



Universidade de Brasília  
Departamento de Estatística

**Associação entre Saúde Cardiovascular e Determinantes Socioeconômicos  
por Situação de Domicílio no Brasil: Uma Aplicação de Modelos de  
Regressão Logística Multinomial**

**Isabela Harumi Lopes Motoki**

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

**Brasília  
2023**

**Isabela Harumi Lopes Motoki**

**Associação entre Saúde Cardiovascular e Determinantes Socioeconômicos  
por Situação de Domicílio no Brasil: Uma Aplicação de Modelos de  
Regressão Logística Multinomial**

Orientador(a): Profa. Joanlise Marco de Leon Andrade

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

**Brasília  
2023**



# Resumo

**Objetivo:** Avaliar a associação entre saúde cardiovascular e determinantes socioeconômicos por situação de domicílio. **Metodologia:** O escore de SCV proposto pela Associação Americana do Coração incluiu 4 metas comportamentais (tabagismo, índice de massa corpórea, exercício e dieta) e 3 metas biológicas (colesterol, pressão arterial e glicose). Prevalências (e intervalos de 95% de confiança) de SCV ideal e suas metas individuais foram estimadas utilizando expansão da amostra com base nos dados da Pesquisa Nacional da Saúde de 2019. Associações entre SCV e determinantes socioeconômicos (Escolaridade, Índice de riqueza e status de ocupação) foram avaliadas por modelos de regressão logística multinomial, ajustando-se por variáveis sociodemográficas. **Resultados:** A prevalência de SCV Ideal (as 7 metas concomitantes) foi de apenas 0,6% (IC<sub>95%</sub>: 0,5 – 0,7) em áreas urbanas e ainda menor em áreas rurais (0,2%, IC<sub>95%</sub>: 0,1 – 0,3). Escolaridade, riqueza e ocupação, além das covariáveis faixa etária, estado civil, presença de doenças crônicas, raça e região apresentaram associações significantes com SCV para o modelo incluindo residentes de áreas urbanas. Para residentes de áreas rurais, os efeitos foram nas mesmas direções porém atenuados e as variáveis riqueza e estado civil não foram significantes. **Conclusão:** Fortes associações entre categorias de SCV e determinantes socioeconômicos e demográficos foram observadas em ambas as situações de domicílio. Devido à baixa prevalência de SCV ideal encontrada, sobretudo em domicílios rurais, é de suma importância a atenção e acompanhamento constante da SCV na população para a prevenção de doenças cardiovasculares e melhor qualidade de vida.

**Palavras-chave:** Estudos Populacionais em Saúde Pública; Pesquisa Nacional de Saúde (PNS) 2019; Escore de Saúde cardiovascular; Regressão logística multinomial; Situação de Domicílio.

# Abstract

**Objective:** Evaluate the association between cardiovascular health and socioeconomic determinants by urban-rural classification. **Methods:** The cardiovascular health score(CVS) proposed by the American Heart Association included 4 behavioral metrics (smoking, body mass index, exercise and diet) and 3 biological metrics (cholesterol, blood pressure and glucose). Prevalences (and 95% confidence intervals) of ideal CVS and their individual metrics were estimated using sample expansion. Associations between CVS and socioeconomic variables (Education, Wealth Index and occupation status) were assessed by multinomial logistic regression models, adjusting for sociodemographic variables based on National Health Survey 2019. **Results:** The prevalence of ideal CVS condition (the 7 concurrent metrics) was only 0,6% (CI<sub>95%</sub>: 0,5 – 0,7) in urban areas and even lower in rural areas (0,2%, CI<sub>95%</sub>: 0,1 – 0,3). Education level, wealth and occupation, in addition to the covariates age group, marital status, chronic diseases status, race and region were significantly associated with CVS for the urban houses model. As for residents of rural areas, the effects were in the same directions but attenuated, and wealth and marital status were not significant. **Conclusions:** Strong associations between CVS categories and socioeconomic and demographic variables were observed in both urban and rural areas. Due to the low prevalence obtained, especially in rural households, constant attention and monitoring of CVS in the population is extremely important in cardiovascular disease prevention and as well as quality of life improvement.

**Keywords:** Public Health Population-based studies, National Health Survey 2019, Ideal cardiovascular health, Life's Simple 7th, Multinomial logistic regression. Urban-rural classification.

# Lista de Figuras e Tabelas

## Lista de Figuras

- 1 Distribuição do Escore de SCV<sup>1</sup> com respectivos Intervalos de 95% de Confiança<sup>2</sup>, estimados para a população adulta brasileira em 2019 (Expansão da amostra da PNS 2019). . . . . 23

## Lista de Tabelas

- 1 Metas de saúde cardiovascular utilizadas na PNS 2019 . . . . . 11
- 2 Prevalências estimadas (e Intervalos de 95% de confiança) de categorias de variáveis sociodemográficas totais e por situação de domicílio - expansão da amostra PNS 2019 . . . . . 21
- 3 Prevalências estimadas (e Intervalos de 95% de confiança) de variáveis sociodemográficas por categorias de saúde cardiovascular - expansão da amostra PNS 2019 . . . . . 24
- 4 Razões de chances, IC95% e p-valores de modelos de regressão multinomial para a variável resposta SCV ideal em 3 categorias (inferior 0-2, intermediária 3-5 e superior ou ideal 6-7) de domicílios Urbanos – Expansão da amostra PNS 2019. . . . . 26
- 5 Razões de chances, IC95% e p-valores de modelos de regressão multinomial para a variável resposta SCV ideal em 3 categorias (inferior 0-2, intermediária 3-5 e superior ou ideal 6-7) de domicílios Rurais – Expansão da amostra PNS 2019. . . . . 27
- 6 Variáveis Finais para construção do escore de riqueza . . . . . 34
- 7 Razões de chances, IC95% e p-valores de modelo de regressão multinomial para a variável resposta SCV em 3 categorias em domicílios Rurais sem as variáveis estado civil e índice de riqueza – Expansão da amostra PNS 2019 34

## Lista de abreviaturas e siglas

DCV	Doenças Cardiovasculares
SCV	Saúde Cardiovascular
AHA	American Heart Association
PNS	Pesquisa Nacional de Saúde
IBGE	Instituto Brasileiro de Geografia e Estatística
UPA	unidades Primárias de Amostragem
n	tamanho de uma amostra
IMC	Índice de Massa Corporal
IC <sub>95%</sub>	Intervalo de Confiança de 95%
RC	Razão de Chances
p-valor	nível descritivo de um teste de hipóteses
log	Logaritmo
exp	Exponencial
MV	Máxima Verossimilhança
MPV	Máxima Pseudo-Verossimilhança
EFI	Ensino Fundamental Incompleto
EFC	Ensino Fundamental Completo
EMI	Ensino Médio Incompleto
EMC	Ensino Médio Completo
ESI	Ensino Superior Incompleto
ESC	Ensino Superior Completo

## Sumário

<b>1 Introdução</b> . . . . .	8
<b>2 Metodologia</b> . . . . .	10
2.1 Conjunto de dados . . . . .	10
2.1.1 Escore de Saúde Cardiovascular Ideal . . . . .	10
2.1.2 Fatores Socioeconômicos . . . . .	11
2.2 Análises Estatísticas . . . . .	12
2.2.1 Análise Fatorial - Método da Componente Principal . . . . .	12
2.2.2 Análise de Regressão Logística Binária e Multinomial . . . . .	14
2.2.3 Regressão Logística para amostragem complexa . . . . .	18
2.2.4 Implementação . . . . .	20
<b>3 Resultados</b> . . . . .	21
<b>4 Discussão</b> . . . . .	28
<b>5 Conclusão</b> . . . . .	31
<b>6 Referências</b> . . . . .	32
<b>7 Apêndice</b> . . . . .	34
7.1 Resultados Índice de Riqueza . . . . .	34
7.2 Resultado modelo final para áreas rurais. . . . .	34
7.3 Códigos em R. . . . .	35
7.4 Códigos em SAS . . . . .	53



# 1 Introdução

As doenças cardiovasculares (DCV) constituem a principal causa de morte no mundo (ROTH et al., 2020) e uma das principais causas para a redução da qualidade de vida da população (ROTH et al., 2018). Segundo a Sociedade Brasileira de Cardiologia (SBC), aproximadamente 14 milhões de brasileiros apresentam alguma DCV e cerca de 400 mil morrem por ano em decorrência dessas enfermidades, o que corresponde a 30% de todas as mortes no país (SBC, 2020). O aumento na incidência de DCV observado nos últimos anos, foi provavelmente impulsionado por uma combinação de mudanças demográficas como o envelhecimento da população e a mudança na exposição a fatores de risco decorrentes da urbanização como dieta pouco saudável, tabagismo, sedentarismo, estresse, hipertensão, entre outros fatores (MORAN et al., 2014).

De acordo com um estudo de 1999 a 2017 nos Estados Unidos, as taxas de mortalidade de DCV são ainda maiores quando medidas em áreas rurais que em áreas urbanas (CROSS et al., 2020). A dificuldade de acesso aos serviços de saúde (ANDRADE et al., 2013), a reduzida escolaridade e o baixo nível socioeconômico dos trabalhadores rurais torna esta população ainda mais vulnerável ao risco cardiometabólico, principalmente porque sabe-se que a baixa posição socioeconômica pode estar associada a uma maior prevalência de fatores de risco para DCVs e a uma maior incidência e mortalidade resultante destas doenças (SCHULTZ et al., 2018). As diferenças entre as áreas rurais e urbanas refletem no desenvolvimento socioeconômico e contribuem para a desigualdade no que se refere aos acessos a itens básicos para qualidade de vida do indivíduos (ARRUDA; MAIA; ALVES, 2018).

Com o objetivo de reduzir as taxas de mortalidade por DCV, a American Heart Association (AHA) estabeleceu o conceito de saúde cardiovascular ideal (SCI) para auxiliar na implementação de políticas de saúde pública e direcionar os esforços para melhorar a saúde cardiovascular da população (LLOYD-JONES et al., 2010). Para tanto, criou-se o escore de saúde cardiovascular (SCV), que é composto por 4 metas de comportamentos saudáveis (ausência de fumo, índice de massa corporal  $< 25\text{kg}/\text{m}^2$ , atividade física regular e alimentação saudável) e 3 metas de fatores saudáveis (níveis de colesterol, pressão arterial e nível de glicose dentro da normalidade). Desse modo, define-se o Escore de SCV como a soma das metas alcançadas, que varia de 0 a 7, sendo o valor 7 considerado como SCV ideal.

Resultados de estudos com a Pesquisa Nacional de Saúde de 2019 mostram que a prevalência de pessoas com SCI é de 0,5% entre os brasileiros (MOTTA et al., 2022). Embora fatores de risco tradicionais para DCVs tenham sido identificados, como por exemplo hipertensão e dislipidemia, a contribuição dos determinantes sociais de saúde na

incidência de DCVs é pouco explorada, sobretudo em países de baixa e média renda como o Brasil. Determinantes sociais, econômicos, culturais, étnicos/raciais, psicológicos e comportamentais influenciam a ocorrência de problemas de saúde e seus fatores de risco na população (HAVRANEK et al., 2015). Entre estudos que avaliam o efeito de determinantes sociais na saúde cardiovascular (realizado em sua maioria em países de alta renda), os resultados mais consistentes são do indicador educação, onde o baixo nível de escolaridade está associado com maior prevalência de fatores de risco para DCVs. Apesar da baixa prevalência de SCI e da grande desigualdade social no Brasil, não se conhece até o momento os resultados da associação entre as metas de saúde cardiovascular e determinantes socioeconômicos separadas por situação de domicílio na população brasileira.

O presente trabalho teve como objetivo estimar a prevalência de SCI e avaliar a associação entre os fatores socioeconômicos e os escores de saúde cardiovascular ideal separadamente por situação de domicílio (rural e urbana) utilizando os dados da Pesquisa Nacional de Saúde de 2019 (PNS). Para isso foi escolhido utilizar análise de regressão logística multinomial por se tratar de um modelo que permite melhores interpretações dos parâmetros e próprio para avaliar uma variável categórica escore de saúde cardiovascular.

## 2 Metodologia

### 2.1 Conjunto de dados

Para esse estudo foi utilizada a amostra da Pesquisa Nacional da Saúde (PNS) do ano de 2019 realizado pelo IBGE e pelo Ministério da Saúde. O questionário coleta informações sobre o desempenho do sistema de saúde nacional com perguntas relacionadas ao uso e acesso aos serviços de saúde, assim como o estilo de vida, a presença de doenças crônicas, entre outros aspectos voltados para a saúde (PNS, 2019).

A amostra da PNS é uma subamostra da Amostra Mestra da Pesquisa Nacional de Amostra por Domicílios (PNAD). Inicialmente, amostras aleatórias simples das unidades primárias de amostragem (UPAs) foram selecionadas por estratos da Amostra Mestra da PNAD. Em seguida, em cada UPA selecionou-se aleatoriamente quantidades fixas de domicílios particulares permanentes a partir do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE). Por fim, selecionou-se um morador para responder à entrevista individual (IBGE, 2020b) com base em uma lista de moradores elegíveis (com idade 15 anos ou mais). Ao todo foram visitados 108.525 domicílios e realizadas 94.114 entrevistas. A maior parte das respostas foram auto-referidas (PNS, 2019).

O cálculo do tamanho da amostra de domicílios e de pessoas levou em consideração: os indicadores de interesse, a estimação de proporções com nível de precisão desejado em intervalos de 95% de confiança, o efeito do plano amostral, o número de domicílios selecionados por UPA, a proporção de domicílios com pessoas na faixa etária de interesse e a possibilidade de estimação dos principais indicadores em níveis estaduais e de capitais

Para o presente trabalho foram incluídos indivíduos que tenham respondido a perguntas relacionadas a diagnóstico de diabetes, de hipertensão e de colesterol alto, importantes para construção do escore de saúde cardiovascular. Além disso, mulheres grávidas foram excluídas da análise devido a possíveis diagnósticos de fatores de risco cardiovascular associados a gravidez que enviesariam o estudo.

#### 2.1.1 Escore de Saúde Cardiovascular Ideal

O escore de SCV ideal, proposto pela American Heart Association (LLOYD-JONES et al., 2010), é constituído pela soma de metas associadas a qualidade de vida da população, com potencial efeito preventivo no desenvolvimento de DCV. Essas metas são compostas por dois grupos: os fatores saudáveis (níveis adequados de colesterol, de pressão arterial e de glicose) e os comportamentos saudáveis (não ser tabagista, apresentar

IMC ideal, realizar atividades físicas regularmente e possuir dieta saudável), resumidos na tabela abaixo:

Tabela 1: Metas de saúde cardiovascular utilizadas na PNS 2019

Meta*	Métrica utilizada	Resultado Favorável 2019**
Comportamentos saudáveis (CS)	Somatório das 4 metas de CS (SCS)	SCS = 4
Tabagismo	Nunca ter fumado ou não fumar há mais de 12 meses.	P050 = 3 ou P052 = 1 ou 2 e P05901 >1
IMC	<25 kg/m <sup>2</sup> (a partir de medidas de peso e altura autorreferidas).	(P00103/P00403) <sup>2</sup> <0,0025
Atividade Física	Prática de exercício físico ou esporte ≥ 150 min/semana.	(P03701*60+P03702) * P035 ≥ 150
Dieta	Frutas e vegetais: consumo ≥ 1 vez/dia;	P018=7 e P00901=7
	Peixe: consumo ≥ 2 vezes/semana;	P015 ≥ 2
	Sódio: consumo de sal adequado, baixo ou muito baixo;	P02601 ≥ 3
	Açúcar: consumo de refrigerantes e/ou sucos industrializados <5 dias/semana.	P02001 + P02002 <5
Fatores saudáveis (FS)	Somatório das 3 metas de FS (SFS)	SFS = 3
Colesterol Total	Ausência de diagnóstico médico de colesterol alto.	Q060 = 2
Pressão arterial	Ausência de diagnóstico médico de hipertensão arterial.	Q00201 = 2
Glicose	Ausência de diagnóstico médico de diabetes.	Q03001 = 3

\*Metas propostas neste estudo foram adaptações de AHA;

\*\*Códigos das variáveis da PNS 2019.

