



Universidade de Brasília
Departamento de Estatística

**Aplicação de Regressão Logística na Associação entre Ocupação e Saúde
Cardiovascular - Pesquisa Nacional de Saúde 2019**

Ana Tércia Freires da Silva

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**

Ana Tércia Freires da Silva

**Aplicação de Regressão Logística na Associação entre Ocupação e Saúde
Cardiovascular - Pesquisa Nacional de Saúde 2019**

Orientador(a): Prof^ª. Dra. Joanlise Marco de Leon Andrade

Coorientador(a): Prof^ª. Dra. Kelb Bousquet Santos

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: Analisar a associação entre ocupação e saúde cardiovascular (SCV) utilizando a Pesquisa Nacional de Saúde de 2019. **Metodologia:** Trata-se de um inquérito de saúde de âmbito nacional, excluindo-se estudantes, aposentados e cuidadores domésticos. Utilizou-se como variável resposta a categorização do escore de SCV em superior e inferior, o qual é construído a partir das métricas comportamentais (tabagismo, índice de massa corporal, atividade física e dieta) e biológicas (colesterol total, pressão arterial e glicose) adaptadas da *American Heart Association*. Como variável explicativa de maior interesse, os grupos de ocupação foram classificados em ocupações de colarinho branco, de colarinho azul e não ocupados (ou mal definidas). Outros determinantes socioeconômicos (escolaridade e renda *per capita*), variáveis sociodemográficas (sexo, faixa etária, raça, estado civil, situação de domicílio e região) e a presença de doenças crônicas foram utilizados em modelos de regressão logística considerando a amostragem complexa da pesquisa. **Resultados:** A análise dos dados indicam que, apenas 9,96% ($IC_{95\%}$: 9,47%; 10,44%) da população brasileira na força de trabalho apresentou SCV superior. Esta prevalência foi maior entre ocupações de colarinho branco, com quase o dobro de indivíduos com SCV superior comparado às de colarinho azul e não ocupados ou de ocupações mal definidas. Apesar de tais diferenças, a associação de ocupação por colarinho não foi significativa quando ajustou-se por covariáveis socioeconômicas e sociodemográficas. **Conclusão:** A ocupação por colarinho não se mostrou associada à saúde cardiovascular. Utilizar as variáveis renda *per capita* e nível de escolaridade como determinantes socioeconômicos parece ser uma melhor estratégia que utilizar ocupação por colarinho.

Palavras-chaves: Pesquisa Nacional de Saúde; Ocupação; Colarinho branco; Colarinho azul; Saúde cardiovascular; Regressão logística; Amostragem complexa.

Abstract

Objective: To analyze the association between occupation and cardiovascular health (CVH) using the 2019 National Health Survey. **Methodology:** This is a nationwide health survey, excluding students, retirees and domestic caregivers. The response variable was the categorization of CVH score into superior and inferior, which is constructed from behavioral metrics (smoking, body mass index, physical activity, and diet) and biological metrics (total cholesterol, blood pressure, and glucose) adapted from the American Heart Association. As the main explanatory variable of interest, occupation groups were classified as white-collar, blue-collar, and not occupied (or poorly defined). Other socioeconomic determinants (education and per capita income), sociodemographic variables (gender, age group, race, marital status, household situation, and region), and the presence of chronic diseases were used in logistic regression models, considering the complex sampling of the survey. **Results:** Data analysis indicates that only 9.96% (95% CI: 9.47%; 10.44%) of the Brazilian working-age population showed superior CVH. This prevalence was higher among white-collar occupations, with nearly twice as many individuals having superior CVH compared to blue-collar and not occupied or poorly defined occupations. Despite these differences, the association of white-collar occupation was not significant when adjusting for socioeconomic and sociodemographic covariates. **Conclusion:** White-collar occupation was not found to be associated with cardiovascular health. Using per capita income and level of education as socioeconomic determinants seems to be a better strategy than relying on white-collar occupation.

Keywords: National Health Survey; Occupation; White-collar; Blue-collar; Cardiovascular health; Logistic regression; Complex sampling.

Lista de Figuras, Quadros e Tabelas

Lista de Figuras

- 1 Curva ROC e AUC dos modelos minimamente ajustado (Modelo 01), moderadamente ajustado (Modelo 02) e modelo completo (Modelo 03) 28

Lista de Quadros

- 1 Métricas de Saúde Cardiovascular Ideal - AHA 2010 11
- 2 Matriz de confusão apresentando as categorias resultantes da classificação de um conjunto de dados em duas classes distintas e comparando-as com as classes reais observadas 16

Lista de Tabelas

- 1 Frequências estimadas (com intervalos de 95% de confiança) de variáveis sociodemográficas e socioeconômicas por categorias de ocupação para a população na força de trabalho - Expansão da amostra da PNS 2019 19
- 2 Frequências estimadas (com intervalos de 95% de confiança) de variáveis sociodemográficas e socioeconômicas por categorias de saúde cardiovascular para a população na força de trabalho - Expansão da amostra da PNS 2019 22
- 3 Razões de chances, IC95% e p-valores para o Modelo 1 de regressão logística com a variável resposta SCV (superior vs inferior) na população brasileira na força de trabalho – Expansão da amostra PNS 2019 25
- 4 Razões de chances, IC95% e p-valores para o Modelo 2 de regressão logística com a variável resposta SCV (superior vs inferior) na população brasileira na força de trabalho – Expansão da amostra PNS 2019 26
- 5 Razões de chances, IC95% e p-valores para o Modelo 3 de regressão logística com a variável resposta SCV (superior vs inferior) na população brasileira na força de trabalho – Expansão da amostra PNS 2019 27

Sumário

1	Introdução	7
2	Objetivos	9
2.1	Objetivo Geral	9
2.2	Objetivos Específicos	9
3	Metodologia	10
3.1	Banco de Dados	10
3.2	Escore e Categorias de Saúde Cardiovascular	10
3.3	Variáveis Explicativas	11
3.4	Modelos de Regressão Logística	12
3.5	Qualidade do Ajuste	15
3.6	Implementação das Análises	17
4	Resultados	18
5	Discussão	29
6	Conclusão	32
	Referências	33
	Apêndice	36

1 Introdução

Segundo a Organização Mundial da Saúde (OMS, 2021), a estimativa de mortes por doenças cardiovasculares (DCV) em 2019 foi de 17,9 milhões, correspondendo a 32% dos óbitos no mundo. Os países que apresentam maior incidência de casos de DCV geralmente são os de baixa e média renda, o que pode ser explicado pela menor capacidade da população em obter acesso a diagnósticos precoces ou a possíveis tratamentos (OMS, 2020).

O desenvolvimento de DCV pode ser influenciado por uma combinação de fatores genéticos, fisiológicos, ambientais e comportamentais (OMS, 2021). Logo, devido a importância de monitorar a saúde cardiovascular a *American Heart Association* (AHA) estabeleceu a definição de saúde cardiovascular ideal, a qual é composta por 7 metas que são divididas em comportamentais (que contemplam a dieta saudável, prática de atividades físicas, não tabagismo e peso corporal adequado) e biológicas (que contemplam o colesterol, pressão arterial e glicemia em níveis normais e adequados), veja Lloyd-Jones et al. (2010) para mais informações. A partir da soma dessas metas é construído o escore de saúde cardiovascular (SCV), o qual varia entre 0 e 7. Um estudo recente (MOTTA et al., 2023) mostrou que a prevalência de saúde cardiovascular superior (alcance de 6 a 7 metas) na população brasileira foi estimada em apenas 8,9% (com intervalo de 95% de confiança: 8,5%; 9,4%). Dessa forma, torna-se perceptível a diferença entre as categorias de saúde dos indivíduos e para uma melhor compreensão das razões subjacentes a essas diferenças, a condução de estudos torna-se crucial, uma vez que alguns determinantes podem não se mostrar estatisticamente tão relevantes quanto outros (CARRAPATO; CORREIA; GARCIA, 2017).

O presente trabalho tem um enfoque especial para a ocupação dos indivíduos que, conforme a OMS (2020) e Havranek et al. (2015), está entre os determinantes importantes associados a saúde cardiovascular. Estudos anteriores mostraram que, a saúde dos indivíduos pode ser influenciada negativamente por suas ocupações, ao considerar os níveis de estresse, força aplicada sobre o trabalho, horas de atuação e exposição a riscos químicos e biológicos (KIVIMÄKI; KAWACHI, 2015; VIRTANEN; KIVIMÄKI, 2018; CARVALHO et al., 2021). Este estudo, por sua vez, traz uma aplicação sobre as características do trabalho dos indivíduos que se divide em áreas denotadas por “colarinho branco” e “colarinho azul” (HARRISON; ROSE, 2006).

Conforme o exposto, torna-se imprescindível a realização de estudos voltados para a saúde cardiovascular, com o intuito de contribuir e aprimorar a promoção da

saúde no país. Em específico, este trabalho irá avaliar a associação entre ocupação e saúde cardiovascular com base na Pesquisa Nacional de Saúde (PNS) de 2019. O modelo de regressão logística será utilizado para este fim.

2 Objetivos

2.1 Objetivo Geral

Analisar a associação entre ocupação e saúde cardiovascular com base na PNS 2019.

2.2 Objetivos Específicos

- Obter o escore de saúde cardiovascular utilizando as métricas da saúde cardiovascular ideal propostas pela AHA;
- Classificar as ocupação dos participantes nas categorias denominadas colarinho azul, colarinho branco e não ocupados;
- Analisar a relação entre as categorias de ocupação e a saúde cardiovascular utilizando métodos descritivos e o modelo logístico ajustando-se por covariáveis socio-demográficas e outros determinantes socioeconômicos;

3 Metodologia

3.1 Banco de Dados

Para a realização deste trabalho foram utilizados os dados públicos da PNS 2019. As informações nesta subseção são baseadas na publicação de Stopa et al. (2020).

A PNS está incluída no Sistema Integrado de Pesquisas Domiciliares (SIPD) do Instituto Brasileiro de Geografia e Estatística (IBGE). Utiliza-se como plano amostral, a amostragem por conglomerados em três estágios, conglomerados com estratificação das unidades primárias de amostragem (UPAs). Em todos os estágios do plano amostral é utilizada a amostragem aleatória simples sendo que, o primeiro estágio é a seleção das UPAs, o segundo estágio envolve a seleção, em cada UPA, de um número fixo de domicílios particulares permanentes e o terceiro estágio é a seleção, em cada domicílio, de um morador com idade maior ou igual a 15 anos. A PNS faz parte do SIPD então, a UPA é considerada como uma subamostra de uma amostra mestre que já existia nesse Sistema. Quanto aos domicílios, não foram incluídos na pesquisa aqueles localizados em setores censitários especiais ou de escassa população.

A coleta dos dados da PNS foi realizada por meio de questionário, o qual é dividido em três seções, com questões voltadas para: o domicílio, todos moradores do domicílio e o morador (de 15 anos ou mais de idade) selecionado. Mais informações sobre o questionário aplicado são encontradas no dicionário disponível no site da PNS (<https://www.pns.iciet.fiocruz.br/bases-de-dados/>).

Foram incluídos nas análises apenas moradores com 18 anos ou mais de idade cujo questionário individual estava completo e foi respondido pelo próprio morador. Participantes do sexo feminino foram excluídas se grávidas ou incertas sobre a possibilidade de gravidez. Uma vez que o propósito deste trabalho é analisar a relação entre a saúde cardiovascular e a ocupação, indivíduos fora da força de trabalho (JÚNIOR et al., 2021; ASSUNÇÃO; CLARO, 2022) foram excluídos. Tais indivíduos correspondem a estudantes que não trabalhavam, aposentados e/ou cuidadores domésticos.

3.2 Escore e Categorias de Saúde Cardiovascular

O escore de saúde cardiovascular proposto pela AHA em 2010, envolve metas descritas em Lloyd-Jones et al. (2010). Algumas dessas variáveis foram autorreferidas e

não diretamente aferidas na PNS 2019. Em razão disso, o escore utilizado neste trabalho representa uma adaptação, conforme apresentado em Motta et al. (2023), cujas métricas componentes são aqui reproduzidas no Quadro 1.

