0$ ou $(\lambda\tau)_{52} \neq 0$
q_t:	6.3
p-valor:	0.2778

O p-valor de 0.2778, do teste de Wald, indica que os parâmetros e interações testados não são significativamente diferentes de zero. Então, serão retirados do modelo, e os resultados comparados.

Modelo 4.2 (Modelo Final): Variável resposta Poisson, função de ligação logarítmica, variáveis explicativas Raça/Cor, Região (Sem Região Centro-Oeste $\lambda_5 = 0$), Sexo e ENSFUND. Interações entre Raça/Cor e Regiões Sudeste, Sul e Centro-Oeste, Raça/Cor e Sexo, Raça/Cor e ENSFUND, Sexo e Regiões Sudeste e Sul, Região (todas) e ENSFUND.

Tabela 43: Resultados dos Parâmetros, Intervalos de Confiança e Testes de Hipóteses para o Modelo 4.2.

Parâmetro	Estimativa	EP	IC(95%)	Valor Z	p-valor
α (Intercepto)	1.92	0.09	(1.75 ; 2.09)	22.43	< 0.0001
β_2 (Raça/Cor - Negra)	2.40	0.09	(2.23 ; 2.57)	27.61	< 0.0001
λ_2 (Região - Nordeste)	1.46	0.03	(1.41 ; 1.51)	57.94	< 0.0001
λ_3 (Região - Sudeste)	1.49	0.09	(1.31 ; 1.67)	16.09	< 0.0001
λ_4 (Região - Sul)	1.81	0.12	(1.58 ; 2.03)	15.61	< 0.0001
τ_2 (Sexo - Masculino)	2.73	0.09	(2.56 ; 2.90)	31.51	< 0.0001
$(\beta\lambda)_{23}$ (Negra $\times$ Sudeste)	-1.49	0.05	(-1.60 ; -1.39)	-28.06	< 0.0001
$(\beta\lambda)_{24}$ (Negra $\times$ Sul)	-3.18	0.06	(-3.30 ; -3.06)	-50.61	< 0.0001
$(\beta\lambda)_{25}$ (Negra $\times$ Centro-Oeste)	-0.91	0.04	(-1.00 ; -0.83)	-21.55	0.0042
$(\beta\tau)_{22}$ (Negra $\times$ Masculino)	0.24	0.09	(0.08 ; 0.41)	2.86	< 0.0001
$(\beta\delta)_{22}$ (Negra $\times$ Estudo ≥ 8)	-0.33	0.04	(-0.41 ; -0.25)	-8.06	< 0.0001
$(\lambda\tau)_{32}$ (Sudeste $\times$ Masculino)	-0.24	0.08	(-0.40 ; -0.08)	-2.89	0.0038
$(\lambda\tau)_{42}$ (Sul $\times$ Masculino)	-0.24	0.11	(-0.46 ; -0.03)	-2.20	0.03
$(\lambda\delta)_{22}$ (Nordeste $\times$ Estudo ≥ 8)	-0.41	0.04	(-0.50 ; -0.33)	-9.47	< 0.0001
$(\lambda\delta)_{32}$ (Sudeste $\times$ Estudo ≥ 8)	0.33	0.05	(0.23 ; 0.43)	6.46	< 0.0001
$(\lambda\delta)_{42}$ (Sul $\times$ Estudo ≥ 8)	0.30	0.06	(0.18 ; 0.43)	4.87	< 0.0001
$(\lambda\delta)_{52}$ (C. Oeste $\times$ Estudo ≥ 8)	0.60	0.06	(0.49 ; 0.72)	10.28	< 0.0001
$(\tau\delta)_{22}$ (Masculino $\times$ Estudo ≥ 8)	-0.44	0.05	(-0.53 ; -0.35)	-9.61	< 0.0001

O desvio do modelo 4.2 $D(\mathbf{y}, \tilde{\boldsymbol{\mu}}) = 25.38$ ($p - valor \approx 0.2791$), que considera como apropriada a aproximação pela $\chi^2_{(22)}$, indica que o modelo se ajusta bem aos dados. Os valores AIC de 326.14 e BIC de 356.54, tiveram uma redução, quando as variáveis explicativas e interações não significativas do modelo 4.1 foram retiradas.

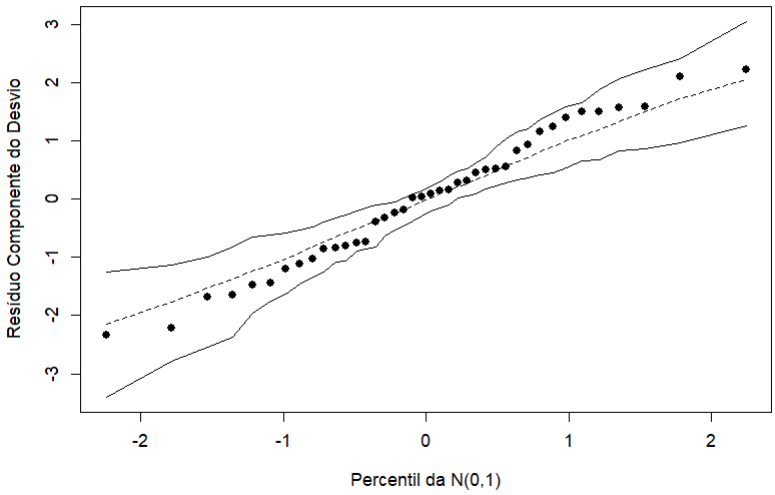


Figura 23: Gráficos de Envelope - Modelo 4.2

O gráfico de envelope do modelo 4.2 (Figura 23) mostra todos os pontos dentro das bandas de confiança, o que indica bom ajuste. Os pontos parecem estar ainda mais centralizados do que no caso do modelo 3.1 (Figura 19).

