



Universidade de Brasília - UnB
Departamento de Estatística - EST

**Fatores Associados ao Diagnóstico de Hipertensão Arterial:
Uma Aplicação de Regressão Logística Multinível**

Bruna Matter dos Santos

Trabalho de Conclusão de Curso apresentado para obtenção de título de Bacharel em Estatística pelo Departamento de Estatística, Instituto de Exatas da Universidade de Brasília.

Brasília, DF

2017

Bruna Matter dos Santos

**Fatores Associados ao Diagnóstico de Hipertensão Arterial:
Uma Aplicação de Regressão Logística Multinível**

Orientadora:

Prof(a). **Maria Teresa Leão Costa**

Trabalho de Conclusão de Curso apresentado para obtenção de título de Bacharel em Estatística pelo Departamento de Estatística, Instituto de Exatas da Universidade de Brasília.

Brasília, DF

2017

Agradecimentos

Primeiramente à Deus, por ter me ajudado, dado forças, capacitado e providenciado tudo que necessitei para realização deste trabalho e conclusão do curso de Estatística nesta instituição.

Aos meus pais, Elenara e Petrônio Felipe, que me deram todo amor, carinho, apoio, suporte e que acreditaram e investiram em mim para chegar até aqui.

Ao meu esposo, Jary, por estar ao meu lado, me dando apoio, amor, carinho e ânimo para conclusão dessa etapa em minha vida.

À minha orientadora, professora Maria Teresa Leão Costa, pela paciência, pelos ensinamentos, compreensão e dedicação e à professora Maria Margarita Urdaneta Gutierrez por permitir a utilização dos dados da Pesquisa Ambiente & Saúde para realização deste trabalho.

Aos meus demais familiares, amigos e todos que de alguma forma contribuíram para conclusão deste trabalho.

“Se não puder se destacar pelo talento, vença pelo esforço.”
(Dave Weinbaum)

Sumário

| | |
|---|-----------|
| Resumo | 9 |
| 1 Introdução | 11 |
| 2 Objetivos | 14 |
| 2.1 Objetivo geral | 14 |
| 2.2 Objetivos Específicos | 14 |
| 3 Regressão Multinível | 15 |
| 3.1 Regressão Linear Multinível | 15 |
| 3.1.1 Formulação do Modelo e Interpretação dos parâmetros | 15 |
| 3.2 Regressão Logística Multinível | 19 |
| 3.2.1 Regressão Logística | 19 |
| 3.2.2 Regressão Logística Multinível | 21 |
| 3.3 Estimação dos Parâmetros do Modelo | 22 |
| 3.4 Testes de Significância e Qualidade do Modelo | 23 |
| 3.5 Seleção de Variáveis | 24 |
| 3.6 Coeficiente de Correlação Intraclasse | 26 |
| 3.7 Coeficiente de Determinação R^2 | 28 |
| 3.8 Diagnóstico dos Resíduos | 28 |
| 4 Aplicação | 31 |
| 4.1 Introdução | 31 |
| 4.2 Dados | 32 |
| 4.2.1 Material e Métodos | 33 |
| 4.2.2 Método de Análise | 36 |
| 4.3 Análise Descritiva | 36 |
| 4.3.1 Nível 1: Indivíduo | 36 |
| 4.3.2 Nível 2: Município | 40 |
| 4.4 Análise Bivariada | 43 |
| 4.4.1 Nível 1: Indivíduo | 43 |

| | |
|--|-----------|
| | 7 |
| 4.4.2 Nível 2: Município | 50 |
| 4.5 Modelagem | 52 |
| 4.6 Diagnóstico dos Resíduos | 57 |
| 4.7 Interpretação do modelo | 59 |
| 5 Conclusão | 61 |
| 6 Referências Bibliográficas | 63 |

Resumo

A hipertensão arterial, por ser identificada como a causa de grande parcela das doenças cardiovasculares que têm levado muitos brasileiros à óbito, tornou-se um fator de risco que pode ser determinante em alavancar diversos problemas de saúde.

Diversos estudos têm mostrado que os determinantes nas condições de saúde de uma população são resultado de características individuais combinadas à influência de fatores ambientais. Dessa forma, neste trabalho, foram estudados os fatores associados ao indivíduo e ao município em que ele vive, na ocorrência de hipertensão arterial, dos moradores da Região dos Pireneus com base nos dados da amostra de domicílios da Pesquisa Ambiente & Saúde.

A análise foi realizada utilizando um modelo logístico multinível dada a estrutura hierárquica dos dados e por buscar identificar as variáveis associadas ao município (maior nível hierárquico) e ao indivíduo (menor nível hierárquico) no modelo que prediz o diagnóstico de pressão alta.

Palavras-chave: Regressão Logística, Regressão Multinível, Regressão Logística Multinível, Hipertensão Arterial, Pressão Alta, Região dos Pireneus.

1 Introdução

A motivação para o estudo da regressão multinível começa em buscar compreender a relação entre indivíduos e grupos. Pesquisas sociais frequentemente envolvem problemas que investigam a relação entre indivíduos e a sociedade com o objetivo de compreender a influência do grupo sobre o indivíduo. A figura abaixo transmite a ideia básica da análise multinível: indivíduos que estão em grupos e a relação existente entre eles.



Figura 1: Indivíduos em grupos

Fonte: <http://tempodegerminar.blogspot.com.br/2015/03/grupos-sociais.html>

O objetivo da pesquisa multinível é estudar a relação existente entre variáveis que caracterizam indivíduos e variáveis que caracterizam grupos e verificar se existe interação entre as variáveis explicativas de diferentes níveis.

Sendo assim, o estudo da regressão multinível está associado ao conceito de sistemas hierárquicos, que consiste em pessoas organizadas em grupos, em diferentes níveis. Conceitualmente, indivíduos e grupos são definidos em níveis separados de um sistema hierárquico, e esses níveis podem ser definidos por variáveis.

Na esfera educacional, por exemplo, alunos estão agrupados em classes, que por sua vez pertencem a escolas. Funcionários de um banco estão alocados em gerências, que por sua vez pertencem a superintendências. Pacientes são cuidados por médicos, que por sua vez trabalham em hospitais. Nos três exemplos mencionados, pode-se identificar a presença de uma estrutura hierárquica e deseja-se entender os efeitos combinados das variáveis individuais como das de grupos.

Segundo Jacob Laros e Katia Puentes-Palacios (2009, p.349), “o reconhecimento da influência do contexto social no comportamento humano destaca a necessidade de adoção de modelos explicativos que favoreçam a identificação desse efeito.” A figura abaixo ilustra a ideia de uma estrutura hierárquica no contexto educacional.

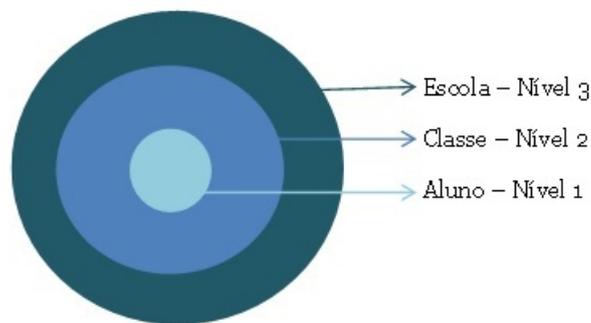


Figura 2: Estrutura hierárquica

O agrupamento de pessoas em diferentes níveis induz a ideia de que estas devem compartilhar atributos similares em decorrência do contexto que lhes é comum. (LAROS e PUENTE-PALACIOS, 2009).

Alunos da mesma escola tendem a ter características similares, como por exemplo: a condição sócio-econômica desses alunos. Assim, quando se trabalha com dados de uma população hierárquica, utilizar o modelo de regressão tradicional, em geral, traz problemas na análise estatística, porque ele não leva em consideração a correlação entre alunos da mesma turma ou da mesma escola. Por terem características semelhantes entre si, pode ocorrer a violação do pressuposto de independência das observações e um possível problema na análise estatística é que a estimação dos erros padrões serão muito menores e os resultados de testes considerados “significativos”, na verdade, estarão errados.

Por isso, quanto maior for a dependência entre membros do mesmo grupo ou nível, maior é a necessidade do uso da análise multinível. (LAROS e MARCIANO, 2009). Logo, essa técnica pode ser mais adequada, por ser uma regressão múltipla que leva em consideração a estrutura hierárquica dos dados.

Alguns critérios para utilização do modelo de regressão multinível, são:

- A população a ser estudada deve ter uma estrutura hierárquica e, conseqüentemente,

as amostras aleatórias em cada nível também;

- Deve haver uma variável resposta medida no menor nível da população em estudo (embora seja possível ter variáveis explicativas em quaisquer um dos níveis).
- Se a variável resposta for quantitativa, o modelo mais adequado é o de regressão linear (multinível), mas se a variável resposta for qualitativa, um modelo mais adequado poderia ser, por exemplo, o de regressão logística (multinível) - por ser mais comumente utilizado, no entanto, existem diversos modelos que poderiam ser aplicados, dependendo do contexto.

Neste trabalho estudou-se a regressão multinível e sua aplicação foi realizada nos dados da amostra de domicílios da Pesquisa Ambiente & Saúde, em que buscou-se identificar os fatores associados, ao indivíduo e ao município, no diagnóstico de pressão alta dos moradores da Região dos Pireneus que foram selecionados.

2 Objetivos

2.1 Objetivo geral

Estudar a regressão multinível e aplicar a técnica aos dados da Pesquisa Ambiente & Saúde para investigar as características individuais e do município associadas ao diagnóstico de pressão alta.

2.2 Objetivos Específicos

- Estudar o modelo multinível;
- Organizar os dados da amostra estudada, que envolve: preparar o banco de dados, tratar adequadamente e fazer a análise descritiva;
- Aplicar o modelo logístico multinível no estudo dos fatores associados ao diagnóstico de hipertensão arterial;
- Estimar os parâmetros, realizar a inferência, predição e analisar os resíduos;
- Interpretar os resultados obtidos;

3 Regressão Multinível

A análise de regressão consiste em expressar a relação entre variáveis explicativas no comportamento de uma variável de interesse, chamada de variável resposta. Essa relação pode ser apresentada na forma de modelos estatísticos. Quando se trabalha com dados de uma população hierárquica, em que indivíduos estão organizados em grupos, em diferentes níveis, é necessário levar em consideração as possíveis similaridades entre indivíduos do mesmo grupo. Nesse caso, o modelo de regressão multinível se torna mais adequado do que o modelo de regressão linear tradicional, por ser projetado para analisar simultaneamente variáveis explicativas de diferentes níveis, por meio de um modelo estatístico que inclua de forma adequada as várias dependências entre indivíduos do mesmo nível.

3.1 Regressão Linear Multinível

3.1.1 Formulação do Modelo e Interpretação dos parâmetros

Considere um exemplo no contexto educacional, em que numa determinada escola deseja-se saber se sexo e idade de alunos do 9º ano, podem influenciar o seu desempenho escolar em matemática e se o número de anos de estudo do professor de matemática (de cada turma) também pode influenciar no desempenho individual dos alunos em matemática.

- O desempenho do aluno é dado pela média das notas finais (realizadas ao longo do 9º ano) nas provas de matemática;
- Sexo do aluno: 1 - Masculino e 2 - Feminino;
- Idade do aluno (medida em anos);
- Anos de estudo do professor de matemática (medido em anos);

Suponha que nessa escola existam 4 turmas do 9º ano, com 40 alunos em cada uma delas. Assim:

- Variável resposta (Y_i): Desempenho em matemática do aluno i pertencente à turma j ;
- Variável explicativa (X_1): Sexo do aluno i pertencente à turma j ;

- Variável explicativa (X_2): Idade do aluno i pertencente à turma j ;
- Variável explicativa (Z) - Nível da turma: Anos de estudo do professor de matemática da turma j ;

Se utilizássemos o modelo de regressão linear tradicional, este seria escrito como:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i \quad (1)$$

O modelo acima assume que as classes são homogêneas entre si, pois não leva em consideração a influência dos anos de estudo do professor de matemática de cada turma no desempenho em matemática dos alunos e outras características. Nesse caso, é interessante aplicar o modelo multinível, que capta a informação de que existe variação no desempenho médio dos alunos, conforme a turma e os anos de estudo do professor da turma.

Um modelo multinível pode ter vários níveis e todos eles podem ter variáveis explicativas. Inicialmente, podem ser feitas equações de regressão separadas por nível. Neste caso, como são dois níveis, cujo o primeiro nível é o aluno e o segundo é a turma, o modelo de regressão multinível (no nível do aluno) pode ser escrito como:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + e_{ij} \quad (2)$$

$$i = 1, \dots, 40 \text{ e } j = 1, \dots, 4$$

Em que:

- Y_{ij} : Desempenho escolar em matemática do aluno i pertencente à turma j ;
- β_{0j} : Intercepto da turma j ;
- $\beta_{1j} X_{1ij}$: Coeficiente de regressão referente ao sexo do aluno i pertencente à turma j ;
- $\beta_{2j} X_{2ij}$: Coeficiente de regressão referente a idade do aluno i pertencente à turma j ;
- Erro aleatório: $e_{ij} \sim (0, \sigma_{ij}^2)$

No intercepto e dentro de cada coeficiente de regressão do modelo apresentado acima, serão inseridas equações de regressão para expressar a relação entre as variáveis no nível da classe e do aluno. Logo, em:

- β_{0j} , será inserida uma equação que inclui a variável Z , para prever o desempenho médio em matemática de uma classe em função dos anos de estudo do professor;

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \quad (3)$$

- β_{1j} , será inserida uma equação que expressa a dependência entre: sexo dos alunos da classe j e anos de estudo do professor dessa turma;

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j} \quad (4)$$

- β_{2j} , será inserida uma equação que expressa a dependência entre: idade dos alunos da classe j e anos de estudo do professor dessa turma;

$$\beta_{2j} = \gamma_{20} + \gamma_{21}Z_j + u_{2j} \quad (5)$$

Essas equações preveem o desempenho médio em matemática de cada turma, tendo os anos de estudo do professor como variável explicativa. Por isso, referente à equação apresentada em (4) por exemplo, pode-se dizer que a variável Z atua como moderadora para a relação entre desempenho em matemática e sexo ou idade do aluno.