Quadro 1: Métricas de Saúde Cardiovascular Ideal - AHA 2010

Metas	Métrica utilizada
Comportamentais	4 metas comportamentais atingidas simultaneamente
Tabagismo	Nunca ter fumado ou não fumar há mais de 12 meses.
Índice de massa corporal (IMC)	$< 25 \text{ kg/m}^2$ (a partir de medidas de peso e altura autorreferidas e imputadas)
Atividade Física	Prática de exercício físico ou esporte ≥ 150 min./ semana.
Dieta adequada	4 metas de dieta atingidas simultaneamente: i) frutas e vegetais: consumo ≥ 1 vez / dia; ii) peixe: consumo ≥ 2 vezes/semana; iii) sódio: consumo de sal na comida na hora e nos alimentos industrializados em quantidade adequada, baixa ou muito baixa; e iv) açúcar: consumo de refrigerantes e/ou sucos industrializados < 5 dias/semana.
Biológicas	3 metas biológicas atingidas simultaneamente
Colesterol total	Ausência de diagnóstico médico autorreferido de colesterol alto.
Pressão arterial	Ausência de diagnóstico médico autorreferido de hipertensão arterial.
Glicose	Ausência de diagnóstico médico autorreferido de diabetes mellitus.

Fonte: Motta et al. (2023), adaptado de Lloyd-Jones et al. (2010).

O escore de SCV foi obtido pela soma das métricas alcançadas por cada indivíduo. Neste trabalho, optou-se pela utilização de uma categorização em dois níveis de SCV com base no referido escore, quais sejam “SCV Superior” (alcançando de 6 a 7 metas) e “SCV Inferior” (alcançando de 0 a 5 metas).

3.3 Variáveis Explicativas

Para classificar a variável ocupação foi utilizada uma adaptação da proposta apresentada por Harrison e Rose (2006), de forma que as ocupações dos indivíduos ficaram separadas em três categorias: colarinho branco, colarinho azul e não ocupados/mal definidas.

Em suma, os trabalhadores classificados como de colarinho branco abrangem profissionais empregados no setor público ou privado, nas forças armadas, no comércio e em

serviços diversos. Já os trabalhadores na categoria colarinho azul referem-se àqueles que desempenham atividades laborais como operários, agricultores ou pescadores. A categoria “não ocupados/mal definidas” engloba indivíduos não ocupados ou cujas ocupações são mal definidas e não puderam ser classificadas como de colarinho branco ou azul. Para auxiliar na separação dos indivíduos nos diferentes tipos de colarinhos foi utilizada a Classificação de Ocupações para Pesquisas Domiciliares (COD), disponível na PNS 2019. A COD, desenvolvida pelo IBGE, é uma versão adaptada da Classificação Internacional Uniforme de Ocupações (CIUO) e tem por finalidade agrupar e organizar as diferentes ocupações existentes dividindo-as em diferentes níveis de acordo com as características e complexidade das tarefas realizadas (IBGE, 2010).

Utilizando o estudo de Bousquet-Santos, Chen e Kubzansky (2021) como referência, considera-se além da ocupação, outros determinantes socioeconômicos foram utilizados como covariáveis neste estudo, incluindo o nível de escolaridade (com diferentes níveis entre indivíduos do ensino fundamental incompleto ao superior completo) e a renda *per capita* (categorizada em diversos níveis entre até 1/4 salário mínimo e mais de 5 salários mínimos). Ressalta-se ainda que a variável renda *per capita* foi obtida excluindo o rendimento das pessoas cuja condição na unidade domiciliar era pensionista, empregado doméstico ou parente do empregado doméstico. Foi considerada também, utilizando o mesmo estudo de referência, a inclusão das covariáveis sociodemográficas: situação de domicílio (urbano e rural), estado civil (solteiro, casado e outros), região geográfica (norte, nordeste, sul, sudeste e centro-oeste), raça/etnia (branca, parda, preta e outras); sexo (masculino e feminino) e faixa etária (com diferentes níveis entre 18 a mais 60 anos de idade). A presença de doenças crônicas (pelo menos uma ou nenhuma) também foi incluída como covariável no modelo.

3.4 Modelos de Regressão Logística

Uma breve descrição de modelos de regressão logística, utilizados para analisar os dados do presente trabalho, é apresentada a seguir, baseada em Hosmer, Lemeshow e Sturdivant (2013), Lee e Forthofer (2006), Agresti (2013) e Agresti (2007).

Para uma variável de resposta binária (atribuindo-se 1 ao “sucesso” e 0 ao “fracasso”) e X uma variável explicativa quantitativa ou qualitativa, a probabilidade de ocorrência de um evento Y dado X pode ser expressa por $\pi(x) = P(Y = 1|X = x) = 1 - P(Y = 0|X = x)$. O modelo de regressão logística pode ser apresentado em forma

linear pelo *logit* (ou logito) dessa probabilidade, dado por:

$$\text{logit}[\pi(x)] = \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x, \quad (3.1)$$

sendo α o intercepto e β o parâmetro estimado da variável explicativa.

Aplicando a função exponencial na Equação 3.1, obtém-se outra forma de expressar o modelo logístico, dado pela chance, do inglês *odds*, expressa como:

$$\text{Chance} = \frac{\pi(x)}{1 - \pi(x)} = e^{\alpha + \beta x}. \quad (3.2)$$

Para facilitar a interpretação do modelo logístico comumente estima-se a razão de chances (RC), calculada por e^β . Se a variável explicativa for categórica com I níveis, escolhe-se um nível como referência, o qual será comparado com os demais. Portanto, o modelo envolverá a estimação de $I - 1$ coeficientes β e suas respectivas RCs.

Caso a variável explicativa seja contínua, a interpretação dos coeficientes estimados passa a variar de acordo com a maneira como ela é incorporada ao modelo e da unidade da mesma. Considerando o *logit* linear nesse tipo de variável (z), tem-se o $\text{logit}[\pi(z)] = \beta_0 + \beta_1 z_1$ que será denotado por $g(z)$. Para se fornecer uma interpretação útil para a covariável contínua, é necessário construir um método e estimar um intervalo para uma mudança arbitrária de “ c ” unidades na covariável. Desta forma, o *log* da razão de chances para a mudança de “ c ” unidades em z é obtido através da diferença de $g(z + c) - g(z) = c\beta_1$ e a RC associada fica $RC(z + c, z) = e^{c\beta_1}$.

O modelo logístico também pode ser estendido de modo a envolver múltiplas variáveis explicativas, ou seja, a equação pode ser apresentada em termos de p preditores, onde $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_p]^T$ é o vetor de covariáveis e $\boldsymbol{\beta} = [\alpha \ \beta_1 \ \beta_2 \ \dots \ \beta_p]^T$ o vetor de parâmetros estimados, como na expressão a seguir:

$$\text{logit}[P(Y = 1|\mathbf{x})] = \mathbf{x}\boldsymbol{\beta} \quad (3.3)$$

As RCs são obtidas a partir da aplicação da exponencial na Equação 3.3, fornecendo a comparação das chances de sucesso entre categorias de cada variável explicativa mantendo-se as demais covariáveis constantes.

As estimativas dos parâmetros em um modelo logístico para amostragens aleatórias simples são comumente obtidas pelo método de máxima verossimilhança. Entretanto, as estimativas para um modelo logístico para amostras com delineamento complexo, como é o

caso da PNS 2019, podem ser obtidas por meio da aplicação do método de máxima pseudo-verossimilhança (MPV). Supondo uma população finita que é dividida em $k = 1, 2, \dots, K$ estratos, sendo que cada estrato é posteriormente subdividido em $j = 1, 2, \dots, M_k$ unidades amostrais primárias, e $i = 1, 2, \dots, N_{kj}$ sendo o elemento na kj -ésima unidade amostral primária. Assim, tem-se a função de pseudo-log-verossimilhança como:

$$\sum_{k=1}^K \sum_{j=1}^{m_k} \sum_{i=1}^{n_{kj}} [w_{kji} \cdot y_{kji}] \cdot \ln[\pi(x_{kji})] + [w_{kji} \cdot (1 - y_{kji})] \cdot \ln[1 - \pi(x_{kji})], \quad (3.4)$$

sendo w_{kji} o peso amostral, y_{kji} a variável de resposta binária e x_{kji} o vetor de covariáveis.

O estimador de MPV é obtido derivando-se a função de pseudo-log-verossimilhança e igualando o resultado a zero.

Para obter os estimadores de variância por MPV em amostragem complexa, recorre-se a métodos como a linearização de Taylor, replicação *Jackknife* e *bootstrap*. Tais métodos são empregados devido à impossibilidade de se calcular os estimadores de variância de forma direta nesse contexto.

Neste estudo é utilizado o método de linearização de Taylor. Com isto, obtém-se o estimador de variância de $\hat{\beta}$ em:

$$\widehat{Var}(\hat{\beta}) = (X'DX)^{-1}S(X'DX)^{-1}, \quad (3.5)$$

em que X é a matriz de covariáveis, $D = WV$ é a matriz diagonal $n \times n$ dos elementos $w_{hjik} \cdot \hat{\pi}(x_{hjik})[1 - \hat{\pi}(x_{hjik})]$, e S é o estimador agrupado dentro do estrato da matriz de covariância.

A avaliação da contribuição das covariáveis em um modelo logístico complexo pode ser realizada pelo teste de Wald e o teste de Wald ajustado, os quais avaliam as hipóteses:

$$\begin{cases} H_0 : \hat{\beta}_j = 0, \quad j = 1, 2, \dots, p \\ H_a : \exists \hat{\beta}_j \neq 0 \end{cases}$$

O teste de Wald, que tem por estatística do teste:

$$W = \frac{\hat{\beta}_j}{\sqrt{\widehat{Var}(\hat{\beta}_j)}}, \quad (3.6)$$

se aproxima de uma distribuição normal padrão e é utilizado neste estudo para avaliar a significância de cada parâmetro estimado, sendo $\sqrt{\widehat{Var}(\hat{\beta}_j)}$ o erro padrão de $\hat{\beta}_j$.

Por outro lado, o teste de Wald ajustado tem por estatística do teste:

$$F = \frac{s - p + 1}{sp} W, \quad (3.7)$$

em que

$$W = \hat{\beta}' [\widehat{Var}(\hat{\beta})]^{-1} \hat{\beta}. \quad (3.8)$$

Tal teste é utilizado para avaliar a significância global de cada variável do estudo, onde $\hat{\beta}$ denota o vetor para os p coeficientes estimados, $s = (\sum_{k=1}^K m_k) - K$ é o número total de unidades amostrais primárias amostradas menos o número de estratos. O p -valor é calculado com base na distribuição F com p e $(s - p + 1)$ graus de liberdade como $Pr[F(p, s - p + 1) \geq F]$.

O intervalo de confiança para cada β_j estimado no modelo logístico é dado por:

$$\hat{\beta}_j \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\beta}_j)}, \quad (3.9)$$

e para a razão de chances:

$$e^{\hat{\beta}_j \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\beta}_j)}}, \quad (3.10)$$

sendo $z_{\alpha/2}$ o percentil da distribuição normal padrão em ambos os casos.

Com isto, neste estudo são apresentados além de p-valores de testes globais para cada variável explicativa, razões de chances, intervalos de confiança e p-valores correspondentes a cada um dos níveis das variáveis explicativas. Para as análises adota-se um nível de significância de $\alpha = 5\%$.

3.5 Qualidade do Ajuste

Testes como Qui-Quadrado de Pearson, de Razão de Verossimilhança (G^2), Pseudo- R^2 e de Hosmer-Lemeshow são frequentemente utilizados para avaliar o ajuste dos modelos logísticos (HOSMER; LEMESHOW; STURDIVANT, 2013; AGRETI, 2007).

Porém, no contexto de amostragens complexas, recomenda-se como uma das alternativas possíveis a utilização do Pseudo- R^2 de McFadden (SHTATLAND; KLEINMAN; CAIN, 2002) que tem por estatística de teste:

$$R^2 = 1 - \frac{\log L(M)}{\log L(0)}, \quad (3.1)$$

onde $\log L(M)$ é o valor da função de log-verossimilhança do modelo sendo estimado e

$\log L(0)$ o valor da função de log-verossimilhança do modelo sem preditores.

A interpretação do R^2 de McFadden em regressão logística, ao contrário de em regressão linear múltipla, denota que valores baixos de R^2 não são necessariamente indicadores de um modelo ruim (TJUR, 2009). Um modelo com R^2 baixo pode ser considerado adequado se o projeto do experimento restringir os valores das variáveis explicativas a um intervalo estreito, resultando em um modelo satisfatório (TJUR, 2009).

Outra medida que pode ser utilizada para avaliar o ajuste do modelo logístico é a curva ROC (*Receive Operating Characteristic*), apresentada em um gráfico que compara a sensibilidade no eixo vertical com 1 menos a especificidade no eixo horizontal, considerando todos os possíveis limites ou pontos de corte de decisão. A medida resumo frequentemente utilizada é a área sob a curva ROC, *area under the curve* (AUC) (MARTINEZ; LOUZADA-NETO; PEREIRA, 2003).