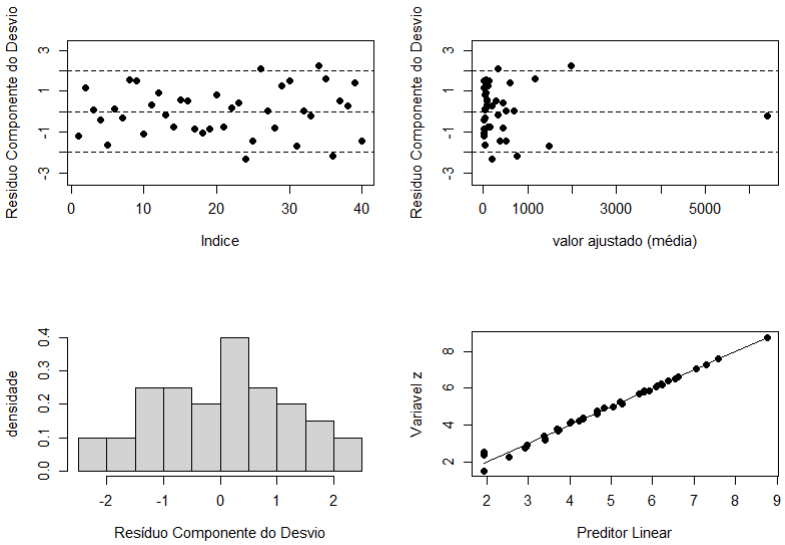


Figura 24: Gráficos de Diagnóstico - Modelo 4.2

Na Figura 24 os gráficos dos resíduos componentes do desvio, pelo índice e pelo valor ajustado, mostram apenas quatro pontos fora dos limites -2 e 2 da distribuição normal. A variável z tem uma relação aproximadamente linear com o preditor linear. Todos os gráficos de diagnóstico seguem o padrão esperado para um modelo bem ajustado.

No gráfico do valor ajustado, pelo resíduo componente do desvio, nota-se um

ponto com o valor ajustado muito acima dos demais, e com o resíduo muito próximo do zero. Se trata do grupo dos indivíduos do sexo masculino, Raça/Cor negra, da região Nordeste e com menos de 8 anos de ensino. Esse grupo teve como valor ajustado 6395.37 e 6391 como observado.

Tabela 44: Quadro Resumo dos Modelos Ajustados.

Modelo	**Faixa Etária**	**Distribuição da Resposta**	**Função de Ligação**	**AIC**	**BIC**	**Variáveis Explicativas**	**Interações**
Modelo 1	Todas	Poisson	Logarítmica	10961.94	10734.35	Raça/Cor, Faixa Etária, ENSFUND, Região e Sexo	-
Modelo 2	15 a 29 anos	Poisson	Inversa	2711.84	2725.35	Raça/Cor, ENSFUND, Região e Sexo	-
Modelo 2.1	15 a 29 anos	Poisson	Logarítmica	4193.75	4207.26	Raça/Cor, ENSFUND, Região e Sexo	-
Modelo 3	15 a 29 anos	Binomial Negativa	Logarítmica	482.18	497.38	Raça/Cor, ENSFUND, Região e Sexo	-
Modelo 3.1	15 a 29 anos	Binomial Negativa	Logarítmica	479.88	491.71	Raça/Cor, Região ($\lambda_5 = 0$) e Sexo	-
Modelo 4	15 a 29 anos	Binomial Negativa	Logarítmica	331.92	372.46	Raça/Cor, ENSFUND, Região e Sexo	Raça/Cor × Região (Todas) Raça/Cor × Sexo Raça/Cor × ENSFUND Região (Todas) × ENSFUND Região (Todas) × Sexo Sexo × ENSFUND
Modelo 4.1	15 a 29 anos	Poisson	Logarítmica	329.92	368.77	Raça/Cor, ENSFUND, Região e Sexo	Raça/Cor × Região (Todas) Raça/Cor × Sexo Raça/Cor × ENSFUND Região (Todas) × ENSFUND Região (Todas) × Sexo Sexo × ENSFUND
Modelo 4.2	15 a 29 anos	Poisson	Logarítmica	329.92	368.77	Raça/Cor, ENSFUND, Região ($\lambda_5 = 0$) e Sexo	Raça/Cor × Sudeste, Sul e Centro-Oeste Raça/Cor × Sexo Raça/Cor × ENSFUND Região (Todas) × ENSFUND Sudeste, Sul × Sexo

5 Valores Preditos

A partir do modelo final (Modelo 4.2), os valores preditos serão analisados.

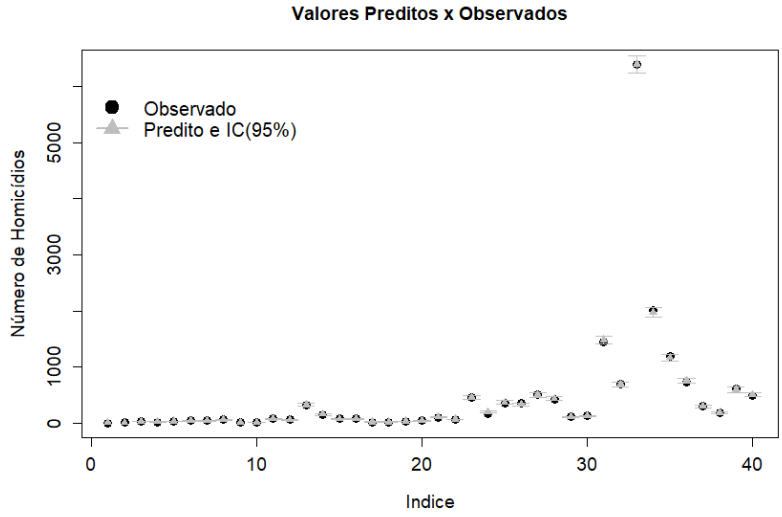


Figura 25: Valores Preditos, IC(95%) e Observados

O gráfico da Figura 25 mostra que todos os valores observados ficaram dentro, ou próximo dos intervalos de confiança preditos. Novamente o ponto referente ao sexo masculino, Raça/Cor negra, região Nordeste e menos de 8 anos de ensino é discrepante.

Apesar disso, seu valor predito também está dentro do intervalo.

Tabela 45: Valores Preditos, Desvio Padrão, IC(95%) e Observados.