Dessa forma, se γ_{01} for positivo, significa que o desempenho médio em matemática de uma turma é maior quando o professor possui mais anos de estudo. Se esse coeficiente for negativo, o desempenho médio em matemática é menor nas turmas que possuem um professor com mais anos de estudo.

Já o coeficiente γ_{11} , se for positivo, implica que o efeito do sexo no desempenho em matemática é maior com um professor que possui mais anos de estudo. Por outro lado, se esse coeficiente for negativo, o efeito do gênero no desempenho em matemática da turma é menor com um professor com mais anos de estudo.

Se γ_{21} for positivo, o efeito da idade sobre o desempenho em matemática é maior nas turmas com um professor com mais anos de estudo. No entanto, se esse coeficiente for negativo, o efeito da idade sobre o desempenho é menor quando o professor tem mais anos de estudo.

Em todas as turmas, os coeficientes de regressão assumem ter distribuição normal multivariada e o termo u faz referência ao erro aleatório no nível da turma. Assim, a diferença entre o modelo multinível para o modelo de regressão usual é que assume-se que cada classe possui um intercepto diferente (β_{0j}) e diferentes coeficientes de regressão (β_{1j} e β_{2j}).

Portanto, valores específicos para o intercepto e para os coeficientes de regressão refletem uma característica da classe. Em geral, a classe que possuir um alto intercepto demonstra ter alunos com um desempenho melhor em matemática do que outra que possua um valor baixo para β_{0j} . Da mesma forma, diferenças no coeficiente de sexo e idade indicam que a relação entre o gênero e idade dos alunos com o seu desempenho em matemática não é a mesma em todas as turmas.

Algumas turmas podem ter um valor elevado para β_{1j} , o que indica que a diferença entre meninos e meninas é relativamente grande. Outras classes podem ter um valor baixo para esse coeficiente, demonstrando que sexo tem um efeito muito pequeno sobre o desempenho dos alunos em matemática e que a diferença entre meninos e meninas é pequena também.

O coeficiente de regressão para idade do aluno pode ser interpretada de forma semelhante: turmas com um alto β_{2j} indicam que a idade do aluno tem um grande impacto em seu desempenho escolar, enquanto turmas que tiverem esse coeficiente baixo indicam que a idade tem um efeito pequeno sobre o desempenho dos alunos em matemática e que a diferença de idade entre alunos da mesma turma é baixa.

Os erros residuais (u_{0j} , u_{1j} , u_{2j}) assumem ter média zero e de serem independentes dos erros e_{ij} no nível do aluno. A tabela abaixo apresenta as variâncias dos erros e a covariância entre os erros (que geralmente não são assumidas como sendo zero) e que deverão ser estimadas.

| Erro aleatório | Variância do erro | Covariância |
|----------------|-------------------|---------------------|
| u_{0j} | $\sigma_{u_0}^2$ | $\sigma_{u_{01}}^2$ |
| u_{1j} | $\sigma_{u_1}^2$ | $\sigma_{u_{02}}^2$ |
| u_{2j} | $\sigma_{u_2}^2$ | $\sigma_{u_{12}}^2$ |

Então, o modelo com duas variáveis explicativas no nível do aluno e uma no nível da classe, pode ser escrito como uma única equação de regressão, quando substituirmos as três equações do nível da classe nos betas da equação no nível do aluno.

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \gamma_{21}X_{2ij}Z_j + u_{1j}X_{1ij} + u_{2j}X_{2ij} + u_{0j} + e_{ij} \quad (6)$$

$$i = 1, \dots, 40 \text{ e } j = 1, \dots, 4$$

- Os termos fixos do modelo são: γ_{00} , $\gamma_{10}X_{1ij}$, $\gamma_{20}X_{2ij}$, $\gamma_{01}Z_j$;
- Os termos $\gamma_{11}X_{1ij}Z_j$ e $\gamma_{21}X_{2ij}Z_j$ são conhecidos como termos de interação;
- Os termos aleatórios são: $u_{1j}X_{1ij}$, $u_{2j}X_{2ij}$, u_{0j} e e_{ij} ;

Assim, por ser um modelos com coeficientes fixos e aleatórios, esse é um caso particular de modelos mistos.

3.2 Regressão Logística Multinível

3.2.1 Regressão Logística

O modelo estatístico de regressão logística é utilizado quando a variável resposta é do tipo binária ou dicotômica, isto é, quando só existem dois possíveis resultados. Novamente, considerando o modelo de regressão linear tradicional, como mostrado em (1), quando a variável resposta (Y_i) é binária, muda a forma de se interpretar a resposta esperada ($E(Y_i)$). A resposta média representa a probabilidade de ($Y_i = 1$) para os níveis das variáveis explicativas, pelo fato de (Y_i) ter distribuição de Bernoulli ($1, \pi$) com probabilidade de sucesso $P(Y_i = 1) = \pi_i$ e de insucesso $P(Y_i = 0) = 1 - \pi_i$.

Sabendo que a distribuição dos erros depende de Y_i , que tem distribuição de Bernoulli, a função logística é apresentada na forma:

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \quad (7)$$

Essa função pode ser linearizada, então, como Y_i é binária e a resposta média representa uma probabilidade ($E(Y_i) = \pi_i$), ao realizar a transformação logarítmica:

$$\pi'_i = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) \quad (8)$$

Obtém-se a transformação logito da probabilidade de π_i :

$$\pi'_i = \beta_0 + \beta_1 X_i \quad (9)$$

Assim, a função resposta transformada é denominada função resposta logito e π'_i é a resposta logito média. A razão de $\frac{\pi_i}{(1-\pi_i)}$, que surge na transformação logito, é conhecida como *odds* ou chance. A razão entre duas chances, conhecida como *odds ratio*, é dada por:

$$\theta = \frac{\text{odds1}}{\text{odds2}} = \frac{\frac{\pi_1}{(1-\pi_1)}}{\frac{\pi_2}{(1-\pi_2)}} \quad (10)$$

O conceito de chance é muito utilizado em regressão logística, pois quando a variável explicativa é quantitativa, as estimativas dos coeficientes são obtidas por meio da sua exponencial ($\exp(\beta_1)$, $\exp(\beta_2)$,...). Mas, quando a variável explicativa é qualitativa, a interpretação da razão de chances revela quantas vezes é mais provável ocorrer sucesso considerando um determinado grupo em relação a outro grupo. Então, se a probabilidade de sucesso é maior que a de fracasso, a *odds* > 1 , caso contrário *odds* < 1 .

A regressão logística multinível é um caso particular de Modelos Lineares Generalizados que é uma classe de modelos caracterizada por:

- Um componente aleatório que especifica a distribuição de probabilidade, Y ;
- Um componente sistemático que especifica as variáveis explicativas (x_1, \dots, x_k) de forma linear, cuja combinação linear dessas variáveis é chamada de preditor linear (η) = $\alpha + \beta_1 X_1 + \dots + \beta_k X_k$;
- Pela função de ligação, que especifica a função $g(\cdot)$, uma função monótona e dife-

renciável, que relaciona $E(Y) = \eta$. Assim, ela conecta os componentes aleatório e sistemático. Logo, $g(\eta)$ é escrito como: $g(\eta) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k$.

Dessa forma, o modelo de regressão linear (tradicional) e o de regressão logística fazem parte dessa classe de Modelos Lineares Generalizados por considerarem, respectivamente, $g(\eta) = \mu$ e Y com distribuição normal, $g(\eta) = \log(\frac{\mu}{1-\mu})$, sendo a distribuição de Y , binomial.

3.2.2 Regressão Logística Multinível

A aplicação da técnica de regressão logística multinível é conveniente quando a variável resposta é binária e os dados possuem uma estrutura hierárquica. A diferença entre o modelo de regressão logística e o modelo de regressão logística multinível é a inserção dos coeficientes aleatórios e das variáveis explicativas em vários níveis. Assim, para construção do modelo para variável resposta dicotômica, com dois níveis, temos que:

- A distribuição de probabilidade de π_{ij} é binomial (μ, η_{ij}) com média (μ) ;
- O preditor linear é a equação de regressão multinível para η , dada por:

$$\eta = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{1j}X_{ij} + u_{0j} \quad (11)$$

- A função de ligação é a função logito dada por: $\eta = \text{logito}(\mu)$.

É importante mencionar que este modelo não inclui o termo da variância do erro. Por ter distribuição binomial, a variância é uma função da média e não pode ser estimada separadamente, dada por:

$$\sigma_e^2 = \frac{\pi_{ij}}{(1 - \pi_{ij})} \quad (12)$$

Então, semelhantemente ao modelo linear multinível, o logístico multinível será escrito como:

$$\text{logito}(\pi_{ij}) = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \gamma_{21}X_{2ij}Z_j + \quad (13)$$

$$u_{1j}X_{1ij} + u_{2j}X_{2ij} + u_{0j} \quad (14)$$

Cuja interpretação dos parâmetros é semelhante a explicada anteriormente, mudando apenas o significado da variável resposta, que representa a probabilidade de sucesso do indivíduo i no grupo j . Consequentemente, o foco da análise está na exponencial dos coeficientes ($\exp(\beta_j)$) das variáveis explicativas do modelo que informarão a chance de sucesso da variável resposta de acordo com o valor estimado do coeficiente do modelo.

3.3 Estimação dos Parâmetros do Modelo

Com o objetivo de estimar os parâmetros do modelo multinível, os estimadores habituais na análise de regressão multinível são os de máxima verossimilhança (MV). Esses estimadores maximizam uma função de verossimilhança e são vantajosos por produzir estimativas mais consistentes, não viesadas e eficientes.

Na análise multinível, podem ser utilizadas duas funções de verossimilhança: a de máxima verossimilhança completa (MVC) e a de máxima verossimilhança restrita (MVR).

- A **MVC** inclui os coeficientes de regressão e os componentes da variância. Ela analisa os coeficientes de regressão como fixos quando os componentes da variância não são estimados. Quando esses componentes são estimados, os coeficientes de regressão são tratados como desconhecidos.
- A **MVR** inclui somente os componentes da variância. A estimativa desses componentes ocorre após remoção dos efeitos fixos do modelo, pois isso a torna menos viesada.

A função de máxima verossimilhança completa é viesada, porém, esse viés geralmente é pequeno. Dessa forma, os dois métodos produzem resultados semelhantes. Além desses métodos, outros comumente utilizados em modelos multiníveis, são: mínimos quadrados generalizados, bootstrap e métodos bayesianos.

No caso do modelo logístico multinível, os métodos de estimação usados são todos numéricos, uma vez que com muitos parâmetros, podem ocorrer problemas de convergência do algoritmo para ser estimado. O método de Quase-Verossimilhança (MQV) é muito utilizado, pois o processo de estimação ocorre de forma iterativa, utilizando expansão de séries de Taylor e, por iniciar com valores aproximados para os parâmetros que são melhorados a cada iteração.

3.4 Testes de Significância e Qualidade do Modelo

A partir da estimação dos parâmetros do modelo pelo método MV (regressão linear multinível) ou MQV (regressão logística multinível), estes com seus respectivos erros-padrão, o teste de significância usualmente utilizado é o Teste de Wald, que é dado pela estatística de significância z , abaixo:

$$z = \frac{\hat{\beta}_j}{EP(\hat{\beta}_j)} \quad (15)$$

Em que $z \sim N(0, 1)$.

Bryk e Raudenbush (1992, p. 50) propõe uma alternativa para testar os componentes da variância:

$$\chi^2 = \sum \frac{(\hat{\beta}_j - \beta)^2}{\hat{V}_j} \quad (16)$$

Com $J - p - 1$ g.l, e:

- β_j é o estimador de mínimos quadrados (calculado separadamente no grupo j);
- β é a média dos estimadores e;
- \hat{V}_j é a variância estimada do grupo j ;

Com objetivo de verificar quão bem o modelo se ajusta a um banco de dados, a partir da função de verossimilhança, utiliza-se a estatística chamada de *deviance*, definida como:

$$deviance = -2 \ln(MV) \quad (17)$$

De modo geral, modelos que possuem um *deviance* menor estão melhor ajustados aos dados em estudo. Essa medida pode ser utilizada para comparação de modelos

aninhados, isto é, quando um modelo é um caso particular de outro mais geral. A diferença entre as *deviances* dos dois modelos aninhados possui distribuição qui-quadrado com d graus de liberdade, sendo d a diferença entre os parâmetros dos modelos.