No Quadro 2 é apresentada a comparação entre uma classe de indivíduos preditos com a classe real, denominada de matriz de confusão a qual é bastante utilizada no contexto da curva ROC (NETO, 2020).

Quadro 2: Matriz de confusão apresentando as categorias resultantes da classificação de um conjunto de dados em duas classes distintas e comparando-as com as classes reais observadas

Resultado	Classe Real	
	Positivo	Negativo
Classe Predita		
Positivo	Verdadeiro-positivo (VP)	Falso-positivo (FP)
Negativo	Falso-negativo (FN)	Verdadeiro-negativo (VN)

Fonte: Neto (2020), adaptado de Martinez, Louzada-Neto e Pereira (2003).

A partir desta matriz, obtém-se a sensibilidade (S_E), também chamada de taxa de verdadeiro positivo (TVP), e a especificidade (E_S), dadas respectivamente pelas equações a seguir (MARTINEZ; LOUZADA-NETO; PEREIRA, 2003):

$$S_E = TVP = \frac{VP}{VP + FN}, \quad (3.2)$$

e,

$$E_S = \frac{VN}{VN + FP}. \quad (3.3)$$

Assim, a taxa de falsos positivos (TFP) ou $1 - E_S$, dada por:

$$TFP = 1 - E_S = \frac{FP}{FP + VN}. \quad (3.4)$$

3.6 Implementação das Análises

O *software* R, na versão 4.2.2, foi utilizado na leitura e tratamento dos dados da PNS 2019 e em todas as análises estatísticas realizadas neste estudo. Os pacotes utilizados foram o `survey`, cuja finalidade foi auxiliar na estimação e avaliação dos parâmetros de modelos com amostragem complexa e o pacote `WeightedROC` para a implementação das curvas ROC.

4 Resultados

Após a aplicação dos critérios de inclusão e exclusão, a amostra do presente estudo ficou composta por 56.950 entrevistas individuais, das 293.725 registradas no banco de dados da PNS 2019. Todas as proporções e respectivos intervalos de 95% de confiança ($IC_{95\%}$) reportados neste capítulo são resultantes de estimações com expansão amostral que considera a amostragem complexa da PNS.

Indivíduos fora da força de trabalho (estudantes que não trabalhavam, aposentados e cuidadores), excluídos das análises, corresponderam à proporção estimada de 23,98% ($IC_{95\%}$: 23,45%; 24,50%) da população de 18 anos ou mais de idade (considerando os critérios de exclusão tratados na Subseção 3.1).

A Tabela 1 apresenta as frequências dos níveis de SCV (variável resposta) e das demais variáveis por categorias de ocupação do tipo colarinho, que é a variável explicativa de maior interesse. Observa-se que indivíduos com ocupações de colarinho azul representam mais da metade da população na força de trabalho (54,12% com $IC_{95\%}$: 53,20%; 55,03%), seguidos por ocupações de colarinho branco com 27,28% ($IC_{95\%}$: 26,43%; 28,13%) e por indivíduos não ocupados ou em ocupações mal definidas 18,60% ($IC_{95\%}$: 17,98%; 19,22%), sendo a quase totalidade desse último grupo composto por indivíduos sem ocupação (99,82% com $IC_{95\%}$: 99,62%; 100,03%). A relação entre as classificações de ocupação COD e de colarinhos é exibida de forma bastante clara nesta tabela, a partir da qual observa-se que ocupações classificadas como de colarinho branco envolvem as 5 primeiras classes da COD, sendo cientistas e intelectuais, atividades técnicas e administrativas as mais frequentes. Já as ocupações classificadas como de colarinho azul são compostas pelas próximas 5 categorias da COD, sendo maior a proporção de indivíduos que trabalhavam com serviços em mercados e comércios, seguida de áreas voltadas para ocupações elementares.

Observa-se ainda que a prevalência de SCV superior em indivíduos com ocupações de colarinho branco foi de aproximadamente o dobro das observadas para as de colarinho azul ou entre não ocupados/mal definidas, diferenças estatisticamente significantes tendo em vista que os intervalos de confiança para o grupo colarinho branco não se sobrepõem com os demais (Tabela 1).

Tabela 1: Frequências estimadas (com intervalos de 95% de confiança) de variáveis sociodemográficas e socioeconômicas por categorias de ocupação para a população na força de trabalho - Expansão da amostra da PNS 2019

Variáveis	(continua)		
	Colarinho branco	Colarinho azul	Não ocupados/ Mal definidas
Total	27,28 (26,43; 28,13)	54,12 (53,20; 55,03)	18,60 (17,98; 19,22)
SCV			
Inferior	84,56 (83,51; 85,61)*	92,44 (91,87; 93,02)	91,11 (90,20; 92,01)
Superior	15,44 (14,39; 16,49)*	7,56 (6,98; 8,13)	8,89 (7,99; 9,80)
COD			
Membros das forças armadas, policiais e bombeiros militares	2,10 (1,74; 2,45)*	0	0
Diretores e gerentes	12,96 (11,84; 14,07)*	0	0
Profissionais das ciências e intelectuais	37,27 (35,61; 38,93)*	0	0
Técnicos e profissionais de nível médio	24,58 (23,33; 25,83)*	0	0
Trabalhadores de apoio administrativo	23,09 (21,76; 24,41)*	0	0
Trabalhadores dos serviços, vendedores dos comércios e mercados	0	34,84 (33,78; 35,90)	0
Trabalhadores qualificados da agropecuária, florestais, da caça e da pesca	0	8,62 (8,06; 9,17)	0
Trabalhadores qualificados, operários e artesãos da construção, das artes mecânicas e outros ofícios	0	19,35 (18,44; 20,27)	0
Operadores de instalações e máquinas e montadores	0	12,69 (11,95; 13,43)	0
Ocupações elementares	0	24,5 (23,58; 25,42)	0
Ocupações mal definidas	0	0	0,18 (0,00; 0,38)
Não ocupados	0	0	99,82 (99,62; 100,03)*
Escolaridade			
Fundamental incompleto	2,51 (2,11; 2,92)*	35,26 (34,26; 36,25)	39,01 (37,38; 40,64)*
Fundamental completo/médio incompleto	4,28 (3,69; 4,86)*	18,61 (17,85; 19,36)	16,60 (15,34; 17,85)
Médio completo/superior incompleto	39,04 (37,48; 40,60)	40,36 (39,38; 41,33)	36,10 (34,38; 37,81)*
Superior completo	54,17 (52,49; 55,85)*	5,78 (5,24; 6,32)	8,29 (7,35; 9,24)*
Renda per capita			
Não definido	0,01 (0,00; 0,02)	0,07 (0,02; 0,13)	0 (0,00; 0,00)*
Até 1/4 salário mínimo	0,33 (0,22; 0,44)*	5,85 (5,45; 6,24)	21,66 (20,42; 22,89)*
Mais de 1/4 até 1/2 salário mínimo	3,03 (2,57; 3,48)*	14,58 (13,94; 15,23)	24,01 (22,50; 25,53)*
Mais de 1/2 até 1 salário mínimo	16,15 (14,99; 17,31)*	32,40 (31,40; 33,39)	29,29 (27,80; 30,77)*
Mais de 1 até 2 salários mínimos	31,68 (30,25; 33,12)	32,97 (31,88; 34,07)	16,15 (14,87; 17,44)*
Mais de 2 até 3 salários mínimos	16,23 (15,17; 17,30)*	8,49 (7,93; 9,06)	4,40 (3,61; 5,18)*
Mais de 3 até 5 salários mínimos	15,78 (14,63; 16,94)*	4,06 (3,65; 4,47)	2,57 (2,00; 3,13)*
Mais de 5 salários mínimos	16,79 (15,39; 18,18)*	1,58 (1,36; 1,80)	1,93 (1,42; 2,44)

Tabela 1: Frequências estimadas (com intervalos de 95% de confiança) de variáveis sociodemográficas e socioeconômicas por categorias de ocupação para a população na força de trabalho - expansão da amostra da PNS 2019

Variáveis	(conclusão)		
	Colarinho branco	Colarinho azul	Não ocupados/ Mal definidas
Sexo			
Masculino	47,65 (46,17; 49,13)*	56,79 (55,59; 57,99)	32,67 (31,01; 34,34)*
Feminino	52,35 (50,87; 53,83)*	43,21 (42,01; 44,41)	67,33 (65,66; 68,99)*
Faixa etária			
18 a 24 anos	11,78 (10,62; 12,95)	11,55 (10,73; 12,38)	21,67 (20,06; 23,29)*
25 a 39 anos	41,42 (40,01; 42,83)*	33,29 (32,35; 34,22)	26,8 (25,3; 28,3)*
40 a 59 anos	39,58 (38,20; 40,97)*	44,54 (43,50; 45,58)	35,23 (33,70; 36,75)*
60 anos ou mais	7,21 (6,51; 7,91)*	10,62 (10,04; 11,20)	16,30 (15,20; 17,40)*
Raça			
Parda	32,44 (31,01; 33,87)*	46,01 (45,01; 47,01)	49,10 (47,40; 50,80)*
Preta	9,44 (8,63; 10,26)*	12,72 (12,05; 13,39)	11,96 (10,89; 13,03)
Branca	56,32 (54,70; 57,94)*	39,97 (38,89; 41,05)	37,58 (35,89; 39,27)
Amarela/Indígena/Ignorado	1,79 (1,30; 2,29)	1,30 (1,10; 1,50)	1,36 (0,94; 1,77)
Estado civil			
Solteiro	40,82 (39,35; 42,30)*	43,57 (42,40; 44,74)	51,21 (49,62; 52,79)*
Casado	49,33 (47,86; 50,81)*	45,33 (44,17; 46,48)	39,02 (37,46; 40,58)*
Separado/Divorciado/Viúvo	9,84 (9,09; 10,59)	11,1 (10,52; 11,68)	9,78 (8,91; 10,65)
Situação de domicílio			
Urbano	96,74 (96,40; 97,09)*	85,07 (84,41; 85,72)	86,71 (85,70; 87,71)
Rural	3,26 (2,91; 3,60)*	14,93 (14,28; 15,59)	13,29 (12,29; 14,30)
Região			
Nordeste	18,69 (17,61; 19,77)*	24,92 (23,97; 25,87)	32,32 (30,87; 33,77)*
Norte	5,96 (5,51; 6,41)*	7,63 (7,2; 8,06)	8,85 (8,12; 9,58)*
Sudeste	50,55 (48,81; 52,29)*	44,57 (43,11; 46,03)	41,41 (39,49; 43,34)
Sul	16,58 (15,56; 17,61)	14,94 (14,20; 15,68)	10,54 (9,57; 11,52)*
Centro-Oeste	8,22 (7,54; 8,89)	7,94 (7,47; 8,41)	6,87 (6,19; 7,55)
Doenças crônicas			
Pelo menos uma	39,87 (38,40; 41,34)	37,68 (36,65; 38,71)	44,89 (43,11; 46,68)*
Não possui	60,13 (58,66; 61,60)	62,32 (61,29; 63,35)	55,11 (53,32; 56,89)*

Fonte: Pesquisa Nacional de Saúde 2019.

*: Variáveis significativas comparando colarinho branco *versus* colarinho azul e não ocupados/ mal definidas *versus* colarinho azul exceto nos casos em que as frequências são iguais a zero como para os níveis da COD; COD: Classificação de Ocupações para Pesquisas Domiciliares; SCV: Saúde cardiovascular.

Quanto às comparações das demais covariáveis entre grupos de ocupação por colarinho, observou-se diferenças significantes em quase todas as variáveis. No grupo classificado com ocupações de colarinho branco (comparado ao de colarinho azul) observou-se maiores frequências de indivíduos mais jovens, com maiores níveis de escolaridade, nas faixas de rendas mais altas, do sexo feminino, de raça/etnia branca, casados e residentes em áreas urbanas. Frequências maiores foram observadas para residentes da região Sudeste e menores para as regiões Norte e Nordeste. Já o grupo de não ocupados (ou com ocupações mal definidas) comparados ao grupo de colarinho azul, foi caracterizado com níveis ainda menores de escolaridade e renda, maiores frequências de mulheres, de indivíduos de raça/etnia parda, solteiros, residentes nas regiões Nordeste e Norte e com pelo menos uma doença crônica.

Analisando os resultados apresentados na Tabela 2, observou-se que apenas 9,96% ($IC_{95\%}$: 9,47%; 10,44%) da população brasileira na força de trabalho apresentou saúde cardiovascular superior, ou seja, alcance de 6 a 7 metas concomitantes. Esta tabela também evidencia que houve diferenças significantes entre as prevalências de diferentes categorias das variáveis explicativas comparando-se entre os grupos com SCV superior e inferior. Entre indivíduos com SCV superior (comparados aos com SCV inferior) observou-se maiores frequências de indivíduos com ocupações do tipo colarinho branco, com maiores níveis de escolaridade, nas faixas de rendas mais altas, do sexo feminino, de raça/etnia branca, solteiros, residentes em áreas urbanas e sem doenças crônicas.

