

**DETERMINANTES DO SOBREPESO E OBESIDADE ENTRE HOMENS E MULHERES NO CEARÁ**

**Natália Cecília de França**

Doutoranda em Economia, CAEN/UFC  
Contato: +1(305)4907631; +55(85)997664043  
[nc.franca@hotmail.com](mailto:nc.franca@hotmail.com)

**João Mário Santos de França**

IPECE e CAEN/UFC  
Contato: [joao.franca@ufc.br](mailto:joao.franca@ufc.br)  
[mario.santos@ipece.ce.gov.br](mailto:mario.santos@ipece.ce.gov.br)

**Diógenes Alberto de França**

Tecnólogo em jogos Digitais, PUC MINAS  
[diogenesaf@yahoo.com.br](mailto:diogenesaf@yahoo.com.br)

Av. da Universidade, 2700  
60020-181 Fortaleza – CE

**Área 2 – Economia Social**

## **DETERMINANTES DO SOBREPESO E OBESIDADE ENTRE HOMENS E MULHERES NO CEARÁ**

### **Resumo**

Sobrepeso e obesidade têm se tornado uma epidemia mundial ao longo dos últimos anos, configurando sério problema de saúde. Neste sentido, o presente estudo tem como objetivo contribuir com a literatura sobre tal agravo entre a população adulta cearense. Em primeiro lugar, fez-se uma avaliação dos determinantes do excesso de peso entre homens e mulheres cearenses por meio da estimação de um modelo Logit ordenado generalizado. Os resultados indicam, por exemplo, que idade, renda domiciliar per capita, consumo elevado de sal, consumo de frutas e mais horas assistindo televisão contribuíram para a prevalência do excesso de peso em ambos os sexos. Entre as mulheres, a escolaridade auxiliou na redução da obesidade, já entre os homens não se observou uma tendência clara. Na segunda etapa do trabalho, foram implementados diferentes algoritmos de mineração de dados na classificação antropométrica da população adulta residente no Ceará, cuja acurácia foi comparada com a do modelo Logit binário. A construção das árvores de decisão indicam que os principais determinantes da prevalência do excesso de peso entre homens e mulheres foram, respectivamente, a situação matrimonial e a condição de ocupação no mercado de trabalho. Ademais, os algoritmos de mineração de dados não apresentaram maior acurácia quando comparados com o modelo Logit binário.

**Palavras-chave:** Obesidade, Determinantes, Classificação, Mineração de Dados.

## **DETERMINANTS OF OVERWEIGHT AND OBESITY BETWEEN MEN AND WOMEN IN CEARÁ**

### **Abstract**

Overweight and obesity have become a worldwide epidemic over the past few years, configuring a serious health problem. In this sense, the present study aims to contribute with the literature on such an aggravation among the adult population of Ceará. Firstly, an evaluation of the determinants of overweight among men and women from Ceará was made through the estimation of a generalized ordered Logit model. The results indicate, for example, that age, per capita household income, high salt intake, fruit consumption and more hours watching television contributed to the prevalence of overweight in both sexes. Among women, schooling contributed to the reduction of obesity; among men, a clear trend was not observed. In the second stage of the work, different algorithms of data mining were implemented in the anthropometric classification of the adult population resident in Ceará, whose accuracy was compared with that of the Logit model. The construction of decision trees indicates that the main determinants of the prevalence of overweight among men and women were, respectively, the marital status and the occupation status in the labor market. In addition, the data mining algorithms were not more accurate when compared with the binary Logit model.

**Keywords:** Obesity, Determinants, Classification, Data Mining.

Classificação JEL: C81, I12

## 1. Introdução

Sobrepeso e obesidade têm se tornado uma epidemia mundial nos últimos anos, além de se configurar um sério problema de saúde pública. Segundo dados da Organização Mundial de Saúde (OMS), a incidência mundial de obesidade dobrou desde 1980<sup>1</sup>. Em 2014, 39% dos adultos com mais de 18 anos de idade estavam com sobrepeso e 13% eram obesos. O Brasil segue essa mesma tendência de aumento da obesidade. Resultados do relatório da Organização das Nações Unidas para Alimentação e Agricultura e da Organização Pan-americana de Saúde (2017) mostram que, em 2010, 17,8% da população brasileira era obesa; passando para 20% em 2014. Dentre os estados brasileiros, o Ceará, em 2013, foi o décimo em termos de prevalência de sobrepeso, conforme dados da Pesquisa Nacional de Saúde (PNS) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Naquele ano, 57,1% dos cearenses acima de 18 anos de idade estava com sobrepeso e 18,7% era o percentual de obesos.