Em se tratando de modelos não aninhados, pode-se utilizar o *Critério de Akaike (AIC)* para medir a qualidade de ajustamento do modelo. O *AIC* é dado por:

$$AIC = deviance + 2q \quad (18)$$

O *AIC* para modelos de regressão multinível é calculado a partir do *deviance* somado ao número de parâmetros estimados (q). Associado à essa quantidade de parâmetros, há uma penalidade para corrigir o viés resultante da comparação de modelos que possuem quantidades diferentes de parâmetros a serem estimados.

Outro índice de ajuste semelhante é o *Critério de Informação Bayesiano (BIC)*, dado por:

$$BIC = deviance + q \ln(N) \quad (19)$$

Semelhante ao *deviance* que deve ser baixo, o modelo que tiver o menor *AIC* ou *BIC* indica estar melhor ajustado para os dados em estudo.

3.5 Seleção de Variáveis

Modelos multiníveis podem facilmente ter um número grande de parâmetros, de acordo com a quantidade de níveis que estão sendo estudados. Ainda que hajam poucas variáveis explicativas, a análise deste tipo de modelo é complexa. Por isso, a forma mais adequada de iniciar a análise, é de forma mais simples, considerando somente as variáveis que podem trazer uma contribuição significativa para o modelo e, gradualmente, serem acrescentadas mais informações (variáveis), até que se atinja o modelo ideal.

Abaixo será descrito o procedimento de seleção de variáveis, dividido em cinco passos, sugere-

rido por Hox (2010):

- **Passo 1:** Considere o modelo sem nenhuma variável explicativa, nele haverá somente o intercepto.

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad (20)$$

Em que γ_{00} é o intercepto, u_{0j} e e_{ij} são os erros aleatórios no nível do grupo e do indivíduo. Por meio desse modelo, pode-se obter uma estimativa para o coeficiente de correlação intraclasse ρ , dado por:

$$\rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_e^2} \quad (21)$$

Em que $\sigma_{u_0}^2$ é a variância do erro no nível do grupo e σ_e^2 é a variância do erro no nível do indivíduo. Obter esse coeficiente é de fundamental importância para identificar a necessidade da análise multinível.

- **Passo 2:** Inserir no modelo todas as variáveis explicativas no menor nível, isto é, dos indivíduos. Assim, pode-se observar a contribuição de cada variável do nível mais baixo.

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + u_{0j} + e_{ij} \quad (22)$$

Em que X_{pij} são as p variáveis explicativas no nível individual.

- **Passo 3:** Adicionar as variáveis explicativas do nível mais alto:

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q}Z_{qj} + u_{0j} + e_{ij} \quad (23)$$

Em que Z_{qj} são as q variáveis explicativas no nível do grupo. Nesse passo, é possível identificar se as variáveis do nível mais alto explicam as diferenças entre os grupos.

- **Passo 4:** Avaliar se os coeficientes de regressão das variáveis explicativas do modelo possuem variância significativa entre os grupos.

$$Y_{ij} = \gamma_{00} + \gamma_{p0}X_{pij} + \gamma_{0q}Z_{qj} + u_{pj}X_{pij} + u_{0j} + e_{ij} \quad (24)$$

Em que u_{pj} são os erros no nível do grupo, dos coeficientes de regressão das variáveis explicativas no nível do indivíduo, X_{pij} .

- **Passo 5:** Adicionar as variáveis de interação entre o nível do grupo e do indivíduo (que tiveram variância significativa no Passo 4).

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{1j}X_{1ij} + u_{0j} + e_{ij} \quad (25)$$

Após seguir esses passos, novamente pode ser realizado o teste qui-quadrado (16) para verificar a melhoria do ajuste do modelo aos dados em estudo.

3.6 Coeficiente de Correlação Intraclasse

O coeficiente de correlação intraclasse apresenta a proporção da variância do erro no nível do grupo comparada à variância total (grupo e indivíduo). Além disso, ela mede a homogeneidade intra grupo e a heterogeneidade entre grupos distintos. Esse coeficiente apresenta valores no intervalo $[0,1]$ e, quanto maior for ρ , maior é a necessidade de utilizar um modelo multinível ao invés de um modelo de regressão clássico.

Em um modelo com dois níveis, ρ é calculado no modelo que possui somente o intercepto, conforme indicado no Passo 1. Mas se houverem três níveis, seguindo os passos propostos acima, o modelo que possui somente o intercepto será escrito como:

$$Y_{ijk} = \gamma_{000} + v_{0k} + u_{0jk} + e_{ijk} \quad (26)$$

cujas variâncias do primeiro, segundo e terceiro níveis serão indicadas por: σ_e^2 , $\sigma_{u_0}^2$ e $\sigma_{v_0}^2$. Dois métodos possíveis para se obter ρ em um estudo com três níveis, são:

- O primeiro método, proposto por (Davis Scott, 1995), define a correlação intraclasse no 2º e 3º níveis, indicado por ρ_{n2} e ρ_{n3} , respectivamente.

$$\rho_{n2} = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{v0}^2 + \sigma_e^2} \quad (27)$$

$$\rho_{n3} = \frac{\sigma_{v0}^2}{\sigma_{v0}^2 + \sigma_{u0}^2 + \sigma_e^2}$$

- O segundo método, proposto por (Siddiqui, Hedeker, Flay, Hu, 1996), define a correlação intraclasse como:

$$\rho_{n2} = \frac{\sigma_{u0}^2 + \sigma_{v0}^2}{\sigma_{u0}^2 + \sigma_{v0}^2 + \sigma_e^2}$$

$$\rho_{n3} = \frac{\sigma_{v0}^2}{\sigma_{v0}^2 + \sigma_{u0}^2 + \sigma_e^2}$$

Ambos os métodos estão corretos, no entanto, (Algina, 2000) recomenda utilizar o primeiro método quando há interesse em decompor a variância em todos os níveis disponíveis, afim de identificar o quanto da variância total pode ser explicada por cada nível. Já o segundo método representa uma estimativa da correlação esperada entre dois indivíduos, aleatoriamente escolhidos, pertencentes ao mesmo nível. Por exemplo: ao escolher dois alunos da mesma turma, pressupõe-se que ambos também pertençam a mesma escola, por isso, o segundo método inclui no numerador a variância entre membros do segundo e terceiro níveis, isto é, a variância entre alunos da mesma classe que pertencem a mesma escola na estimativa da correlação esperada do nível 2 (entre alunos da mesma turma).

Já na abordagem logística multinível, o coeficiente de correlação intraclasse possui fator de escala igual a 1, assim a variância que deve ser considerada é de $\frac{\pi^2}{3} \approx 3,29$ (Hox, 2010). Dessa forma, ρ é dado por:

$$\rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0+3,29}^2}$$

E terá a mesma interpretação do modelo de regressão linear multinível.

3.7 Coeficiente de Determinação R^2

Na análise de regressão tradicional, uma maneira de avaliar a qualidade do ajuste do modelo é através do coeficiente de determinação. Ele indica a quantidade de variabilidade nos dados que é explicada pelo modelo de regressão. Mas, se tratando de modelos multiníveis, utilizar o R^2 pode não ser o mais adequado, pela possibilidade de existir uma grande quantidade da variância não explicada em vários níveis, e também, pois se houverem coeficientes aleatórios no modelo, ele se tornará extremamente complexo.

Assim, para calcular a estatística análoga ao coeficiente de determinação, (Hox, 2010) sugere o R^2 para modelos multiníveis, separado por nível:

- No menor nível, pode-se utilizar:

$$R_1^2 = \frac{\sigma_{e|b}^2 - \sigma_{e|m}^2}{\sigma_{e|b}^2}$$

Em que $\sigma_{e|b}^2$ é a variância do erro no nível mais baixo no modelo que possui apenas o intercepto e $\sigma_{e|m}^2$ é a variância do erro no nível mais baixo a ser comparado.

- No segundo nível, utiliza-se:

$$R_1^2 = \frac{\sigma_{u_0|b}^2 - \sigma_{u_0|m}^2}{\sigma_{u_0|b}^2}$$

Em que $\sigma_{u_0|b}^2$ é a variância do erro no segundo nível no modelo que possui apenas o intercepto e $\sigma_{u_0|m}^2$ é a variância do erro no nível mais baixo a ser comparado.

3.8 Diagnóstico dos Resíduos

A análise de resíduos é importante no estudo de regressão no intuito de verificar os pressupostos de normalidade e linearidade dos dados. Modelos multiníveis também assumem esses pressupostos e, por serem mais complexos do que os modelos de regressão tradicionais, a verificação dessas suposições ainda se faz necessária, visto que o não atendimento

delas compromete a qualidade do modelo. Assim, no caso multinível, tem-se um resíduo para cada efeito aleatório do modelo e cada β possui um resíduo que deve ser estudado.

Por meio do gráfico dos quantis da distribuição normal e os quantis dos resíduos padronizados, pode-se verificar se há normalidade. Se os resíduos tiverem distribuição normal então o gráfico deve apresentar uma linha reta na diagonal, conforme será apresentado na Figura 3.1. Com objetivo de verificar linearidade e variância constante, pode-se utilizar o gráfico de dispersão dos resíduos e os valores preditos, que fornece informações sobre possíveis falhas nessas suposições. Se elas não são violadas, os pontos devem estar dispersos de forma aleatória em torno do zero, como mostra a Figura 3.2.

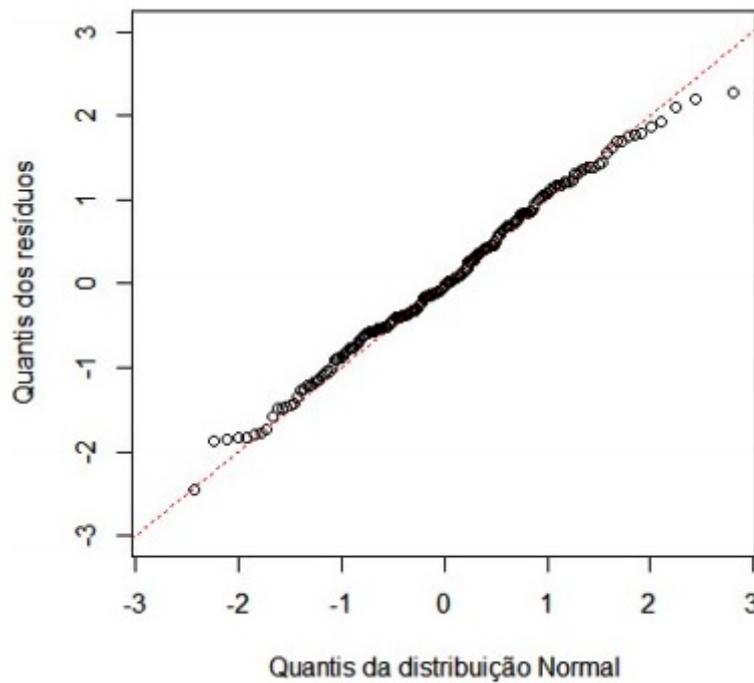


Figura 3: Gráfico de Normalidade

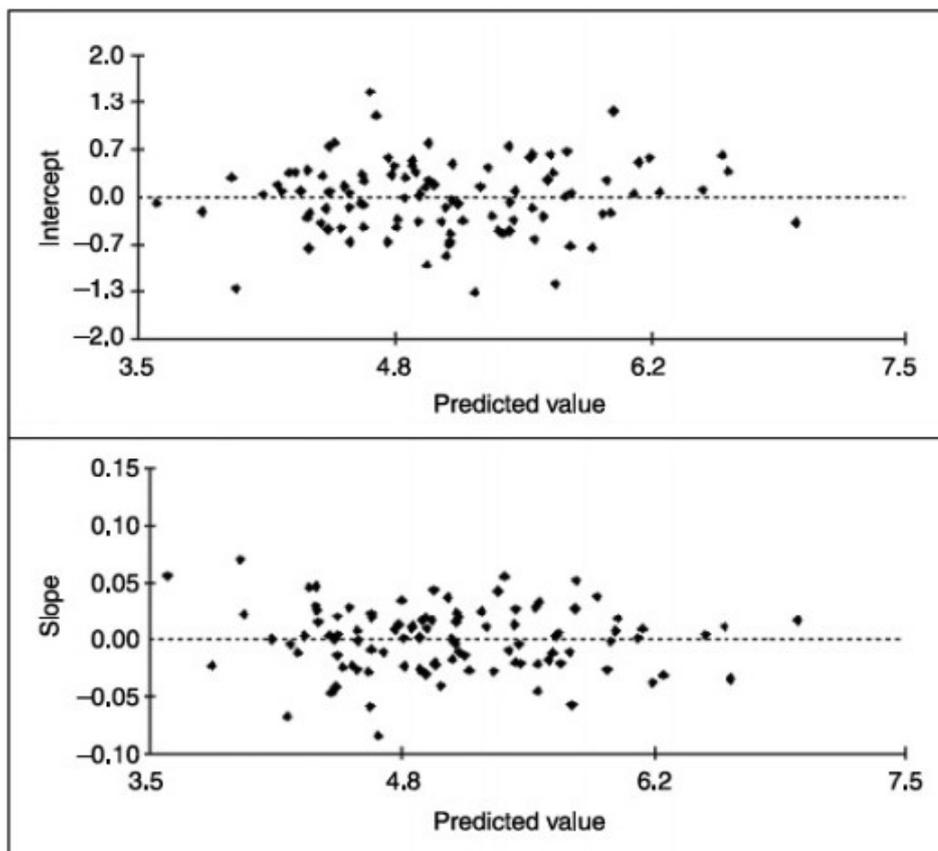


Figura 4: Gráfico de Linearidade

Fonte: HOX, J. J.. Multilevel analysis: techniques and applications. Second Edition, 2010.