### 2.1.2 Fatores Socioeconômicos

As variáveis explicativas que foram utilizadas no modelo de regressão logística incluem nível de escolaridade, status de ocupação ou empregabilidade e riqueza. Para escolaridade e empregabilidade o questionário disponibiliza informações sobre o grau de instrução e de ocupação dos indivíduos. A variável latente riqueza foi estimada conforme o índice de riqueza estabelecido pelo programa Demographic and Health Surveys (DHS), índice válido internacionalmente. Para tanto uma análise fatorial foi conduzida para a construção de um escore que representa níveis de riqueza acumulada baseados em bens de consumo.

## 2.2 Análises Estatísticas

### 2.2.1 Análise Fatorial - Método da Componente Principal

A análise fatorial pode ser utilizada para descrever dimensões latentes de um conjunto maior de variáveis correlacionadas em termos de poucos fatores que serão representativos sem muito prejuízo (JOHNSON; WICHERN et al., 2002). Descreve-se a seguir a análise fatorial com base em Rencher (2002).

Seja um vetor  $\mathbf{X} = [X_1, X_2, \dots, X_p]$  com média  $\boldsymbol{\mu}$  e matriz de covariância  $\boldsymbol{\Sigma}$ , o modelo fatorial assume que  $\mathbf{X}$  é linearmente dependente de variáveis aleatórias não observáveis  $F_1, F_2, \dots, F_m$ , denominadas fatores comuns e de  $p$  fontes de variação aleatória  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$  denominadas erros. Dessa forma, o modelo é expresso em notação matricial por:

$$\underset{(p \times 1)}{\mathbf{X}} - \underset{(p \times 1)}{\boldsymbol{\mu}} = \underset{(p \times m)}{\mathbf{LF}} + \underset{(m \times 1)}{\boldsymbol{\varepsilon}}, \quad (2.2.1)$$

onde o coeficiente  $l_{ij}$  é a carga (*loading*) da  $i$ -ésima variável sobre o  $j$ -ésimo fator, sendo  $\mathbf{L}$  denominada matriz de cargas fatoriais.

Além disso, assume-se que

$$E(\mathbf{L}) = \mathbf{0}_{m1},$$

$$Cov(\mathbf{F}) = \mathbf{I}_{mm},$$

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}, Cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}, \text{ onde } \boldsymbol{\Psi} \text{ é uma matriz diagonal,}$$

e  $\mathbf{F}$  e  $\boldsymbol{\varepsilon}$  são independentes.

A partir dessas suposições, o modelo fatorial é ortogonal e implica numa estrutura de covariância  $\boldsymbol{\Sigma}$  para  $\mathbf{X}$ . Da equação 2.2.1, tem-se que

$$\begin{aligned} (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' &= (\mathbf{LF} + \boldsymbol{\varepsilon})(\mathbf{LF} + \boldsymbol{\varepsilon})' \\ &= \mathbf{LF}(\mathbf{LF})' + \boldsymbol{\varepsilon}(\mathbf{LF})' + \mathbf{LF}\boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'. \end{aligned}$$

Assim, a matriz de covariância é dada por

$$\begin{aligned} \boldsymbol{\Sigma} = Cov(\mathbf{X}) &= E((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})') \\ &= \mathbf{LE}(\mathbf{FF}')(\mathbf{L})' + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')(\mathbf{L})' + \mathbf{LE}(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \\ &= \mathbf{LL}' + \boldsymbol{\Psi}. \end{aligned} \quad (2.2.2)$$

Como  $(\mathbf{X} - \boldsymbol{\mu})\mathbf{F}'$  pode ser expresso por  $\mathbf{L}\mathbf{F}\mathbf{F}' + \boldsymbol{\varepsilon}\mathbf{F}'$ , então,

$$Cov(\mathbf{X}, \mathbf{F}) = \mathbf{L}E(\mathbf{F}\mathbf{F}') + E(\boldsymbol{\varepsilon}\mathbf{F}') = \mathbf{L}. \quad (2.2.3)$$

Por fim, dada a equação 2.2.2,

$$\begin{aligned} Var(X_i) &= \sigma_{ii} \\ &= l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \psi_i \\ &= h_i^2 + \psi_i, \end{aligned} \quad (2.2.4)$$

onde  $h_i^2$  representa a comunalidade, isto é, a variância de termos comuns associada aos  $F_i$  das equações do modelo e  $\psi_i$  representa a unicidade, caracterizando a variância associada a  $\varepsilon_i$  específica da  $i$ -ésima equação do modelo.

A estrutura de covariância é expressa, então, pelos resultados das equações 2.2.2, 2.2.3 e 2.2.4. Existem alguns métodos de estimação das cargas  $l_{ij}$  e das variâncias específicas  $\varepsilon_i$ . Neste trabalho, foi adotado o método da componente principal. Uma vez que estimadores são obtidos é possível identificar os fatores e valores estimados para os próprios fatores chamados de escores fatoriais que ajudarão a definir a variável de interesse **índice de riqueza**.

A análise fatorial do componente principal da matriz de covariância da amostra, comumente denotada pela estatística  $S$ , é especificada em termos de seus pares de autovalores-autovetores  $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \dots, (\hat{\lambda}_p, \hat{e}_p)$  onde  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$ . Sendo  $m$  o número de fatores comuns menor que  $p$ , a matriz de cargas  $\mathbf{L}$  é dada por:

$$\tilde{\mathbf{L}} = \left[ \sqrt{\hat{\lambda}_1 \hat{e}_1} : \sqrt{\hat{\lambda}_2 \hat{e}_2} : \dots : \sqrt{\hat{\lambda}_m \hat{e}_m} \right],$$

e as variâncias específicas estimadas são fornecidas pelos elementos diagonais da matriz  $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$  tal que  $\tilde{\psi}_i = s_{ii} - \sum_{m=1}^{j=1} \tilde{l}_{ij}^2$ . As comunalidades, portanto, são estimadas por  $\tilde{h}_i^2 = \tilde{l}_{i1}^2 + \tilde{l}_{i2}^2 + \dots + \tilde{l}_{im}^2$ .

No presente trabalho, essa técnica foi utilizada em busca de uma variável latente que seja capaz de definir o nível riqueza do indivíduo por meio de um conjunto de variáveis disponíveis no questionário com perguntas relacionadas às condições do domicílio em que moram.

A construção do índice de riqueza foi realizada com base nas instruções detalhadas por Rutstein (2015) e WFP (2017). O primeiro passo sugerido envolve a seleção das variáveis do questionário para construção do índice. No presente trabalho foram escolhidas as seguintes informações sobre os domicílios: fonte de água potável; quantidade

de banheiros; material do piso; material das paredes externas; material do telhado; tipo de rede de distribuição de água; componente usado para acender o fogão; destino do lixo; número de pessoas por quantidade de cômodos e se possui televisão, carro, máquina de lavar roupa, telefone fixo, celular, microondas, computador, motocicleta, internet e empregado(a) doméstico(a).

O segundo passo descrito em Rutstein (2015) e WFP (2017) envolve a categorização das variáveis escolhidas em variáveis binárias com base no que se considera favorável para condições de moradia do domicílio, atribuindo 1 para favorável e 0 caso contrário. A partir de então, para selecionar variáveis capazes de discriminar os domicílios entre ricos e pobres analisa-se a frequência relativa em cada uma e elimina-se aquelas que apresentarem frequência maior que 95% em um dos níveis por serem consideradas pouco discriminantes. Além disso, analisa-se a correlação tetracórica do conjunto de dados, medida de correlação utilizada para variáveis binárias (OXFORD, 2008), identificando variáveis que tenham correlação alta entre si ( $> 0,95$ ), com o objetivo de excluir aquelas que já são explicadas por meio de outra, ou identificando variáveis que possuem correlação baixa entre si ( $< 0,5$ ) retirando aquelas que não agregariam à construção do escore.

Em sequência, realiza-se a análise fatorial considerando apenas 1 fator com as variáveis selecionadas na etapa anterior e retira-se a variável que apresentar a menor comunalidade. Depois, executa-se a análise fatorial sem a variável excluída retirando novamente a variável que apresentar a menor comunalidade  $\tilde{h}^2_i$ . Repete-se esse processo até que todas as comunalidades sejam maiores que 0,3. Com o modelo final utiliza-se o escore fatorial baseado em apenas 1 componente, já que este explica a maior proporção total da variância e é o mais usado nesse contexto (WFP,2017). Cria-se então uma nova variável para representar a riqueza do indivíduos residentes do domicílio com base em um escore de 1 a 5 definido por quintis.

### 2.2.2 Análise de Regressão Logística Binária e Multinomial

Modelos de regressão logística foram adotados para a predição do escore de saúde cardiovascular no presente trabalho. Tais modelos são utilizados para relacionar variáveis respostas categóricas com variáveis explicativas categorias e/ou contínuas e estão descritos a seguir com base em Agresti (2013) e Agresti (2019). Por simplicidade, descreve-se inicialmente um modelo de regressão logística binária e, em sequência, um modelo, multinomial.

Considerando uma variável resposta binária  $Y$  em que o interesse é modelar a probabilidade de “sucesso”  $\pi(x)$  (em que “sucesso” representa a ocorrência de algum evento de interesse), assumindo que as observações são variáveis binomiais independentes define-se a chance como:

$$\frac{\pi(x)}{1 - \pi(x)}. \quad (2.2.5)$$

Para que o modelo tenha uma forma linear utiliza-se o logaritmo da chance expressa na Expressão 2.2.5, conhecido como a função de ligação logito, definindo, então, o modelo tal como:

$$\log\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n, \quad (2.2.6)$$

em que  $\alpha$  representa o intercepto da equação,  $\beta_i$  é a taxa de crescimento ou redução na chance de ocorrência do evento de interesse dada pelo acréscimo de uma unidade da variável explicativa correspondente  $x_i$  quando as demais variáveis estão constantes e  $n$  é o número de variáveis explicativas do modelo.

Para facilitar a interpretação, aplica-se a exponencial em ambos os lados da Equação 2.2.6 de forma que a chance de sucesso (Expressão 2.2.5) fique na escala original e seja possível interpretar o impacto do crescimento de cada unidade de  $x$  sobre essa chance, definida agora por:

$$\frac{\pi(x)}{1 - \pi(x)} = \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n).$$

Pode-se ainda dizer que a chance de sucesso é  $e^{\beta_i}$  vezes maior (se  $e^{\beta_i} > 1$ ) ou  $(1 - e^{\beta_i})\%$  menor (caso contrário) para cada aumento de unidade em  $x_i$ , mantendo-se as demais variáveis explicativas constantes.

Em particular, para variáveis explicativas categóricas com  $k$  níveis, define-se  $k - 1$  variáveis binárias no seguinte formato:

- $x_1 = 1$  se nível = 1, 0 caso contrário;
- $x_2 = 1$  se nível = 2, 0 caso contrário;
- ...
- $x_{k-1} = 1$  se nível =  $k - 1$ , 0 caso contrário.

Nesse caso, o modelo envolvendo 1 variável explicativa categórica será dado por:

$$\text{logito}[P(Y = 1)] = \alpha + \beta_1 x_1 + \dots + \beta_{k-1} x_{k-1},$$

onde a categoria  $k$  é o nível de referência a ser comparado com os demais. Isso implica que indivíduos na categoria de nível  $k$  possuem logito =  $\alpha$ , na categoria de nível 1 possuem o logito =  $\alpha + \beta_1$ , na categoria 2 possuem o logito =  $\alpha + \beta_2$  e assim por diante até indivíduos da  $k - 1$  com logito =  $\alpha + \beta_{k-1}$ .



Assim, dado o mesmo raciocínio da aplicação da exponencial para facilitar a interpretação, a razão de chances para a comparação de um categoria qualquer  $j$  com a categoria de referência  $k$  é dada por  $e^{\beta_j}$ . Desse modo, é possível definir o quanto a chance de alcançar o “sucesso” ( $Y = 1$ ) aumenta ou diminui quando comparado a indivíduos que estão em níveis diferentes na variável explicativa  $x$ .

No caso do presente estudo foi definido como variável resposta o escore de saúde cardiovascular (SCV) em 3 categorias sendo: classe superior a de indivíduos que atingem 6 a 7 metas, classe intermediária para os que atingiram 3 a 5 metas e classe inferior para 0 a 2 metas. Para as variáveis explicativas foram definidas três variáveis categóricas representando fatores socioeconômicos quais sejam escolaridade, índice de riqueza e status de ocupação (se ocupado ou não), além de outras covariáveis sociodemográficas que são importantes para predição do escore de SCV. O trabalho envolveu a construção de 2 modelos (um para cada situação de domicílio) a partir dos quais foram estimadas razões de chances com respeito às variáveis explicativas.

Para tanto, utilizou-se os modelos de regressão logística multinomial que são uma generalização da regressão logística binária (descritos na Expressão 2.2.6), os quais podem ser aplicados para a variáveis-resposta com mais de 2 categorias. Tais modelos permitem a estimação da chance de ocorrência de uma categoria em relação à categoria de referência estimada em função das variáveis explicativas.

Para uma variável resposta  $Y$  com  $c$  categorias e suas respectivas probabilidades  $(\pi_1, \dots, \pi_c)$ , em que  $\sum_j \pi_j = 1$  e as observações são independentes, o modelo multinomial com uma categoria de referência fixa é equivalente a utilizar, simultaneamente,  $c - 1$  modelos logísticos binários para todos os pares de categorias com uma categoria de referência fixa  $c$  que representa o desfecho de interesse. No presente estudo tal modelo foi utilizado adotando-se a categoria inferior do escore de SCV (0 a 2 metas) como a de referência.