Tabela 2: Frequências estimadas (com intervalos de 95% de confiança) de variáveis sociodemográficas e socioeconômicas por categorias de saúde cardiovascular para a população na força de trabalho - Expansão da amostra da PNS 2019

Variáveis	SCV Superior	SCV Inferior	Total
Total	9,96 (9,47; 10,44)	90,05 (89,56; 90,53)	
Ocupação			
Colarinho branco	42,31 (39,89; 44,73)*	25,62 (24,77; 26,46)	27,28 (26,43; 28,13)
Colarinho azul	41,07 (38,63; 43,51)*	55,56 (54,63; 56,49)	54,12 (53,20; 55,03)
Não ocupados/ Mal definidas	16,62 (14,99; 18,25)	18,82 (18,16; 19,48)	18,60 (17,98; 19,22)
COD			
Membros das forças armadas, policiais e bombeiros militares	0,57 (0,33; 0,81)	0,57 (0,47; 0,67)	0,57 (0,48; 0,67)
Diretores e gerentes	5,05 (4,00; 6,10)*	3,37 (3,02; 3,71)	3,54 (3,19; 3,88)
Profissionais das ciências e intelectuais	20,30 (18,35; 22,26)*	9,05 (8,50; 9,60)	10,17 (9,58; 10,76)
Técnicos e profissionais de nível médio	9,53 (8,02; 11,04)*	6,39 (6,02; 6,77)	6,71 (6,34; 7,07)
Trabalhadores de apoio administrativo	6,86 (5,68; 8,04)	6,24 (5,83; 6,64)	6,30 (5,91; 6,69)
Trabalhadores dos serviços, vendedores dos comércios e mercados	18,25 (16,35; 20,15)	18,92 (18,27; 19,57)	18,86 (18,23; 19,48)
Trabalhadores qualificados da agropecuária, florestais, da caça e da pesca	2,09 (1,56; 2,61)*	4,95 (4,62; 5,27)	4,66 (4,37; 4,96)
Trabalhadores qualificados, operários e artesãos da construção, das artes mecânicas e outros ofícios	7,65 (6,33; 8,97)*	10,79 (10,19; 11,38)	10,47 (9,92; 11,03)
Operadores de instalações e máquinas e montadores	4,89 (3,91; 5,88)*	7,09 (6,65; 7,53)	6,87 (6,45; 7,29)
Ocupações elementares	8,19 (6,87; 9,51)*	13,82 (13,23; 14,40)	13,26 (12,71; 13,81)
Ocupações mal definidas	0,03 (0,00; 0,09)	0,03 (0,00; 0,08)	0,03 (0,00; 0,07)
Não ocupado	16,59 (14,96; 18,22)	18,79 (18,13; 19,45)	18,57 (17,95; 19,19)
Escolaridade			
Fundamental incompleto	9,54 (8,26; 10,81)*	28,96 (28,17; 29,74)	27,02 (26,28; 27,77)
Fundamental completo/médio incompleto	11,10 (9,63; 12,57)*	14,68 (14,14; 15,23)	14,32 (13,81; 14,84)
Médio completo/superior incompleto	44,57 (42,24; 46,9)*	38,61 (37,82; 39,40)	39,21 (38,45; 39,96)
Superior completo	34,80 (32,43; 37,17)*	17,75 (16,96; 18,55)	19,45 (18,62; 20,27)
Renda per capita			
Não definido	0,02 (0,00; 0,05)	0,04 (0,01; 0,08)	0,04 (0,01; 0,07)
Até 1/4 salário mínimo	4,75 (3,88; 5,62)*	7,56 (7,20; 7,92)	7,28 (6,94; 7,62)
Mais de 1/4 até 1/2 salário mínimo	9,26 (8,10; 10,42)*	13,62 (13,05; 14,18)	13,18 (12,65; 13,72)
Mais de 1/2 até 1 salário mínimo	23,20 (21,19; 25,21)*	27,85 (27,07; 28,62)	27,39 (26,63; 28,14)
Mais de 1 até 2 salários mínimos	26,52 (24,29; 28,74)*	29,82 (28,97; 30,67)	29,49 (28,69; 30,30)
Mais de 2 até 3 salários mínimos	13,14 (11,51; 14,76)*	9,48 (9,00; 9,96)	9,84 (9,38; 10,31)
Mais de 3 até 5 salários mínimos	11,52 (9,91; 13,13)*	6,48 (6,05; 6,9)	6,98 (6,54; 7,42)
Mais de 5 salários mínimos	11,59 (10,13; 13,06)*	5,15 (4,69; 5,62)	5,792 (5,29; 6,29)

(continua)

Tabela 2: Frequências estimadas (com intervalos de 95% de confiança) de variáveis sociodemográficas e socioeconômicas por categorias de saúde cardiovascular para a população na força de trabalho - expansão da amostra da PNS 2019

Variáveis	SCV Superior	SCV Inferior	Total (conclusão)
Sexo			
Masculino	46,07 (43,68; 48,45)*	50,22 (49,36; 51,09)	49,81 (49,00; 50,63)
Feminino	53,93 (51,55; 56,32)*	49,78 (48,91; 50,64)	50,19 (49,37; 51,00)
Faixa etária			
18 a 24 anos	25,16 (22,81; 27,52)*	12,21 (11,56; 12,85)	13,50 (12,86; 14,14)
25 a 39 anos	40,37 (38,08; 42,66)*	33,63 (32,87; 34,38)	34,30 (33,59; 35,01)
40 a 59 anos	29,54 (27,49; 31,6)*	42,77 (41,99; 43,55)	41,46 (40,71; 42,20)
60 anos ou mais	4,92 (4,06; 5,78)*	11,39 (10,91; 11,87)	10,75 (10,31; 11,19)
Raça			
Parda	39,61 (37,25; 41,96)*	43,25 (42,43; 44,07)	42,89 (42,09; 43,69)
Preta	10,06 (8,68; 11,45)	11,86 (11,34; 12,39)	11,69 (11,19; 12,18)
Branca	48,77 (46,20; 51,34)*	43,46 (42,59; 44,33)	43,99 (43,11; 44,86)
Amarela/Indígena/Ignorado	1,56 (1,05; 2,07)	1,43 (1,23; 1,64)	1,44 (1,25; 1,64)
Estado civil			
Solteiro	55,73 (53,46; 58,01)*	42,97 (42,09; 43,86)	44,24 (43,39; 45,10)
Casado	36,48 (34,26; 38,70)*	46,21 (45,34; 47,09)	45,245 (44,41; 46,08)
Separado/Divorciado/Viúvo	7,78 (6,80; 8,77)*	10,81 (10,36; 11,26)	10,51 (10,09; 10,93)
Situação de domicílio			
Urbano	92,78 (91,82; 93,73)*	88,09 (87,63; 88,55)	88,56 (88,12; 88,99)
Rural	7,22 (6,27; 8,18)*	11,91 (11,45; 12,37)	11,44 (11,01; 11,88)
Região			
Nordeste	26,63 (24,71; 28,55)	24,37 (23,68; 25,06)	24,60 (23,93; 25,26)
Norte	7,70 (6,90; 8,49)	7,37 (7,04; 7,70)	7,40 (7,08; 7,72)
Sudeste	43,71 (41,04; 46,39)	45,82 (44,78; 46,87)	45,61 (44,63; 46,60)
Sul	13,30 (11,69; 14,9)	14,71 (14,15; 15,27)	14,57 (14,04; 15,11)
Centro-Oeste	8,67 (7,56; 9,77)	7,73 (7,38; 8,07)	7,82 (7,47; 8,17)
Doenças crônicas			
Pelo menos uma	31,05 (28,85; 33,24)*	40,57 (39,73; 41,41)	39,62 (38,81; 40,43)
Não possui	68,95 (66,76; 71,15)*	59,43 (58,59; 60,27)	60,38 (59,57; 61,19)

Fonte: Pesquisa Nacional de Saúde 2019.

SCV: Saúde cardiovascular; *: Variáveis significativas comparando SCV superior *versus* SCV inferior; COD: Classificação de Ocupações para Pesquisas Domiciliares.

As Tabelas 3 a 5 apresentam resultados de análises de 3 modelos de regressão logística (denominados **Modelos 1, 2 e 3**, respectivamente) da variável resposta binária SCV (com SCV inferior como grupo de referência), ajustados por ocupação do tipo colarinho (com colarinho azul como referência) em adição à diferentes grupos de covariáveis, elencadas posteriormente.

O **Modelo 1**, também chamado de minimamente ajustado e cujos resultados são apresentados na Tabela 3, incluiu apenas as covariáveis faixa etária, sexo e raça, além de ocupação por colarinho (presente em todos os modelos por ser a variável explicativa de interesse principal). Observou-se que indivíduos pertencentes à categoria de ocupação colarinho branco possuem 2,07 ($IC_{95\%}$: 1,85; 2,33) vezes mais chance de apresentarem SCV superior que os de colarinho azul, ajustando-se pelas covariáveis. Não houve diferença estatisticamente significativa entre os grupos de não ocupados (ou de ocupações mal definidas) e de colarinho branco. Todas as covariáveis apresentaram comparações significantes, sendo a chance de apresentar SCV superior maior entre mulheres, mais jovens e de raça/etnia branca. Destaca-se a relação inversa entre SCV e faixa etária pelo decréscimo das RCs (todas < 1) indicando que quanto mais velhos eram os indivíduos, menores as chances de apresentarem SCV superior. Ressalta-se ainda que as comparações envolvendo níveis de raça/etnia “preta” e “amarela/indígena/ignorados” (comparadas à parda) não foram estatisticamente significantes.

O **Modelo 2** (moderadamente ajustado) difere do **Modelo 1** por incluir também as variáveis estado civil, região, situação de domicílio e nível de escolaridade (Tabela 4). A magnitude do efeito da variável ocupação por colarinho diminui de modo que a chance de se apresentar SCV superior entre indivíduos do grupo colarinho branco passa a ser apenas 23% maior que para o grupo de colarinho azul, ajustando-se pelas demais covariáveis. Resultados similares aos do **Modelo 1** são observados para as variáveis faixa etária e raça mas a variável sexo deixa de ser significativa. Nota-se que nível de escolaridade apresenta uma relação positiva e forte com SCV, observada pelas estimativas crescentes de RCs variando entre 1,72 a 4,51. O maior efeito foi observado para indivíduos com ensino superior completo, que apresentaram 4,51 ($IC_{95\%}$: 3,71; 5,47) vezes mais chances estar no grupo de SCV superior que o grupo de referência, que inclui os indivíduos com ensino fundamental incompleto. Além disso, na presença das demais variáveis, solteiros e residentes de regiões urbanas apresentaram maiores chances de estar no grupo de SCV superior. Indivíduos residentes nas regiões Sul e Sudeste apresentaram menores chances de SCV superior que os na região Nordeste (referência). Não houve diferença da região Nordeste com as regiões Norte e Centro-oeste (Tabela 4).

O Modelo 3, cujos resultados são apresentados na Tabela 5, foi o modelo que envolveu o ajuste de mais covariáveis. Além das incluídas no Modelo 2, foram incorporadas as variáveis faixas de renda *per capita* e presença de pelo menos uma doença crônica (*versus* nenhuma). Ao adicionar tais variáveis ao modelo (ambas significantes), ocupação por colarinho e raça deixaram de ser significantes e os efeitos de escolaridade, estado civil e situação de domicílio diminuíram e os efeitos de região e de sexo aumentaram. Esse modelo apresenta como variáveis de maior destaque, tal como nos modelos anteriores, a faixa etária e a escolaridade, incluindo-se também para este caso a variável renda *per capita*. Indivíduos com maiores renda *per capita* possuem maiores chances de apresentarem SCV superior. Por exemplo, indivíduos que possuem mais de 5 salários mínimos têm cerca de 2,66 ($IC_{95\%}$: 2,00; 3,55) vezes mais chances estarem no grupo de SCV superior que de menor grupo de renda (até 1/4 de salário mínimo *per capita*). A presença de doenças crônicas é a única variável clínica utilizada como explicativa. Apesar de seu efeito não ser tão acentuado, verifica-se que indivíduos sem doenças crônicas possuem 43% ($IC_{95\%}$: 1,28; 1,60) mais chances de estarem na classe superior de SCV.