4 Aplicação

4.1 Introdução

Diversos estudos têm mostrado que os determinantes das condições de saúde de uma população são resultado de características individuais combinadas à influência de fatores ambientais. De acordo com Diex-Roux (2007), o local de residência constitui uma unidade de análise importante, por possuir atributos físicos e sociais que, em tese, afetam as condições de vida de uma população.

As doenças que mais têm causado morte entre os brasileiros são as cerebrovasculares (como AVC) e as cardiovasculares (como infarto do miocárdio), sendo esta última responsável por 33% dos óbitos com causas conhecidas. Um dos maiores fatores de risco que tem levado a ocorrência dessas doenças, sendo estimado entre 25% a 40% dos casos, é a hipertensão arterial (FUCHS FD, 2004).

Vários estudos epidemiológicos e ensaios clínicos já demonstraram a drástica redução da mortalidade por problemas cardiovasculares com o tratamento da hipertensão arterial. Dessa forma, existem boas evidências médicas de que medidas de pressão arterial podem identificar adultos com maior risco para o desenvolvimento de doenças cardiovasculares, em razão da hipertensão.

Em discussões acerca dos fatores sociais que podem influenciar as condições de saúde de uma população, Whitehead et al (2001) comenta sobre a importância de se influenciar os diferentes elementos que fazem parte do cotidiano do indivíduo com objetivo de prevenir doenças. Assim, elementos relacionados ao ambiente social, às condições de vida, ao acesso a saúde e lazer são necessários para que sejam identificados processos socioculturais que influenciem as condições de saúde da população.

Sabe-se que a ocorrência da hipertensão está associada a comportamentos e hábitos tais como a falta de atividade física, alimentação não saudável, fumo e uso de álcool. Por isso, um dos lugares privilegiados para observação dos determinantes sociais é o espaço em que as pessoas vivem, trabalham e estabelecem suas relações sócio-afetivas. São nos locais onde as pessoas vivem o seu cotidiano que é possível identificar os elementos que produzem determinados comportamentos ou estilo de vida.

Com o objetivo de estudar os fatores associados ao diagnóstico de hipertensão arterial, relacionados às características do indivíduo e do ambiente, isto é, do local de residência, será utilizado o modelo de regressão logística multinível por este conseguir incorporar em sua formulação a natureza de agrupamento da população em estudo.

4.2 Dados

O presente trabalho utilizou os dados provenientes da amostra de domicílios obtida pela Pesquisa Ambiente & Saúde – realizada pelo Departamento de Saúde Coletiva (DSC) da Faculdade de Ciências da Saúde (FS) da Universidade de Brasília (UnB) em parceria com o Núcleo de Estudos em Saúde Coletiva da Universidade Federal de Goiás (NESC/UFG) e o Ministério da Saúde, Instituto de Comunicação e Informação Científica e Tecnológica da Fundação Oswaldo Cruz (MS/ICICT/FIOCRUZ). Essa pesquisa foi realizada na Região dos Pireneus e seu principal objetivo foi descrever e estudar as associações entre as características físicas e sociais do ambiente e as condições de vida e saúde da população nessa Região do Entorno do Distrito Federal (RIDE-DF). Nela foi trabalhada a combinação de diversos métodos e técnicas para a caracterização do ambiente relacionado com a disponibilidade e qualidade de alimentos, espaços para atividade física, consumo de fumo e álcool, assim como outras características da esfera social, econômica e relacionadas a disponibilidade de serviços de saúde.

A Região dos Pireneus é composta por sete municípios do estado de Goiás - Abadiânia, Alexânia, Cocalzinho de Goiás, Corumbá de Goiás, Mimoso de Goiás, Padre Bernardo e Pirenópolis e que fazem parte da Região Integrada de Desenvolvimento do Distrito Federal e Entorno (RIDE-DF). De acordo com a Pesquisa, estima-se que a Região dos Pireneus representa 28% do total da população da RIDE-DF, com 459.246 habitantes (IBGE – Censo Demográfico, 2010).

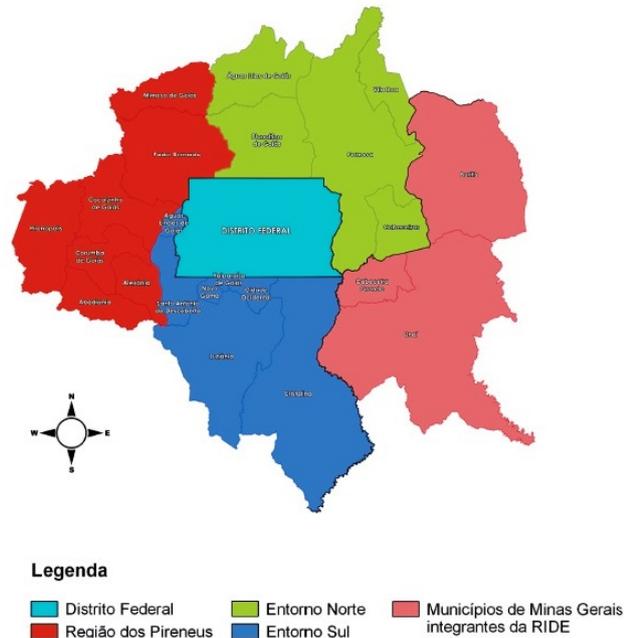


Figura 5: Mapa de localização da RIDE-DF

Fonte: GUTIERREZ, M.; PEREIRA, E.; STEINKE, V.. Ambiente e saúde: uma abordagem multidisciplinar. Goiânia: Gráfica UFG, 2017.

4.2.1 Material e Métodos

Segundo Gutierrez et. al (2017, p.19), por se tratar de uma Pesquisa que apresenta a articulação entre diferentes componentes, o ponto de partida, e no qual os demais se somam para produção de respostas sobre as condições de saúde da população da Região dos Pireneus, RIDE/DF, a Pesquisa se trata de um estudo epidemiológico analítico com um componente descritivo tipo inquérito domiciliar de base populacional e componentes qualitativos aninhados. O projeto apresenta seis componentes de pesquisa, sendo:

- Caracterização do ambiente físico por meio de técnicas de observação sistemática e georreferenciamento de bens e serviços;
- Descrição da disponibilidade de alimentos saudáveis;
- Caracterização do ambiente social por meio de técnicas etnográficas e métodos qualitativos;
- Inquérito de saúde de base populacional;

- Avaliação cognitiva do questionário do inquérito;
- Monitoramento e avaliação com o propósito de avaliar o possível impacto nas condições de saúde por mudanças no ambiente físico e social ao longo do tempo produto da implementação de políticas públicas por meio do monitoramento periódico e sistemático das condições de saúde dos participantes do estudo e a população em geral.

O questionário utilizado para coletar os dados correspondeu à primeira versão do questionário utilizado na Pesquisa Nacional de Saúde (PNS), uma pesquisa de base domiciliar, de âmbito nacional, realizada em parceria com o Instituto Brasileiro de Geografia e Estatística (IBGE) com periodicidade de 5 anos. O questionário é modular e foi dividido em duas partes, em que, na primeira, o entrevistado é o responsável pelo domicílio. São investigadas no domicílio a sua composição, características, gastos dos moradores com saúde, entre outros fatores relacionados. Na segunda parte do questionário, o entrevistado foi selecionado entre os moradores elegíveis do domicílio (idade igual ou superior a 18 anos) para responder ao questionário individual. Este questionário individual, por sua vez, conta com 10 módulos (A a J). Neste relatório foram considerados somente os módulos de A a D e M, listados a seguir, por trazerem informações dos moradores relacionados ao tema deste estudo:

- A) Características sócio-demográficas e apoio social;
- B) Auto-avaliação do estado de saúde nos seus vários domínios;
- C) Estilo de vida, autoferidos;
- D) Morbidade e;
- M) Características de saúde dos moradores do domicílio.

Na Pesquisa Ambiente & Saúde, a amostra estudada foi constituída por indivíduos adultos com idade igual ou superior a 18 anos, residentes na área urbana e rural dos 7 municípios do Goiás que compõem a Região dos Pireneus da Região Integrada de Desenvolvimento do Distrito Federal e Entorno (RIDE-DF). Foi realizada amostragem por conglomerados em três estágios (sem substituição).

- No primeiro estágio, foram selecionados 50 setores censitários, com probabilidade proporcional ao tamanho, dado pelo número de domicílios do setor. Para a seleção dos setores censitários foram considerados, a partir da lista de setores obtida na Fundação

Instituto de Geografia e Estatística (CD ROM - Contagem da população 2010 Agregado por Setores Censitários), primeiramente, todos os setores com domicílios particulares permanentes, isto é, aqueles que servem como residência de famílias e são construídos com material durável (permanente). Para a seleção da amostra de setores, foram eliminados todos os setores com menos de 30 domicílios particulares permanentes e os setores classificados como especiais (penitenciárias, instalações militares, asilos, reservas indígenas, entre outros).

- No segundo estágio, foram selecionados 18 domicílios por setor censitário por meio de uma amostragem sistemática, onde o primeiro domicílio foi selecionado aleatoriamente entre o primeiro do setor e a razão entre o número total de domicílios no setor e o número de domicílios a serem selecionados (denominado pulo). A seleção dos 17 domicílios restantes foi realizada de forma sistemática pela contagem do pulo até completar os 18 domicílios.
- No terceiro estágio, em cada domicílio, um indivíduo foi selecionado com equiprobabilidade entre os moradores adultos residentes (18 anos ou mais de idade), para responder ao questionário individual.

O procedimento de seleção do adulto no domicílio ocorreu da seguinte forma:

- 1) Foram listados todos os indivíduos com idade igual ou superior a 18 anos residentes no domicílio selecionado para participar da pesquisa;
- 2) Cada um recebeu uma numeração de 1 em diante, começando pelos homens do domicílio, do mais velho para o mais novo e, em seguida, listou-se as mulheres seguindo a mesma ordem, da mais velha para a mais nova;
- 3) Contagem de quantos adultos moram no domicílio para, em seguida, de acordo com o número do domicílio dentro do setor censitário e do número de adultos naquela residência, selecionou-se aquele que responderia as perguntas do módulo individual, seguindo uma tabela numérica aleatória para cada setor censitário.

4.2.2 Método de Análise

Inicialmente, neste estudo, foram consideradas as 770 respostas, no entanto, após o tratamento dos dados foram retiradas as observações com respostas em branco das variáveis sexo e ocorrência de pressão alta, por isso o banco de dados contou com 715 observações.

Na modelagem foram consideradas 677 observações, em virtude da eliminação dos registros com respostas em branco para as variáveis em estudo.

Utilizou-se modelo de regressão logístico multinível em que foram estudadas quais características associadas aos indivíduos (considerado menor nível) e quais características associadas ao ambiente em que estes vivem (considerado maior nível) influenciam na ocorrência de pressão alta.

Para análise dos resultados, foi utilizado o software estatístico SAS. Nele, a análise dessa técnica é feita através do PROC GLIMMIX, procedimento para análise de modelos lineares generalizados mistos, incluindo a regressão logística multinível, sendo obrigatório colocar a classe de referência para as variáveis qualitativas e avaliar o grau de ajustamento do modelo.

4.3 Análise Descritiva

Com objetivo de conhecer sobre os indivíduos selecionados na amostra de domicílios da Região dos Pireneus da Pesquisa Ambiente & Saúde foi realizada a análise descritiva para se ter uma ideia das características dessas pessoas e fazer uma pré-seleção de possíveis variáveis que estejam associadas ao diagnóstico de pressão alta.

4.3.1 Nível 1: Indivíduo

- Características Sócio-Demográficas

Pela tabela abaixo nota-se que, na amostra em estudo, 53,8% são mulheres, o que indica certa predominância do sexo feminino. Mais da metade das pessoas é casada ou vive com companheiro, representando 57,2% da amostra em estudo. A média de idade é de 46,3 anos de idade, o que indica este ser um grupo mais adulto. Com relação à educação,

observa-se que 55,2% das pessoas possuem grau de instrução até o Elementar Completo e Fundamental Incompleto e apenas 9%, aproximadamente, possui nível superior completo ou mais. Além disso, a média de anos de estudo com aprovação é, em cerca, de 8 anos. Essas informações revelam o baixo grau de escolaridade da maior parte das pessoas que fizeram parte da amostra estudada. Quanto à ocupação, foi obtido que 53,7% da amostra trabalha e está em atividade atualmente e 30,6% já trabalhou, mas não trabalha mais.