Com  $\pi_j(\mathbf{x}) = P(Y = j|\mathbf{x})$ , a função de ligação, conhecida como logito de categoria de referência, é definida por

$$\log \frac{\pi_j(\mathbf{x})}{\pi_c(\mathbf{x})}, \quad j = 1, \dots, c - 1. \quad (2.2.7)$$

Considerando o vetor de variáveis explicativas do modelo para  $i$ -ésima observação  $\mathbf{x}_i = (x_{i1}, \dots, x_{ic})$  e  $\boldsymbol{\beta}_c = (\beta_{j1}, \dots, \beta_{jp})'$ , o vetor de parâmetros para o  $j$ -ésimo logito de categoria de referência (Expressão 2.2.7) o modelo pode ser expresso por:

$$\log \frac{\pi_j(\mathbf{x})}{\pi_c(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}'_j \mathbf{x}, \quad j = 1, \dots, c - 1, \quad (2.2.8)$$

Dessa maneira, o modelo tem  $c - 1$  equações com parâmetros separados para

cada uma com base na categoria de referência, mas sendo possível utilizá-lo para estimar o efeito em todos os  $c(c-1)$  logitos de pares de categorias da seguinte forma:

$$\log \frac{\pi_a(\mathbf{x})}{\pi_b(\mathbf{x})} = \log \frac{\pi_a(\mathbf{x})}{\pi_c(\mathbf{x})} - \log \frac{\pi_b(\mathbf{x})}{\pi_c(\mathbf{x})}.$$

Ainda, é possível escrever o modelo em termos da probabilidade de se observar a  $j$ -ésima categoria da variável resposta  $\pi_j(\mathbf{x})$  como:

$$\pi_j(\mathbf{x}) = \frac{e^{\alpha_j + \beta'_j \mathbf{x}}}{1 + \sum_{h=1}^{c-1} e^{\alpha_h + \beta'_h \mathbf{x}}}. \quad (2.2.9)$$

Sabendo que  $\mathbf{y}_i = (y_{i1}, \dots, y_{ic})$  representa uma observação do processo multinomial pertencente à categoria  $j$ , com apenas  $y_{ij} = 1$  e os demais  $y_{ik} = 0$ ,  $\forall (k \neq j)$ , então  $\sum_j y_{ij} = 1$ . Desse modo, como  $\pi_c(\mathbf{x})$  pode ser escrito como  $1 - (\pi_1 + \dots + \pi_{c-1})$  e  $y_{ic}$  como  $1 - (y_{i1} + \dots + y_{i,c-1})$ , então a função de verossimilhança para a  $i$ -ésima observação é dada por:

$$\begin{aligned} \log \left[ \prod_{j=1}^c \pi_j(\mathbf{x}_i)^{y_{ij}} \right] &= \sum_{j=1}^{c-1} y_{ij} \log \pi_j(\mathbf{x}_i) + \left( 1 - \sum_{j=1}^{c-1} y_{ij} \right) \log \left[ 1 - \sum_{j=1}^{c-1} \pi_j(\mathbf{x}_i) \right] \\ &= \sum_{j=1}^{c-1} y_{ij} \log \frac{\pi_j(\mathbf{x}_i)}{1 - \sum_{j=1}^{c-1} \pi_j(\mathbf{x}_i)} + \log \left[ 1 - \sum_{j=1}^{c-1} \pi_j(\mathbf{x}_i) \right]. \end{aligned} \quad (2.2.10)$$

Ao substituir as probabilidades de sucesso  $\pi_j(\mathbf{x}_i)$  da Equação 2.2.10 pela Expressão 2.2.9, a função de log-verossimilhança para  $n$  observações independentes é dada por:

$$\begin{aligned} \log \prod_{i=1}^n \left[ \prod_{j=1}^c \pi_j(\mathbf{x}_i)^{y_{ij}} \right] &= \sum_{i=1}^n \left[ \sum_{j=1}^{c-1} y_{ij} (\alpha_j + \beta'_j \mathbf{x}_i) - \log \left( 1 + \sum_{j=1}^{c-1} \exp(\alpha_j + \beta'_j \mathbf{x}_i) \right) \right] \\ &= \sum_{j=1}^{c-1} \left[ \alpha_j \left( \sum_{i=1}^n y_{ij} \right) + \sum_{k=1}^p \beta_{jk} \left( \sum_{i=1}^n x_{ik} y_{ij} \right) \right] \\ &\quad - \sum_{i=1}^n \log \left[ 1 + \sum_{j=1}^{c-1} \exp(\alpha_j + \beta'_j \mathbf{x}_i) \right]. \end{aligned}$$

A estatística suficiente para  $\beta_{jk}$  é  $\sum_i x_{ik} y_{ij}$  para  $j = 1, \dots, c-1$  e  $k = 1, \dots, p$  e para  $\alpha_j$  sendo o número total de observações na categoria  $j$ , ou seja,  $\sum_i y_{ij} = \sum_i x_{i0} y_{ij}$ , com  $x_{i0} = 1$ .

Em geral, o método de Newton-Raphson pode fornecer as estimativas dos parâmetros de MV, uma vez que a função de log-verossimilhança é côncava. Para amostras grandes, os estimadores seguem distribuições normais com desvios-padrões obtidos pela raiz quadrada dos elementos da diagonal da matriz de informação inversa.

### 2.2.3 Regressão Logística para amostragem complexa

Como já mencionado, o conjunto de dados do estudo tem um amostragem complexa. Por tal razão, o método de máxima pseudo-verossimilhança (MPV) é comumente utilizado para estimar os parâmetros do modelo a partir de correções que levam em consideração pesos amostrais (SILVA, 2018). Considerando  $Y$  uma variável resposta binária (por simplicidade),  $\mathbf{x}$  o conjunto de variáveis explicativas,  $\pi(\mathbf{x})$  a probabilidade de sucesso dado  $\mathbf{x}$  e  $w_i$  peso amostral da  $i$ -ésima observação da amostra então, o vetor de parâmetros  $\boldsymbol{\beta}$  é dado pela equação:

$$\sum_{ics} w_i \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{ics} w_i [y_i - \pi(\mathbf{x}_i)] \mathbf{x}_i = 0, \quad (2.2.11)$$

em que  $L(\boldsymbol{\beta})$  é a função de log verossimilhança.

Para situações em que os dados são coletados por planos amostrais complexos, não existe normalmente uma forma analítica direta para o cálculo do erro padrão de estimadores e geralmente são utilizados métodos numéricos tais como a linearização de Taylor e os métodos de reamostragem de *Jackknife* e *bootstrap*. Para os modelos de regressão multinomiais, o cálculo dessas estimativas é obtidos de forma análoga.

Estudos com amostragem complexa limitam a utilização de métodos usuais de diagnóstico de ajuste de modelos e outros procedimentos da inferência clássica tais como testes estatísticos de Razões de Verossimilhança  $G^2$ , os quais nem sempre são apropriados quando os modelos avaliados são logísticos. Alternativas como pseudo- $R^2$  podem ser utilizadas para avaliar a adequação do modelo, como o  $R^2$  de McFadden citado na Expressão 2.2.15.

Para testar hipóteses lineares sobre os coeficientes do modelo, descrito por SAS... (2019), as hipóteses nulas podem ser expressas por:

$$H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{c},$$

onde  $\mathbf{L}$  é a matriz de coeficientes para hipótese linear,  $\mathbf{c}$  é o vetor de constantes cujo rank é  $r$  e  $\boldsymbol{\theta}$  vetor de coeficientes incluindo  $\beta$  e  $\alpha$ . Sendo  $\hat{\boldsymbol{\theta}}$  o estimador de máxima verossimilhança e  $\hat{V}(\hat{\boldsymbol{\theta}})$  a estimativa da matriz de covariância dada por:

$$\hat{V}(\hat{\theta}) = \hat{Q}^{-1} \hat{G} \hat{Q}^{-1}, \quad (2.2.12)$$

em que  $\hat{Q} = \sum_{h=1}^H \sum_{n_h}^{i=1} \sum_{m_{hi}}^{j=1} w_{hij} \hat{\mathbf{D}}_{hij} (\text{diag}(\hat{\pi}_{hij}) - \hat{\pi}_{hij} \hat{\pi}'_{hij})^{-1} \hat{\mathbf{D}}'_{hij}$ ,

$$\hat{\mathbf{G}} = \frac{n-1}{n-p} \sum_{h=1}^H \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_{h\cdot})(e_{hi} - \bar{e}_{h\cdot})',$$

$$e_{hi} = \sum_{j=1}^{m_{hi}} w_{hij} \hat{\mathbf{D}}_{hij} (\text{diag}(\hat{\pi}_{hij}) - \hat{\pi}_{hij} \hat{\pi}'_{hij})^{-1} (y_{hij} - \pi_{hij}) \text{ e}$$

$$\bar{e}_{h\cdot} = \frac{1}{n_h} \sum_{i=1}^{n_h} e_{hi},$$

sendo  $\mathbf{D}_{hij}$  a matriz de derivadas parciais da função de ligação  $g$  em relação a  $\theta$  e  $\hat{\mathbf{D}}_{hij}$  e as probabilidades de resposta  $\hat{\pi}_{hij}$  são avaliadas em  $\hat{\theta}$ .

Para o método de linearização de Taylor usada nesse trabalho a estatística do teste é:

$$W_F = \frac{f-r+1}{fr} (\mathbf{L}\hat{\theta} - \mathbf{c})' [\mathbf{L}\hat{\mathbf{V}}(\hat{\theta})\mathbf{L}']^{-1} (\mathbf{L}\hat{\theta} - \mathbf{c})$$

Intervalos de  $(1 - \alpha)\%$  confiança de Wald podem ser calculados por:

$$\hat{\beta}_j \pm z_{\alpha/2} V(\beta_j). \quad (2.2.13)$$

onde  $V(\beta_j)$  é a variância estimatimada dada por:

$$V(\beta_j) = \sum_{r=1}^R \alpha_r (\beta_r - \beta)(\beta_r - \beta)'. \quad (2.2.14)$$

em que  $R$  o número total de pesos de replicação de bootstrap e  $\alpha_r$  é o coeficiente de replicação para a  $r$ -ésima replicação igual a  $\frac{1}{R}$ .

Ainda, a transformação para  $e^{\hat{\beta} \pm z_{\alpha/2} V(\beta_j)}$  permite a obtenção de intervalos de confiança para as razões de chances.

Para se verificar a adequação do modelo, pode-se utilizar os testes de  $\chi_2$  (Qui-Quadrado de Pearson) ou  $G^2$  (Qui-Quadrado da Razão de Verossimilhança) de qualidade de ajuste caso os preditores sejam categóricos e os dados não sejam esparsos. Na regressão logística é comum o uso de pseudo- $R^2$  como, por exemplo, o coeficiente  $R^2$  de McFadden:

$$R^2 = 1 - \frac{\log L(M)}{\log L(0)}, \quad (2.2.15)$$

onde  $\log L(M)$  e  $\log L(0)$  são, respectivamente, a verossimilhança logarítmica maximizada para o modelo ajustado e o modelo nulo contendo apenas o intercepto.

#### 2.2.4 Implementação

As análises foram implementadas por meio do *software* R versão 4.1.2. Utilizando os pacotes *survey* para ajuste por amostragem complexa do delineamento do estudo da PNS 2019, *psych* para análise fatorial e *SAScii* para a leitura dos dados que são disponíveis na página eletrônica do IBGE (IBGE, 2020a). Além disso, também utilizou-se o comando *surveylogistic* do *software* SAS para o ajuste de modelos de regressão logística multinomial com amostragem complexa.

### 3 Resultados

A amostra final do estudo foi constituída por 77.494 indivíduos dos quais 61.199 são indivíduos cujo domicílio pertence à área urbana e 16.295 à área rural. A Tabela 2 apresenta prevalências estimadas de níveis de variáveis sócio-demográficas por situação de domicílio com base na expansão da amostra utilizada. Observou-se que 87,4% (IC<sub>95%</sub>: 87,0 – 87,8) dos indivíduos eram residentes em domicílios localizados em áreas urbanas e apenas 12,6% (IC<sub>95%</sub>: 12,2 – 13,0), em áreas rurais. Entre os indivíduos residentes em áreas urbanas houve uma maior proporção estimada de pessoas com ensino médio completo ou superior incompleto (37,7%, IC<sub>95%</sub>: 37,2 – 38,4), empregados (63,0%, IC<sub>95%</sub>: 62,3 – 63,7), do sexo feminino (55,5%, IC<sub>95%</sub>: 54,8 – 56,2), da raça branca (46,0%, IC<sub>95%</sub>: 45,1 – 46,8) e da região sudeste (48,3%, IC<sub>95%</sub>: 47,4 – 49,3). Já nas áreas rurais houve uma maior predominância de indivíduos sem instrução ou com ensino fundamental incompleto (62,7%, IC<sub>95%</sub>: 61,4 – 64,1), do sexo masculino (50,5%, IC<sub>95%</sub>: 49,2 – 51,8), da raça parda (53,8%, IC<sub>95%</sub>: 52,2 – 55,4) e da região nordeste (47,0%, IC<sub>95%</sub>: 45,4 – 48,6). As demais variáveis (faixa etária, estado civil e presença de doenças crônicas) apresentaram proporções semelhantes entre as situações de domicílio.

Tabela 2: Prevalências estimadas (e Intervalos de 95% de confiança) de categorias de variáveis sociodemográficas totais e por situação de domicílio - expansão da amostra PNS 2019

	Urbana	Rural	Total
<b>Escolaridade<sup>a</sup></b>			
Até EFI	29,1 (28,4 - 29,9)	62,7 (61,4 - 64,1)	33,4 (32,7 - 34,1)
EFC a EMI	13,9 (13,4 - 14,4)	13,4 (12,6 - 14,3)	13,8 (13,4 - 14,3)
EMC a ESI	37,7 (37,2 - 38,4)	20,3 (19,1 - 21,4)	35,5 (34,9 - 36,2)
ESC	19,2 (18,4 - 20,1)	3,5 (3,1 - 4,7)	17,2 (16,6 - 18,7)
<b>Situação de Ocupação</b>			
Empregado	63,0 (62,3 - 63,7)	54,2 (53,0 - 55,4)	61,9 (61,3 - 62,5)
Sem ocupação	37,0 (36,3 - 37,7)	45,8 (44,6 - 47,0)	38,1 (37,5 - 38,7)
<b>Índice de Riqueza</b>			
1° Nível	13,6 (13,1 - 14,1)	13,0 (12,2 - 13,9)	11,3 (10,9 - 11,7)
2° Nível	15,8 (15,2 - 16,3)	18,3 (17,1 - 19,4)	15,7 (15,2 - 16,2)
3° Nível	20,3 (19,7 - 20,1)	20,5 (19,3 - 21,7)	18,8 (18,2 - 19,3)
4° Nível	22,4 (21,7 - 23,2)	21,4 (20,3 - 22,5)	24,5 (23,7 - 25,3)
5° Nível	27,8 (26,7 - 28,8)	26,7 (25,2 - 28,2)	29,6 (28,7 - 30,6)
<b>Sexo</b>			
Masculino	44,5 (43,8 - 45,2)	50,5 (49,2 - 51,8)	45,2 (44,6 - 45,9)
Feminino	55,5 (54,8 - 56,2)	49,5 (48,2 - 50,8)	54,8 (54,1 - 55,4)
<b>Faixa Etária</b>			
18 a 24 anos	12,0 (11,4 - 12,6)	11,8 (10,8 - 12,9)	12,0 (11,5 - 12,5)
25 a 39 anos	28,8 (28,1 - 29,5)	27,1 (26,0 - 28,1)	28,6 (28,0 - 29,2)
40 a 59 anos	37,1 (36,4 - 37,8)	36,3 (35,2 - 37,5)	37,0 (36,4 - 37,6)

Continua

	Conclusão		
	Urbana	Rural	Total
60 ou mais anos	22,1 (21,4 - 22,7)	24,8 (23,7 - 25,8)	22,4 (21,8 - 23,0)
<b>Raça/Etnia</b>			
Parda	41,1 (40,3 - 41,9)	53,8 (52,2 - 55,4)	42,7 (42,0 - 43,4)
Branca	46,0 (45,1 - 46,8)	35,0 (33,5 - 36,6)	44,6 (43,8 - 45,4)
Preta	11,4 (10,9 - 11,9)	10,3 (9,4 - 11,3)	11,2 (10,8 - 11,7)
Outros <sup>b</sup>	1,6 (1,4 - 1,8)	0,9 (0,7 - 1,1)	1,5 (1,3 - 1,7)
<b>Estado Civil</b>			
Solteiro	39,4 (38,7 - 40,2)	41,7 (40,3 - 43,2)	39,7 (39,0 - 40,4)
Casado	45,5 (44,7 - 46,2)	47,9 (46,6 - 49,3)	45,8 (45,1 - 46,5)
Outros <sup>c</sup>	15,1 (14,6 - 15,6)	10,3 (9,6 - 11,0)	14,5 (14,1 - 14,9)
<b>Doença Crônica</b>			
Pelo menos 1	42,8 (42,0 - 43,6)	42,9 (41,4 - 44,4)	42,8 (42,1 - 43,5)
Sem doença	57,2 (56,4 - 58,0)	57,1 (55,6 - 58,6)	57,2 (56,5 - 57,9)
<b>Região</b>			
Norte	6,8 (6,5 - 7,1)	10,5 (9,7 - 11,3)	7,3 (7,0 - 7,6)
Nordeste	22,2 (21,6 - 22,7)	47,0 (45,4 - 48,6)	25,3 (24,7 - 25,9)
Sudeste	48,3 (47,4 - 49,3)	21,8 (20,4 - 23,2)	45,0 (44,1 - 45,9)
Sul	14,8 (14,3 - 15,3)	15,5 (14,5 - 16,6)	14,9 (14,4 - 15,3)
Centro-Oeste	7,9 (7,6 - 8,2)	5,2 (4,6 - 5,7)	7,6 (7,3 - 7,9)
<b>Total, N (%)</b>	<b>122.412.763 (87,4%)</b>	<b>17.624.183 (12,6%)</b>	<b>140.036.946</b>

<sup>a</sup> EFI: Ensino Fundamental Incompleto; EFC: Ensino Fundamental Completo; EMI: Ensino Médio Incompleto; EMC: Ensino Médio Completo; ESI: Ensino Superior Incompleto; ESC: Ensino Superior Completo.