Excesso de peso corporal se configura como um dos fatores de risco às doenças crônicas não transmissíveis (DCNT)<sup>2</sup>, responsáveis por 72,6% das mortes prematuras entre os brasileiros em 2013 (MS, 2015). A expansão da obesidade pode ter contribuído, por exemplo, para o aumento da incidência de diabetes e hipertensão no Brasil na última década. Dados do Inquérito Telefônico de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas (Vigitel) indicam que, entre 2006 e 2016, o diagnóstico médico de diabetes no Brasil foi de 5,5% para 8,9%; e o de hipertensão, de 22,5% para 25,7%.

Além de ser fator de risco às DCNT, sobrepeso e obesidade também geram perdas econômicas. Os custos mundiais com obesidade representavam 2,8% do Produto Interno Bruto (PIB) global em 2012, de acordo com um estudo realizado pelo McKinsey Global Institute (2014). O mesmo estudo aponta que a obesidade era o terceiro maior encargo social no Brasil, com gastos que correspondiam a 2,4% do PIB nacional, ficando atrás apenas dos custos com mortes violentas e alcoolismo.

O recente *boom* da obesidade pode ser explicado pelo aumento da inatividade física, fatores genéticos e hormonais, crescimento econômico, urbanização, mudança nos hábitos alimentares reduzindo o consumo de alimentos saudáveis e aumentando o de alimentos ultraprocessados, etc. Dados do Vigitel mostram, por exemplo, que o consumo regular (ao menos 5 dias da semana) de feijão, alimento típico da dieta do brasileiro, caiu entre 2012 e 2016, passando de 67,5% para 61,3%.

Uma das medidas adotadas pelo Ministério da Saúde (MS) no enfrentamento da obesidade foi o lançamento, em 2014, do novo Guia Alimentar para a População Brasileira, que trás recomendações para uma alimentação saudável. Nesta mesma linha, o Governo Federal também lançou a campanha “Brasil Saudável e Sustentável” para mostrar as vantagens de uma alimentação saudável. O controle da obesidade também está presente entre as metas do “Plano de Ações Estratégicas para o Enfrentamento das Doenças Crônicas Não Transmissíveis (DCNT) no Brasil 2011 – 2022”, elaborado pelo MS em parceria com diversas outras instituições.

Neste sentido, este estudo visa contribuir com a literatura sobre os determinantes do sobrepeso e da obesidade entre a população adulta residente no Ceará. Para tal, foram estimados modelos Logit ordenados desagregados por sexo. A prevalência do excesso de peso aumentou com a idade e a renda domiciliar per capita em ambos os sexos. Alimentação saudável foi fator protetor contra obesidade entre homens e mulheres. Ainda em relação ao comportamento dos indivíduos, o consumo de frutas, o

<sup>1</sup> <http://www.who.int/mediacentre/factsheets/fs311/en/>. Acesso em 27 de setembro de 2017.

<sup>2</sup> Exemplos de DCNT são doenças cardiovasculares, diabetes, câncer e doenças respiratórias crônicas.

consumo excessivo de sal e o alto número de horas assistindo televisão contribuíram para o excesso de peso. Entre as mulheres, a escolaridade apresentou relação negativa com a prevalência do excesso de peso, o que não se observou entre os homens.

Ademais, foram utilizados algoritmos de mineração de dados na classificação antropométrica de homens e mulheres residentes no Ceará. Um bom algoritmo de classificação se mostra relevante na formulação de políticas públicas de forma mais eficiente e eficaz. De acordo com as árvores de decisão, a situação matrimonial foi o principal determinante do excesso de peso entre os homens; e a condição de ocupação no mercado de trabalho foi o principal determinante entre as mulheres. Os algoritmos utilizados neste trabalho não apresentaram ganhos em termos de acurácia quando comparados ao modelo Logit.

O presente estudo, além desta introdução é composto por uma seção sobre os determinantes da obesidade e outra que trata dos algoritmos de mineração de dados utilizados na classificação da obesidade. Mais adiante são mostrados os aspectos metodológicos, os resultados e as considerações finais encerram o trabalho.