Tabela 1: Variáveis relacionadas às características sócio-demográficas

| Variável Explicativa | % |
|--|------|
| Sexo | |
| Feminino | 53,8 |
| Masculino | 46,1 |
| Estado civil | |
| Nunca foi casado (a) | 25,0 |
| Casado ou vive com companheiro (a) | 57,2 |
| Separado ou divorciado (a) | 10,5 |
| Viúvo (a) | 7,2 |
| Grau de Instrução | |
| Analfabeto/Menos de um ano de instrução | 11,6 |
| Elementar Incompleto | 19,8 |
| Elementar Completo e Fundamental Incompleto | 23,8 |
| Fundamental Completo e Ensino Médio Incompleto | 13,1 |
| Ensino Médio Completo e Superior Incompleto | 22,8 |
| Superior Completo ou mais | 8,8 |
| Raça | |
| Branca | 34,0 |
| Preta | 9,8 |
| Amarela | 7,3 |
| Parda | 47,1 |
| Indígena | 1,8 |
| Ocupação | |
| Trabalha e em atividade atualmente | 53,7 |
| Trabalha, mas não em atividade atualmente | 6,4 |
| Já trabalhou, mas não trabalha mais | 30,6 |
| Nunca trabalhou | 9,2 |

- Auto-Avaliação do Estado de Saúde

Da amostra em estudo, mais da metade dos moradores pesquisados da Região dos Pireneus, avaliaram sua saúde como boa ou muito boa, cerca de 58% dos respondentes, conforme pode-se observar pela figura abaixo.

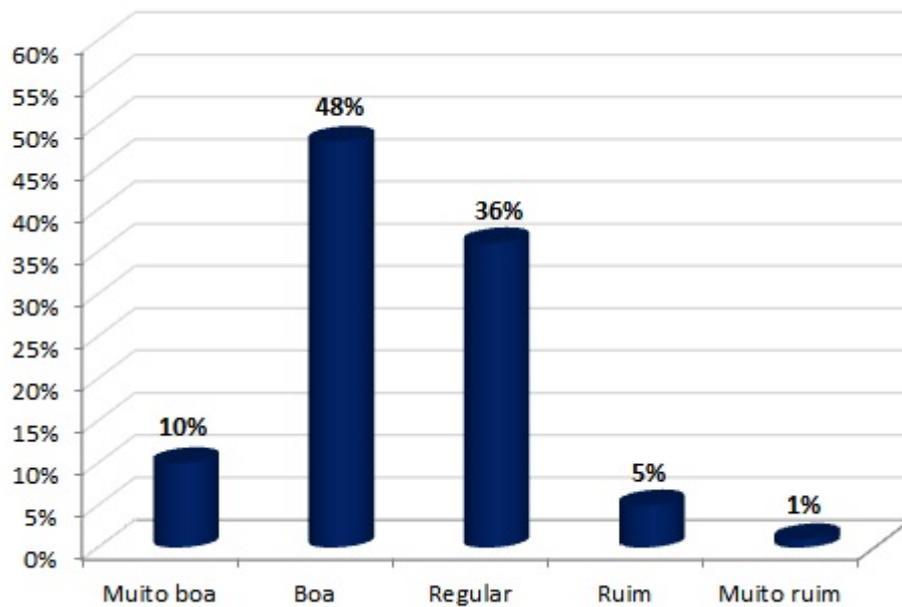


Figura 6: Auto-Avaliação do Estado de Saúde

A tabela abaixo apresenta os resultados das variáveis relacionadas à auto-avaliação do estado de saúde.

Tabela 2: Variáveis relacionadas à auto-avaliação do estado de saúde

| Variável Explicativa | % |
|---|------|
| Grau em que sente dores no corpo | |
| Nenhum | 36,2 |
| Leve | 22,9 |
| Médio | 21,7 |
| Intenso | 16,0 |
| Muito intenso | 3,0 |
| Grau em que sente dores de cabeça ou enxaqueca | |
| Nenhum | 46,7 |
| Leve | 21,2 |
| Médio | 14,2 |
| Intenso | 14,0 |
| Muito intenso | 3,8 |

Assim, nota-se que mais de 40% das pessoas sente dores no corpo acima do grau leve de intensidade. E quanto ao grau de dores de cabeça ou enxaqueca, quase 32% sentem dor acima do grau leve de intensidade.

- Estilo de vida

O Índice de Massa Corporal (IMC) médio é de 26,13 classificado como acima do peso pela Organização Mundial de Saúde (OMS). Foi visto também que as pessoas praticam algum esporte ou atividade física em média três vezes por semana. Com relação à alimentação, aproximadamente, 55% tira o excesso da gordura visível quando come carne vermelha, 58% nunca ingere bebida alcoólica e quase 16% costuma fumar diariamente.

Tabela 3: Variáveis relacionadas ao estilo de vida

| Variável Explicativa | % |
|--|------|
| Ao ingerir carne vermelha, costuma: | |
| Tirar o excesso da gordura visível | 55,5 |
| Comer com a gordura | 44,4 |
| Não responderam | 5,3 |
| Frequência com que costuma ingerir bebidas alcoólicas | |
| Não bebo nunca | 57,9 |
| Menos de uma vez por mês | 13,0 |
| Uma vez ou mais por mês | 29,0 |
| Fuma atualmente? | |
| Sim, diariamente | 15,5 |
| Sim, menos que diariamente | 1,4 |
| Não fumo atualmente | 83,0 |

- Morbidade

Observou-se que, aproximadamente 26% dos respondentes já receberam o diagnóstico de pressão alta, além disso, cerca de 8% receberam o diagnóstico de diabetes e 17% tem colesterol alto.

Tabela 4: Variáveis relacionadas à morbidade

| Variável Explicativa | % |
|---|------|
| Já recebeu o diagnóstico de pressão alta? | |
| Sim | 26,4 |
| Não | 73,6 |
| Já recebeu o diagnóstico de diabetes? | |
| Sim | 8,2 |
| Não | 83,3 |
| Não responderam | 8,5 |
| Já recebeu o diagnóstico de colesterol alto? | |
| Sim | 17,0 |
| Não | 69,5 |
| Não responderam | 13,4 |

4.3.2 Nível 2: Município

Os residentes da amostra da Região dos Pireneus participantes da Pesquisa estão distribuídos em 41 setores censitários. Ao fazer o agrupamento dessas pessoas pelos municípios, observa-se que a maior parte da população reside nos municípios de Pirenópolis e Padre Bernardo, conforme pode ser ilustrado pela figura abaixo.

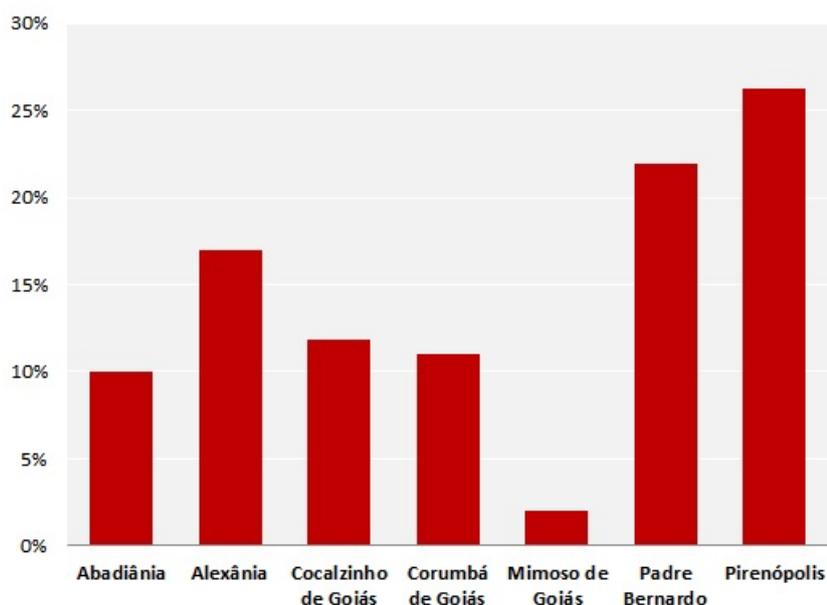


Figura 7: Distribuição por município dos moradores da Região dos Pireneus

As principais características da vizinhança obtidas são que:

- 37% afirma ter muito lixo ou entulho nas ruas da vizinhança;
- 25% acha que há muito barulho na vizinhança;
- 75% não acha que há área de lazer para praticar esportes ou atividade física;
- 50% afirma que há locais para comprar frutas e verduras;
- 48% se sente segura ao caminhar pela vizinhança de dia ou de noite;
- 11% afirma que durante os últimos seis meses frequentemente ou às vezes houve uma briga em que uma arma foi utilizada;
- 42% diz ter ocorrido roubo ou furto nos últimos seis meses;

Ao fazer essa análise por município, observou-se que:

- Os municípios de Cocalzinho de Goiás e Padre Bernardo foram os que tiveram a maior frequência na resposta afirmativa sobre a existência de muito lixo ou entulho nas ruas da vizinhança, sendo de 48,7% no município de Padre Bernardo;
- Pirenópolis foi o município que obteve a maior frequência de respostas quanto à ocorrência de muito barulho nas ruas da vizinhança (30,3%);
- O município de Mimoso de Goiás foi o que obteve a maior frequência de respostas quanto à não existência de áreas de lazer para prática de atividades físicas ou de esportes e quanto à não existência de locais próximos para compra de frutas e legumes (100% em ambos os casos);
- Abadiânia foi o município que obteve o maior percentual de respostas com relação ao sentimento de não se sentir seguro caminhando na vizinhança (62,5%), seguido por Padre Bernardo (60,7%);
- Em Alexânia, quase 20% dos respondentes dizem que ocorreu, frequentemente ou às vezes, brigas na vizinhança na qual uma arma foi utilizada;
- Em Padre Bernardo, pouco mais de 50% dos respondentes afirma a ocorrência de roubo ou furto, frequentemente ou às vezes, na vizinhança.

Embora hajam alguns municípios que tiveram destaque na frequência de respostas, notou-se pouca variabilidade entre a maioria. No banco de dados constam apenas 12 respostas de pessoas selecionadas do município de Mimoso do Goiás, dessa forma os percentuais obtidos não podem ser considerados tão significativos.

A tabela a seguir apresenta as variáveis relacionadas às características da vizinhança, por município:

Tabela 5: Variáveis relacionadas às características da vizinhança, por município

| Variável Explicativa | Abadiânia | Alexânia | Cocalzinho de Goiás | Corumbá de Goiás | Mimoso de Goiás | Padre Bernardo | Pirenópolis |
|--|-----------|----------|---------------------|------------------|-----------------|----------------|-------------|
| Há muito lixo ou entulho nas ruas da sua vizinhança? | | | | | | | |
| Sim | 43,06 | 35,83 | 46,43 | 20,25 | 28,57 | 48,73 | 28,19 |
| Não | 56,94 | 64,17 | 52,38 | 79,75 | 71,43 | 51,27 | 70,21 |
| Não responderam | 0,00 | 0,00 | 1,19 | 0,00 | 0,00 | 0,00 | 1,60 |
| Há muito barulho na sua vizinhança? | | | | | | | |
| Sim | 20,83 | 27,50 | 21,43 | 6,33 | 0,00 | 25,95 | 30,32 |
| Não | 79,17 | 72,50 | 77,38 | 93,67 | 100,00 | 74,05 | 68,09 |
| Não responderam | 0,00 | 0,00 | 1,19 | 0,00 | 0,00 | 0,00 | 1,60 |
| Há alguma área de lazer para praticar esportes ou fazer caminhada na sua vizinhança? | | | | | | | |
| Sim | 19,44 | 17,50 | 50,00 | 24,05 | 0,00 | 21,52 | 23,40 |
| Não | 80,56 | 82,50 | 48,81 | 75,95 | 100,00 | 78,48 | 75,00 |
| Não responderam | 0,00 | 0,00 | 1,19 | 0,00 | 0,00 | 0,00 | 1,60 |
| Há algum local na vizinhança para comprar frutas e legumes? | | | | | | | |
| Sim | 65,28 | 66,67 | 47,62 | 17,72 | 0,00 | 44,30 | 55,85 |
| Não | 34,72 | 33,33 | 51,19 | 82,28 | 100,00 | 55,70 | 42,55 |
| Não responderam | 0,00 | 0,00 | 1,19 | 0,00 | 0,00 | 0,00 | 1,60 |
| Se sente seguro caminhando na sua vizinhança, de dia e de noite? | | | | | | | |
| Sim | 37,50 | 43,33 | 57,14 | 56,96 | 64,29 | 39,24 | 52,13 |
| Não | 62,50 | 56,67 | 41,67 | 43,04 | 35,71 | 60,76 | 46,28 |
| Não responderam | 0,00 | 0,00 | 1,19 | 0,00 | 0,00 | 0,00 | 1,60 |
| Com que frequência aconteceu uma briga na sua vizinhança na qual uma arma tenha sido usada? | | | | | | | |
| Frequentemente | 5,56 | 9,17 | 0,00 | 1,27 | 0,00 | 6,33 | 1,06 |
| Às vezes | 8,33 | 10,00 | 10,71 | 1,27 | 0,00 | 7,59 | 5,32 |
| Raramente | 18,06 | 15,83 | 27,38 | 18,99 | 0,00 | 11,39 | 14,89 |
| Nunca | 68,06 | 65,00 | 60,71 | 78,48 | 100,00 | 74,68 | 77,13 |
| Não responderam | 0,00 | 0,00 | 1,19 | 0,00 | 0,00 | 0,00 | 1,60 |
| Com que frequência aconteceu roubo ou furto na sua vizinhança? | | | | | | | |
| Frequentemente | 23,61 | 13,33 | 17,86 | 6,33 | 0,00 | 26,58 | 17,55 |
| Às vezes | 23,61 | 23,33 | 30,95 | 27,85 | 7,14 | 24,68 | 22,34 |
| Raramente | 25,00 | 26,67 | 19,05 | 30,38 | 0,00 | 15,82 | 19,15 |
| Nunca | 27,78 | 36,67 | 30,95 | 35,44 | 92,86 | 32,91 | 39,36 |
| Não responderam | 0,00 | 0,00 | 1,19 | 0,00 | 0,00 | 0,00 | 1,60 |

4.4 Análise Bivariada

Com objetivo de identificar as possíveis variáveis preditoras do modelo de regressão logística multinível, foi analisada a associação entre o conjunto de possíveis variáveis explicativas e a variável resposta, ter diagnóstico de pressão alta, utilizando o Teste qui-quadrado de independência.