<sup>b</sup> Outras raças/etnias incluem Amarela, Indígena e Ignorada;

<sup>c</sup> Outros estados civis incluem Separado, Divorciado e Viúvo.

A Figura 1 apresenta a distribuição do escore de SCV por situação de domicílio. Nota-se que apenas 0,6% (IC<sub>95%</sub>: 0,4 - 0,6) dos residentes de áreas urbanas atingiram a condição de SCV Ideal (as 7 metas concomitantes) e tal prevalência é ainda menor em áreas rurais (0,2%, IC<sub>95%</sub>: 0,1 - 0,3). Quando o escore foi categorizado em 3 níveis (SCV superior, intermediária e inferior), a proporção de domicílios que atingiram 6 a 7 metas (classe superior) foi de 9,4% (IC<sub>95%</sub>: 9,0 - 9,9) em áreas urbanas e quase metade desta prevalência, 5,2% (IC<sub>95%</sub>: 4,7 - 5,8), em áreas rurais. A prevalência de SCV intermediária (3 a 5 metas) foi um pouco menor em áreas urbanas (80,8%, IC<sub>95%</sub>: 80,3 - 81,3) que em, rurais (85,7%, IC<sub>95%</sub>: 84,8 - 86,5). Prevalências de SCV inferior (0 a 2 metas) não foram muito distintas entre áreas urbanas (9,7%, IC<sub>95%</sub>: 9,3 - 10,1) e rurais (9,0%, IC<sub>95%</sub>: 8,3 - 9,7). Isto é, os domicílios urbanos apresentaram maior prevalência de SCV superior enquanto os domicílios rurais apresentaram maior proporção de indivíduos na classe intermediária de SCV.

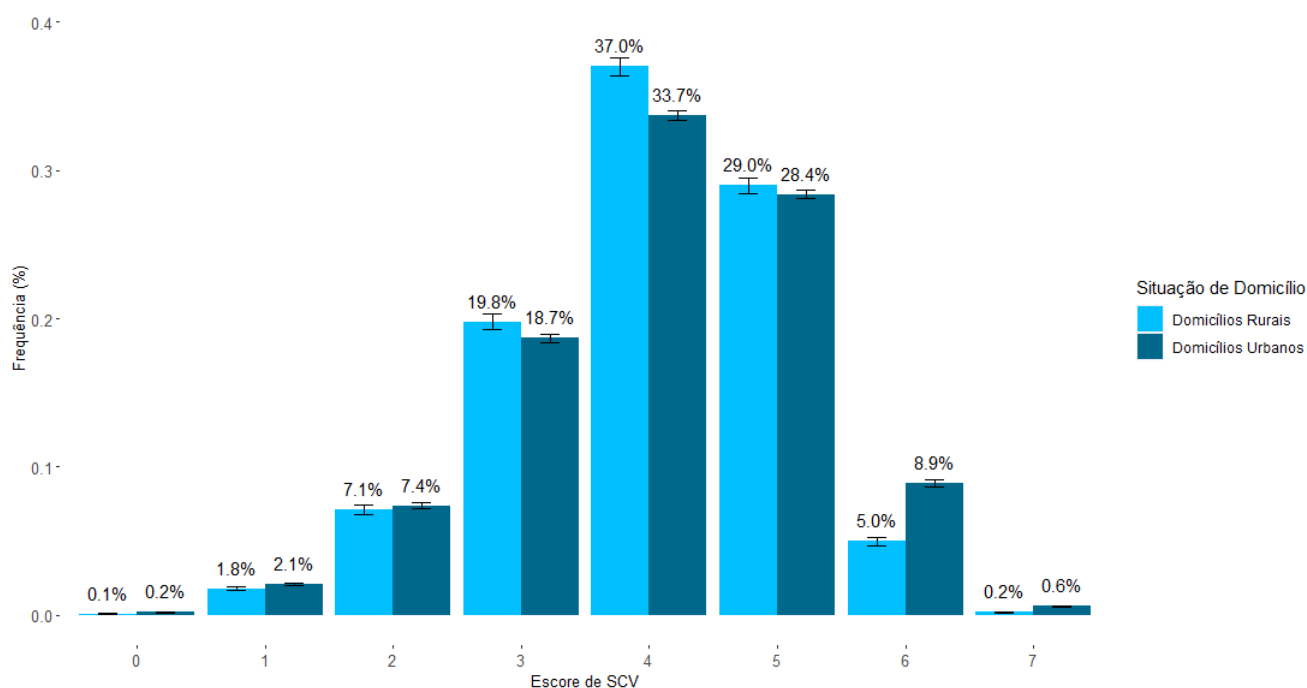


Figura 1: Distribuição do Escore de SCV<sup>1</sup> com respectivos Intervalos de 95% de Confiança<sup>2</sup>, estimados para a população adulta brasileira em 2019 (Expansão da amostra da PNS 2019).

Notas:

<sup>1</sup>SCV: Saúde Cardiovascular

<sup>2</sup>Os respectivos intervalos de 95% de confiança para as prevalências de cada valor do escore de SCV ( $Y$ ) foram:

Para área rural,

$IC_{Y=0}$ : (0,05-0,2);  $IC_{Y=1}$ : (1,5-2,1);  $IC_{Y=2}$ : (6,4-7,6);  $IC_{Y=3}$ : (18,8-20,7);  $IC_{Y=4}$ : (35,7-38,1);  $IC_{Y=5}$ : (27,9-29,9);

$IC_{Y=6}$ : (4,4-5,5);  $IC_{Y=7}$ : (0,1,0,3).

Para a área urbana,

$IC_{Y=0}$ : (0,1-0,2);  $IC_{Y=1}$ : (1,9-2,3);  $IC_{Y=2}$ : (7,0-7,7);  $IC_{Y=3}$ : (18,1-19,2);  $IC_{Y=4}$ : (33,0-34,3);  $IC_{Y=5}$ : (27,7-29,0);

$IC_{Y=6}$ : (8,4-9,3);  $IC_{Y=7}$ : (0,5-0,7).

A Tabela 3 apresenta a distribuição conjunta entre o escore de SCV em 3 categorias e variáveis sócio-demográficas. Indivíduos com maiores níveis de índice de escolaridade e de riqueza, empregados, mais jovens, solteiros e sem doenças crônicas apresentam maiores prevalências de SCV superior para ambas as situações de domicílio sendo tais relações mais pronunciadas em áreas urbanas. A prevalência de SCV superior foi muito similar entre homens e mulheres para ambas as situações de domicílio. Um percentual um pouco maior de mulheres residentes em áreas urbanas apresentou SCV inferior, ao contrário da classe intermediária de SCV. Tais diferenças foram mais acentuadas em áreas rurais. Residentes de áreas urbanas de raça/etnia branca apresentaram as maiores prevalências de SCV superior. Em áreas rurais, as prevalências de SCV foram próximas, exceto para raças/etnias amarela, indígena e ignorada. Destaca-se que tais grupos etnico-raciais foram os menos representados na amostra. As regiões Nordeste, Norte e Centro-Oeste apresentaram as maiores prevalências de SCV superior com diferenças menores entre regiões para áreas rurais.



Tabela 3: Prevalências estimadas (e Intervalos de 95% de confiança) de variáveis sociodemográficas por categorias de saúde cardiovascular - expansão da amostra PNS 2019

Variáveis	Urbana				Rural			
	Inférieur	Intermediária	Superior	Inférieur	Intermediária	Superior	Superior	
<b>Total</b>	9,7 (9,3; 10,1)	80,8 (80,3; 81,4)	9,4 (9,0; 9,9)	9,0 (8,3; 9,7)	85,7 (84,8; 86,5)	5,2 (4,7; 5,8)		
<b>Escolaridade</b>								
Até Fundamental Incompleto	17,6 (16,7; 18,5)	79,1 (78,1; 80,1)	3,3 (2,9; 3,7)	12,3 (11,3; 13,3)	85,1 (84,1; 86,2)	2,6 (2,1; 3,0)		
Fundamental completo a médio incompleto	9,9 (8,8; 11,1)	82,7 (81,3; 84,2)	7,3 (6,4; 8,3)	4,4 (3,3; 5,5)	89,4 (87,3; 91,5)	6,2 (4,6; 7,8)		
Médio completo a superior incompleto	5,7 (5,2; 6,2)	83,3 (82,4; 84,2)	11,0 (10,2; 11,7)	2,7 (1,9; 3,4)	85,7 (83,7; 87,6)	11,7 (9,8; 13,5)		
Superior completo	5,4 (4,8; 6,0)	77,2 (76,0; 78,5)	17,3 (16,2; 18,5)	5,1 (3,1; 7,1)	82,2 (77,8; 86,6)	12,7 (8,6; 16,9)		
<b>Situação de Ocupação</b>								
Empregado	6,6 (6,1; 7,0)	82,7 (82,0; 83,4)	10,7 (10,1; 11,3)	5,8 (5,2; 6,5)	88,0 (86,9; 89,0)	6,2 (5,4; 7,0)		
Sem ocupação	15,0 (14,3; 15,8)	77,6 (76,8; 78,4)	7,3 (6,8; 7,9)	12,8 (11,5; 14,0)	83,0 (81,7; 84,4)	4,2 (3,5; 4,9)		
<b>Índice de Riqueza</b>								
1º nível	13,7 (12,6; 14,7)	80,4 (79,3; 81,6)	5,9 (5,2; 6,6)	9,3 (7,8; 10,8)	86,8 (85,1; 88,5)	3,9 (2,9; 4,9)		
2º nível	10,2 (9,1; 11,2)	82,7 (81,6; 83,9)	7,1 (6,4; 7,8)	9,8 (8,1; 11,5)	86,6 (84,8; 88,4)	3,6 (2,5; 4,7)		
3º nível	10,6 (9,7; 11,5)	82,2 (81,0; 83,4)	7,2 (6,5; 8,0)	8,9 (7,5; 10,3)	84,6 (82,9; 86,3)	6,5 (5,1; 7,8)		
4º nível	8,4 (7,6; 9,3)	80,9 (79,6; 82,2)	10,7 (9,7; 11,7)	9,7 (8,2; 11,2)	85,1 (83,3; 86,9)	5,2 (4,1; 6,3)		
5º nível	7,9 (7,1; 8,7)	78,9 (77,7; 80,1)	13,2 (12,1; 14,2)	7,9 (6,7; 9,1)	85,9 (84,3; 87,5)	6,2 (5,0; 7,4)		
<b>Sexo</b>								
Masculino	8,1 (7,5; 8,6)	82,6 (81,8; 83,4)	9,3 (8,6; 10,0)	6,8 (6,1; 7,5)	88,0 (87,0; 89,1)	5,2 (4,4; 6,0)		
Feminino	11,0 (10,5; 11,6)	79,4 (78,7; 80,1)	9,6 (9,0; 10,1)	11,3 (10,1; 12,5)	83,4 (82,1; 84,6)	5,3 (4,6; 6,1)		
<b>Faixa Etária (anos)</b>								
18 a 24	0,6 (0,3; 0,9)	80,5 (78,6; 82,5)	18,9 (17,0; 20,8)	0,5 (0,0; 1,0)	85,3 (82,1; 88,6)	14,2 (11,0; 17,4)		
25 a 39	3,0 (2,5; 3,5)	85,0 (84,1; 86,0)	12,0 (11,1; 12,8)	2,3 (1,7; 2,9)	90,5 (89,3; 91,8)	7,1 (6,0; 8,3)		
40 a 59	10,9 (10,1; 11,6)	81,7 (80,9; 82,5)	7,4 (6,8; 8,0)	10,4 (9,2; 11,5)	86,2 (84,9; 87,5)	3,4 (2,7; 4,1)		
60 ou mais	21,4 (20,4; 22,5)	74,1 (73,0; 75,2)	4,4 (3,9; 5,0)	18,4 (16,6; 20,2)	79,9 (78,1; 81,7)	1,7 (1,2; 2,2)		
<b>Raça/Etnia</b>								
Parda	9,8 (9,1; 10,4)	81,4 (80,6; 82,2)	8,8 (8,2; 9,4)	9,0 (8,1; 10,0)	85,8 (84,7; 87,0)	5,2 (4,4; 5,9)		
Preta	10,8 (9,7; 11,9)	81,2 (79,7; 82,7)	8,0 (6,9; 9,1)	9,0 (6,9; 11,1)	85,7 (83,1; 88,2)	5,3 (3,7; 6,8)		
Amarela/Indígena/Ignorada	11,6 (7,8; 15,4)	78,7 (74,0; 83,4)	9,7 (6,8; 12,5)	8,4 (3,7; 13,1)	90,5 (85,8; 95,3)	1,1 (0,0; 2,2)		
Branca	9,3 (8,7; 9,9)	80,3 (79,5; 81,2)	10,4 (9,7; 11,1)	9,0 (7,9; 10,2)	85,4 (84,0; 86,9)	5,5 (4,6; 6,5)		
<b>Estado Civil</b>								
Solteiro	5,7 (5,3; 6,2)	81,6 (80,6; 82,5)	12,7 (11,9; 13,5)	5,7 (4,8; 6,7)	86,9 (85,6; 88,1)	7,4 (6,4; 8,4)		
Casado	10,5 (9,9; 11,2)	81,6 (80,8; 82,3)	7,9 (7,3; 8,5)	10,1 (9,1; 11,2)	85,9 (84,7; 87,1)	4,0 (3,3; 4,6)		
Separado/Divorciado/Viúvo	17,5 (16,3; 18,7)	76,8 (75,5; 78,1)	5,7 (5,0; 6,3)	17,1 (14,7; 19,5)	80,2 (77,7; 82,7)	2,7 (1,6; 3,8)		
<b>Doença Crônica</b>								
Pelo menos 1	14,3 (13,6; 15,0)	78,6 (77,8; 79,4)	7,1 (6,6; 7,7)	13,8 (12,5; 15,1)	82,3 (80,9; 83,7)	4,0 (3,2; 4,7)		
Sem doença	6,3 (5,9; 6,7)	82,5 (81,8; 83,2)	11,2 (10,6; 11,8)	5,5 (4,8; 6,1)	88,3 (87,3; 89,2)	6,3 (5,5; 7,0)		
<b>Região</b>								
Nordeste	9,0 (8,4; 9,5)	80,5 (79,6; 81,3)	10,6 (9,8; 11,3)	7,7 (6,7; 8,7)	86,4 (85,1; 87,7)	5,9 (5,0; 6,8)		
North	6,7 (6,1; 7,4)	83,1 (82,0; 84,2)	10,2 (9,3; 11,1)	7,0 (5,5; 8,5)	86,5 (84,5; 88,6)	6,5 (5,0; 8,0)		
Sudeste	10,2 (9,5; 10,9)	80,8 (79,9; 81,8)	9,0 (8,2; 9,7)	10,8 (9,1; 12,6)	84,8 (82,9; 86,7)	4,4 (3,2; 5,5)		
Sul	10,6 (9,7; 11,4)	80,8 (79,6; 82,0)	8,6 (7,6; 9,6)	11,9 (10,2; 13,6)	84,4 (82,6; 86,3)	3,7 (2,6; 4,8)		
Centro-Oeste	9,7 (8,7; 10,6)	80,0 (78,8; 81,3)	10,3 (9,2; 11,4)	8,9 (7,1; 10,6)	85,4 (82,1; 88,7)	5,8 (2,7; 8,8)		

Razões de chances estimadas, intervalos de 95% de confiança e p-valores calculados por modelos de regressão logística multinomial para áreas urbanas e rurais são apresentados, respectivamente, nas Tabelas 4 e 5. A categoria de referência da variável resposta foi definida por indivíduos da classe inferior de SCV (0 a 2 metas).