Tabela 3: Razões de chances, IC95% e p-valores para o Modelo 1 de regressão logística com a variável resposta SCV (superior vs inferior) na população brasileira na força de trabalho – Expansão da amostra PNS 2019

Variáveis	P-valor ^a	RC	IC _{95%}	P-valor
(Intercepto)		0,16	(0,14; 0,19)	< 0,001
Ocupação (Ref: Colarinho azul)	< 0,001			
Colarinho branco		2,07	(1,85; 2,33)	< 0,001
Não ocupados/ Mal definidas		1,06	(0,92; 1,22)	0,422
Sexo (Ref: Masculino)	0,005*			
Feminino		1,16	(1,05; 1,29)	0,005*
Faixa etária (Ref: 18 a 24 anos)	< 0,001			
25 a 39 anos		0,54	(0,46; 0,63)	< 0,001
40 a 59 anos		0,32	(0,28; 0,38)	< 0,001
60 anos ou mais		0,21	(0,17; 0,27)	< 0,001
Raça (Ref: Parda)	0,023*			
Preta		0,93	(0,78; 1,10)	0,368
Branca		1,15	(1,03; 1,29)	0,011*
Amarela/Indígena/Ignorado		1,12	(0,73; 1,71)	0,597

Fonte: Pesquisa Nacional de Saúde 2019.

Ref: Referência; RC: Razão de Chances; *a*: P-valor do teste global; *: P-valor < 0,05.

Tabela 4: Razões de chances, IC95% e p-valores para o Modelo 2 de regressão logística com a variável resposta SCV (superior vs inferior) na população brasileira na força de trabalho – Expansão da amostra PNS 2019

Variáveis	P-valor ^a	RC	IC _{95%}	P-valor
(Intercepto)		0,11	(0,09; 0,15)	< 0,001
Ocupação (Ref: Colarinho azul)	< 0,001			
Colarinho branco		1,23	(1,08; 1,41)	0,002*
Não ocupados/ Mal definidas		1,01	(0,87; 1,17)	0,934
Escolaridade (Ref: Fundamental incompleto)	< 0,001			
Fundamental completo/médio incompleto		1,72	(1,39; 2,12)	< 0,001
Médio completo/superior incompleto		2,37	(2,01; 2,80)	< 0,001
Superior completo		4,51	(3,71; 5,47)	< 0,001
Sexo (Ref: Masculino)	0,12			
Feminino		1,09	(0,98; 1,21)	0,122
Faixa etária (Ref: 18 a 24 anos)	< 0,001			
25 a 39 anos		0,53	(0,45; 0,62)	< 0,001
40 a 59 anos		0,38	(0,32; 0,44)	< 0,001
60 anos ou mais		0,29	(0,23; 0,37)	< 0,001
Raça (Ref: Parda)	0,05			
Preta		0,93	(0,78; 1,11)	0,439
Branca		1,15	(1,02; 1,29)	0,022*
Amarela/Indígena/Ignorado		1,09	(0,73; 1,63)	0,683
Estado civil (Ref: Solteiro)	< 0,001			
Casado		0,78	(0,70; 0,88)	< 0,001
Separado/Divorciado/Viúvo		0,84	(0,71; 0,99)	0,041*
Situação de domicílio (Ref: Urbano)	< 0,001			
Rural		0,80	(0,69; 0,93)	0,004*
Região (Ref: Nordeste)	< 0,001			
Norte		0,90	(0,79; 1,02)	0,103
Sudeste		0,71	(0,63; 0,81)	< 0,001
Sul		0,67	(0,57; 0,79)	< 0,001
Centro-Oeste		0,88	(0,76; 1,02)	0,093

Fonte: Pesquisa Nacional de Saúde 2019.

Ref: Referência; RC: Razão de Chances; *a*: P-valor do teste global; *: P-valor < 0,05.

Tabela 5: Razões de chances, IC95% e p-valores para o Modelo 3 de regressão logística com a variável resposta SCV (superior vs inferior) na população brasileira na força de trabalho – Expansão da amostra PNS 2019

Variáveis	P-valor ^a	RC	IC _{95%}	P-valor
(Intercepto)		0,08	(0,06; 0,11)	< 0,001
Ocupação (Ref: Colarinho azul)	0,263			
Colarinho branco		1,11	(0,97; 1,27)	0,134
Não ocupados/ Mal definidas		1,08	(0,93; 1,26)	0,289
Escolaridade (Ref: Fundamental incompleto)	< 0,001			
Fundamental completo/médio incompleto		1,61	(1,3; 1,98)	< 0,001
Médio completo/superior incompleto		2,11	(1,78; 2,51)	< 0,001
Superior completo		3,31	(2,67; 4,11)	< 0,001
Renda per capita (Ref: Até 1/4 de salário mínimo)	< 0,001			
Não definido		0,37	(0,03; 4,51)	0,432
Mais de 1/4 até 1/2 salário mínimo		0,98	(0,78; 1,24)	0,883
Mais de 1/2 até 1 salário mínimo		1,21	(0,97; 1,50)	0,084
Mais de 1 até 2 salários mínimos		1,30	(1,04; 1,64)	0,024*
Mais de 2 até 3 salários mínimos		1,92	(1,49; 2,48)	< 0,001
Mais de 3 até 5 salários mínimos		2,19	(1,63; 2,93)	< 0,001
Mais de 5 salários mínimos		2,66	(2,00; 3,55)	< 0,001
Sexo (Ref: Masculino)	0,002*			
Feminino		1,19	(1,06; 1,32)	0,002*
Faixa etária (Ref: 18 a 24 anos)	< 0,001			
25 a 39 anos		0,53	(0,45; 0,63)	< 0,001
40 a 59 anos		0,37	(0,31; 0,43)	< 0,001
60 anos ou mais		0,26	(0,2; 0,33)	< 0,001
Raça (Ref: Parda)	0,534			
Preta		0,94	(0,79; 1,12)	0,521
Branca		1,07	(0,95; 1,20)	0,285
Amarela/Indígena/Ignorado		1,06	(0,72; 1,58)	0,762
Estado civil (Ref: Solteiro)	< 0,001			
Casado		0,79	(0,7; 0,88)	< 0,001
Separado/Divorciado/Viúvo		0,85	(0,72; 1,01)	0,063
Situação de domicílio (Ref: Urbano)	0,05			
Rural		0,86	(0,74; 1,00)	0,048*
Região (Ref: Nordeste)	< 0,001			
Norte		0,87	(0,77; 1,00)	0,045*
Sudeste		0,66	(0,58; 0,75)	< 0,001
Sul		0,61	(0,52; 0,72)	< 0,001
Centro-Oeste		0,79	(0,69; 0,92)	0,002*
Doenças crônicas (Ref: Pelo menos uma)	< 0,001			
Não possui		1,43	(1,28; 1,60)	< 0,001

Fonte: Pesquisa Nacional de Saúde 2019.

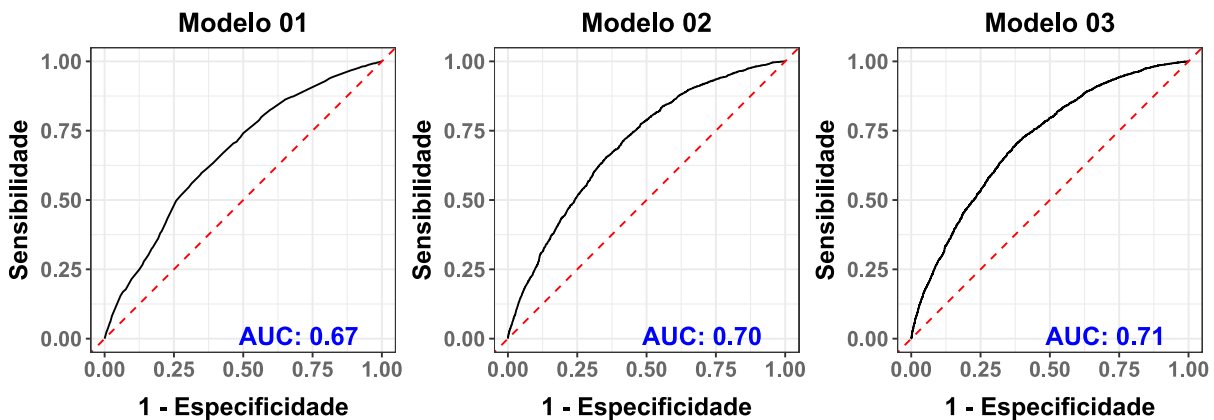
Ref: Referência; RC: Razão de Chances; a: P-valor do teste global; *: P-valor < 0,05.

Para realizar uma análise global das variáveis, optou-se por empregar o teste de Wald aproximado com uma distribuição F. Esse teste produziu interpretações semelhantes às anteriores, ou seja, a variável ocupação por colarinho apresentou significância estatística somente nos Modelos 01 e 02.

A fim de avaliar a adequação do modelo, calculou-se o coeficiente R^2 de McFadden, que resultou em valores baixos dados por 0,048 (Modelo 1), 0,072 (Modelo 2) e 0,083 (Modelo 3). Esses resultados indicam uma melhora no ajuste à medida que as variáveis explicativas foram incluídas, ou seja, o Modelo 1 apresenta um ajuste relativamente pior, enquanto os Modelo 2 e 3 indicam melhores ajustes.

Uma outra abordagem utilizada para avaliar os modelos foi por meio da curva ROC e sua respectiva área sob a curva (AUC), apresentadas em Figura 1. As AUCs obtidas para os Modelos 1 a 3 foram respectivamente: $AUC_1 = 0,67$, $AUC_2 = 0,70$ e $AUC_3 = 0,71$. Seguindo as interpretações propostas por Hosmer, Lemeshow e Sturdivant (2013), constata-se que o Modelo 1 apresenta uma capacidade de discriminação fraca ($0,5 < ROC < 0,7$). Já os Modelos 2 e 3 exibem uma capacidade de discriminação aceitável ($0,7 \leq ROC < 0,8$).

Figura 1: Curva ROC e AUC dos modelos minimamente ajustado (Modelo 01), moderadamente ajustado (Modelo 02) e modelo completo (Modelo 03)



5 Discussão

Este estudo apresenta frequências estimadas por expansão da amostra e avalia a associação entre saúde cardiovascular e ocupação ajustando-se por outros determinantes sociodemográficos e socioeconômicos na população adulta brasileira que estava na força de trabalho com base em dados da PNS de 2019. Identificou-se associação entre saúde cardiovascular e ocupação por tipo de colarinho apenas nos modelos logísticos minimamente e moderadamente ajustados (Modelos 1 e 2). Tal relação perde sua significância no modelo mais completo (Modelo 3) com a inclusão das variáveis renda *per capita* e presença de doenças crônicas.

Observou-se que apenas 7,56% ($IC_{95\%}$: 6,98%; 8,13%) dos indivíduos da categoria colarinho azul foram classificados como tendo saúde cardiovascular superior, um resultado bastante próximo ao grupo de desocupados/mal definidos. Em contrapartida, os trabalhadores de colarinho branco possuem quase o dobro de indivíduos com saúde cardiovascular superior em comparação com os trabalhadores de colarinho azul. Foram identificados poucos estudos sobre a população brasileira nesse contexto, mas em outros países há pesquisas que exploram o impacto adicional da ocupação na saúde. Com isto, foi identificado que os trabalhadores de colarinho azul, geralmente envolvidos em trabalhos mais braçais, mostram-se menos propensos a adotar comportamentos de estilo de vida saudáveis em comparação com os trabalhadores de colarinho branco (MORRIS et al., 1999).

A natureza do trabalho coloca demandas diretas sobre os indivíduos e há uma variação substancial nas demandas físicas entre diferentes ocupações, com o trabalho de colarinho azul sendo mais exigente fisicamente (SHINYA, 2015). Com isto, estudos identificam que o trabalho de colarinho azul está associado a uma redução mais rápida da saúde com o avanço da idade, resultando em uma deterioração mais acelerada da saúde desses trabalhadores em comparação com os de colarinho branco (CASE; DEATON, 2005; CHOO; DENNY et al., 2006). Além disso, a exposição ocupacional a condições de trabalho insalubres e prejudiciais possuem uma contribuição para a depreciação do estoque de saúde, sendo que em contraste, os trabalhadores de colarinho branco, devido a ocupações melhor remuneradas, têm maior capacidade financeira para investir em saúde e maior incentivo para fazê-lo (SHINYA, 2015; MOREFIELD; RIBAR; RUHM, 2012).