Sexo	Raça/Cor	Região	Anos de Estudo	Observado	Predito	D.P.	IC(95%)
Feminino	Branca	Norte	Não	4	6.84	0.59	(5.69 ; 7.99)
Feminino	Branca	Norte	Sim	10	6.84	0.59	(5.69 ; 7.99)
Feminino	Branca	Nordeste	Não	30	29.55	2.57	(24.52 ; 34.58)
Feminino	Branca	Nordeste	Sim	18	19.60	1.73	(16.21 ; 22.98)
Feminino	Branca	Sudeste	Não	23	30.42	2.87	(24.79 ; 36.05)
Feminino	Branca	Sudeste	Sim	43	42.29	3.89	(34.66 ; 49.92)
Feminino	Branca	Sul	Não	40	41.67	4.03	(33.77 ; 49.56)
Feminino	Branca	Sul	Sim	65	56.44	5.28	(46.09 ; 66.78)
Feminino	Branca	Centro-Oeste	Não	11	6.84	0.59	(5.69 ; 7.99)
Feminino	Branca	Centro-Oeste	Sim	9	12.52	1.23	(10.12 ; 14.92)
Feminino	Negra	Norte	Não	78	75.52	3.30	(69.06 ; 81.98)
Feminino	Negra	Norte	Sim	61	54.49	2.67	(49.26 ; 59.72)
Feminino	Negra	Nordeste	Não	324	326.27	13.59	(299.64 ; 352.9)
Feminino	Negra	Nordeste	Sim	149	156.16	7.84	(140.8 ; 171.53)
Feminino	Negra	Sudeste	Não	79	75.52	5.97	(63.81 ; 87.23)
Feminino	Negra	Sudeste	Sim	79	75.77	6.00	(64.01 ; 87.52)
Feminino	Negra	Sul	Não	16	19.17	2.24	(14.78 ; 23.55)
Feminino	Negra	Sul	Sim	15	18.73	2.20	(14.43 ; 23.03)
Feminino	Negra	Centro-Oeste	Não	26	30.32	1.67	(27.04 ; 33.59)
Feminino	Negra	Centro-Oeste	Sim	45	40.06	2.43	(35.3 ; 44.81)
Masculino	Branca	Norte	Não	98	105.04	3.89	(97.42 ; 112.66)
Masculino	Branca	Norte	Sim	69	67.74	3.13	(61.6 ; 73.88)
Masculino	Branca	Nordeste	Não	460	453.81	16.11	(422.24 ; 485.38)
Masculino	Branca	Nordeste	Sim	169	194.14	8.55	(177.38 ; 210.9)
Masculino	Branca	Sudeste	Não	354	369.22	16.15	(337.56 ; 400.88)
Masculino	Branca	Sudeste	Sim	353	331.07	14.97	(301.72 ; 360.41)
Masculino	Branca	Sul	Não	502	501.68	20.47	(461.55 ; 541.81)
Masculino	Branca	Sul	Sim	431	438.22	18.89	(401.19 ; 475.25)
Masculino	Branca	Centro-Oeste	Não	117	105.04	3.89	(97.42 ; 112.66)
Masculino	Branca	Centro-Oeste	Sim	137	124.05	7.24	(109.86 ; 138.23)
Masculino	Negra	Norte	Não	1451	1480.26	34.33	(1412.97 ; 1547.55)
Masculino	Negra	Norte	Sim	689	688.83	23.21	(643.33 ; 734.32)
Masculino	Negra	Nordeste	Não	6391	6395.37	77.71	(6243.06 ; 6547.67)
Masculino	Negra	Nordeste	Sim	2008	1974.10	41.73	(1892.31 ; 2055.89)
Masculino	Negra	Sudeste	Não	1189	1169.84	32.01	(1107.1 ; 1232.58)
Masculino	Negra	Sudeste	Sim	731	756.88	24.86	(708.14 ; 805.61)
Masculino	Negra	Sul	Não	299	294.49	14.67	(265.74 ; 323.24)
Masculino	Negra	Sul	Sim	188	185.61	10.53	(164.97 ; 206.26)
Masculino	Negra	Centro-Oeste	Não	608	594.25	22.32	(550.51 ; 638)
Masculino	Negra	Centro-Oeste	Sim	492	506.37	20.16	(466.87 ; 545.88)

As informações da Tabela 45 mostram que, das 40 possíveis combinações de níveis das variáveis respostas, 30 (75%) tiveram o valor observado dentro do intervalo

de confiança estimado. As 10 restantes não tiveram o valor observado muito distante do intervalo, como mostra a Figura 25. Então, para a maioria das combinações possíveis, o modelo obteve um valor predito próximo do observado, corroborando com os resultados de diagnóstico, e testes de hipóteses avaliados.

6 Conclusão

O modelo 4.2, com resposta de Poisson, ligação logarítmica, e algumas interações se adequou bem aos dados, como mostra a análise de diagnóstico, testes de hipóteses, e os valores preditos. Para modelos que não consideram as interações entre as variáveis explicativas, foi observado o fenômeno de sobredispersão. Ao implementar o modelo Binomial Negativo, esse problema foi contornado. Porém, quando as interações foram consideradas, algumas se mostraram significativas para o modelo, e a estimativa do parâmetro de dispersão ϕ tendeu ao infinito. Quando isso ocorre, a distribuição Binomial Negativa converge assintoticamente para a distribuição de Poisson.

A partir da classificação do IBGE para a variável Raça/Cor, Pretos e Pardos foram agrupados na categoria "Negros", e apenas negros e brancos foram filtrados, uma vez que a soma dessas categorias representa 99% do total de mortes e homicídios no ano. Esse estudo mostrou que os dados de mortalidade, do ano de 2020, apontam para uma maior proporção de negros mortos por homicídio, em relação aos brancos. Uma maior proporção também para os jovens de 15 a 29 anos. A fim de direcionar as análises para buscar consequencias claras da desigualdade racial, os jovens mortos por homicídio, no ano de 2020, foram a população definida para entrar na modelagem.

O modelo implementado (4.2), considerado como modelo final, toma a distribuição de Poisson para a variável resposta, as variáveis explicativas: Raça/Cor, Região, Sexo e Anos de Estudo. As interações presentes no modelo são: Raça/Cor e Regiões Sudeste, Sul e Centro-Oeste, Raça/Cor e Sexo, Raça/Cor e Anos de Estudo, Sexo e Regiões Sudeste e Sul, Região (todas) e Anos de Estudo. Função de Ligação logarítmica. Toma como valores de referência, os seguintes níveis das variáveis explicativas: Raça/Cor Branca, Região Norte ou Centro-Oeste, Sexo Feminino e Menos de 8 anos de Estudo. Os valores estimados dos parâmetros mostram a diferença em relação a essa classificação.