Em áreas urbanas (Tabela 4), comparando-se SCV superior e inferior observa-se que quanto maior era a escolaridade, maior a chance de se alcançarem SCV superior, com destaque a indivíduos de ensino superior completo que apresentaram 6,6 (IC<sub>95%</sub>: 5,2 – 8,3%) vezes maior chance de estarem na categoria superior de SCV em relação à indivíduos com ensino fundamental incompleto. Índice de riqueza também foi positivamente associado à SCV chegando a uma chance 2,24 maior de SCV superior para indivíduos do 5<sup>o</sup> comparados aos do 1<sup>o</sup> nível. Idade foi inversamente associada à SCV superior, sendo os indivíduos mais jovens os que apresentaram maiores chances de SCV superior. Indivíduos empregados, de raça branca (vs. parda), solteiros (vs. casados), sem doenças crônicas e da região nordeste (sem diferença significativa com a região norte) apresentaram as maiores chances de SCV superior vs. inferior. A variável sexo não foi significativa em tal comparação.

Com respeito aos resultados da comparação entre as categorias de SCV intermediária e inferior, ainda em áreas urbanas, observou-se as mesmas tendências porém com efeitos de menores magnitudes nas associações entre SCV e variáveis explicativas, com exceção de índice de riqueza e estado civil, que não foram significantes, e de sexo que foi significativa indicando que homens possuíam maior chance de apresentarem SCV intermediária que inferior. Além disso, indivíduos da região norte apresentaram maior chance de estarem na categoria intermediária de SCV que a nordeste.

Em áreas rurais, (Tabela 5), índice de riqueza, estado civil e a diferença entre regiões norte e nordeste não foram associados à SCV em ambas as comparações entre níveis de SCV (superior vs. inferior ou intermediária vs. inferior). Razões de chances entre faixas etárias foram bastante similares às observadas em áreas urbanas em todas as comparações. Em contrapartida, escolaridade, ocupação, sexo, raça e diferenças entre regiões exerceram, em geral, efeitos ainda mais marcantes nas chances de se alcançar maiores níveis de SCV que os observados em áreas urbanas. Indivíduos de raça parda comparada à categoria que incluiu indivíduos de raças amarela, indígena ou sem declaração de raça (“Amarela/Indígena/Ignorada”) apresentaram maior chance de estarem na classe superior de SCV. Porém, tal categoria agregada representou menos de 1% da população residente em áreas rurais (Tabela 2). Implementou-se ainda um modelo para áreas rurais sem ajustar pelas variáveis estado civil e índice de riqueza e os resultados mantiveram estimativas semelhantes (Tabela 7 no Apêndice).

Tabela 4: Razões de chances, IC95% e p-valores de modelos de regressão multinomial para a variável resposta SCV ideal em 3 categorias (inferior 0-2, intermediária 3-5 e superior ou ideal 6-7) de domicílios Urbanos – Expansão da amostra PNS 2019.

Variáveis	p-valor <sup>a</sup>	Categorias de Saúde Cardiovascular					
		Superior vs. Inferior			Intermediária vs. Inferior		
		RC <sup>b</sup>	IC95%	p-valor	RC	IC95%	p-valor
<b>Escolaridade<sup>c</sup></b>	<0,001						
EFC a EMI vs. Até EFI		1,61	(1,25-2,07)	<0,001	1,05	(0,90-1,23)	0,549
EMC a ESI vs. Até EFI		2,95	(2,41-3,62)	<0,001	1,49	(1,31-1,71)	<0,001
ESC vs. Até EFI		6,60	(5,21-8,37)	<0,001	1,78	(1,52-2,10)	<0,001
<b>Situação de Ocupação</b>	<0,001						
Empregado		1,35	(1,15-1,58)	<0,001	1,33	(1,19-1,49)	<0,001
<b>Índice de Riqueza</b>	<0,001						
2 <sup>o</sup> vs. 1 <sup>o</sup> nível		1,14	(0,92-1,42)	0,229	1,11	(0,95-1,29)	0,183
3 <sup>o</sup> vs. 1 <sup>o</sup> nível		1,30	(1,03-1,63)	0,025	1,14	(0,98-1,33)	0,094
4 <sup>o</sup> vs. 1 <sup>o</sup> nível		1,74	(1,38-2,18)	<0,001	1,17	(0,99-1,36)	0,052
5 <sup>o</sup> vs. 1 <sup>o</sup> nível		2,24	(1,75-2,86)	<0,001	1,25	(1,05-1,49)	0,012
<b>Sexo</b>	0,003						
Feminino vs. Masculino		0,97	(0,84-1,12)	0,643	0,86	(0,78-0,96)	0,005
<b>Faixa Etária (anos)</b>	<0,001						
25 a 39 vs. 18 a 24		0,11	(0,06-0,19)	<0,001	0,20	(0,11-0,34)	<0,001
40 a 59 vs. 18 a 24		0,03	(0,02-0,05)	<0,001	0,06	(0,04-0,11)	<0,001
60 ou mais vs. 18 a 24		0,01	(0,01-0,03)	<0,001	0,04	(0,02-0,07)	<0,001
<b>Raça/Etnia</b>	0,002						
Branca vs. Parda		1,28	(1,09-1,50)	0,002	1,18	(1,05-1,32)	0,006
Preta vs. Parda		0,85	(0,68-1,06)	0,144	0,92	(0,79-1,07)	0,260
Outros <sup>d</sup> vs. Parda		0,88	(0,53-1,45)	0,604	0,84	(0,57-1,24)	0,378
<b>Estado Civil</b>	<0,001						
Casado vs. Solteiro		0,73	(0,62-0,85)	<0,001	0,95	(0,85-1,07)	0,415
Outros <sup>e</sup> vs. Solteiro		0,69	(0,57-0,84)	<0,001	0,88	(0,77-1,01)	0,064
<b>Doença Crônica</b>	<0,001						
Pelo menos 1		2,46	(2,13-2,83)	<0,001	1,77	(1,61-1,95)	<0,001
<b>Região</b>	<0,001						
Norte vs. Nordeste		1,11	(0,93-1,33)	0,244	1,23	(1,07-1,40)	0,003
Sudeste vs. Nordeste		0,51	(0,43-0,60)	<0,001	0,79	(0,70-0,89)	<0,001
Sul vs. Nordeste		0,48	(0,39-0,60)	<0,001	0,78	(0,68-0,89)	<0,001
Centro-Oeste vs. Nordeste		0,62	(0,51-0,75)	<0,001	0,79	(0,69-0,91)	<0,001

Fonte: Pesquisa Nacional de Saúde 2019.

<sup>a</sup> Análise do efeito global da variável;

<sup>b</sup> RC: Razão de chances;

<sup>c</sup> EFI: Ensino Fundamental Incompleto; EFC: Ensino Fundamental Completo; EMI: Ensino Médio Incompleto; EMC: Ensino Médio Completo; ESI: Ensino Superior Incompleto; ESC: Ensino Superior Completo.

<sup>d</sup> Outras raças/etnias incluem Amarela, Indígena e Ignorada;

<sup>e</sup> Outros estados civis incluem Separado, Divorciado e Viúvo.

Para fins de avaliação de qualidade de ajuste, o coeficiente  $R^2$  de McFadden foi obtido. Tal coeficiente foi relativamente baixo no modelo para áreas urbanas ( $R^2 = 0,10$ ) e em ambos os modelos para áreas rurais ( $R^2 = 0,12$  para os modelos cujos resultados são apresentados nas Tabelas 5 e 7).

Tabela 5: Razões de chances, IC95% e p-valores de modelos de regressão multinomial para a variável resposta SCV ideal em 3 categorias (inferior 0-2, intermediária 3-5 e superior ou ideal 6-7) de domicílios Rurais – Expansão da amostra PNS 2019.

Variáveis	p-valor <sup>a</sup>	Categorias de Saúde Cardiovascular					
		Superior vs, Inferior			Intermediária vs, Inferior		
		RC <sup>b</sup>	IC95%	p-valor	RC	IC95%	p-valor
<b>Escolaridade<sup>c</sup></b>	<0,001						
EFC a EMI vs, Até EFI		1,87	(1,19-2,93)	0,007	1,33	(0,98-1,81)	0,069
EMC a ESI vs, Até EFI		5,49	(3,50-8,61)	<0,001	2,01	(1,46-2,78)	<0,001
ESC vs, Até EFI		6,38	(3,38-12,05)	<0,001	1,52	(0,96-2,41)	0,072
<b>Situação de Ocupação</b>	0,006						
Empregado		1,54	(1,07-2,20)	0,019	1,37	(1,12-1,66)	0,002
<b>Índice de Riqueza</b>	0,265						
2 <sup>o</sup> vs, 1 <sup>o</sup> nível		0,78	(0,49-1,24)	0,288	0,92	(0,71-1,19)	0,531
3 <sup>o</sup> vs, 1 <sup>o</sup> nível		1,32	(0,87-2,02)	0,198	0,98	(0,77-1,26)	0,886
4 <sup>o</sup> vs, 1 <sup>o</sup> nível		1,00	(0,65-1,54)	0,984	0,94	(0,72-1,22)	0,644
5 <sup>o</sup> vs, 1 <sup>o</sup> nível		1,16	(0,69-1,95)	0,588	1,01	(0,75-1,35)	0,965
<b>Sexo</b>	<0,001						
Feminino vs, Masculino		0,55	(0,40-0,76)	<0,001	0,60	(0,50-0,72)	<0,001
<b>Faixa Etária (anos)</b>	<0,001						
25 a 39 vs, 18 a 24		0,13	(0,04-0,36)	<0,001	0,25	(0,09-0,67)	0,006
40 a 59 vs, 18 a 24		0,03	(0,01-0,07)	<0,001	0,07	(0,03-0,19)	<0,001
60 ou mais vs, 18 a 24		0,01	(0,00-0,04)	<0,001	0,05	(0,02-0,12)	<0,001
<b>Raça/Etnia</b>	0,004						
Branca vs, Parda		1,61	(1,16-2,23)	0,005	1,28	(1,04-1,56)	0,018
Preta vs, Parda		1,06	(0,67-1,68)	0,792	1,00	(0,76-1,33)	0,998
Outros <sup>d</sup> vs, Parda		0,20	(0,06-0,71)	0,013	0,97	(0,50-1,88)	0,919
<b>Estado Civil</b>	0,775						
Casado vs, Solteiro		0,94	(0,68-1,31)	0,716	1,05	(0,84-1,32)	0,667
Outros <sup>e</sup> vs, Solteiro		0,87	(0,50-1,50)	0,613	0,94	(0,70-1,25)	0,673
<b>Doença Crônica</b>	<0,001						
Pelo menos 1		2,14	(1,60-2,88)	<0,001	1,92	(1,63-2,25)	<0,001
<b>Região</b>	<0,001						
Norte vs, Nordeste		0,99	(0,67-1,46)	0,962	0,95	(0,73-1,25)	0,730
Sudeste vs, Nordeste		0,41	(0,26-0,64)	<0,001	0,62	(0,48-0,80)	<0,001
Sul vs, Nordeste		0,26	(0,16-0,44)	<0,001	0,53	(0,40-0,69)	<0,001
Centro-Oeste vs, Nordeste		0,64	(0,34-1,19)	0,156	0,73	(0,55-0,97)	0,029

Fonte: Pesquisa Nacional de Saúde 2019.

<sup>a</sup> Análise do efeito global da variável;

<sup>b</sup> RC: Razão de chances;

<sup>c</sup> EFI: Ensino Fundamental Incompleto; EFC: Ensino Fundamental Completo; EMI: Ensino Médio Incompleto; EMC: Ensino Médio Completo; ESI: Ensino Superior Incompleto; ESC: Ensino Superior Completo.

<sup>d</sup> Outras raças/etnias incluem Amarela, Indígena e Ignorada;

<sup>e</sup> Outros estados civis incluem Separado, Divorciado e Viúvo.

## 4 Discussão

O presente estudo teve por objetivo estimar a prevalência de SCV ideal e identificar associações entre SCV e fatores socioeconômicos na população adulta brasileira separadamente em domicílios rurais e urbanos.

Os resultados evidenciaram a baixíssima prevalência de SCV ideal na população brasileira sobretudo em domicílios rurais (0,2% e 0,6% em domicílios rurais e urbanos, respectivamente, Figura 1). Uma menor aderência a fatores de proteção cardiovascular associada a uma situação mais desfavorável socioeconômica podem levar a maiores taxas de mortalidade em áreas rurais (ARRUDA; MAIA; ALVES, 2018). Além disso, a falta de acesso e baixa qualidade dos serviços de saúde disponíveis podem contribuir para um aumento nas prevalências de condições crônicas, prejudicando a prevenção e o tratamento das mesmas (FELISBINO-MENDES et al., 2014). Em um estudo no Brasil (KASSOUF, 2005), por exemplo, observou-se que áreas urbanas possuem uma maior oferta de serviços de saúde e melhor acesso aos centros de saúde que áreas rurais, o que levaria a menores prevalências de SCI em áreas rurais comparado a áreas urbanas.

O modelo de regressão logística multinomial permitiu identificar, com significância estatística que, em ambas as áreas urbanas e rurais, possuir maiores níveis de escolaridade e estar empregado aumentam as chances de indivíduos estarem em classes superiores de SCV (6 a 7 metas), ajustando-se por covariáveis (Tabelas 4 e 5). Além disso, nota-se que indivíduos mais jovens, da raça branca, solteiros, sem doenças crônicas e das regiões norte e nordeste tendem a ter melhores condições de SCV. Observou-se ainda que, em geral, os efeitos das variáveis significantes nos modelos para áreas rurais foram ainda mais acentuados que, para áreas urbanas.

Destaca-se que a variável preditora de maior impacto na SCV foi a escolaridade, especialmente em áreas rurais, o que pode ter contribuído para uma menor prevalência de SCV ideal em indivíduos residentes de áreas rurais tendo em vista os menores níveis de escolaridade observados. O percentual de indivíduos com ensino superior completo em áreas rurais foi de apenas 3,5% (IC<sub>95%</sub>: 3,1 – 4,7%) e o de ensino médio completo, de 20,3 % (IC<sub>95%</sub>: 19,1 – 21,4%). Tais percentuais foram muito mais elevados em áreas urbanas, com 19,2% (IC<sub>95%</sub>: 18,4 – 20,1%) para ensino superior completo e com 37,7% (IC<sub>95%</sub>: 37,2 – 38,4%) para ensino médio completo. Tais resultados evidenciam a maior vulnerabilidade social e de saúde cardiovascular da população residente em áreas rurais.