Modelos de regressão logística foram utilizados para avaliar a associação entre saúde cardiovascular (variável resposta) e ocupação por colarinho na presença de covariáveis sociodemográficas e socioeconômicas. Observou-se que tal associação se tornou

mais fraca à medida que mais covariáveis foram sendo adicionadas aos modelos (Modelos 1 a 3). Em particular observou-se que ao se incluir a variável renda *per capita* ao modelo, ocupação deixou de ser significativa, o que sugere a informação adicionada pela variável ocupação é redundante à de renda. É possível ainda que renda seja um fator socioeconômico ainda mais relevante na discriminação entre grupos de saúde cardiovascular. Isso pode ser explicado pela capacidade da renda em captar a variação de informações relacionadas ao colarinho, ou seja, à posição socioeconômica dos indivíduos. Os dados da PNS 2019 mostram um relação forte entre ocupação por colarinho e renda *per capita*, tendo os indivíduos de ocupações do tipo colarinho branco frequências muito mais altas nas faixas mais elevadas de renda *per capita* que os de colarinho azul (Tabela 1). Frequências ainda mais baixas de renda são observadas entre os desocupados.

A literatura aponta a renda como uma variável importante associada à saúde cardiovascular sendo que esta se relaciona com ao acesso a bens e serviços, incluindo o acesso a uma educação de qualidade e aos cuidados e tratamentos médicos (KAPLAN; KEIL, 1993). Por outro lado, a ocupação é frequentemente utilizada em estudos com diferentes categorizações, o que pode resultar em impactos variados em relação à sua importância para a saúde cardiovascular (KAPLAN; KEIL, 1993). Conforme Schreuder et al. (2008), as diferenças socioeconômicas podem exercer um impacto maior na qualidade da saúde dos indivíduos do que as condições de trabalho. Logo, avaliando a renda dos indivíduos de acordo com o colarinho, neste estudo é constatado que mais de 30% dos brasileiros de colarinho branco recebem acima de 3 salários mínimos, enquanto menos de 6% dos trabalhadores de colarinho azul têm a mesma remuneração. Essa disparidade também é evidente na análise da escolaridade, em que mais de 50% dos indivíduos de colarinho branco possuem ensino superior completo, enquanto menos de 6% dos trabalhadores de colarinho azul alcançaram o mesmo nível educacional.

Nos resultados obtidos identifica-se também que indivíduos do sexo feminino são relacionados a mais chances de apresentarem saúde cardiovascular superior (Tabela 5). De modo geral, as mulheres tendem a utilizar mais os serviços de saúde do que os homens e essa diferença pode ser atribuída, em parte, às necessidades específicas das mulheres relacionadas à gravidez e ao parto (PINHEIRO et al., 2002). A qualidade da saúde desses grupos pode se diferenciar devido a fatores como a maiores frequências de tabagismo e de consumo excessivo de álcool e às desvantagens em situações relacionadas ao trabalho, resultando em um significativo aumento dos riscos de problemas de saúde a longo prazo entre homens (PINHEIRO et al., 2002).

Além da variável sexo, observa-se pelo Modelo 3 que indivíduos com idades mais

avançadas possuem mais chances de desenvolverem problemas cardiovasculares o que pode ser explicado por maiores períodos de exposição aos fatores de risco, tais como poluição, tabagismo, alimentação inadequada e falta de atividade física (COSTANTINO; PANENI; COSENTINO, 2016). Constata-se também que as regiões Norte e Nordeste do Brasil são apontadas em estudos como os locais com maiores frequências de doenças cardiovasculares (LEITE et al., 2015) e piores prevalências sobre indicadores de saúde cardiovascular ideal (MATOZINHOS et al., 2017). Entretanto, ajustando-se pelas demais covariáveis, observou-se que os residentes no Nordeste apresentam mais chances de possuírem saúde cardiovascular superior comparada com as demais regiões do Brasil (Tabela 5).

A utilização da PNS 2019 proporciona uma série de vantagens e benefícios. Dentre esses, destacam-se a sua representatividade nacional, a diversidade de temas relacionados à saúde, a possibilidade de compará-la com edições anteriores e a contribuição no embasamento de políticas públicas de saúde com base em evidências. Essa pesquisa também fornece informações atualizadas e detalhadas sobre a saúde da população brasileira, permitindo uma compreensão mais abrangente das condições de saúde, acesso aos serviços e estilos de vida. Com isto, a PNS auxilia na tomada de decisões e no planejamento de ações para promover a saúde e o bem-estar da população.

No entanto, este estudo possui uma limitação a qual reside no uso de medidas construídas a partir de respostas autorreferidas, o que pode resultar em uma subestimação dos diagnósticos e, por consequência, uma superestimação da prevalência da SCV superior. Contudo, é possível assumir que os resultados obtidos neste estudo são relevantes para a sociedade, devido à ampla quantidade de informações disponíveis na PNS 2019, bem como ao sólido planejamento amostral dos dados. Além disso, é válido destacar que estudos anteriores com uma perspectiva de aplicação similar apresentaram bons resultados, como evidenciado por Moreira et al. (2020).

6 Conclusão

Conclui-se que a variável ocupação por colarinho não se mostrou associada à saúde cardiovascular ajustando-se por outras variáveis, em particular pela variável renda *per capita*. Tais resultados sugerem que seria mais produtivo se trabalhar com renda *per capita* como variável socioeconômica em conjunto com a escolaridade. No entanto, é necessário salientar que a relevância da renda *per capita* e da escolaridade não exclui a possibilidade de outras variáveis socioeconômicas apresentarem significância estatística. Portanto, ao abordar a relação entre as variáveis socioeconômicas e a saúde cardiovascular dos indivíduos, é essencial adotar uma abordagem abrangente, considerando múltiplos fatores sociodemográficos e socioeconômicos, a fim de se obter uma compreensão mais completa e precisa dos determinantes envolvidos.

Referências

- AGRESTI, A. **An introduction to categorical data analysis**. [S.l.]: John Wiley & Sons, 2007.
- AGRESTI, A. **Categorical Data Analysis**. 3rd. ed. EUA: John Wiley & Sons, 2013.
- ASSUNÇÃO, A. A.; CLARO, R. M. **Characteristics of work and employment related to leisure-time physical activity: results of the National Health Survey, Brazil, 2013**. *Annals of Work Exposures and Health*, Oxford University Press UK, v. 66, n. 1, p. 102–112, 2022.
- BOUSQUET-SANTOS, K.; CHEN, R.; KUBZANSKY, L. D. **A sad heart: Depression and favorable cardiovascular health in Brazil**. *Preventive Medicine*, v. 142, p. 106378, 2021. ISSN 0091-7435.
- CARRAPATO, P.; CORREIA, P.; GARCIA, B. **Health determinants in Brasil: searching for health equity**. *Saúde e Sociedade*, v. 26, n. 3, p. 676–689, 2017. DOI <https://doi.org/10.1590/s0104-12902017170304>.
- CARVALHO, T. G. S. de et al. **Exposure to chemical and biological agents at work and cardiovascular disease in Brazil: a population-based study**. *Journal of Occupational and Environmental Medicine*, LWW, v. 63, n. 6, p. e341–e345, 2021.
- CASE, A.; DEATON, A. S. Broken down by work and sex: How our health declines. In: **Analyses in the Economics of Aging**. [S.l.]: University of Chicago Press, 2005. p. 185–212.
- CHOO, E.; DENNY, M. et al. **Wearing out the decline in health**. University of Toronto, Department of Economics, Working Papers, 2006.
- COSTANTINO, S.; PANENI, F.; COSENTINO, F. **Ageing, metabolism and cardiovascular disease**. *The Journal of physiology*, Wiley Online Library, v. 594, n. 8, p. 2061–2073, 2016.
- HARRISON, E.; ROSE, D. **The European Socio-Economic Classification (ESeC)-Draft User Guide**. Institute for Social and Economic Research. University of Essex, Colchester, UK, 2006.
- HAVRANEK, E. P. et al. **Social Determinants of Risk and Outcomes for Cardiovascular Disease**. *Circulation*, v. 132, n. 9, p. 873–898, 2015. Disponível em: <https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000000228>.
- HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied Logistic Regression**. EUA: John Wiley & Sons, 2013.
- IBGE. **Principais diferenças entre a COD e a CBO - Domiciliar**. Instituto Brasileiro de Geografia e Estatística (IBGE), 2010. Disponível em: https://www.ibge.gov.br/arquivo/projetos/sipd/oitavo_forum/COD.pdf. Acesso em: 04 jan. 2023.

- JÚNIOR, P. R. B. d. S. et al. **Cobertura de plano de saúde no Brasil: análise dos dados da Pesquisa Nacional de Saúde 2013 e 2019**. *Ciência e Saúde Coletiva*, SciELO Brasil, v. 26, p. 2529–2541, 2021.
- KAPLAN, G. A.; KEIL, J. E. **Socioeconomic factors and cardiovascular disease: a review of the literature**. *Circulation*, v. 88, n. 4, p. 1973–1998, 1993. Disponível em: <https://www.ahajournals.org/doi/abs/10.1161/01.CIR.88.4.1973>.
- KIVIMÄKI, M.; KAWACHI, I. **Work stress as a risk factor for cardiovascular disease**. *Current cardiology reports*, Springer, v. 17, p. 1–9, 2015.
- LEE, E. S.; FORTHOFFER, R. N. **Analyzing Complex Survey Data**. 2nd. ed. EUA: Sage Publications, 2006.
- LEITE, I. d. C. et al. **Carga de doença no Brasil e suas regiões, 2008**. *Cadernos de Saúde Pública*, SciELO Brasil, v. 31, p. 1551–1564, 2015.
- LLOYD-JONES, D. M. et al. **Defining and Setting National Goals for Cardiovascular Health Promotion and Disease Reduction**. *Circulation*, v. 121, n. 4, p. 586–613, 2010. Disponível em: <https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.109.192703>.
- MARTINEZ, E. Z.; LOUZADA-NETO, F.; PEREIRA, B. d. B. **A curva ROC para testes diagnósticos**. *Cad. saúde colet.*, (Rio J.), p. 7–31, 2003.
- MATOZINHOS, F. P. et al. **Cardiovascular health in Brazilian state capitals 1**. *Revista Latino-Americana de Enfermagem*, SciELO Brasil, v. 25, 2017.
- MOREFIELD, B.; RIBAR, D. C.; RUHM, C. J. **Occupational status and health transitions**. *The BE Journal of Economic Analysis & Policy*, De Gruyter, v. 11, n. 3, 2012.
- MOREIRA, A. D. et al. **Saúde cardiovascular e validação do escore autorreferido no Brasil: uma análise da Pesquisa Nacional de Saúde**. *UCiência & Saúde Coletiva*, SciELO Public Health, v. 25, p. 4259–4268, 2020.
- MORRIS, W. R. et al. **Do Blue-Collar Workers Perceive the Worksite Health Climate Differently Than White-Collar Workers?**. *American Journal of Health Promotion*, v. 13, n. 6, p. 319–324, 1999. PMID: 10557504. Disponível em: <https://doi.org/10.4278/0890-1171-13.6.319>.
- MOTTA, A. C. S. V. et al. **Prevalência de saúde cardiovascular ideal na população adulta brasileira-Pesquisa Nacional de Saúde 2019**. *Epidemiologia e Serviços de Saúde*, SciELO Brasil, v. 32, p. e2022669, 2023.
- NETO, Z. G. d. S. **Curva ROC para comparação de modelos de predição para variáveis dicotômicas**. Universidade Federal de Viçosa, 2020.
- OMS. **World Health Organization package of essential noncommunicable (PEN) disease interventions for primary health care**. Organização Mundial da Saúde (OMS), 2020. Disponível em: <https://www.who.int/publications/i/item/9789240009226>. Acesso em: 23 nov. 2022.

- OMS. **Cardiovascular diseases (CVDs)**. Organização Mundial da Saúde (OMS), 2021. Disponível em: ([https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))). Acesso em: 23 nov. 2022.
- 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.
- SCHREUDER, K. et al. **Job demands and health complaints in white and blue collar workers**. *Work*, IOS Press, v. 31, n. 4, p. 425–432, 2008.
- SHINYA, K. **Which is worse for your long-term health, a white-collar or a blue-collar job?** *Journal of the Japanese and International Economies*, v. 38, p. 228–243, 2015.
- SHTATLAND, E. S.; KLEINMAN, K.; CAIN, E. M. **One more time about R² measures of fit in logistic regression**. *NESUG 15 proceedings*, v. 15, p. 222–226, 2002.
- STOPA, S. R. et al. **Pesquisa Nacional de Saúde 2019: histórico, métodos e perspectivas**. *Epidemiologia e Serviços de Saúde*, SciELO Brasil, v. 29, 2020.
- TJUR, T. **Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination**. *The American Statistician*, Taylor Francis, v. 63, n. 4, p. 366–372, 2009.
- VIRTANEN, M.; KIVIMÄKI, M. **Long working hours and risk of cardiovascular disease**. *Current cardiology reports*, Springer, v. 20, p. 1–7, 2018.