## **2. Determinantes socioeconômicos da obesidade**

O entendimento sobre os determinantes do sobrepeso e da obesidade é muito importante ao ser capaz de gerar informações úteis para os formuladores de políticas públicas. Inatividade física, mudanças nos padrões de consumo, escolaridade, genética, dentre outros são fatores capazes de afetar a classificação antropométrica das pessoas.

Roemling e Qaim (2012) fizeram uma análise sobre a evolução da transição nutricional vivenciada na Indonésia entre 1993 e 2007. Os autores constataram que, seguindo a tendência mundial, os níveis de obesidade aumentaram no País, principalmente entre as mulheres. Além disso, seus resultados indicam que mudanças nos padrões alimentares e reduções nos níveis de atividade física foram fatores importantes para o aumento no índice de massa corporal (IMC) da população.

Ao fazer uma análise multinível dos determinantes da obesidade entre a população adulta chinesa em 2006, Zhang e Liao (2013) ressaltam que a renda tem impacto positivo na incidência de obesidade. Por outro lado, os autores destacam que escolaridade e ocupação manual se relacionam de maneira negativa com obesidade. Eles definiram como ocupação manual os cargos de fazendeiros, pescadores, caçadores, soldados e trabalhadores não qualificados. Em geral, pessoas alocadas em tais funções exercem algum grau de atividade física durante o trabalho, contribuindo, assim, para a menor chance de serem obesas em relação aos demais trabalhadores.

Em âmbito nacional, Andrade et al. (2015) constataram que a obesidade entre homens e mulheres residentes na área urbana de Belo Horizonte (BH) é influenciada por fatores distintos. Por exemplo, maior nível de escolaridade reduziu a prevalência de sobrepeso entre as mulheres e foi fator de risco entre os homens. Além disso, o consumo de refrigerantes dietéticos exerceu impacto positivo na prevalência de excesso de peso para ambos os sexos. Em outra análise sobre a população adulta de BH, Ward et al. (2015) também destacam a diferenciação dos determinantes da obesidade entre homens e mulheres. Segundo os autores, escolaridade se relacionou de maneira negativa com o IMC e a obesidade entre as mulheres, enquanto que as rendas do domicílio e da vizinhança tiveram impacto positivo entre os homens.

Ainda sobre a distinção dos fatores determinantes da obesidade entre os sexos, Siqueira et al. (2015) fizeram estudo sobre a obesidade abdominal na população com mais de 40 anos de idade residente no município de Cambé, Paraná. Os resultados indicam que a prevalência da obesidade se revelou bem mais acentuada entre as mulheres do que entre os homens. Fatores como idade, inatividade física no lazer e

hipertensão arterial tiveram impacto positivo na obesidade em ambos os sexos. Por sua vez, classe econômica, tabagismo e diabetes se mostraram significativos apenas entre as mulheres.

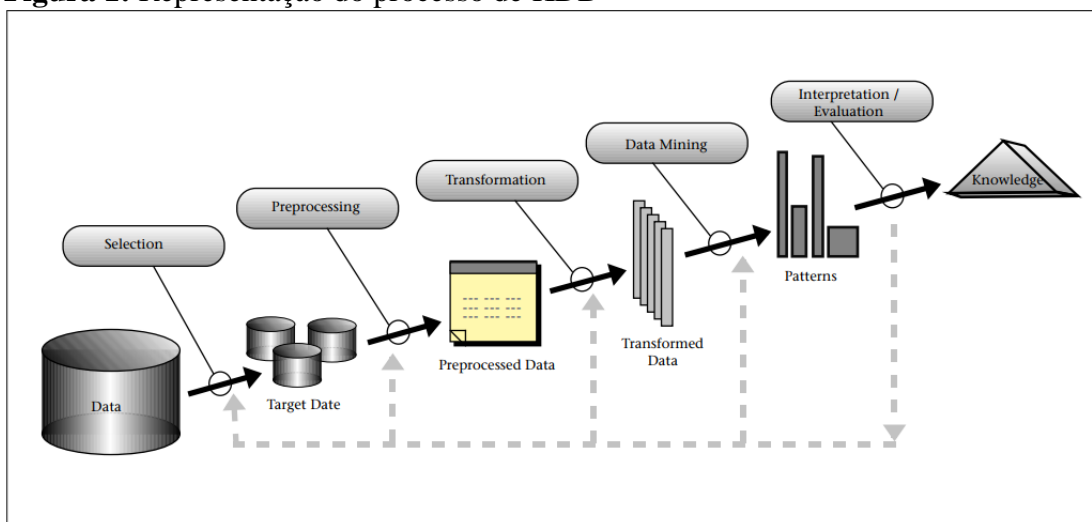
Correia et al. (2011) analisaram os determinantes da obesidade entre as mulheres em idade reprodutiva residentes no Ceará. Por meio da utilização de análise multivariada, os autores salientam que fatores como ter idade acima de trinta anos, ser casada ou utilizar contraceptivos tiveram impacto positivo na obesidade. Ainda em relação às mulheres em idade reprodutiva, Costa et al. (2014) fizeram uma análise da evolução, entre 1997 e 2006, do excesso de peso e seus determinantes entre aquelas residentes em Pernambuco. Mais uma vez, a baixa escolaridade foi apontada como fator de risco para a prevalência de tal agravo. Idade e maior número de gestações também contribuíram positivamente para a obesidade.