4.4.1 Nível 1: Indivíduo

– Características Sócio-Demográficas:

Tabela 6: Variáveis explicativas (relacionadas às características sócio-demográficas) e ocorrência de hipertensão arterial

| Variável Explicativa | Possui Pressão Alta (%) | χ^2 | P-valor |
|--|-------------------------|----------|---------|
| Sexo | | 2,4659 | 0,1163 |
| Feminino | 28,8 | | |
| Masculino | 23,6 | | |
| Grau de Instrução | | 56,9961 | <,0001 |
| Analfabeto/Menos de um ano de instrução | 40,1 | | |
| Elementar Incompleto | 44,4 | | |
| Elementar Completo e Fundamental Incompleto | 25,8 | | |
| Fundamental Completo e Ensino Médio Incompleto | 9,3 | | |
| Ensino Médio Completo e Superior Incompleto | 20,7 | | |
| Superior Completo ou mais | 23,5 | | |
| Ocupação | | 40,8839 | <,0001 |
| Trabalha e em atividade atualmente | 16,9 | | |
| Trabalha, mas não em atividade atualmente | 28,2 | | |
| Já trabalhou, mas não trabalha mais | 62,3 | | |
| Nunca trabalhou | 39,4 | | |

Observou-se que, entre as mulheres pesquisadas, aproximadamente, 29% possui pressão alta, enquanto cerca de 24% dos homens também recebeu esse diagnóstico. Assim, nota-se uma prevalência um pouco maior de pressão alta entre as mulheres do que entre os homens. Com relação ao grau de instrução, as categorias que tiveram maior destaque foram dos analfabetos ou que possuem menos de um ano de instrução e dos que disseram ter grau elementar incompleto. Foi visto que existem mais pessoas com pressão alta nestes graus de escolaridade do que nos demais níveis de instrução apresentados.

Quanto à ocupação, nota-se que a categoria dos que já trabalharam, mas não mais o fazem atualmente são os que possuem a maior frequência de pessoas com pressão alta (62,3%).

Com exceção de sexo, as variáveis citadas acima possuem p-valor baixo - inferior a 5% (grau de significância escolhido para análise) indicando possível associação com o diagnóstico de pressão alta.

A figura abaixo apresenta os boxplots das variáveis idade e diagnóstico de pressão alta. Nele, é possível identificar que a doença é mais frequente nos indivíduos da amostra com idade mais avançada, estando a média de idade dessas pessoas próxima dos 60 anos.

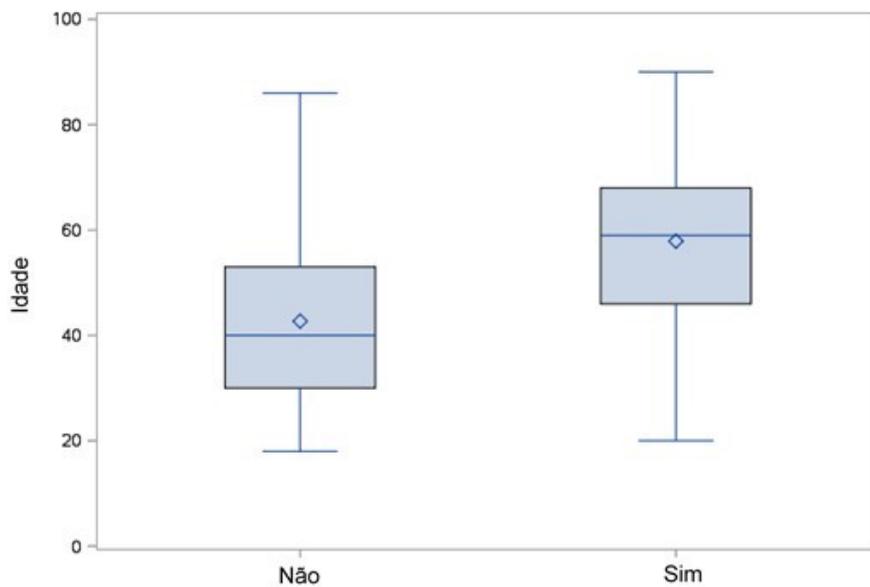


Figura 8: Box-plot: Idade x Pressão Alta

– Auto-Avaliação do Estado de Saúde

A tabela abaixo indica que somente a variável “Grau em que sente dores de cabeça ou enxaqueca” não foi significativa para ocorrência de pressão alta. Entre os que auto avaliaram sua saúde como muito boa há uma incidência baixa de pessoas com pressão alta (9,8%), enquanto os que auto avaliaram sua saúde como muito ruim, quase 73% possui pressão alta. Assim, quanto pior a avaliação da saúde, maior é a frequência de pessoas com pressão alta. Semelhante aos resultados dessa variável, entre os que sentem dores no corpo, quanto maior o nível da dor, maior é a frequência de pessoas

com pressão alta, sendo 50% a frequência de indivíduos com hipertensão que dizem sentir dores no corpo em um grau muito intenso.

Tabela 7: Variáveis Explicativas (relacionadas a auto-avaliação de saúde) e ocorrência de hipertensão arterial

| Variável Explicativa | Possui Pressão Alta (%) | χ^2 | P-valor |
|---|--------------------------------|----------|----------------|
| Auto Avaliação de Saúde | | 46,3744 | <,0001 |
| Muito boa | 9,8 | | |
| Boa | 15,3 | | |
| Regular | 64,7 | | |
| Ruim | 55,2 | | |
| Muito ruim | 72,7 | | |
| Grau em que sente dores de cabeça ou enxaqueca | | 7,0157 | 0,1351 |
| Nenhum | 26,9 | | |
| Leve | 21,7 | | |
| Média | 22,5 | | |
| Intenso | 32,0 | | |
| Muito intenso | 40,7 | | |
| Grau em que sente dores no corpo | | 46,3744 | <,0001 |
| Nenhum | 15,8 | | |
| Leve | 25,0 | | |
| Médio | 27,1 | | |
| Intenso | 47,0 | | |
| Muito intenso | 50,0 | | |

O Índice de Massa Corporal (IMC) mostra em qual faixa de peso o indivíduo se encontra de acordo com sua altura. Pela figura abaixo, nota-se que entre os que tem hipertensão o IMC é maior, quando comparado aos que não tem pressão alta.

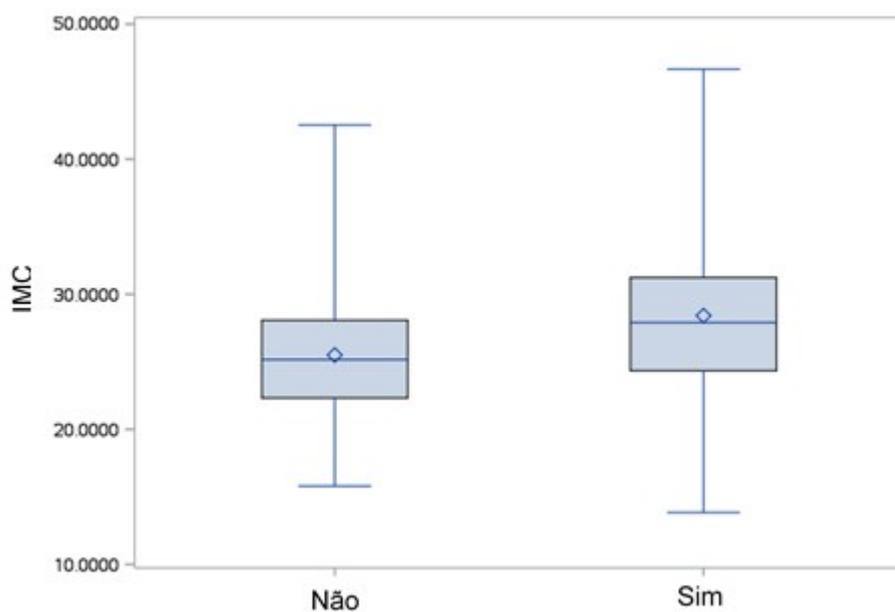


Figura 9: IMC x Pressão Alta

A tabela abaixo indica que somente a variável relacionada ao hábito de fumar não será considerada como uma possível preditora para ocorrência de hipertensão, uma vez que o resultado do p-valor foi elevado. Notou-se também que a variável relacionada ao consumo de carne vermelha, dos que tiram o excesso da gordura visível, quase 43% tem pressão alta enquanto, aproximadamente, dos que comem a carne com a gordura, “apenas” 20% tem a doença. Quanto à frequência no consumo de bebida alcoólica, quase 33% dos que dizem não beber nunca possuem hipertensão, enquanto, dos que bebem pelo menos uma vez por mês, 16,3% tem a doença.

Tabela 8: Variáveis explicativas (características relacionadas ao estilo de vida) e ocorrência de hipertensão arterial

| Variável Explicativa | Possui Pressão Alta (%) | χ^2 | P-valor |
|---|-------------------------|----------|---------|
| Costume ao ingerir carne vermelha: | | 9,8347 | 0,0017 |
| Tirar o excesso da gordura visível | 42,4 | | |
| Comer com a gordura | 19,3 | | |
| Frequência no consumo de bebida alcoólica: | | 21,3733 | <,0001 |
| Não bebo nunca | 32,8 | | |
| Menos de uma vez por mês | 20,4 | | |
| Uma vez ou mais por mês | 16,3 | | |
| Hábito de fumar: | | 0,3492 | 0,8398 |
| Sim, diariamente | 27,9 | | |
| Sim, menos que diariamente | 20,0 | | |
| Não fumo atualmente | 26,2 | | |

– Morbidade

Pode-se observar, pela tabela abaixo, que entre os que possuem pressão alta, 60% vai ao médico regularmente, enquanto 40% só vai quando há algum problema ou nunca vai ao médico.

Referente às mudanças no estilo de vida adotadas pelos hipertensos, foi identificado que:

- 50% faz dieta;
- 76% não pratica atividades físicas;
- 81% faz uso de medicamentos;

Além disso, 56% recebeu assistência médica nos últimos 6 meses e na última vez em que receberam assistência médica para hipertensão, 38,6% foram atendidos em uma unidade de saúde pública seguido por 22% que receberam atendimento em um pronto-socorro ou emergência de hospital público.

Tabela 9: Variáveis relacionadas à morbidade

| Variável explicativa | % |
|---|------|
| Ida regular ao médico por causa da hipertensão: | |
| Sim | 59,8 |
| Não, só quando tem algum problema | 30,7 |
| Nunca vai | 9,5 |
| Faz dieta por causa da hipertensão? | |
| Sim | 50,2 |
| Faz atividades físicas por causa da hipertensão? | |
| Sim | 23,3 |
| Toma medicamentos por causa da hipertensão? | |
| Sim | 81,0 |
| Última vez em que recebeu assistência médica: | |
| Há menos de 6 meses | 56,6 |
| Entre 6 meses e 1 ano | 8,4 |
| Entre 1 ano e menos de 2 anos | 9,0 |
| Entre 2 anos e menos de 3 anos | 7,4 |
| 3 anos ou mais atrás | 10,0 |
| Nunca recebeu | 8,5 |
| Onde recebeu atendimento na última vez que recebeu assistência médica? | |
| Unidade de saúde pública | 38,6 |
| UPA (Unidade de Pronto Atendimento) | 4,7 |
| Outro tipo de Ponto de Atendimento Público (24 horas) | 0,5 |
| Pronto-socorro ou emergência de hospital público | 22,2 |
| Ambulatório de hospital público | 5,3 |
| Consultório particular | 11,6 |
| Pronto-socorro ou emergência de hospital privado | 6,3 |
| Outro | 2,1 |

A tabela abaixo apresenta os resultados do diagnóstico de outras doenças e hipertensão:

Tabela 10: Variáveis explicativas (relacionadas à outras doenças) e a ocorrência de hipertensão arterial

| Variável explicativa | Possui Pressão Alta (%) | χ^2 | P-valor |
|---|-------------------------|----------|---------|
| Diabetes | | 44,0506 | <,0001 |
| Sim | 64,1 | | |
| Não | 22,6 | | |
| Colesterol Alto | | 31,3959 | <,0001 |
| Sim | 46,0 | | |
| Não | 23,7 | | |
| Doenças do coração | | 26,7945 | <,0001 |
| Sim | 56,6 | | |
| Não | 24,0 | | |
| Já recebeu o diagnóstico de AVC? | | 12,4712 | 0,0004 |
| Sim | 69,0 | | |
| Não | 25,6 | | |
| Já recebeu o diagnóstico de asma? | | 1,4300 | 0,2318 |
| Sim | 33,3 | | |
| Não | 25,8 | | |
| Já recebeu o diagnóstico de artrite? | | 6,5921 | 0,0102 |
| Sim | 46,6 | | |
| Não | 25,5 | | |
| Já recebeu o diagnóstico de depressão? | | 6,4922 | 0,0108 |
| Sim | 37,2 | | |
| Não | 24,8 | | |

Assim, a única variável a ser retirada será relacionada ao diagnóstico de asma, devido ao seu alto p-valor, enquanto as demais apresentadas permanecerão. Observou-se que entre os que possuem cada uma das doenças listadas, há uma frequência elevada de pessoas que também foram diagnosticadas com pressão alta. Entre os que possuem diabetes, por exemplo, 64% também possui pressão alta.