O valor positivo para o parâmetro referente a Raça/Cor "Negra", mostra que, a Raça/Cor tem interferência no número de homicídios de jovens com 15 a 29 anos, sendo maior o número esperado de assassinatos de pessoas negras, em relação as brancas. Como mostra a Tabela 43, a variável Raça/Cor, individualmente, quando é classificada como "Negra"(Pretos e Pardos) tem um incremento multiplicativo de $e^{\beta_2} = 11.04$ na média predita de homicídios, ou seja, um aumento relativo de 1104%. As interações, uma a uma, entre Raça/Cor e as regiões Sudeste, Sul e Centro-oeste têm as estimativas de parâmetros $(\beta\lambda)_{23}, (\beta\lambda)_{24}$ e $(\beta\lambda)_{25}$ negativos. Isso indica que, nessas regiões, o impacto da Raça/Cor nos números de homicídios de jovens é mais ameno do que Norte e Nordeste. O mesmo

ocorre para jovens que tem mais de 8 anos de estudo, uma vez que $(\beta\delta)_{22}$ também é negativo. A interação $(\beta\tau)_{22}$ indica que o impacto da Raça/Cor na contagem de homicídios é maior para indivíduos do sexo masculino. E o sexo masculino, individualmente, também aponta para um impacto positivo, com o incremento multiplicativo de e^{β_2}. Ou seja, jovens negros do sexo masculino têm um incremento multiplicativo de $e^{\beta 2+\tau 2+(\beta\tau)_{22}} = 216.45$, isto é, um aumento de 21645% no valor predito de homicídios de homens negros, em relação ao valor de referência. Esse número pode ser amenizado, se consideradas pessoas que possuem mais de 8 anos de estudo.

Os resultados obtidos na análise descritiva corroboram com os valores estimados para os parâmetros do modelo final. Existe uma concentração de homicídios de negros, em relação a brancos, no grupo de pessoas com menos de 8 anos de estudo, da região Nordeste, e do sexo masculino. E todas essas variáveis, quando analisadas separadamente, em relação a raça/cor, mostraram essas tendências.

A título de comparação, o modelo 2.1, que teve um ajuste inadequado, tem um incremento multiplicativo para negros de, aproximadamente, apenas 5.07, muito menor do que o estimado para o modelo final 11.04. Se considerados homens negros, o incremento multiplicativo do modelo 2.1 é, aproximadamente, de 75.41. Por isso, o maior valor predito por esse modelo é de 5062.77, muito distante do maior valor observado, 6391. O modelo 3.1, mostra bom ajuste, segundo a análise de diagnóstico realizada. Porém, seu valor ajustado para o maior valor observado foi de, aproximadamente, 2432.22. Para o modelo 3.1, homens negros tiveram um aumento de 5019.55% no valor predito de homicídios, comparado ao valor de referência. Esse aumento, se considerado apenas a Raça/Cor negra, foi de apenas 390.29%. A entrada das interações se mostrou muito importante para permitir o modelo de captar o impacto da Raça/Cor nos números de homicídios.

A desigualdade racial mostrou ter impacto significativo nos resultados de números de homicídios. A associação entre pertencer a Raça/Cor dos pretos e pardos e o número de mortes, se mostrou estatísticamente significante.

Referências

ARAÚJO, E. e. a. A utilização da variável raça/cor em saúde pública: possibilidades e limites. *Interface - Comunic., Saude, Educ.*, 2009.

CHOR, D.; LIMA, C. R. d. A. Aspectos epidemiológicos das desigualdades raciais em saúde no brasil. *Caderno de Saúde Pública do Rio de Janeiro*, 2005.

FERREIRA, R. F.; CAMARGO, A. C. As relações cotidianas e a construção da identidade negra. *Psicologia: Ciência e profissão*, 2011.

IBGE. *Pesquisa Nacional por Amostra de Domicílios - PNAD*. 2018. Disponível em: ⟨https://sidra.ibge.gov.br/tabela/6403⟩.

JøRGENSEN, B. *The Theory of Dispersion Models*. Canadá: Chapman and Hall/CRC; 1ª edição, 1996.

OSORIO, R. G. A classificação de cor ou raça do ibge revisitada. *Características Étnico-raciais da População: Classificações e identidades.*, 2013.

PAULA, A. G. *Modelos de Regressão*. São Paulo, Brasil: Instituto de Matemática e Estatística - USP, 2013.

SALDANHA, R. d. F.; BASTOS, R. R.; BARCELLOS, C. *Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS)*. 2019. Disponível em: ⟨http://ref.scielo.org/dhcq3y⟩.

SAÚDE, M. da. *Manual de Procedimentos sobre o Sistema de Informações de Mortalidade*. Brasília, Brasil: Assessoria de Comunicação e Educação em Saúde - Ascom, 2001.

SOARES, S. S. D. *Educação: Um Escudo Contra o Homicídio?* Brasília, Brasil: IPEA, 2007.

WAISELFISZ, J. J. *Mapa da Violência 2012: Os Novos Padrões da Violência Homicida no Brasil*. Brasília, Brasil: Instituto Sangari, 2011.

WAISELFISZ, J. J. *Mapa da Violência 2012: A Cor dos Homicídios no Brasil*. São Paulo, Brasil: CEBELA, FLACSO, 2012.

Apêndice

A Apêndice A - Mortalidade, no período de 1996 a 2019

A título de comparação com o ano de 2020, ano considerado para as análises do estudo, os resultados dos anos anteriores, desde o primeiro ano disponível no SIM, foram analisados.

A.1 Mortalidade de 1996 a 2019 por Faixa Etária e Raça/Cor

Mortalidade, por todas as causas registradas, segundo Faixa Etária e Raça/Cor.

Tabela 46: Mortalidade por faixa etária e raça/cor de 1996 até 2019

Faixa Etária	Amarela	Branca	Ignorada	Indígena	Parda	Preta
0 a 4 anos	7309	505175	378319	16488	459920	34323
5 a 14 anos	1135	88903	44499	2191	96426	12247
15 a 29 anos	6298	583732	231130	6734	807051	132295
30 a 49 anos	16799	1426390	504623	8778	1358141	319767
50 a 69 anos	47075	3634820	947796	11966	2214468	598401
70 ou mais	107823	6689847	1397890	19557	2829485	689981

Tabela 47: Mortalidade por faixa etária e raça/cor binária de 1996 até 2019

Faixa Etária	Branca	Negra
0 a 4 anos	505.175	494.243
5 a 14 anos	88.903	108.673
15 a 29 anos	583.732	939.346
30 a 49 anos	1.426.390	1.677.908
50 a 69 anos	3.634.820	2.812.869
70 ou mais	6.689.847	3.519.466

O número de mortes é maior para pessoas negras para as faixa etárias de 5 a 14 anos, 15 a 29 anos e 30 a 49 anos.