Observou-se ainda que no modelo para áreas rurais, indivíduos do sexo masculino apresentaram maiores chances de estarem nas categorias superior e intermediária (comparadas à inferior) e no modelo de área urbana a estimativa foi significativa apenas na comparação da classe intermediária vs. inferior. Tal resultado nos surpreendeu ini-

cialmente, pois indivíduos do sexo masculino possuem maiores riscos de desenvolverem doenças cardiovasculares (GAO et al., 2019). Especulamos que tal achado possa ser explicado, em parte, por maiores taxas de detecção de condições de saúde entre mulheres, o que é corroborado por um estudo brasileiro (PINHEIRO et al., 2002) no qual a auto-referência de condição de saúde foi declarada pior entre as mulheres, sendo ainda mais evidente em mulheres residentes em áreas rurais. Ainda, tal estudo mostrou que as mulheres procuram mais os serviços de saúde para exames de rotina ou prevenção, enquanto os homens buscam mais cuidados curativos, o que poderia justificar maiores frequências de diagnóstico de doenças em indivíduos do sexo feminino e, por consequência, piores escores de SCV.

Outro resultado a ser ressaltado foi a falta de associação entre SCV e o índice de Riqueza no modelo de regressão logística multinomial apenas para áreas rurais. Tais achados podem ser explicados pelos menores percentuais de domicílios rurais (comparados aos de domicílios urbanos) com os bens de consumo utilizados na construção do índice de riqueza (resultados não apresentados). Além disso, os cinco níveis do índice de riqueza são obtidos pela ordenação de escores de uma variável latente, separados por quintis. No presente estudo, o índice de riqueza foi construído separadamente para cada situação de domicílio. Quando tal índice foi calculado considerando toda a população (urbana e rural), a maior parte dos indivíduos de áreas rurais estavam nos menores níveis de índice de riqueza (resultados não apresentados). Ou seja, o índice de riqueza parece não ter discriminado bem os indivíduos quanto à riqueza acumulada em áreas rurais, sendo a vasta maioria pertencente às classes com menor riqueza.

A principal limitação do presente estudo envolve o uso de medidas autorreferidas, o que pode levar a prevalências subestimadas de diagnósticos e, conseqüentemente, superestimadas de SCV ideal. Metas biológicas seriam melhor avaliadas se aferidas, pois a auto declaração depende da memória e esclarecimento do participante e do acesso a serviços de saúde (para diagnóstico médico de diabetes, hipercolesterolemia e hipertensão arterial). É possível que o impacto da utilização de medidas autoreferidas no lugar de aferidas seja ainda maior em áreas rurais (ANDRADE et al., 2013). Contudo, tais tipos de informação já foram utilizadas em outros estudos nacionais e internacionais apresentando boa correlação com as medidas biológicas (MOREIRA et al., 2020).

Além disso, a amostra tão inferior de indivíduos da área rural, seguindo as frequências populacionais, causa um aumento nos comprimentos dos intervalos de confiança das prevalências apresentadas nesse trabalho, o que pode ter contribuído para a observada redução no número de preditores estatisticamente significantes nas análises de regressão logística. Ainda, alguns níveis de variáveis sociodemográficas podem não ter quantidade suficiente de indivíduos na amostra, como por exemplo para a raça/etnia amarela/indígena/ignorada estimada em apenas 1,5% da população, sendo ainda menos

expressiva na área rural (0,9%).

Mesmo com as limitações citadas, o presente estudo apresenta resultados de suma importância, pois são representativos da população adulta brasileira devido ao cuidadoso planejamento amostral e validação da metodologia desenvolvida pelo Instituto Brasileiro de Geografia e Estatística juntamente com o Ministério da Saúde. A amostra apresenta informações atualizadas sobre a saúde e fatores sociais, econômicos e demográficos no Brasil, sendo um conjunto de dados valiosos que podem ajudar a direcionar e melhorar as intervenções para alcançar uma maior prevalência de saúde cardiovascular ideal na população adulta brasileira.

## 5 Conclusão

Conclui-se que a prevalência de saúde cardiovascular ideal no Brasil é muito baixa, sobretudo quando observada em indivíduos que residem em domicílios de áreas rurais. A escolaridade exerceu o maior impacto na SCV, após ajuste pelas demais variáveis explicativas, sendo ainda mais importante em áreas rurais.

Tais resultados apresentam evidência irrefutável da importância de divulgação e acompanhamento constante da SCV da população, uma vez que uma boa saúde cardiovascular (simplificada pelas metas de SCV) previne ou retarda o desenvolvimento de doenças cardiovasculares assim como melhoram a qualidade de vida da população. As fortes associações entre SCV e determinantes socioeconômicos sinalizam a necessidade de investimento e apoio de políticas públicas em fatores socioeconômicos como caminho indireto para a prevenção das doenças cardiovasculares, sobretudo em áreas com maiores desigualdades sociais, como as regiões rurais.

Sugere-se, para estudos futuros o acompanhamento das prevalências de SCV ao longo do tempo e a verificação dos resultados quanto à prevenção das doenças cardiovasculares e taxas de mortalidade associadas. Ainda, técnicas de avaliação de qualidade de ajuste também podem ser exploradas de forma mais aprofundada no contexto de análises de modelos de regressão logística multinomial em estudos com delineamentos envolvendo amostragem complexa.



## 6 Referências

### Referências

- AGRESTI, A. Categorical Data Analysis. [S.l.]: John Wiley & Sons, Inc, 2013.
- AGRESTI, A. An Introduction to Categorical Data Analysis. [S.l.]: John Wiley & Sons, Inc, 2019.
- ANDRADE, L. d. et al. Regional disparities in mortality after ischemic heart disease in a brazilian state from 2006 to 2010. PloS one, Public Library of Science San Francisco, USA, v. 8, n. 3, p. e59363, 2013.
- ARRUDA, N. M.; MAIA, A. G.; ALVES, L. C. Desigualdade no acesso à saúde entre as áreas urbanas e rurais do brasil: uma decomposição de fatores entre 1998 a 2008. Cadernos de Saúde Pública, SciELO Brasil, v. 34, 2018.
- CROSS, S. H. et al. Rural-urban differences in cardiovascular mortality in the us, 1999-2017. JAMA, American Medical Association, v. 323, n. 18, p. 1852–1854, 2020.
- FELISBINO-MENDES, M. S. et al. Avaliação dos fatores de risco cardiovasculares em uma população rural brasileira. Cadernos de Saúde Pública, SciELO Public Health, v. 30, p. 1183–1194, 2014.
- GAO, Z. et al. Gender differences in cardiovascular disease. Medicine in Novel Technology and Devices, Elsevier, v. 4, p. 100025, 2019.
- HAVRANEK, E. P. et al. Social determinants of risk and outcomes for cardiovascular disease: a scientific statement from the american heart association. Circulation, Am Heart Assoc, v. 132, n. 9, p. 873–898, 2015.
- IBGE. 2020. Acessado em: 2022-09-05. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/saude/29540-2013-pesquisa-nacional-de-saude.html?edicao=9177&t=microdados>.
- IBGE, B. Pesquisa nacional de saúde: 2019; informações sobre domicílios, acesso e utilização dos serviços de saúde: Brasil, grandes regiões e unidades da federação/ibge, coordenação de trabalho e rendimento. Rio de Janeiro: IBGE, 2020.
- JOHNSON, R. A.; WICHERN, D. W. et al. Applied multivariate statistical analysis. [S.l.]: Prentice hall Upper Saddle River, NJ, 2002. v. 5.
- KASSOUF, A. L. Acesso aos serviços de saúde nas áreas urbana e rural do brasil. Revista de economia e Sociologia Rural, SciELO Brasil, v. 43, p. 29–44, 2005.
- LLOYD-JONES, D. M. et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the american heart association’s strategic impact goal through 2020 and beyond. Circulation, Am Heart Assoc, v. 121, n. 4, p. 586–613, 2010.

- MORAN, A. E. et al. Temporal trends in ischemic heart disease mortality in 21 world regions, 1980 to 2010: the global burden of disease 2010 study. Circulation, Am Heart Assoc, v. 129, n. 14, p. 1483–1492, 2014.
- MOREIRA, A. D. et al. Saúde cardiovascular e validação do escore autorreferido no brasil: uma análise da pesquisa nacional de saúde. Ciência & Saúde Coletiva, SciELO Public Health, v. 25, p. 4259–4268, 2020.
- MOTTA, A. C. S. V. et al. Prevalência de saúde cardiovascular ideal na população adulta brasileira – pesquisa nacional de saúde 2019. 2022.
- OXFORD. tetrachoric correlation coefficient. 2008. Acessado em: 2023-03-14. Disponível em: <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803103253423>.
- PINHEIRO, R. S. et al. Gênero, morbidade, acesso e utilização de serviços de saúde no brasil. Ciência & saúde coletiva, SciELO Brasil, v. 7, p. 687–707, 2002.
- PNS. 2019. Acessado em: 2022-07-21. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html?=&t=o-que-e>.
- RENCHER, A. C. Methods of Multivariate Analysis. [S.l.]: John Wiley Sons, Inc, 2002. v. 2.
- ROTH, G. A. et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. The Lancet, Elsevier, v. 392, n. 10159, p. 1736–1788, 2018.
- ROTH, G. A. et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the gbd 2019 study. Journal of the American College of Cardiology, American College of Cardiology Foundation Washington DC, v. 76, n. 25, p. 2982–3021, 2020.
- RUTSTEIN, S. O. Steps to constructing the new dhs wealth index. Rockville, MD: ICF International, 2015.
- SAS proc surveylogistic. 2019. Acessado em: 2023-03-06. Disponível em: [https://documentation.sas.com/doc/en/pgmsascdc/9.4.3.3/statug/statug\\_surveylogistic\\_details40.htm#statug\\_surveylogistic\\_surveylogistic1h](https://documentation.sas.com/doc/en/pgmsascdc/9.4.3.3/statug/statug_surveylogistic_details40.htm#statug_surveylogistic_surveylogistic1h).
- SBC. Reduzir os óbitos por doenças cardiovasculares é o maior desafio da cardiologia. 2020. Acessado em: 2022-09-16. Disponível em: <https://www.portal.cardiol.br/post/reduzir-os-bitos-por-doenas-cardiovasculares--o-maior-desafio-da-cardiologia>.
- SCHULTZ, W. M. et al. Socioeconomic status and cardiovascular outcomes: challenges and interventions. Circulation, Am Heart Assoc, v. 137, n. 20, p. 2166–2178, 2018.
- SILVA, D. P. e P. N. Análise de dados amostrais complexos. <https://djalmapessoa.github.io/adac/index.html>, 2018.
- WFP. VAM Guidance Paper: Creation of a Wealth Index. 2017. Acessado em: 2022-09-05. Disponível em: <https://www.wfp.org/publications/creation-wealth-index-june-2017>.

## 7 Apêndice

### 7.1 Resultados Índice de Riqueza

Tabela 6: Variáveis Finais para construção do escore de riqueza

Urbana		Rural	
Variáveis Retiradas	Variáveis Finais	Variáveis Retiradas	Variáveis Finais
possuir motocicleta	tipo de piso	possuir motocicleta	tipo de piso
tipo de abastecimento de água	tipo de telhado	tipo de abastecimento de água	tipo de telhado
número de pessoas por cômodo	possuir carro	número de pessoas por cômodo	possuir carro
número de banheiros	possuir celular	número de banheiros	possuir celular
possuir empregado(a) doméstico(o)	possuir computador	possuir empregado(a) doméstico(o)	possuir computador
tipo de parede	possuir microondas	tipo de parede	possuir microondas
fonte de água potável	máquina de lavar roupa	fonte de água potável	máquina de lavar roupa
	destino do lixo	possuir telefone fixo	destino do lixo
	escoamento sanitário	escoamento sanitário	possuir internet
	possuir telefone fixo		
	possuir internet		

### 7.2 Resultado modelo final para áreas rurais

Tabela 7: Razões de chances, IC95% e p-valores de modelo de regressão multinomial para a variável resposta SCV em 3 categorias em domicílios Rurais sem as variáveis estado civil e índice de riqueza – Expansão da amostra PNS 2019

Variáveis	p-valor geral <sup>a</sup>	Comparações de Categorias de Saúde Cardiovascular					
		Superior vs. Inferior		Intermediária vs. Inferior			
		RC <sup>b</sup>	IC95%	p-valor	RC	IC95%	p-valor
<b>Escolaridade<sup>c</sup></b>	<0,001						
EFC a EMI vs. Até EFI		1,94	(1,24-3,04)	0,004	1.34	(0.99-1.81)	0.058
EMC a ESI vs. Até EFI		5,83	(3,75-9,06)	<0,001	2.05	(1.49-2.81)	<0,001
ESC vs. Até EFI		6,82	(3,67-12,69)	<0,001	1.57	(1.00-2.46)	0.050
<b>Situação de Ocupação</b>	0,005						
Empregado		1,55	(1,09-2,22)	0,016	1.38	(1.13-1.67)	0.001
<b>Sexo</b>	<0,001						
Feminino vs. Masculino		0,54	(0,40-0,75)	<0,001	0.59	(0.50-0.71)	<0,001
<b>Faixa Etária (anos)</b>	<0,001						
25 a 39 vs. 18 a 24		0,13	(0,04-0,36)	<0,001	0.25	(0.09-0.68)	0.007
40 a 59 vs. 18 a 24		0,02	(0,01-0,07)	<0,001	0.07	(0.03-0.19)	<0,001
60 ou mais vs. 18 a 24		0,01	(0,00-0,03)	<0,001	0.05	(0.02-0.12)	<0,001
<b>Raça/Etnia</b>	0,003						
Branca vs. Parda		1,64	(1,18-2,27)	0,003	1.29	(1.05-1.58)	0.014
Preta vs. Parda		1,08	(0,68-1,71)	0,745	1.00	(0.75-1.32)	0.975
Outras <sup>d</sup> vs. Parda		0,20	(0,06-0,70)	0,012	0.98	(0.50-1.90)	0.941
<b>Doença Crônica</b>	<0,001						
Pelo menos 1		2,14	(1,59-2,88)	<0,001	1.91	(1.63-2.24)	<0,001
<b>Região</b>	<0,001						
Norte vs. Nordeste		0,98	(0,66-1,46)	0,929	0.96	(0.73-1.26)	0.760
Sudeste vs. Nordeste		0,41	(0,27-0,63)	<0,001	0.62	(0.49-0.79)	<0,001
Sul vs. Nordeste		0,27	(0,17-0,43)	<0,001	0.54	(0.42-0.68)	<0,001
Centro-Oeste vs. Nordeste		0,65	(0,35-1,22)	0,179	0.74	(0.57-0.96)	0.023

Fonte: Pesquisa Nacional de Saúde 2019.

<sup>a</sup> Análise do efeito global da variável;

<sup>b</sup> RC: Razão de chances;

<sup>c</sup> EFI: Ensino Fundamental Incompleto; EFC: Ensino Fundamental Completo; EMI: Ensino Médio Incompleto; EMC: Ensino Médio Completo; ESI: Ensino Superior Incompleto; ESC: Ensino Superior Completo.

<sup>d</sup> Outras raças/etnias incluem Amarela, Indígena e Ignorada.