Apêndice

Código em linguagem R

```
# LEITURA DOS DADOS
# PACOTES
# install.packages("pacman")
library(pacman)
p_load(survey, ggplot2, dplyr, foreign, tidyverse, data.table, purrr,
       readxl, psych, reshape2, stringr, openxlsx, jtools, WeightedROC,
       ggpubr, patchwork, magrittr)

# pns.rdata baixado do site da PNS
# https://www.pns.icict.fiocruz.br/bases-de-dados/
load("../pns.rdata") # incluir o diretório

# CRITERIOS DE INCLUSÃO E EXCLUSÃO
#-----
## V0025A: SELEÇÃO DO MORADOR COM 15 ANOS OU MAIS
pns <- pns2019[pns2019$V0025A==1,]

## C008: IDADE DO MORADOR
pns <- pns[pns$C008 >= 18,]

## M001: ENTREVISTA DO MORADOR
pns <- pns[pns$M001 == 1,]

## M00203: APTIDÃO DO MORADOR SELECIONADO A RESPONDER
pns <- pns[pns$M00203 == 1,]

## P005: A SR. ESTA GRÁVIDA NO MOMENTO?
class(pns$P005)
table(pns$P005)
pns <- pns %>% mutate(P005 = replace_na(as.character(P005), "-"))
pns <- pns[pns$P005 %in% c("-", "2"),]
```

```
## DIABETES, COLESTEROL E HIPERTENSÃO
# -----
# Q03001: DIAGNÓSTICO DE DIABETES?
# Q060: DIAGNÓSTICO DE COLESTEROL ALTO?
# Q00201: DIAGNÓSTICO DE HIPERTENSÃO?
# -----
pns <- pns %>%
  mutate(Q03001 = replace_na(as.character(Q03001), "-"),
         Q060 = replace_na(as.character(Q060), "-"),
         Q00201 = replace_na(as.character(Q00201), "-")) %>%
  filter(Q03001 != "-" & Q060 != "-" & Q00201 != "-") #77495

# VARIÁVEIS
## SEXO
pns <- pns %>% mutate(sexo = case_when(
  (C006 == '1') ~ '0 Masculino',
  (C006 == '2') ~ '1 Feminino'))

## FAIXA ETÁRIA
pns <- pns %>% mutate(idade = case_when(
  (C008>= 18 & C008 <= 24) ~ '0 18 a 24 anos',
  (C008>= 25 & C008 <= 39) ~ '1 25 a 39 anos',
  (C008>= 40 & C008 <= 59) ~ '2 40 a 59 anos',
  (C008>= 60) ~ '3 60 anos ou mais'))

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

pns <- pns %>% mutate(raca = case_when(
```

```
raca == '1' ~ '2 Branca',
raca == '2' ~ '1 Preta',
raca == '3' ~ '3 Amarela/Indígena/Ignorado',
raca == '0' ~ '0 Parda'))

## ESTADO CIVIL
# agrupando Separado ou Desquitado(a) judicialmente +
# Divorciado(a) + Viúvo(a)
pns <- pns %>% mutate(estado_civil = case_when(
  C011 == '1' ~ 1,
  C011 == '2' ~ 2,
  C011 == '3' ~ 2,
  C011 == '4' ~ 0))

pns <- pns %>% mutate(estado_civil = case_when(
  (estado_civil == '0') ~ '0 Solteiro',
  (estado_civil == '1') ~ '1 Casado',
  (estado_civil == '2') ~ '2 Separado/Divorciado/Viúvo'))

## REGIÕES - agrupadas por Estado(V0001)
pns <- pns %>% 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))

pns <- pns %>% 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' ))

## SITUAÇÃO DE DOMICÍLIO
table(pns$V0026,useNA = "always")
```

```
pns <- pns %>% mutate(Sit_domicilio = case_when(
  V0026 == 1 ~ "0 Urbano",
  V0026 == 2 ~ "1 Rural"))

## DOENÇAS CRÔNICAS (DC)
#-----
# Q074: ASMA OU BRONQUITE ASMÁTICA
# Q079: ARTRITE OU REUMATISMO
# Q084: PROBLEMA CRÔNICO DE COLUNA
# Q088: DORT
# Q116/ Q11604: DOENÇA NO PULMÃO
# Q120: CÂNCER
# Q124: INSUFICIÊNCIA RENAL CRÔNICA
# Q128: OUTRA DC OU DOENÇA COM MAIS DE 6 MESES DE DURAÇÃO
# Q092: DEPRESSÃO
# Q110/ Q11006: OUTRA DOENÇA MENTAL
#-----

pns <- pns %>% 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))

pns <- pns %>% mutate(doenca_cronica = case_when(
  (doenca_cronica == '0') ~ "1 Não possui",
  (doenca_cronica == '1') ~ "0 Pelo menos uma"))

## DOENÇA CARDIOVASCULAR (DCV)
#-----
# Q06306: DIAGNÓSTICO DE DCV(INFARTO, ANGINA, INS. CARDIACA OU OUTRA)
# Q068: DIAGNÓSTICO DE AVC OU DERRAME?
#-----

pns <- pns %>% mutate(dcv = case_when(
```

```
(Q06306 == '1' | Q068 == '1' ) ~ 1,
(Q06306 == '2' & Q068 == '2') ~ 0))
```

```
# ESCORE DE SAUDE CARDIOVASCULAR (VARIÁVEL RESPOSTA)
```

```
#-----
```

```
## FATORES BIOLÓGICOS
```

```
# Q00201: DIAGNOSTICO DE HIPERTENSAO ARTERIAL?
```

```
pns <- pns %>% mutate(fav_pa = case_when((Q00201 == 2) ~ 1, TRUE ~ 0))
```

```
# Q03001: DIAGNÓSTICO DE DIABETES?
```

```
pns <- pns %>% mutate(fav_glu = case_when( Q03001 == 2 ~ 1, TRUE ~ 0))
```

```
# Q060: DIAGNÓSTICO DE COLESTEROL ALTO?
```

```
pns <- pns %>% mutate(fav_col = case_when( Q060 == 1 ~ 0,
  Q060 == 2 ~ 1))
```

```
# SOMA DOS FATORES DE SAÚDE
```

```
pns <- pns %>% mutate(fatores_sau = fav_pa + fav_glu + fav_col)
```

```
## FATORES COMPORTAMENTAIS
```

```
# IMC (<25kg/m2)
```

```
#-----
```

```
# P00103: Peso em kg autorreferido
```

```
# P00403: Altura em cm autorreferido
```

```
# P00404: Altura + imputação
```

```
# P00104: Peso + imputação
```

```
#-----
```

```
imc <- pns$P00104/(pns$P00404^2)
```

```
pns <- pns %>% mutate(fav_imc = case_when(
  imc < 0.0025 ~ 1, imc >= 0.0025 ~ 0 ))
```

```
# EXERCÍCIO FÍSICO (maior ou igual a 150 min/week)
```



```
#-----  
# P03701: Duração da atividade física (parte inteira horas)  
# P03702: Duração da atividade física (parte minutos)  
# P035: Dias por semana da prática de esporte  
#-----  
  
pns <- pns %>% mutate(fav_exerc = case_when(  
  (P03701*60 + P03702)*P035 >=150 ~ 1, TRUE ~ 0 ))  
  
# DIETA  
#-----  
# P015: Dias na semana que costuma consumir peixe  
# P02001+P02002: Quantos dias na semana toma refrigerante/  
# suco artificial  
# P00901: Em quantos dias da semana, come pelo menos um tipo  
# de verdura/legume  
# P018: Quantas vezes na semana consome frutas?  
# P02601: O que acha do consumo do sal?  
#-----  
  
pns <- pns %>% mutate(  
  fav_peixe = case_when(P015 < 2 ~ 0, P015 >= 2 ~ 1),  
  fav_acucar = case_when(P02001 + P02002 < 5 ~ 1,  
    P02001 + P02002 >= 5 ~ 0, TRUE ~ 1),  
  fav_veg = case_when(P00901 == 7 & P018 == 7 ~ 1, TRUE ~ 0),  
  fav_sal = case_when(P02601 <= 2 ~ 0, P02601 >= 3 ~ 1))  
  
# ESCORE DA DIETA  
pns <- pns%>% mutate(  
  dieta_score = fav_peixe + fav_acucar + fav_veg + fav_sal)  
  
# DIETA FAVORÁVEL  
pns <- pns %>% mutate(fav_dieta= case_when(  
  dieta_score == 4 ~ 1, TRUE ~ 0))
```

```

# TABAGISMO
# P050: Atualmente fuma?
# 1 = Sim, diariamente, 2 = Sim, menos que diariamente,
# 3 = Nao fumo atualmente
pns <- pns %>% mutate(fumante = case_when(P050 == 3 ~ 1,
  P050 != 3 ~ 0))

# P052: No passado, ja fumou?
# 1 = Sim, diariamente, 2 = Sim, menos que diariamente,
# 3 = Nao, nunca fumei
pns <- pns %>% mutate(testes = case_when(P052 > 1 ~ 1,
  P052 <= 1 ~ 0))

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

#
pns <- pns %>% mutate(fav_fumo = case_when(case_when(
  P050 == 3 ~ 1,
  P050 != 3 ~ 0) == 1 &
  case_when(
    P052 == 1 & P05901 > 1 ~ 1,
    P052 == 2 & P05901 > 1 ~ 1,
    P052 == 3 ~ 1,
    P052 == 1 & P05901 <= 1 ~ 0,
    P052 == 2 & P05901 <= 1 ~ 0) == 1 ~ 1,
  TRUE ~ 0))

# SOMA DE COMPORTAMENTOS SAUDÁVEIS
pns <- pns%>% mutate(comp_score = fav_imc+

```

```
fav_exerc + fav_fumo + fav_dieta)

# SCV BINÁRIO
pns <- pns %>% mutate(cvh_score = fatores_sau + comp_score)
pns <- pns %>% mutate(cvh_score_bin = case_when(
  (cvh_score == 6 | cvh_score == 7) ~ "1 Superior",
  TRUE ~ "0 Inferior"))

## FATORES SOCIOECONÔMICOS
#-----
## ESCOLARIDADE
pns$escolaridade <- 1 # fundamental
pns$escolaridade[pns$VDD004A %in% c(3,4)] <- 2 #medio
pns$escolaridade[pns$VDD004A %in% c(5,6)] <- 3 # superior incompleto
pns$escolaridade[pns$VDD004A >= 7] <- 4 # superior completo

pns <- pns %>% mutate(escolaridade_nome = case_when(
  escolaridade == 1 ~ "0 Ensino Fundamental incompleto ou equivalente",
  escolaridade == 2 ~ "1 Ensino Médio incompleto ou equivalente",
  escolaridade == 3 ~ "2 Ensino Superior Incompleto ou equivalente",
  escolaridade == 4 ~ "3 Ensino Superior Completo"))

## OCUPAÇÃO
# FORA DA FORÇA
# F001011: recebia pensão/aposentadoria
# E024011: Motivo de não procurar emprego
# E01201: COD
table(pns$F001011,pns$E024011)

pns = pns %>% mutate(foradafc = case_when(
  is.na(E01201) & F001011 == 1 & !(E024011 %in% c(7,8)) ~ "aposentado",
  is.na(E01201) & E024011 == 7 & F001011 == 1 ~ "cuidador/aposentado",
  is.na(E01201) & E024011 == 8 & F001011 == 1 ~ "estudante/aposentado",
  is.na(E01201) & E024011 == 7 & F001011 == 2 ~ "cuidador",
```

```
is.na(E01201) & E024011 == 8 & F001011 == 2 ~ "estudante",
!is.na(E01201) ~ "ocupado",
TRUE ~ "não ocupado"))
```

COLARINHOS

```
ajustar = c("110","2","411","412","511","512","999","299","599")
pns$cod <- as.character(pns$E01201)
pns = pns %>% mutate(
  CODIGO = case_when(cod == "0" ~ "xxxx",!(cod %in% c("0")) ~ cod),
  CODIGO = ifelse(CODIGO %in% ajustar, paste0(0,CODIGO),CODIGO),
  CODIGO1 = substr(CODIGO, 1, 1),
  colarinho = case_when(
    CODIGO1 %in% as.character(seq(0,4)) ~ "1 colarinho branco",
    CODIGO1 %in% as.character(seq(5,9)) ~ "0 colarinho azul",
    TRUE ~ "2 nao ocupado/mal definidos"))
```

COM A COD E SEM A COD

```
pns <- pns %>% mutate(conf_cod = case_when(
  !is.na(E01201) ~ "1Tem código", TRUE ~ "0Não tem código"))
```