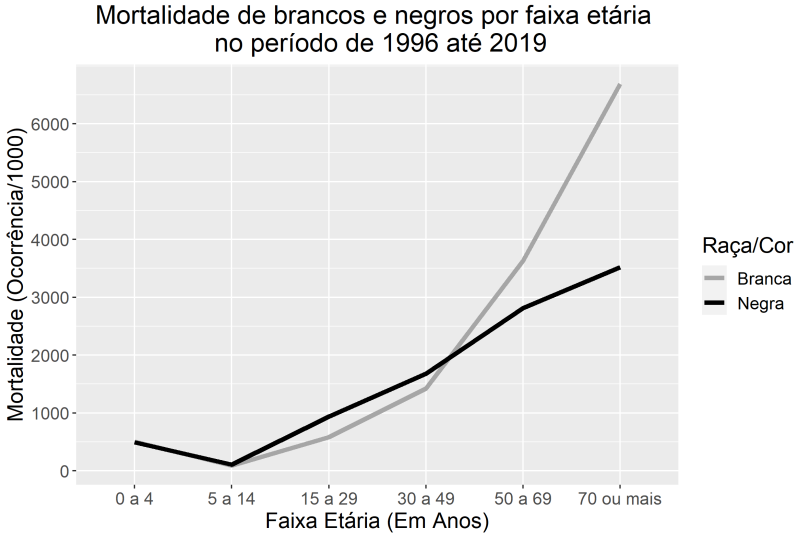


Figura 26: Mortalidade de Brancos e Negros segundo Faixa Etária no período entre 1996 e 2019

A.2 Mortalidade de 1996 a 2019 por Capítulo CID - 10

Os dados relativos de mortalidade, segundo capítulos da CID-10 para o período de 1996 até 2019, são:

Tabela 48: Mortalidade relativa por Capítulo CID-10 e raça/cor no período de 1996 a 2019

Capítulo CID-10	Branca	Preta	Amarela	Parda	Indígena	Ignorado
I	1.94%	0.39%	0.03%	1.43%	0.02%	0.7%
II	8.65%	0.95%	0.12%	3.74%	0.02%	1.53%
III	0.24%	0.04%	0%	0.16%	0%	0.07%
IV	2.74%	0.46%	0.04%	1.69%	0.02%	0.66%
V	0.41%	0.09%	0.01%	0.31%	0%	0.09%
VI	1.34%	0.1%	0.02%	0.44%	0%	0.19%
VII	0%	0%	0%	0%	0%	0%
VIII	0.01%	0%	0%	0%	0%	0%
IX	14.51%	2.12%	0.21%	7.59%	0.04%	3.48%
X	5.98%	0.61%	0.09%	2.6%	0.03%	1.25%
XI	2.45%	0.33%	0.03%	1.46%	0.01%	0.61%
XII	0.15%	0.02%	0%	0.09%	0%	0.03%
XIII	0.2%	0.02%	0%	0.09%	0%	0.04%
XIV	1.2%	0.16%	0.02%	0.57%	0%	0.22%
XV	0.05%	0.02%	0%	0.07%	0%	0.02%
XVI	0.89%	0.05%	0.01%	0.86%	0.02%	0.72%
XVII	0.45%	0.02%	0%	0.28%	0.01%	0.18%
XVIII	3.21%	0.64%	0.06%	2.89%	0.03%	2.23%
XX	4.74%	0.78%	0.05%	5.31%	0.04%	1.47%

Desconsiderando os ignorados, soma de brancos e negros (pardos e pretos), representam a proporção de 85.56% do total de mortes por Capítulo CID-10.

Tabela 49: Mortalidade relativa de brancos e negros por Capítulo CID-10 no período de 1996 a 2019

Capítulo CID-10	Brancos	Negros
I	1.94%	1.82%
II	8.65%	4.69%
III	0.24%	0.20%
IV	2.74%	2.15%
V	0.41%	0.40%
VI	1.34%	0.54%
VII	0%	0%
VIII	0.01%	0%
IX	14.51%	9.71%
X	5.98%	3.21%
XI	2.45%	1.79%
XII	0.15%	0.11%
XIII	0.2%	0.11%
XIV	1.2%	0.73%
XV	0.05%	0.09%
XVI	0.89%	0.91%
XVII	0.45%	0.30%
XVIII	3.21%	3.53%
XX	4.74%	6.09%

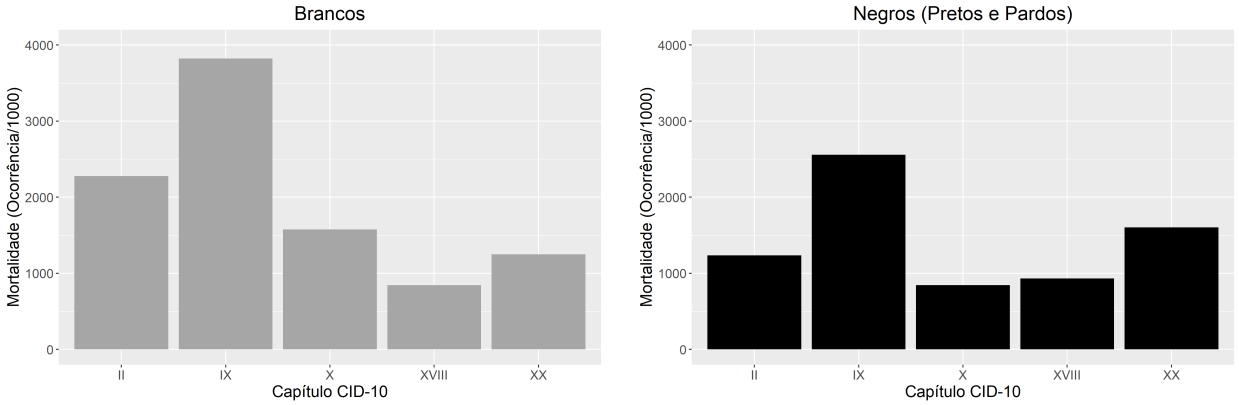
Os 5 capítulos CID-10 com mais registros de mortes são os mesmos para brancos e negros, porém em ordem diferentes: "Neoplasias (tumores)", "Doenças do aparelho circulatório", "Doenças do aparelho respiratório", sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte"e "Causas externas de morbidade e mortalidade". Respectivamente os capítulos II, IX, X, XVIII e XX.

Tabela 50: 5 maiores mortalidades de brancos e negros por Capítulo CID-10 no período de 1996 a 2019

Capítulo CID-10	Brancos
IX	3822895
II	2278781
X	1576284
XX	1248393
XVIII	844879

Capítulo CID-10	Negros
IX	2558322
XX	1604270
II	1237227
XVIII	931556
X	844295

Na Tabela 50, os Capítulos CID-10 estão ordenados por quantidade de mortes. Destaca-se que para as pessoas consideradas negras, o capítulo XX, "Causas externas de morbidade e mortalidade", é o segundo com mais mortes enquanto que entre os brancos é apenas o quarto. Existe também grande diferença no número de mortes registradadas como consequência do capítulo X, "Doenças do aparelho respiratório", que é a terceira causa com mais mortes entre brancos, e apenas a quinta entre negros.