### 7.3 Códigos em R

```
## Pacotes Usados —————
library(dplyr)
library(survey)
library(psych)
library(ggplot2)
library(MASS)
require(gridExtra)
library(fabricatr)
# carregando PNS 2019
pns <- read.SAScii(fn = "PNS_2019.txt", sas_ri = "input_PNS_2019.sas"
, beginline = 1)

## # Criterios De Inclus o —————
table(pns$V0025A)
pns_ci <- pns[pns$V0025A==1,]

### 1- (idade)
class(pns_ci$C008)
summary(pns_ci$C008)
table(pns_ci$C008<18)
pns_ci <- pns_ci[pns_ci$C008 >= 18,] #91683

summary(pns_ci$C008)

### 2- M001 = questionario realizado
class(pns$M001)
table(pns$M001)
pns_ci <- pns_ci[pns_ci$M001 == "1" ,] #90846

### 3-
## Morador apto a aquele capaz de responder
(falar) por si pr prio); 1=sim
class(pns_ci$M00203)
table(pns_ci$M00203)
pns_ci <- pns_ci[pns_ci$M00203 == "1" ,] #89135
```

```
### 4- P005 = 2 ou missing - gravida
```

```
class(pns_ci$P005)
```

```
table(pns_ci$P005)
```

```
pns_ci <- pns_ci[pns_ci$P005 %in% c("1", "2"),] #86004
```

```
table(pns_ci$P005)
```

```
### 5- Q03001 nao missing e Q060 nao missing e Q00201 nao missing
```

```
## Q03001 -Algun medico ja lhe deu o diagnostico de diabetes?
```

```
## Q060 - Algun medico ja lhe deu o diagnostico de colesterol alto?
```

```
## Q00201 - Algun medico ja lhe deu o diagnostico de hipertensao?
```

```
class(pns_ci$Q03001)
```

```
table(pns_ci$Q03001)
```

```
class(pns_ci$Q060)
```

```
table(pns_ci$Q060)
```

```
class(pns_ci$Q00201)
```

```
table(pns_ci$Q00201)
```

```
pns_ci <- pns_ci[pns_ci$Q03001 != "1" & pns_ci$Q060 != "1" & pns_ci$Q00201 !=
```

```
which(is.na(pns_ci$P00404));which(is.na(pns_ci$P00104))
```

```
pns_ci <- pns_ci[-64235,]
```

```
# Covari veis -----
```

```
dados19 <- pns_ci
```

```
rm(pns_ci)
```

```
rm(pns)
```

```
### Sexo
```

```
dados19 <- dados19 %>% mutate(sexo = case_when(
```

```
  (C006 == '1') ~ '0Mas',
```

```
  (C006 == '2') ~ '1Fem'))
```

```
dados19$sexo <- as.factor(dados19$sexo)
```

```
### Idade
```

```
dados19 <- dados19 %>% mutate(idade = case_when(
```

```
  (C008>= 18 & C008 <= 24) ~ '0_18_a_24',
```

```
  (C008>= 25 & C008 <= 39) ~ '1_25_a_39',
```

```
  (C008>= 40 & C008 <= 59) ~ '2_40_a_59',
```

```
  (C008>= 60) ~ '3_>=60')
```

```
))
dados19$idade <- as.factor(dados19$idade)

### Raça - agrupando amarela + indígena + ignorado
dados19 <- dados19 %>% mutate(raca = case_when(
  C009 == '1' ~ 1,
  C009 == '2' ~ 2,
  C009 == '3' ~ 3,
  C009 == '4' ~ 0,
  C009 == '5' ~ 3,
  C009 == '9' ~ 3
))

dados19 <- dados19 %>% mutate(raca = case_when(
  raca == '1' ~ '3_Branca',
  raca == '2' ~ '1_Preta',
  raca == '3' ~ '2_Ama/Ind/Ign',
  raca == '0' ~ '0_Parda'
))
dados19$raca <- as.factor(dados19$raca)

### Estado Civil - agrupando Separado ou desquitado(a) judicialmente
+ Divorciado(a) + Viúvo(a)
dados19 <- dados19 %>% mutate(estado_civil = case_when(
  C011 == '1' ~ 1,
  C011 == '2' ~ 2,
  C011 == '3' ~ 2,
  C011 == '4' ~ 0,
))

dados19 <- dados19 %>% mutate(estado_civil = case_when(
  (estado_civil == '0') ~ '0_solteiro',
  (estado_civil == '1') ~ '1_Casado',
  (estado_civil == '2') ~ '2_Sepa/Divo/Viu'
))
dados19$estado_civil <- as.factor(dados19$estado_civil)
round(prop.table(table(dados19$estado_civil)),4)*100
addmargins(table(dados19$estado_civil))
```

```

### Regi es agrupadas por Estado (V0001)
dados19$V0001 <- as.numeric(dados19$V0001)
dados19 <- dados19 %>% mutate(regioes = case_when(
  (V0001 >= 11 & V0001 <= 17) ~ 0,
  (V0001 >= 21 & V0001 <= 29) ~ 1,
  (V0001 >= 31 & V0001 <= 35) ~ 2,
  (V0001 >= 41 & V0001 <= 43) ~ 3,
  (V0001 >= 50 & V0001 <= 53) ~ 4
))

dados19 <- dados19 %>% mutate(regioes = case_when(
  (regioes == '0') ~ '1_Norte',
  (regioes == '1') ~ '0_Nordeste',
  (regioes == '2') ~ '2_Sudeste',
  (regioes == '3') ~ '3_Sul',
  (regioes == '4') ~ '4_Centro-Oeste'
))
dados19$regioes <- as.factor(dados19$regioes)
round(prop.table(table(dados19$regioes)),4)*100
addmargins(table(dados19$regioes))

### Doen a Cronica
dados19 <- dados19 %>% mutate(doenca_cronica = case_when(
  (Q074 == '1' | Q079 == '1' | Q084 == '1' | Q088 == '1' | Q11604 == '1'
  | Q120 == '1' | Q124 == '1' | Q128 == '1' | Q092 == '1' | Q11006 == '1') ~ 1,
  TRUE ~ 0
))
round(prop.table(table(dados19$doenca_cronica)),4)*100
addmargins(table(dados19$doenca_cronica))
dados19 <- dados19 %>% mutate(doenca_cronica = case_when(
  (doenca_cronica == '0') ~ '1_Nao',
  (doenca_cronica == '1') ~ '0_Sim'
))
dados19$doenca_cronica <- as.factor(dados19$doenca_cronica)

### Situa o domicilio
dados19$Sit_domicilio <- dados19$V0026
dados19$Sit_domicilio <- ordered(dados19$Sit_domicilio ,

```

```

                                levels = c("1", "2"),
                                labels = c("Urbano", "Rural"))
table(dados19$Sit_domicilio)

### Doen a Cardiovascular
dados19 <- dados19 %>% mutate(dcv = case_when(
  (Q06306 == '1' | Q068 == '1') ~ 1,
  (Q06306 == '2' & Q068 == '2') ~ 0
))
round(prop.table(table(dados19$dcv)), 4) * 100

# Constru o do SCI —————

##### Fatores Saud veis

### Favor veis para press o arterial
dados19 <- dados19 %>% mutate(fav_pa = case_when(
  (Q00201 == "2") ~ 1,
  TRUE ~ 0
))
round(prop.table(table(dados19$fav_pa)), 4) * 100

### Favor vel para glicose
dados19 <- dados19 %>% mutate(fav_glu = case_when(
  Q03001 == '2' ~ 1,
  TRUE ~ 0
))
round(prop.table(table(dados19$fav_glu)), 4) * 100

### Favor vel para colesterol
dados19 <- dados19 %>% mutate(fav_col = case_when(
  Q060 == '1' ~ 0,
  Q060 == '2' ~ 1
))
round(prop.table(table(dados19$fav_col)), 4) * 100

### Soma dos fatores saud veis
dados19 <- dados19 %>% mutate(fatores_sau = fav_pa +

```



```

                                fav_glu + fav_col)
round(prop.table(table(dados19$fatores_sau)),4)*100

### Fatores saudaveis binarios
dados19 <- dados19 %>% mutate(factors_bin = case_when(
  fatores_sau <= 2 | dev == '1' ~ 0,
  fatores_sau == 3 & dev == '0' ~ 1
))

##### Comportamentos Saud veis

### Fatores favor veis IMC

imc <- dados19$P00104/(dados19$P00404^2)
dados19<- dados19%>% mutate(fav_imc = case_when(
  imc < 0.0025 ~ 1,
  imc >= 0.0025 ~ 0
))

### Fatores favor veis exerc cios
dados19<- dados19%>% mutate(fav_exerc = case_when(
  (P03701*60 + P03702)*P035 >=150 ~ 1,
  TRUE ~ 0
))

### Fatores favor veis dieta

# PEIXE
dados19<- dados19%>% mutate(fav_peixe = case_when(
  P015 < 2 ~ 0,
  P015 >= 2 ~ 1
))

# A UCAR
dados19<- dados19%>% mutate(fav_acucar = case_when(
  P02001 + P02002 < 5 ~ 1,
  P02001 + P02002 >= 5 ~ 0,

```

```
TRUE ~ 1
))

# FRUTAS E VEGETAIS
dados19<- dados19%>% mutate(fav_veg = case_when(
  P00901 == 7 & P018 == 7 ~ 1,
  TRUE ~ 0
))

# SAL
dados19<- dados19%>% mutate(fav_sal = case_when(
  P02601 <= 2 ~ 0,
  P02601 >= 3 ~ 1
))

# Score Dieta
dados19<- dados19%>% mutate(dieta_score = fav_peixe+
                             fav_acucar + fav_veg + fav_sal)

# Favoravel Dieta
dados19<- dados19%>% mutate(fav_dieta= case_when(
  dieta_score == '4' ~ 1,
  TRUE ~ 0
))

### Fumante atual
dados19<- dados19%>% mutate(fumante = case_when(
  P050 == '3' ~ 1, #n o fumo atualmente
  P050 != '3' ~ 0
))

### ex fumante
dados19<- dados19 %>% mutate(testes = case_when(
  P052 > 1 ~ 1,
  P052 <= 1 ~ 0,
))
```

```

dados19<- dados19%>% mutate(ex_fumante = case_when(
  P052 == '1' & P05901 > '01' ~ 1,
  P052 == '2' & P05901 > '01' ~ 1,
  P052 == '3' ~ 1,
  P052 == '1' & P05901 <= '01' ~ 0,
  P052 == '2' & P05901 <= '01' ~ 0,
))

### Favoraveis a fumo
dados19 <- dados19 %>% mutate(fav_fumo = case_when(case_when(
  P050 == '3' ~ 1,
  P050 != '3' ~ 0) == '1' &
  case_when(
    P052 == '1' & P05901 > '01' ~ 1,
    P052 == '2' & P05901 > '01' ~ 1,
    P052 == '3' ~ 1,
    P052 == '1' & P05901 <= '01' ~ 0,
    P052 == '2' & P05901 <= '01' ~ 0) == '1' ~ 1,
  TRUE ~ 0
))

### Score Comportamentos saud veis

dados19<- dados19%>% mutate(comp_score = fav_imc+
  fav_exerc + fav_fumo + fav_dieta)

# ### Fatores favor veis saude em categorias
#
# ### Comportamentos saudaveis em fatores binarios
dados19 <- dados19 %>% mutate(comp_bin = case_when(
  comp_score <= 2 ~ 0,
  TRUE ~ 1))

##### ESCORE DE SAUDE CARD IDEAL

dados19<- dados19%>% mutate(cvh_score = fatores_sau +
  comp_score)

```

```

dados19$cvh_score <- as.factor(dados19$cvh_score)
##### SCI binario
dados19$cvh_score_bin <- 0
dados19$cvh_score_bin[dados19
                        $cvh_score ==6] <- 1
dados19$cvh_score_bin[dados19
                        $cvh_score ==7] <- 1
table(dados19$cvh_score_bin)
table(dados19$cvh_score_bin)/77494

##### SCI categorico
dados19<- dados19 %>% mutate(cvh_cat = case_when(
  cvh_score < 4 ~ '0Baixo',
  cvh_score >= 4 & cvh_score < 7 ~ '1Inter',
  cvh_score >= 7 ~ '2Alto'
))

# Escolaridade e Ocupação - Fatores Socioeconomicos
# Construção Wealth Index #####
wealth <- dados19

### water_supply - Fonte de Agua - A00501
wealth <- wealth %>% mutate(water_supply = case_when(
  A005010 %in% c(2,3,4,5,6,7,8) ~ 0,
  A005010 == 1 ~ 1))

### floor - Material do piso principal- A004010
wealth <- wealth %>% mutate(floor = case_when(
  A004010 %in% c(3,4,5,9) ~ 0,
  A004010 %in% c(1,2) ~ 1))

### walls - Principal parede - A002010
wealth <- wealth %>% mutate(walls = case_when(
  A002010 %in% c(2,3,5,6,7) ~ 0,
  A002010 %in% c(1,4) ~ 1))

### roof - Principal telhado - A003010

wealth <- wealth %>% mutate(roof = case_when(

```

```

A003010 == 3 & A001 == 2 ~ 1,
A003010 == 3 & A001 != 2 ~ 0,
A003010 %in% c(1,3,5,6,9) ~ 0,
A003010 %in% c(2,4) ~ 1))
### housekeeper - Servi  os Dom  sticos - A02101
wealth <- wealth %>% mutate(housekeeper = case_when(
  A02101 == 1 ~ 1,
  A02101 == 2 ~ 0))

### car - Autom  veis - A018027
wealth <- wealth %>% mutate(car = case_when(
  A018027 == 2 ~ 0,
  A018027 == 1 ~ 1))

### cell_phone - Telefone Celular - A018019
wealth <- wealth %>% mutate(cell_phone = case_when(
  A018019 == 1 ~ 1,
  A018019 == 2 ~ 0))

### computer - Computador - A018023
wealth <- wealth %>% mutate(computer = case_when(
  A018023 == 1 ~ 1,
  A018023 == 2 ~ 0))

### microwave - Microondas - A018021
wealth <- wealth %>% mutate(microwave = case_when(
  A018021 == 1 ~ 1,
  A018021 == 2 ~ 0))

### washing - Maquina de lavar roupa - A018015
wealth <- wealth %>% mutate(washing = case_when(
  A018015 == 1 ~ 1,
  A018015 == 2 ~ 0))

### water_drink - Agua para beber - A009010
wealth <- wealth %>% mutate(water_drink = case_when(
  A009010 %in% c(2,3,4,6) ~ 0,
  A009010 %in% c(1,5) ~ 1))