4.4.2 Nível 2: Município

A tabela abaixo mostra a ocorrência de pressão alta da amostra entre os municípios da Região dos Pireneus. Nota-se que a maior incidência de pressão alta ocorre nos municípios de Padre Bernardo e Pirenópolis, representando quase 60% dos indivíduos pesquisados que possuem pressão alta nessa região.

Tabela 11: Variável explicativa (relacionadas ao município) e ocorrência de hipertensão arterial

| Municípios | Sim | χ^2 | P-valor |
|---------------------|------------|----------|----------------|
| Abadiânia | 23,6 | 4,6958 | 0,5834 |
| Alexânia | 25,0 | | |
| Cocalzinho de Goiás | 29,7 | | |
| Corumbá de Goiás | 29,1 | | |
| Mimoso de Goiás | 21,4 | | |
| Padre Bernardo | 31,0 | | |
| Pirenópolis | 28,4 | | |

Entre as principais variáveis relacionadas ao ambiente e ocorrência de hipertensão arterial, decidiu-se ser mais liberal nessa primeira análise (ver Hosmer e Lemeshow, 1989), por isso foram consideradas aquelas com p-valor abaixo de 0,25 como pode ser visto na tabela a seguir. No entanto, como houve problemas na estimação do nível de significância da maior parte das variáveis, todas da tabela abaixo foram testadas no modelo, com exceção da relacionada à ocorrência de briga entre gangues, que não teve problemas na estimação e pode-se observar que ela não foi significativa.

A partir de cada variável relacionada ao ambiente (nível do município) foi criada uma nova variável, informando a frequência, por município, referente à resposta “Sim”. Dessa forma, o percentual dessas respostas foi atribuído a cada indivíduo, de acordo com seu município. Com isso, desejava-se observar a variabilidade das respostas relacionadas ao ambiente, entre os municípios, para se identificar se essas variáveis associadas ao local em que o indivíduo vive podem estar associadas à ocorrência de pressão alta.

Tabela 12: Variáveis explicativas (relacionadas às características da vizinhança) e ocorrência de hipertensão arterial

| Variável explicativa | Possui Pressão Alta (%) | χ^2 | P-valor |
|--|-------------------------|------------|---------|
| Há muito lixo ou entulho nas ruas da vizinhança? | | (1) 2,2682 | 0,3217 |
| Sim | 23,2 | | |
| Não | 28,3 | | |
| Há muito barulho na sua vizinhança? | | (1) 0,1178 | 0,9428 |
| Sim | 25,4 | | |
| Não | 26,7 | | |
| Há áreas de lazer para praticar esportes ou fazer caminhadas? | | (1) 0,3584 | 0,8359 |
| Sim | 24,7 | | |
| Não | 27,0 | | |
| Se sente seguro ao caminhar na sua vizinhança? | | (1) 3,4057 | 0,1822 |
| Sim | 29,6 | | |
| Não | 23,5 | | |
| Ocorrência de brigas com o uso de armas? | | (1) 0,2665 | 0,9919 |
| Frequentemente | 25,0 | | |
| Às vezes | 24,0 | | |
| Raramente | 27,6 | | |
| Nunca | 26,5 | | |
| Ocorrência de roubo ou furto? | | (1) 6,0341 | 0,1966 |
| Frequentemente | 22,6 | | |
| Às vezes | 25,7 | | |
| Raramente | 21,8 | | |
| Nunca | 31,5 | | |

(1) Os resultados dos testes χ^2 podem não ser válidos, pois 30% das células tiveram o número de observações esperados menor do que 5.

Nota-se pouca variabilidade entre as categorias apresentadas, o que pode justificar a não significância dos p-valores. Aparentemente, não há evidências de que essas variáveis relacionadas ao município tem associação com o diagnóstico de pressão alta, uma vez que as frequências dos indivíduos que possuem a doença são similares entre as categorias.

4.5 Modelagem

Considerando as possíveis variáveis explicativas apresentadas na análise bivariada, foi realizada a modelagem cuja variável resposta é o diagnóstico de hipertensão arterial de um indivíduo residente na Região dos Pireneus que foi pesquisado. Assim, neste estudo, a categoria de sucesso é o indivíduo ter pressão alta. Para a modelagem, as mulheres que responderam que só tiveram pressão alta durante a gravidez foram analisadas como não tendo o diagnóstico da doença.

Foram seguidos os passos propostos por Hox (2010) na construção do modelo de regressão logística multinível, por isso, inicialmente foi ajustado o modelo nulo para analisar o coeficiente de correlação intra-classe e realizado o teste de homogeneidade das variâncias entre os municípios.

– Passo 1: Modelo com apenas o intercepto:

Tabela 13: Modelo Nulo: Apenas com o intercepto

| Modelo | Valor | Erro padrão | P-valor |
|------------------|---------|-------------|---------|
| Intercepto | -1,0898 | 0,09866 | <,0001 |
| $\sigma_{u_0}^2$ | 0,01059 | 0,03113 | 0,3668 |
| ρ | 0,3215% | - | - |
| Homogeneidade | 0,66 | - | 0,9954 |

O resíduo no nível do município tem (u_0) tem distribuição normal com média 0 e variância de 0,01059, o que implicou no coeficiente de correlação intra-classe extremamente baixo, indicando que 0,3215% da ocorrência de pressão alta nos indivíduos da Região dos Pireneus pode ser atribuída ao nível do município. Já o resíduo no nível do indivíduo tem, por construção, média zero e variância 1. Para verificar a significância de $\sigma_{u_0}^2$ foi realizado o teste de Wald, com aproximação de Satterthwaite e o p-valor obtido foi de 0,3668 sendo não considerado como significativo. Além disso, o teste de homogeneidade das variâncias indica que não se rejeita a hipótese de igualdade de variâncias em cada município, assim, não há evidências de que o modelo multinível se faz necessário nesse estudo, devido à aparente falta de heterogeneidade entre os municípios da Região dos Pireneus.

Nesse caso, a solução mais adequada a ser dada é a utilização de um modelo de regressão logística (tradicional) uma vez que não se faz necessário o uso do modelo multinível. No entanto, como o objetivo desse trabalho é a apresentação da técnica da regressão multinível e sua aplicação a um caso real, será dada a continuidade na construção do Modelo Multinível e posteriormente, será apresentada uma solução utilizando o modelo de regressão logística. É importante mencionar também a *deviance* obtida nessa construção inicial do modelo. Ela representa uma medida do grau de ajuste do modelo e que pode ser usada para comparar modelos. Então, quanto menor a *deviance*, maior o ajuste do modelo. No caso do modelo nulo definido acima, tem-se que a *deviance* é igual a 3054,89.

– Passo 2: Modelo com as variáveis explicativas no menor nível (indivíduo)

Nessa etapa são incluídas as variáveis explicativas fixas no nível do indivíduo. Para a modelagem foram testadas todas as variáveis que tiveram p-valor até 0,25 na análise bivariada e o modelo final conta com as variáveis que foram significativas, isto é, com p-valor até 0,05. A categoria “Boa”, da variável “Auto avaliação da saúde” foi retirada do modelo por ter apresentado p-valor elevado. Assim, após o processo de seleção das variáveis explicativas, o intercepto obtido foi de -7,8128 com erro padrão associado de 0,9609.

A tabela abaixo apresenta os resultados obtidos para o modelo final no nível do indivíduo.

Tabela 14: Modelo final com as variáveis explicativas no menor nível

| Variável explicativa | $\hat{\beta}$ | Erro Padrão | P-valor |
|--|---------------|-------------|---------|
| Idade | 0,05776 | 0,008453 | <,0001 |
| Auto avaliação da saúde | | | |
| Muito boa | - | - | - |
| Regular | 0,9590 | 0,4920 | 0,05 |
| Ruim ou muito ruim | 1,8926 | 0,6046 | 0,0018 |
| IMC | 0,1041 | 0,02465 | <,0001 |
| Ao comer carne vermelha, costuma: | | | |
| Tirar o excesso de gordura visível | 0,5523 | 0,2565 | 0,0318 |
| Comer com a gordura | - | - | - |
| Diagnóstico de diabetes | | | |
| Sim | 0,8086 | 0,3985 | 0,0430 |
| Não | - | - | - |

O *deviance* obtido desse modelo, com as variáveis citadas acima, foi de 422,52 indicando que houve um ajuste significativo quando comparado ao modelo nulo apresentado no Passo 1.

– Passo 3: Modelo com as variáveis explicativas no maior nível (município)

Nessa etapa são incluídas as variáveis explicativas fixas no nível do município. Dessa forma, foram testadas as variáveis mencionadas na análise bivariada no nível do município. Tais variáveis são referentes a: quantidade de lixo nas ruas, barulho na vizinhança, área de lazer na vizinhança, se existe local para comprar frutas e verduras próximo, segurança ao andar de dia ou de noite, ocorrência de roubo ou furto e briga entre gangues. No entanto, nenhuma dessas variáveis teve coeficientes significativos para o modelo, apesar disso, verificou-se entre elas qual teria maior inclinação à significância (as que tiveram o menor p-valor) sendo selecionadas as variáveis relacionadas a existência de muito lixo e barulho na vizinhança. Dessa forma, optou-se por incluir essas variáveis para que tivessem variáveis no nível do município, no modelo que explica o diagnóstico de hipertensão arterial. O *deviance* obtido com a inclusão delas foi de 390,98, havendo um pequena melhora no ajuste do modelo quando comparado ao modelo apresentado no Passo 2.

A tabela abaixo apresenta os resultados obtidos para o modelo com a adição das variáveis explicativas relacionadas a ocorrência de lixo e barulho na vizinhança. O intercepto resultante foi de -8,0246 com erro padrão associado de 1,1209.

Tabela 15: Modelo final com as variáveis explicativas no menor e no maior nível

| Variável explicativa | $\hat{\beta}$ | Erro Padrão | P-valor |
|--|---------------|-------------|---------|
| Idade | 0,05770 | 0,008866 | <,0001 |
| Auto avaliação da saúde | | | |
| Muito boa | - | - | - |
| Regular | 1,1076 | 0,5319 | 0,0378 |
| Ruim ou muito ruim | 1,8236 | 0,6441 | 0,0048 |
| IMC | 0,1059 | 0,02580 | <,0001 |
| Ao comer carne vermelha, costuma: | | | |
| Tirar o excesso da gordura visível | 0,5644 | 0,2705 | 0,0375 |
| Comer com a gordura | - | - | - |
| Diagnóstico de diabetes | | | |
| Sim | 0,7805 | 0,4073 | 0,05 |
| Não | - | - | - |
| Há muito lixo na sua vizinhança? | | | |
| Sim | 1,8745 | 1,3719 | 0,1725 |
| Não | - | - | - |
| Há muito barulho na sua vizinhança? | | | |
| Sim | -2,4797 | 1,8427 | 0,1790 |
| Não | - | - | - |

É importante mencionar que os modelos apresentados nos passos 2 e 3 são chamados modelos de componentes de variância, por decomporem a variância do intercepto em componentes distintos da variância para cada nível hierárquico. Nesses modelos, considera-se que o intercepto varia entre os municípios, ou seja, esperava-se que a variância do intercepto mostrasse que a ocorrência média de hipertensão arterial não é igual em todos os municípios, considerando os coeficientes de regressão fixos. No entanto, a variância $\sigma_{u_{0j}}^2$ estimada nos passos 2 e 3 foram iguais a zero, indicando homogeneidade entre os municípios. Dessa forma, não seria necessário o próximo passo referente à adição das variáveis de interação entre os níveis.

Portanto, nem a variável referente a existência de muito lixo ou entulho na vizinhança, nem a variável relativa ao barulho na vizinhança poderão ser consideradas no modelo por não terem sido significativas e com isso, o modelo que se torna mais adequado para explicar a ocorrência de pressão alta é o apresentado no passo anterior (Passo 2), em que só existem as variáveis significativas no nível do indivíduo (o que já era esperado com base no coeficiente de correlação obtido no modelo nulo). Dessa forma, o modelo final com as variáveis explicativas no nível do indivíduo é um modelo logístico, por não ter variáveis de outros níveis em sua composição.

A tabela abaixo apresenta os resultados obtidos para o modelo final (apenas com variáveis no nível do indivíduo).