(a) Raça/Cor = Branca

(b) Raça/Cor = Preta ou Parda

Figura 27: Mortalidade da população de brancos e negros por Capítulo CID-10 no período de 1996 até 2019

B Apêndice B - Valores Preditos

Alguns resultados obtidos a partir dos valores preditos pelo modelo final (Modelo 4.2). Contém tabela com o valor predito, segundo os diferentes parâmetros e interações estimados, e gráficos de perfis com valores e intervalosd e confiança preditos, para diferentes variáveis explicativas.

Tabela 51: Valores Preditos Segundo Níveis das Variáveis Explicativas

Sexo	Raça/Cor	Região	Estudo	Valor Predito μ_{iklm}
Feminino	Branca	Norte	< 8	e^{α}
Feminino	Branca	Norte	≥ 8	e^{α}
Feminino	Branca	Nordeste	< 8	$e^{\alpha+\lambda_2}$
Feminino	Branca	Nordeste	≥ 8	$e^{\alpha+\lambda_2+(\lambda\delta)_{22}}$
Feminino	Branca	Sudeste	< 8	$e^{\alpha+\lambda_3}$
Feminino	Branca	Sudeste	≥ 8	$e^{\alpha+\lambda_3+(\lambda\delta)_{32}}$
Feminino	Branca	Sul	< 8	$e^{\alpha+\lambda_4}$
Feminino	Branca	Sul	≥ 8	$e^{\alpha+\lambda_4+(\lambda\delta)_{42}}$
Feminino	Branca	Centro-Oeste	< 8	e^{α}
Feminino	Branca	Centro-Oeste	≥ 8	$e^{\alpha+(\lambda\delta)_{52}}$
Feminino	Negra	Norte	< 8	$e^{\alpha+\beta_2}$
Feminino	Negra	Norte	≥ 8	$e^{\alpha+\beta_2+(\beta\delta)_{22}}$
Feminino	Negra	Nordeste	< 8	$e^{\alpha+\beta_2+\lambda_2}$
Feminino	Negra	Nordeste	≥ 8	$e^{\alpha+\beta_2+\lambda_2+(\lambda\delta)_{22}}$
Feminino	Negra	Sudeste	< 8	$e^{\alpha+\beta_2+\lambda_3+(\beta\lambda)_{23}}$
Feminino	Negra	Sudeste	≥ 8	$e^{\alpha+\beta_2+\lambda_3+(\beta\lambda)_{23}+(\lambda\delta)_{32}}$
Feminino	Negra	Sul	< 8	$e^{\alpha+\beta_2+\lambda_4+(\beta\lambda)_{24}}$
Feminino	Negra	Sul	≥ 8	$e^{\alpha+\beta_2+\lambda_4+(\beta\lambda)_{24}+(\lambda\delta)_{42}}$
Feminino	Negra	Centro-Oeste	< 8	$e^{\alpha+\beta_2+(\beta\lambda)_{25}+(\lambda\delta)_{52}}$
Feminino	Negra	Centro-Oeste	≥ 8	$e^{\alpha+\beta_2+(\beta\lambda)_{25}+(\lambda\delta)_{52}}$
Masculino	Branca	Norte	< 8	$e^{\alpha+\tau_2}$
Masculino	Branca	Norte	≥ 8	$e^{\alpha+\tau_2+(\tau\delta)_{22}}$
Masculino	Branca	Nordeste	< 8	$e^{\alpha+\tau_2+\lambda_2}$
Masculino	Branca	Nordeste	≥ 8	$e^{\alpha+\tau_2+\lambda_2+(\lambda\delta)_{22}+(\tau\delta)_{22}}$
Masculino	Branca	Sudeste	< 8	$e^{\alpha+\tau_2+\lambda_3+(\lambda\tau)_{32}}$
Masculino	Branca	Sudeste	≥ 8	$e^{\alpha+\tau_2+\lambda_3+(\lambda\tau)_{32}+(\lambda\delta)_{32}+(\tau\delta)_{22}}$
Masculino	Branca	Sul	< 8	$e^{\alpha+\tau_2+\lambda_4+(\lambda\tau)_{42}}$
Masculino	Branca	Sul	≥ 8	$e^{\alpha+\tau_2+\lambda_4+(\lambda\tau)_{42}+(\lambda\delta)_{42}+(\tau\delta)_{22}}$
Masculino	Branca	Centro-Oeste	< 8	$e^{\alpha+\tau_2}$
Masculino	Branca	Centro-Oeste	≥ 8	$e^{\alpha+\tau_2+(\lambda\delta)_{52}+(\tau\delta)_{22}}$
Masculino	Negra	Norte	< 8	$e^{\alpha+\beta_2+\tau_2+(\beta\tau)_{22}}$
Masculino	Negra	Norte	≥ 8	$e^{\alpha+\beta_2+\tau_2+(\beta\tau)_{22}+(\beta\delta)_{22}+(\tau\delta)_{22}}$
Masculino	Negra	Nordeste	< 8	$e^{\alpha+\beta_2+\lambda_2+\tau_2+(\beta\tau)_{22}}$
Masculino	Negra	Nordeste	≥ 8	$e^{\alpha+\beta_2+\lambda_2+\tau_2+(\beta\tau)_{22}+(\beta\delta)_{22}+(\lambda\delta)_{22}+(\tau\delta)_{22}}$
Masculino	Negra	Sudeste	< 8	$e^{\alpha+\beta_2+\lambda_3+\tau_2+(\beta\lambda)_{23}+(\beta\tau)_{22}}$
Masculino	Negra	Sudeste	≥ 8	$e^{\alpha+\beta_2+\lambda_3+\tau_2+(\beta\lambda)_{23}+(\beta\tau)_{22}+(\beta\delta)_{22}+(\lambda\delta)_{32}+(\tau\delta)_{22}}$
Masculino	Negra	Sul	< 8	$e^{\alpha+\beta_2+\lambda_4+\tau_2+(\beta\lambda)_{24}+(\beta\tau)_{22}}$
Masculino	Negra	Sul	≥ 8	$e^{\alpha+\beta_2+\lambda_4+\tau_2+(\beta\lambda)_{24}+(\beta\tau)_{22}+(\beta\delta)_{22}+(\lambda\delta)_{42}+(\tau\delta)_{22}}$
Masculino	Negra	Centro-Oeste	< 8	$e^{\alpha+\beta_2+\tau_2+(\beta\lambda)_{25}+(\beta\tau)_{22}}$
Masculino	Negra	Centro-Oeste	≥ 8	$e^{\alpha+\beta_2+\tau_2+(\beta\lambda)_{25}+(\beta\tau)_{22}+(\beta\delta)_{22}+(\lambda\delta)_{52}+(\tau\delta)_{22}}$

B.1 Raça/Cor e Sexo

Tabela 52: Análise Descritiva dos Valores Preditos e Comparação com Valores Observados. Raça/Cor x Sexo.