NOMEANDO A COD

```
table(pns$CODIGO1,useNA = "always")
pns <- pns %>% mutate(nome_cod = case_when(
  CODIGO1 == 0 ~ "1 Membros das forças armadas, policiais
    e bombeiros militares",
  CODIGO1 == 1 ~ "2 Diretores e gerentes",
  CODIGO1 == 2 ~ "3 Profissionais das ciências e intelectuais",
  CODIGO1 == 3 ~ "4 Técnicos e profissionais de nível médio",
  CODIGO1 == 4 ~ "5 Trabalhadores de apoio administrativo",
  CODIGO1 == 5 ~ "6 Trabalhadores dos serviços, vendedores dos
    comércios e mercados",
  CODIGO1 == 6 ~ "7 Trabalhadores qualificados da agropecuária,
    florestais, da caça e da pesca",
  CODIGO1 == 7 ~ "8 Trabalhadores qualificados, operários e
```

```
      artesãos da construção, das artes mecânicas e outros ofícios",
CODIGO1 == 8 ~ "9 Operadores de instalações e máquinas e montadores",
CODIGO1 == 9 ~ "91 Ocupações elementares",
CODIGO1 == "x" ~ "92 Ocupações mal definidas",
TRUE ~ "0 Não ocupado"))

## RENDA PER CAPITA
pns <- pns %>% mutate(renda_per_capita = case_when(
  VDF004 == 1 ~ "1 Até 1/4 salário mínimo",
  VDF004 == 2 ~ "2 Mais de 1/4 até 1/2 salário mínimo",
  VDF004 == 3 ~ "3 Mais de 1/2 até 1 salário mínimo",
  VDF004 == 4 ~ "4 Mais de 1 até 2 salários mínimos",
  VDF004 == 5 ~ "5 Mais de 2 até 3 salários mínimos",
  VDF004 == 6 ~ "6 Mais de 3 até 5 salários mínimos",
  VDF004 == 7 ~ "7 Mais de 5 salários mínimos",
  TRUE ~ "8 Não definido"))

# MODELOS
#-----
## RETIRAR GRUPO FORA DA FORÇA
pns = pns %>% filter(foradafc %in% c("ocupado","não ocupado"))
pns <- pns %>%mutate(across(
  c(sexo, idade, raca, regioes, Sit_domicilio, estado_civil,
  escolaridade_nome, cvh_score_bin, cvh_score,
  comp_score, fatores_sau, doenca_cronica,
  CODIGO1, colarinho, foradafc, conf_cod,
  renda_per_capita, nome_cod), factor))

## EXPANSÃO DA AMOSTRA
glimpse(pns)
desPNS = svydesign(id = ~UPA_PNS, strat = ~V0024,
  weight = ~V00291, nest = TRUE, data = pns)

## MODELO MINIMAMENTE AJUSTADO
mod1 = svyglm(cvh_score_bin ~ idade + sexo + raca + colarinho,
```

```

family = quasibinomial(), design = desPNS)

## MODELO MODERADAMENTE AJUSTADO
mod2 = svyglm(cvh_score_bin ~ idade + sexo + raca + colarinho +
  estado_civil + regioes + Sit_domicilio + escolaridade_nome,
  family = quasibinomial(), design = desPNS)

## MODELO COMPLETO
mod3 = svyglm(cvh_score_bin ~ idade + sexo + raca + colarinho +
  estado_civil + regioes + Sit_domicilio + escolaridade_nome +
  renda_per_capita + doenca_cronica, family = quasibinomial(),
  design = desPNS)

# TESTAR SIGNIFICÂNCIA
#-----
r = c(~idade, ~sexo, ~raca, ~colarinho, ~estado_civil, ~regioes,
  ~Sit_domicilio, ~escolaridade_nome, ~renda_per_capita,
  ~doenca_cronica)

## mod1
regTermTest(mod1, ~ idade + sexo + raca + colarinho)$p # pvalor
for (i in 1:4) {
  print(r[[i]])
  print(regTermTest(mod1, r[[i]])$p)}

## mod2
regTermTest(mod2, ~ idade + sexo + raca + colarinho +
  estado_civil + regioes + Sit_domicilio + escolaridade_nome)$p

for (i in 1:8) {
  print(r[[i]])
  print(regTermTest(mod2, r[[i]])$p)}

## mod3
regTermTest(mod3, ~idade + sexo + raca + colarinho +

```

```
estado_civil + regioes + Sit_domicilio + escolaridade_nome +
renda_per_capita + doenca_cronica)$p

for (i in 1:10) {
  print(r[[i]])
  print(regTermTest(mod3,r[[i]])$p)}

# TESTES PARA AJUSTE
#-----
## R^2 McFadden
data.frame(Modelo = seq(1,3),
RMcFadden = c(attr(summ(mod1), "rsqmc"),
  attr(summ(mod2), "rsqmc"),
  attr(summ(mod3), "rsqmc")))

## CURVA ROC
# FPRT = taxa poderada de falsos positivos
# TPRT = taxa ponderada de verdadeiros positivos
# modelo 01
y = mod1$y
w = weights(desPNS)
y.hat = fitted(mod1)
roc <- WeightedROC(y.hat, y, w)
auc = paste0("AUC: ", round(WeightedAUC(roc),2))
a = ggplot()+
  geom_path(aes(FPR, TPR), data=roc)+
  coord_equal() +
  geom_abline(intercept = 0, slope = 1, size = 0.5,linetype ="dashed",
  color = "red") +
  annotate("text", x=0.7, y=0.01, label= auc,size = 5,col = "blue",
  fontface = "bold")+
  xlab("1 - Especificidade") + ylab("Sensibilidade") + theme_bw()+
  labs(title = "Modelo 01") +
  theme(plot.title = element_text(hjust = 0.5,face = "bold",size = 15),
  axis.text.y = element_text(face = "bold",size = 12,
```

```
    colour = "#696969"),
  axis.text.x = element_text(face = "bold",size = 12,
    colour = "#696969"),
  axis.title.x = element_text(face = "bold",size = 14,vjust = -1),
  axis.title.y = element_text(face = "bold",size = 14,vjust = 1.5))

# modelo 02
y = mod2$y
w = weights(desPNS)
y.hat = fitted(mod2)
roc <- WeightedROC(y.hat, y, w)
auc = paste0("AUC: ", round(WeightedAUC(roc),2),"0")
b = ggplot() +
  geom_path(aes(FPR, TPR), data=roc) +
  coord_equal() +
  geom_abline(intercept = 0, slope = 1, size = 0.5,
    linetype ="dashed",color = "red") +
  annotate("text", x=0.7, y=0.01, label= auc,size = 5,
    col = "blue",fontface = "bold")+
  xlab("1 - Especificidade") + ylab("Sensibilidade") + theme_bw() +
  labs(title = "Modelo 02") +
  theme(plot.title = element_text(hjust = 0.5,face = "bold",size = 15),
    axis.text.y = element_text(face = "bold",size = 12,
      colour = "#696969"),
    axis.text.x = element_text(face = "bold",size = 12,
      colour = "#696969"),
    axis.title.x = element_text(face = "bold",size = 14,vjust = -1),
    axis.title.y = element_text(face = "bold",size = 14,vjust = 1.5))

# modelo 03
y = mod3$y
w = weights(desPNS)
y.hat = fitted(mod3)
roc <- WeightedROC(y.hat, y, w)
(auc = paste0("AUC: ", round(WeightedAUC(roc),2)))
c = ggplot() +
```



```

geom_path(aes(FPR, TPR), data=roc) +
coord_equal() +
geom_abline(intercept = 0, slope = 1, size = 0.5,
  linetype = "dashed", color = "red") +
annotate("text", x=0.7, y=0.01, label= auc, size = 5,
  col = "blue", fontface = "bold")+
labs(title = "Modelo 03") +
xlab("1 - Especificidade") + ylab("Sensibilidade") + theme_bw() +
theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 15),
  axis.text.y = element_text(face = "bold", size = 12,
    colour = "#696969"),
  axis.text.x = element_text(face = "bold", size = 12,
    colour = "#696969"),
  axis.title.x = element_text(face = "bold", size = 14, vjust = -1),
  axis.title.y = element_text(face = "bold", size = 14, vjust = 1.5))

# PLOTAR AS TRES CURVAS ROC + AUC EM UMA FIGURA
ggpubr::ggarrange(a,b,c, nrow = 1, ncol = 3)

# GERAR TABELAS
#-----
modelos = list(mod1,mod2,mod3)
tabela_modelos <- vector(mode = "list", length = length(modelos))
for (i in 1:length(modelos)) {
  mod = modelos[[i]]
  sum1 = summary(mod)
  conf_est = confint.default(mod)
  conf <- exp(conf_est)
  tab_modelo = data.frame(round(sum1$coefficients,3)) %>%
    rownames_to_column("variaveis") %>%
    mutate(Estimate = paste0(round(Estimate,2), " (",
      round(conf_est[,1],2), "; ",
      round(conf_est[,2],2), ")"),

  odds = round(exp(sum1$coefficients[,1]),2),

```

```

ic = paste0(" (",round(conf[,1],2),"; ",
            round(conf[,2],2),")"),

pvalor = case_when(sum1$coefficients[,4] < 0.001 ~ "< 0.001",
                  sum1$coefficients[,4] < 0.05 ~ paste0(as.character(
                    round(sum1$coefficients[,4],3)),"*"),
                  sum1$coefficients[,4] < 0.01 ~ paste0(as.character(
                    round(sum1$coefficients[,4],3)), "**"),

                  TRUE ~ as.character(round(sum1$coefficients[,4],3)))
tabela_modelos[[i]] <- tab_modelo %>% select(1,6,7,8)}

names(tabela_modelos) = c("mod1", "mod2", "mod3")
# SALVAR A TABELA NO COMPUTADOR
write.xlsx(tabela_modelos, file= ".../modelos.xlsx")

# TABELA COLARINHO POR COLUNA
#-----
## TABELA DO COLARINHO
resumo = c(~cvh_score_bin,~nome_cod, ~escolaridade_nome ,
           ~renda_per_capita , ~sexo , ~idade , ~raca, ~estado_civil,
           ~Sit_domicilio, ~regioes , ~doenca_cronica)

lista <- vector(mode = "list", length = length(resumo))

for (i in 1:length(resumo)) {
  tabela = data.frame(t(svyby(resumo[[i]], ~colarinho, desPNS,
                             svymean, vartype = "ci"))) %>%
  slice(-1) %>%
  stats::setNames(c("azul", "branco", "nao_ocupado")) %>%
  mutate(across(where(is.character),as.numeric))
  tabela = tabela*100
  tabela = tibble::rownames_to_column(tabela , "variaveis")
  lista[[i]] <- tabela}

```

```

names(lista) = as.character(resumo)
for (i in 1:3) {
  lista[[i]][,2:4] <- round(lista[[i]][,2:4],2)}
# SALVAR A TABELA NO COMPUTADOR
write.xlsx(tabela,".../tabela_colarinho.xlsx")

## TABELA TOTAL
resumo = c(~cvh_score_bin, ~idade ,~sexo ,~raca , ~nome_cod,
  ~colarinho, ~estado_civil ,~regioes ,~Sit_domicilio,
  ~escolaridade_nome ,~renda_per_capita ,~doenca_cronica)

tabela = data.frame(variavel = NA, freq = NA)
for (i in 1:length(resumo)){
  a = data.frame(round(svymean(resumo[[i]],design = desPNS)*100,3))
  b = data.frame(round(confint(svymean(resumo[[i]],desPNS))*100,3))
  tabela = tabela %>% add_row(variavel = rownames(a),
    freq = paste0(a[,1], " (" ,b[,1],"; " ,b[,2],")")) }

tabela = tabela %>% slice(-1)

# SALVAR A TABELA NO COMPUTADOR
write.xlsx(tabela,".../tabela_resumo.xlsx")

## TABELA POR SCV
#-----
lista <- vector(mode = "list", length = length(resumo))
for (i in 1:length(resumo)) {
  tabela = data.frame(t(svyby(resumo[[i]],~cvh_score_bin,
    desPNS,svymean, vartype = "ci"))) %>% slice(-1) %>%
  stats::setNames(c("inferior", "superior")) %>%
  mutate(across(where(is.character),as.numeric))
  tabela = tabela*100 # inferior = < 6, superior >= 6
  tabela = tibble::rownames_to_column(tabela , "variaveis")
  lista[[i]] <- tabela}

```

```
names(lista) = as.character(resumo)
# arredondar
for (i in 1:3) {
  lista[[i]][,2:3] <- round(lista[[i]][,2:3],2)}
# SALVAR A TABELA NO COMPUTADOR
write.xlsx(lista,".../tabela_scv.xlsx")
```