Tabela 16: Modelo final

| Variável explicativa | $\hat{\beta}$ | Erro Padrão | P-valor |
|--|---------------|-------------|---------|
| Idade | 0,05776 | 0,008453 | <,0001 |
| Auto avaliação da saúde | | | |
| Muito boa | - | - | - |
| Regular | 0,9590 | 0,4920 | 0,05 |
| Ruim ou muito ruim | 1,8926 | 0,6046 | 0.0018 |
| IMC | 0,1041 | 0,02465 | <,0001 |
| Ao comer carne vermelha, costuma: | | | |
| Tirar o excesso de gordura visível | 0,5523 | 0,2565 | 0,0318 |
| Comer com a gordura | - | - | - |
| Diagnóstico de diabetes | | | |
| Sim | 0,8086 | 0,3985 | 0,0430 |
| Não | - | - | - |

Antes de interpretar os parâmetros do modelo e a razão de chances com seus respectivos intervalos de confiança, é necessário fazer o diagnóstico dos resíduos para verificar os pressupostos do modelo.

4.6 Diagnóstico dos Resíduos

A análise dos resíduos é necessária para verificação dos pressupostos do modelo e valores discrepantes e influentes. Como o interesse do trabalho é a regressão logística multinível, foi realizado o diagnóstico dos resíduos no nível do município, em que notou-se que o pressuposto de normalidade para o efeito aleatório dos municípios pode ser facilmente verificado, assim como a existência de valores discrepantes. Porém não se verificou se esses valores discrepantes são influentes nas variâncias ou nas estimativas dos coeficientes do modelo.

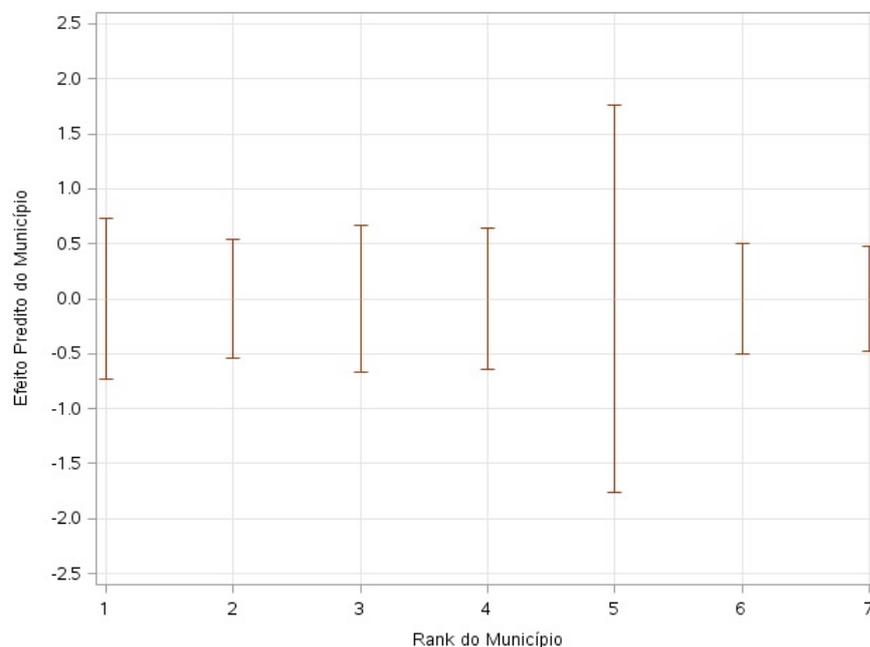


Figura 10: Resíduos do Nível do Município

A figura acima apresenta o efeito predito de cada município pelo seu rank, isto é, ordenou-se do menor para o maior os efeitos preditos de cada município, com barras do erro de predição. O intervalo que intercepta o zero, mostra que a ocorrência de pressão alta daquele município não é significativamente diferente do resultado geral dos municípios. Se o intervalo de confiança fosse inteiramente abaixo da linha pontilhada, implica que o diagnóstico de pressão alta é significativamente menor para esse município, já se o intervalo de confiança fosse totalmente acima da linha pontilhada, o diagnóstico de pressão alta seria significativamente maior para aquele município. Conclui-se, por-

tanto, que não há diferença significativa entre os municípios na ocorrência de hipertensão arterial.

Para verificar o pressuposto de normalidade do efeito aleatório do município, basta verificar se os efeitos preditos padronizados destes são aproximadamente normais. A figura abaixo apresenta o gráfico quantil-quantil, em relação aos quantis teóricos da distribuição normal, para essa medida.

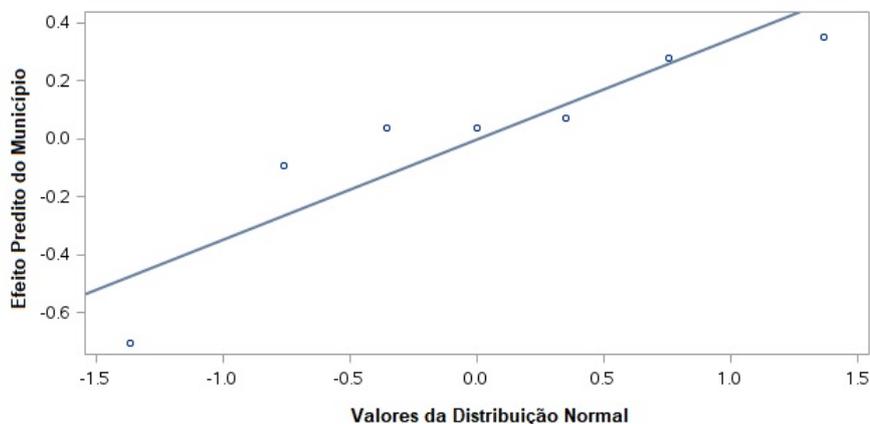


Figura 11: Resíduos Padronizados do Nível do Município em relação aos respectivos valores da distribuição normal

Os resíduos deveriam estar distribuídos ao longo de uma linha reta, mas apesar de isso não ocorrer, o que certamente se deve à existência de poucas observações (sete municípios) é possível observar que a violação ocorre mais intensamente nas primeiras observações e não chega a inviabilizar a utilização do modelo escolhido anteriormente.

4.7 Interpretação do modelo

Tabela 17: Razão de chances e intervalos de confiança

| Variável explicativa | Razão de Chances | Intervalo de Confiança |
|--|------------------|------------------------|
| Idade | 1,059 | (1,042; 1,077) |
| Auto avaliação da saúde | | |
| Muito boa | - | - |
| Regular | 2,609 | (0,992; 6,859) |
| Ruim ou muito ruim | 6,637 | (2,023; 21,767) |
| IMC | 1,110 | (1,057; 1,165) |
| Ao comer carne vermelha, costuma: | | |
| Tirar o excesso da gordura visível | 1,737 | (1,050; 2,876) |
| Comer com a gordura | - | - |
| Diagnóstico de diabetes | | |
| Sim | 2,245 | (1,026; 4,911) |
| Não | - | - |

Os resultados indicam que as variáveis idade e IMC foram altamente significativas enquanto as demais também foram, mas nem tanto quanto elas. A razão de chances dessas variáveis mostra que:

- Idade: A chance de um indivíduo ter pressão alta aumenta 5,9% para cada aumento de um ano na sua idade (mantidas constantes as demais variáveis);
- IMC: A chance de um indivíduo ter pressão alta aumenta 11% para cada aumento de uma unidade no seu IMC (mantidas constantes as demais variáveis).

A categoria de referência para a variável relacionada à auto avaliação do estado de saúde é “Muito boa” e foram agrupadas as categorias “Ruim” e “Muito Ruim” em uma só, para uma melhor estimação dos coeficientes. Assim, a chance de um indivíduo que avaliou a sua saúde como ruim ou muito ruim ter o diagnóstico de pressão alta é 6,6 vezes superior quando comparada aos que auto avaliaram o seu estado de saúde como muito bom e os que avaliaram seu estado de saúde como regular tem 2,6 vezes mais chance de ter pressão alta, quando comparados com os que auto avaliaram o seu estado de saúde como muito bom. Já a categoria de referência quanto a variável relacionada à ingestão de carne vermelha, é a “Comer com a gordura”. Observa-se que os que tiram o excesso de gordura visível da carne possuem 73% mais chance de ter pressão alta do

que os que não tiram. Esse resultado chamou a atenção por ser “diferente” do que se espera, no entanto, foi verificado que 59% das pessoas com pressão alta tiram o excesso da gordura visível da carne vermelha, enquanto os demais 41% que tem hipertensão, não retiram. Essa informação parece demonstrar que aqueles que tem a consciência de que tem essa doença já tomam um cuidado maior com a sua alimentação, retirando o excesso da gordura visível da carne e por isso apresentam “mais chance de ter pressão alta”, enquanto os que comem com a gordura tem “menos chance”, uma vez que (ainda) não tem a doença. Nota-se também que aqueles que foram diagnosticados com diabetes possuem 2,2 vezes mais chances de receber o diagnóstico de pressão alta do que os que não tem diabetes.

5 Conclusão

Com este Relatório foi possível perceber que a abordagem multinível, considerando o segundo nível os municípios, não foi adequada para estudar os fatores associados no diagnóstico de pressão alta dos indivíduos pesquisados que vivem na Região dos Pireneus. O baixíssimo coeficiente de correlação intraclasse obtido no modelo nulo já indicou a baixa variabilidade entre os municípios quanto à ocorrência de pressão alta.

O modelo final que explica o diagnóstico de pressão alta na Região dos Pireneus está associado apenas a características do indivíduo, ao seu estilo de vida, alimentação e a forma como ele auto avalia sua saúde. O modelo final (apresentado no Passo 2) indicou que quanto maior a idade, maior a chance do indivíduo ter pressão alta, bem como o aumento no seu IMC, também implica num aumento na chance dele desenvolver essa doença. A forma como as pessoas auto avaliam sua saúde é um indicativo de problemas de saúde que elas têm e que também possam vir a desenvolver. Viu-se que aqueles que autoavaliaram sua saúde como ruim ou muito ruim tem 6,6 vezes mais chance de ter hipertensão arterial do que aqueles que autoavaliaram sua saúde como muito boa. Neste caso, certamente muitos que possuem a doença autoavaliaram seu estado de saúde como ruim ou muito ruim justamente pelo fato de possuírem pressão alta. A variável que chamou a atenção nesse estudo foi referente ao comportamento ao comer carne vermelha, em que foi visto que os que tiram a gordura da carne possuem 73,7% mais chance de ter pressão alta quando comparados aos que comem a carne com a gordura. Conforme dito anteriormente, uma possível explicação para isso seria o fato do indivíduo que já tem a consciência de que tem pressão alta, toma um cuidado maior em sua alimentação, enquanto os que não tem (ou não sabem que tem a doença), não se preocupam e ingerem carne com a gordura.

Algumas limitações nesse estudo podem ser referentes ao número de observações pequeno (em relação ao número de variáveis estudadas) e que se reduziu ainda mais com o tratamento dos dados para realização da análise estatística e por não ter sido realizado um estudo de variáveis relacionadas à características sócio-econômicas e de renda. Estudos posteriores, incluindo outras variáveis sócios-ambientais, ainda em fase de con-

clusão no processo de coleta e crítica dos dados na Pesquisa Ambiente & Saúde, poderão evidenciar os fatores associados ao ambiente no diagnóstico de pressão alta.

6 Referências Bibliográficas

BATISTA. G.M.. Fatores associados ao desempenho de alunos do 3º ano do ensino médio do Distrito Federal : uma análise multinível aplicada a dados do Saeb 2013 [**Trabalho de conclusão de curso**]. Brasília, Distrito Federal: Universidade de Brasília, 2016.

DIEX-HOUX A.V.. Neighborhoods and health: where are we and were do we go from here? **Revue d'Épidémiologie et de Santé Publique**. 2007; 55: 13-21.

FUCHS FD.. **Hipertensão arterial sistêmica**. In: Duncan BB, Schmidt MI, Giugliani ERJ, et al. Medicina ambulatorial: condutas de atenção primária baseada em evidências. Porto Alegre: Artmed; 2004. p.641-56.

GUTIERREZ, M.; PEREIRA, E.; STEINKE, V.. **Ambiente e saúde: uma abordagem multidisciplinar**. Goiânia: Gráfica UFG, 2017.

HOX, J. J.. **Multilevel analysis: techniques and applications**. Second Edition. Great Britain: Routledge, 2010.

LAROS, J.; MARCIANO, J.. Análise multinível aplicada aos dados do NELS:88. **Estudos em avaliação educacional**. Brasília, Distrito Federal; 19 (40): 263-278, maio 2008.

LAROS, J.; PUENTE-PALACIOS, K.. Análise multinível: contribuições para estudos sobre efeito do contexto social no comportamento individual. **Estudos de psicologia**. Campinas; 26 (3): 349-361, julho - setembro 2009.

MARTINS, A.; Rocha, M.. Regressão Logística Multinível: Uma aplicação de Modelos

Lineares Generalizados Mistos [**Trabalho de conclusão de curso**]. Brasília, Distrito Federal: Universidade de Brasília, 2014.

TEIXEIRA, A. L.. Fatores associados ao desempenho escolar em Matemática do SAEB 2013 para alunos do 3º ano do Ensino Médio no Distrito Federal: Uma Aplicação de Regressão Multinível [**Trabalho de conclusão de curso**]. Brasília, Distrito Federal: Universidade de Brasília, 2015.

The sixth report of the Joint National Committee on Prevention, detection evaluation and treatment of high blood pressure. **Archives of Internal Medicine** **1997**;157:2413-46.

Veterans Administration Coopertative Study Group on Antyhipertensive Agents. **Effects of treatment on morbidity in hipertension: results in patients with diastolic blood pressure averagings 115 through 129 mmHg.** **JAMA** 1997;202:1028-34.