Raça/Cor	Sexo	Predito	Observado	D.P.	Variância	C.V.
Branca	Feminino	253.00	253	17.60	309.84	69.57
Branca	Masculino	2690.00	2690	167.24	27968.31	62.17
Negra	Feminino	872.00	872	93.19	8684.29	106.87
Negra	Masculino	14046.00	14046	1838.07	3378508.17	130.86

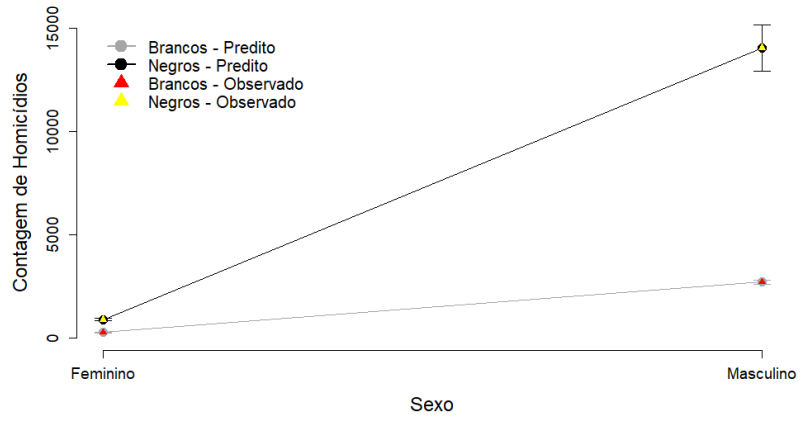


Figura 28: Valores Preditos Segundo Raça/Cor e Sexo.

B.2 Raça/Cor x Região

Tabela 53: Análise Descritiva dos Valores Preditos e Comparação com Valores Observados. Raça/Cor x Região.

Raça/Cor	Região	Predito	Observado	D.P.	Variância	C.V.
Branca	Norte	186.46	181	48.39	2341.33	103.80
Branca	Nordeste	697.10	677	202.82	41135.95	116.38
Branca	Sudeste	773.00	773	181.90	33086.98	94.13
Branca	Sul	1038.00	1038	244.46	59758.89	94.20
Branca	Centro-Oeste	248.45	274	61.08	3730.89	98.34
Negra	Norte	2299.10	2279	671.53	450955.11	116.83
Negra	Nordeste	8851.90	8872	2906.29	8446538.77	131.33
Negra	Sudeste	2078.00	2078	539.54	291100.81	103.86
Negra	Sul	518.00	518	135.17	18271.25	104.38
Negra	Centro-Oeste	1171.00	1171	299.59	89754.45	102.34

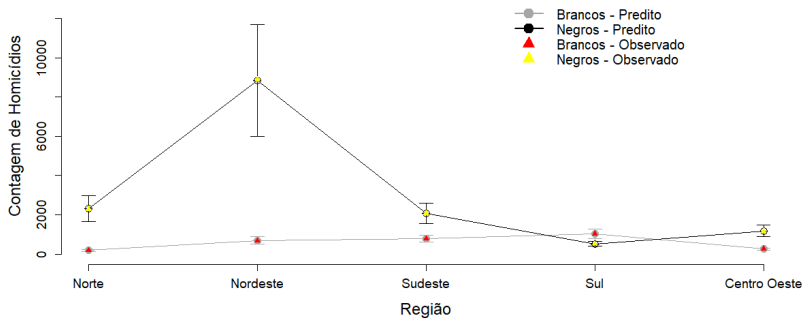


Figura 29: Valores Preditos Segundo Raça/Cor e Região.

B.3 Raça/Cor e Ensino Fundamental

Tabela 54: Análise Descritiva dos Valores Preditos e Comparação com Valores Observados. Raça/Cor x ENSFUND.

Raça/Cor	Anos de Estudo	Predito	Observado	D.P.	Variância	C.V.
Branca	< 8	1650.10	1639	196.46	38595.94	119.06
Branca	≥ 8	1292.90	1304	148.21	21967.63	114.64
Negra	< 8	10461.00	10461	1945.83	3786243.85	186.01
Negra	≥ 8	4457.00	4457	604.05	364872.28	135.53

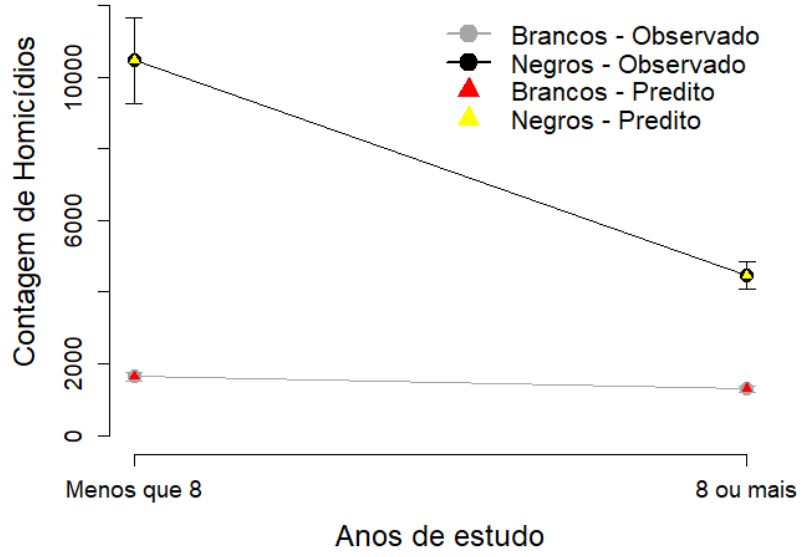


Figura 30: Valores Preditos Segundo Raça/Cor e ENSFUND.