```

```
### sanitary_drain - Esgoto - A01501
```

```
wealth <- wealth %>% mutate(sanitary_drain = case_when(  
  A01501 %in% c(1,2) ~ 1,  
  TRUE ~ 0))
```

```
### garbage - LIXO - A016010
```

```
wealth <- wealth %>% mutate(garbage = case_when(  
  A016010 %in% c(3,4,5,6) ~ 0,  
  A016010 %in% c(1,2) ~ 1))
```

```
### home_phone - Telefone Fixo - A018017
```

```
wealth <- wealth %>% mutate(home_phone = case_when(  
  A018017 == 1 ~ 1,  
  A018017 == 2 ~ 0))
```

```
### motorcycle - Motocicleta - A018025
```

```
wealth <- wealth %>% mutate(motorcycle = case_when(  
  A018025 == 1 ~ 1,  
  A018025 == 2 ~ 0))
```

```
### internet - Internet - A01901
```

```
wealth <- wealth %>% mutate(internet = case_when(  
  A01901 == 1 ~ 1,  
  A01901 == 2 ~ 0))
```

```
### crowding - Pessoas/Comodo - C001/A011<3
```

```
wealth <- wealth %>% mutate(crowding = case_when(  
  (C001/A011) < 3 ~ 1,  
  TRUE ~ 0))
```

```
### bathrooms - Pessoas/Banheiro - C001/A01401 <=4
```

```
wealth <- wealth %>% mutate(bathrooms = case_when(  
  (C001/A01401) <= 3 ~ 1,  
  (C001/A01401) > 3 ~ 0 ))
```

```
# Separa o em Urbana e Rural —
```

```

urbana <- subset(dados19, V0026 == '1')
wealthu <- subset(wealth, V0026 == '1')
wealthr <- subset(wealth, V0026 == '2')
rural <- subset(dados19, V0026 == '2')

# Analisa Fatorial URBANA #####
### Dados wealth index - 18 vari veis
wealthu$id <- NA
wealthu$id <- 1:nrow(wealthu)
wealth_index <- subset(wealthu, select = c(id, V0026, water_supply, floor, walls,
roof, housekeeper, car, cell_phone,
computer, microwave, washing,
water_drink, sanitary_drain, garbage,
home_phone, motorcycle, internet,
crowding, bathrooms))

### Criterio para adequa o da base de dados: Correla es tetracoricas
, teste de Barlett e Teste KMO
(corrTetra<- tetrachoric(wealth_index[c(-1,-2)]))
corPlot(corrTetra$rho, numbers=TRUE, upper=FALSE, diag=FALSE)

# Teste de Bartlett
cortest.bartlett(corrTetra$rho, n=nrow(wealth_index[c(-1,-2)]))

#Teste de Kaiser-Meyer-Olkin (KMO)
KMO(corrTetra$rho)
#2. Numero de fatores
# Calculo dos autovalores
round(eigen(corrTetra$rho)$values, 2)
sum(eigen(corrTetra$rho)$values > 1) #4
# Scree-plot
plot(eigen(corrTetra$rho)$values, type="b", ylab="Autovalores") #1
# Calculo da proporcao explicada por cada fator
proporcao_explicacao<-eigen(corrTetra$rho)$values/
sum(eigen(corrTetra$rho)$values)
# Calculo da proporcao de explicacao acumulada
proporcao_acumulada <- cumsum(proporcao_explicacao)
# Imprime a proporcao acumulada com 2 casas decimais

```

```
round(proporcao_acumulada, 2)
```

```
#PCA do psych
```

```
sub0 = subset(wealth_index, select = -c(id, V0026))
```

```
pca0<-psych::principal(sub0, rotate="varimax", nfactors=1,scores=TRUE  
, cor="tet")
```

```
sub1 = subset(sub0, select = -c(motorcycle)) #0,0061
```

```
pca2<-psych::principal(sub1, rotate="varimax", nfactors=1,scores=TRUE  
, cor="tet")
```

```
sub2 = subset(sub1, select = -c(water_supply))
```

```
pca3<-psych::principal(sub2, rotate="varimax", nfactors=1,scores=TRUE  
, cor="tet")
```

```
sub3 = subset(sub2, select = -c(crowding))
```

```
pca4<-psych::principal(sub3, rotate="varimax", nfactors=1,scores=TRUE  
, cor="tet")
```

```
sub4 = subset(sub3, select = -c(bathrooms))
```

```
pca5<-psych::principal(sub4, rotate="varimax", nfactors=1,scores=TRUE  
, cor="tet")
```

```
sub5 = subset(sub4, select = -c(walls))
```

```
pca6<-psych::principal(sub5, rotate="varimax", nfactors=1,scores=TRUE  
, cor="tet")
```

```
sub6 = subset(sub5, select = -c(housekeeper))
```

```
pca7<-psych::principal(sub6, rotate="varimax", nfactors=1,scores=TRUE  
, cor="tet")
```

```
sub7 = subset(sub6, select = -c(water_drink))
```

```
pca8<-psych::principal(sub7, rotate="varimax", nfactors=1,scores=TRUE  
, cor="tet")
```



```

# Construindo o índice
urbana$pcascore_wealth_index=pca8$scores[,1]

urbana <- urbana %>% mutate(wealth_index = ntile(pcascore_wealth_index, 5))
urbana$wealth_index <- as.factor(urbana$wealth_index)

# Analisa Fatorial RURAL #####
### Dados wealth index - 18 variáveis
wealthr$id <- NA
wealthr$id <- 1:nrow(wealthr)
wealth_index <- subset(wealthr, select = c(id, V0026, water_supply, floor, walls,
roof, housekeeper, car, cell_phone,
computer, microwave, washing,
water_drink, sanitary_drain, garbage,
home_phone, motorcycle, internet,
crowding, bathrooms))

### Critérios para adequação da base de dados:
Correlações tetracóricas
, teste de Barlett e Teste KMO
(corrTetra <- tetrachoric(wealth_index[c(-1, -2)]))
corPlot(corrTetra$rho, numbers=TRUE, upper=FALSE, diag=FALSE)

# Teste de Bartlett
cortest.bartlett(corrTetra$rho, n=nrow(wealth_index[c(-1, -2)]))

# Teste de Kaiser-Meyer-Olkin (KMO)
KMO(corrTetra$rho)

#2. Número de fatores
# Cálculo dos autovalores
round(eigen(corrTetra$rho)$values, 2)
sum(eigen(corrTetra$rho)$values > 1) #4
# Scree-plot
plot(eigen(corrTetra$rho)$values, type="b", ylab="Autovalores") #1
# Cálculo da proporção explicada por cada fator
proporcao_explicacao <- eigen(corrTetra$rho)$values/

```

```
sum(eigen(corrTetra$rho)$values)
# Calculo da proporcao de explicacao acumulada
proporcao_acumulada <- cumsum(proporcao_explicacao)
# Imprime a proporcao acumulada com 2 casas decimais
round(proporcao_acumulada, 2)

#PCA do psych
sub0 = subset(wealth_index, select = -c(id, V0026))
pca0<-psych::principal(sub0, rotate="varimax", nfactors=1,scores=TRUE
, cor="tet")

sub1 = subset(sub0, select = -c(motorcycle)) #0,0061
pca2<-psych::principal(sub1, rotate="varimax", nfactors=1,scores=TRUE
, cor="tet")

sub2 = subset(sub1, select = -c(water_supply))
pca3<-psych::principal(sub2, rotate="varimax", nfactors=1,scores=TRUE
, cor="tet")

sub3 = subset(sub2, select = -c(crowding))
pca4<-psych::principal(sub3, rotate="varimax", nfactors=1,scores=TRUE
, cor="tet")

sub4 = subset(sub3, select = -c(water_drink))
pca5<-psych::principal(sub4, rotate="varimax", nfactors=1,scores=TRUE
, cor="tet")

sub5 = subset(sub4, select = -c(housekeeper))
pca6<-psych::principal(sub5, rotate="varimax", nfactors=1,scores=TRUE
, cor="tet")

sub6 = subset(sub5, select = -c(home_phone))
pca7<-psych::principal(sub6, rotate="varimax", nfactors=1,scores=TRUE
, cor="tet")

sub7 = subset(sub6, select = -c(bathrooms))
pca8<-psych::principal(sub7, rotate="varimax", nfactors=1,scores=TRUE
, cor="tet")
```

```

sub8 = subset(sub7, select = -c(walls))
pca9<-psych::principal(sub8, rotate="varimax", nfactors=1,scores=TRUE
, cor="tet")

sub9 = subset(sub8, select = -c(sanitary_drain))
pca10<-psych::principal(sub9, rotate="varimax", nfactors=1,scores=TRUE
, cor="tet")

# Construindo do indice
rural$pca_score_wealth_index=pca10$scores[,1]
rural <- rural %>% mutate(wealth_index = ntile(pca_score_wealth_index, 5))
rural$wealth_index <- as.factor(rural$wealth_index)

# DESIGN -----
desURB=svydesign(id=~UPA_PNS, strat=~V0024,
                weight=~V00291, nest=TRUE, data=urbana)
desRUR=svydesign(id=~UPA_PNS, strat=~V0024,
                weight=~V00291, nest=TRUE, data=rural)

# DESCRITIVA URBANA -----
x1 <- cbind(svytotal(~escolaridade,desURB),
            svymean(~escolaridade,desURB))
x11 <- cbind(confint(svytotal(~escolaridade,desURB)),
             confint(svymean(~escolaridade,desURB)))

x2 <- cbind(svytotal(~regioes,desURB),
            svymean(~regioes,desURB))
x21 <- cbind(confint(svytotal(~regioes,desURB)),
             confint(svymean(~regioes,desURB)))

x3 <- cbind(svytotal(~estado_civil,desURB),
            svymean(~estado_civil,desURB))
x31 <- cbind(confint(svytotal(~estado_civil,desURB)),
             confint(svymean(~estado_civil,desURB)))

x4 <- cbind(svytotal(~raca,desURB),
            svymean(~raca,desURB))
x41 <- cbind(confint(svytotal(~raca,desURB)),

```

```

      confint (svymean(~raca ,desURB)))

x5 <- cbind(svytotal(~ocupacao ,desURB) ,
            svymean(~ocupacao ,desURB))
x51 <- cbind(confint (svytotal(~ocupacao ,desURB)) ,
             confint (svymean(~ocupacao ,desURB)))

x6 <- cbind(svytotal(~idade ,desURB) ,
            svymean(~idade ,desURB))
x61 <- cbind(confint (svytotal(~idade ,desURB)) ,
             confint (svymean(~idade ,desURB)))

x7 <- cbind(svytotal(~sexo ,desURB) ,
            svymean(~sexo ,desURB))
x71 <- cbind(confint (svytotal(~sexo ,desURB)) ,
             confint (svymean(~sexo ,desURB)))

x8 <- cbind(svytotal(~doenca_cronica ,desURB) ,
            svymean(~doenca_cronica ,desURB))
x81 <- cbind(confint (svytotal(~doenca_cronica ,desURB)) ,
             confint (svymean(~doenca_cronica ,desURB)))

x9 <- cbind(svytotal(~wealth_index ,desURB) ,
            svymean(~wealth_index ,desURB))
x91 <- cbind(confint (svytotal(~wealth_index ,desURB)) ,
             confint (svymean(~wealth_index ,desURB)))

##### Distribui o variaveis por saude cardiovascular Urbana

a1 <- svyby(~cvh_cat , ~escolaridade , desURB , svymean ,vartype="ci")
a2 <- svyby(~cvh_cat , ~ocupacao , desURB , svymean ,vartype="ci")
a3 <- svyby(~cvh_cat , ~wealth_index , desURB , svymean ,vartype="ci")
a4 <- svyby(~cvh_cat , ~sexo , desURB , svymean ,vartype="ci")
a5 <- svyby(~cvh_cat , ~idade , desURB , svymean ,vartype="ci")
a6 <- svyby(~cvh_cat , ~raca , desURB , svymean ,vartype="ci")
a7 <- svyby(~cvh_cat , ~estado_civil , desURB , svymean ,vartype="ci")
a8 <- svyby(~cvh_cat , ~regioes , desURB , svymean ,vartype="ci")
a9 <- svyby(~cvh_cat , ~doenca_cronica , desURB , svymean ,vartype="ci")

```

```
# DESCRITIVA RURAL ———
x1 <- cbind(svytotal(~escolaridade,desRUR),
            svymean(~escolaridade,desRUR))
x11 <- cbind(confint(svytotal(~escolaridade,desRUR)),
            confint(svymean(~escolaridade,desRUR)))

x2 <- cbind(svytotal(~regioes,desRUR),
            svymean(~regioes,desRUR))
x21 <- cbind(confint(svytotal(~regioes,desRUR)),
            confint(svymean(~regioes,desRUR)))

x3 <- cbind(svytotal(~estado_civil,desRUR),
            svymean(~estado_civil,desRUR))
x31 <- cbind(confint(svytotal(~estado_civil,desRUR)),
            confint(svymean(~estado_civil,desRUR)))

x4 <- cbind(svytotal(~raca,desRUR),
            svymean(~raca,desRUR))
x41 <- cbind(confint(svytotal(~raca,desRUR)),
            confint(svymean(~raca,desRUR)))

x5 <- cbind(svytotal(~ocupacao,desRUR),
            svymean(~ocupacao,desRUR))
x51 <- cbind(confint(svytotal(~ocupacao,desRUR)),
            confint(svymean(~ocupacao,desRUR)))

x6 <- cbind(svytotal(~idade,desRUR),
            svymean(~idade,desRUR))
x61 <- cbind(confint(svytotal(~idade,desRUR)),
            confint(svymean(~idade,desRUR)))

x7 <- cbind(svytotal(~sexo,desRUR),
            svymean(~sexo,desRUR))
x71 <- cbind(confint(svytotal(~sexo,desRUR)),
            confint(svymean(~sexo,desRUR)))

x8 <- cbind(svytotal(~doenca_cronica,desRUR),
            svymean(~doenca_cronica,desRUR))
x81 <- cbind(confint(svytotal(~doenca_cronica,desRUR)),
```

```

confint(svymean(~doenca_cronica,desRUR)))

x9 <- cbind(svytotal(~wealth_index,desRUR),
            svymean(~wealth_index,desRUR))
x91 <- cbind(confint(svytotal(~wealth_index,desRUR)),
             confint(svymean(~wealth_index,desRUR)))

##### Distribuição de variáveis por saúde cardiovascular Rural

a1 <- svyby(~cvh_cat, ~escolaridade, desRUR, svymean, vartype="ci")
a2 <- svyby(~cvh_cat, ~ocupacao, desRUR, svymean, vartype="ci")
a3 <- svyby(~cvh_cat, ~wealth_index, desRUR, svymean, vartype="ci")
a4 <- svyby(~cvh_cat, ~sexo, desRUR, svymean, vartype="ci")
a5 <- svyby(~cvh_cat, ~idade, desRUR, svymean, vartype="ci")
a6 <- svyby(~cvh_cat, ~raca, desRUR, svymean, vartype="ci")
a7 <- svyby(~cvh_cat, ~estado_civil, desRUR, svymean, vartype="ci")
a8 <- svyby(~cvh_cat, ~regioes, desRUR, svymean, vartype="ci")
a9 <- svyby(~cvh_cat, ~doenca_cronica, desRUR, svymean, vartype="ci")

```

## 7.4 Códigos em SAS

```

PROC SURVEYLOGISTIC data=pns.urbana;
weight V00291;
strata v0024;
cluster upa_pns;
class escolaridade
      ocupacao
      wealth_index
      sexo
      idade
      raca
      estado_civil
      doenca_cronica
      regioes
      / ref=first param=reference;
model cvh_cat(descending) = escolaridade ocupacao wealth_index
      sexo idade raca estado_civil doenca_cronica regioes / LINK=glogit;
OUTPUT OUT = a PRED = yhat;

```

```
run;

PROC SURVEYLOGISTIC data=pns.rural;
weight V00291;
strata v0024;
cluster upa_pns;
class escolaridade
      ocupacao
      wealth_index
      sexo
      idade
      raca
      estado_civil
      doenca_cronica
      regioes
      / ref=first param=reference;
model cvh_cat(descending) = escolaridade ocupacao wealth_index
sexo idade raca estado_civil doenca_cronica regioes / LINK=glogit;
run;
```

```
PROC SURVEYLOGISTIC data=pns.rural;
weight V00291;
strata v0024;
cluster upa_pns;
class escolaridade
      ocupacao
      sexo
      idade
      raca
      doenca_cronica
      regioes
      / ref=first param=reference;
model cvh_cat(descending) = escolaridade ocupacao
sexo idade raca doenca_cronica regioes / LINK=glogit;
run;
```