B.4 Sexo e Região

Tabela 55: Análise Descritiva dos Valores Preditos e Comparação com Valores Observados. Sexo x Região.

Sexo	Região	Predito	Observado	D.P.	Variância	C.V.
Feminino	Norte	143.69	153	34.66	1201.43	96.49
Feminino	Nordeste	531.58	521	143.12	20484.06	107.70
Feminino	Sudeste	224.00	224	23.20	538.01	41.42
Feminino	Sul	136.00	136	18.40	338.45	54.11
Feminino	Centro-Oeste	89.74	91	15.43	238.05	68.78
Masculino	Norte	2341.87	2307	660.86	436729.97	112.88
Masculino	Nordeste	9017.42	9028	2870.13	8237662.94	127.32
Masculino	Sudeste	2627.00	2627	392.44	154008.79	59.75
Masculino	Sul	1420.00	1420	142.35	20264.35	40.10
Masculino	Centro-Oeste	1329.71	1354	254.26	64646.01	76.48

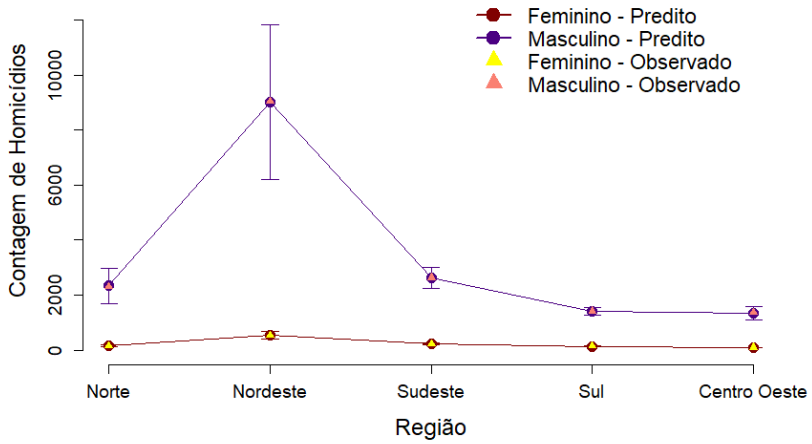


Figura 31: Valores Preditos Segundo Sexo e Região.

B.5 Sexo e Ensino Fundamental

Tabela 56: Análise Descritiva dos Valores Preditos e Comparação com Valores Observados. Sexo x Anos de Estudo.

SEXO	Anos de Estudo	Predito	Observado	D.P.	Variância	C.V.
Feminino	< 8	642.10	631	95.19	9060.38	148.24
Feminino	≥ 8	482.90	494	43.84	1921.62	90.78
Masculino	< 8	11469.00	11469	1896.27	3595836.08	165.34
Masculino	≥ 8	5267.00	5267	559.81	313384.47	106.29

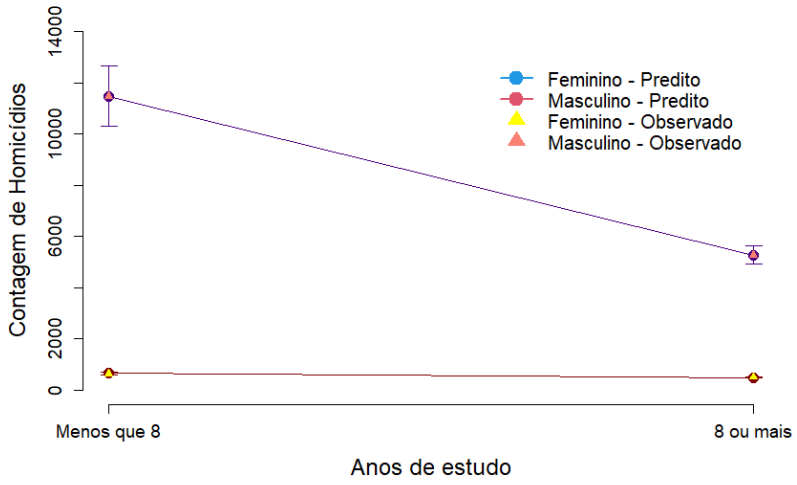


Figura 32: Valores Preditos Segundo Sexo e ENSFUND.

B.6 Ensino Fundamental e Região

Tabela 57: Análise Descritiva dos Valores Preditos e Comparação com Valores Observados. Anos de Estudo x Região.

Anos de Estudo	Região	Predito	Observado	D.P.	Variância	C.V.
< 8	Norte	1667.66	1631	710.09	504228.14	170.32
< 8	Nordeste	7205.00	7205	3067.90	9412006.56	170.32
< 8	Sudeste	1645.00	1645	527.56	278322.11	128.28
< 8	Sul	857.00	857	228.69	52298.61	106.74
< 8	Centro-Oeste	736.45	762	276.61	76515.22	150.24
≥ 8	Norte	817.90	829	323.96	104949.89	158.44
≥ 8	Nordeste	2344.00	2344	928.43	861982.46	158.44
≥ 8	Sudeste	1206.00	1206	329.84	108795.02	109.40
≥ 8	Sul	699.00	699	189.63	35958.03	108.51
≥ 8	Centro-Oeste	683.00	683	228.72	52313.86	133.95

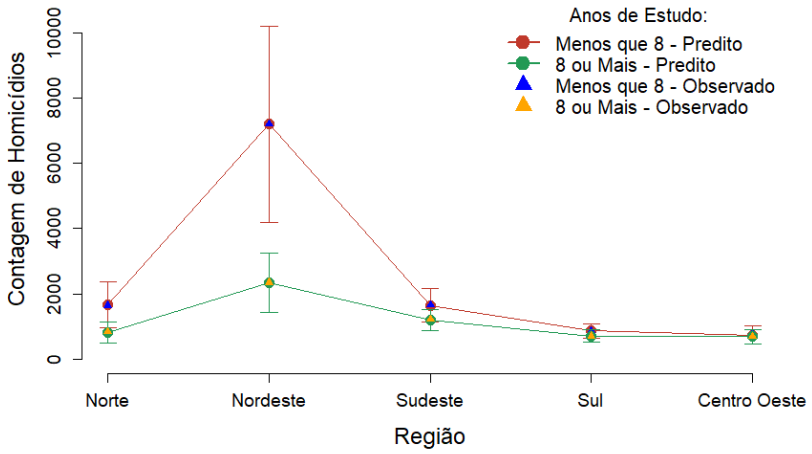


Figura 33: Valores Preditos Segundo Anos de Estudo e Região.