### 3. Mineração de dados na classificação da obesidade

Atualmente, existe um enorme volume de informações disponíveis no dia-a-dia das pessoas. Torna-se, pois, necessária a obtenção de conhecimento útil em meio a tantas informações disponíveis. Fayyad, Piatetsky-Shapiro e Smyth (1996) definem o processo de Descoberta de Conhecimento nas Bases de Dados (do inglês *Knowledge Discovery in Databases – KDD*) como "um processo não trivial de identificação de novos padrões válidos, úteis e compreensíveis" nos dados.

A Figura 1 representa as etapas do processo de KDD segundo Fayyad, Piatetsky-Shapiro e Smyth (1996). Em primeiro lugar, selecionam-se os dados de interesse em meio às informações disponíveis. Tais dados passam por um tratamento que consiste nas etapas de pré-processamento e transformação. Posteriormente, são aplicadas ferramentas de mineração de dados para a extração de padrões e geração de conhecimento relevante.

**Figura 1:** Representação do processo de KDD



Fonte: Fayyad, Piatetsky-Shapiro e Smyth (1996).

A mineração de dados (*data mining*), uma das etapas do processo de KDD, é o processo de descobrir padrões em bases de dados (WITTEN & FRANK, 2005). Esses padrões devem constituir conhecimento útil capaz de gerar vantagens para quem os utiliza. Além disso, conforme Hand, Mannila e Smyth (2001), os padrões obtidos devem ser, preferencialmente, de fácil entendimento. Algoritmos como árvores de decisão, classificadores bayesianos, redes neurais artificiais e máquinas de vetores de suporte (SVM) têm sido utilizados na classificação da obesidade, principalmente obesidade infantil.

Zhang et al. (2009) fizeram uma comparação entre o modelo Logit e seis algoritmos de mineração de dados na classificação da obesidade infantil na Inglaterra entre 1988 e 2003. Os resultados indicam que, em geral, máquinas de vetores de suporte e classificadores bayesianos obtiveram um melhor desempenho na classificação do sobrepeso e da obesidade. Além disso, tais ferramentas mostraram uma maior acurácia em comparação ao modelo Logit.

Pochini, Wu e Hu (2014), ao construírem uma árvore de decisão e estimarem um modelo Logit, constataram que fazer atividade física e tomar café da manhã todos os dias reduziu a prevalência de obesidade entre adolescentes residentes nos Estados Unidos em 2011. Por sua vez, o consumo de cigarro e de bebidas doces elevou o risco de ser obeso. Ainda sobre os principais fatores de risco da obesidade, Qendraj e Xhafaj (2015) construíram uma árvore de decisão a partir dos dados de Tirana Intermedical Centre para a população adulta. Os autores constataram que ser fumante foi o fator que mais contribuiu para a obesidade.

#### 4. Aspectos metodológicos

##### 4.1 Base de dados e descrição das variáveis

As informações utilizadas neste estudo foram obtidas na Pesquisa Nacional de Saúde - PNS 2013, realizada pelo IBGE em convênio com o Ministério da Saúde. A Pesquisa abrange todo o território nacional e tem como objetivo gerar informações relevantes sobre as condições de saúde da população brasileira. Pelo fato de a PNS ser uma pesquisa amostral, considerou-se os devidos pesos para expansão da amostra nas análises.

As aferições de peso, altura, circunferência da cintura e pressão arterial, bem como a realização de exames laboratoriais foram feitas em adultos com idade superior a 18 anos selecionados aleatoriamente em cada domicílio com entrevista realizada<sup>3</sup>. Diante da natureza do presente estudo, apenas esses indivíduos foram selecionados para as análises, totalizando 6.223.101 pessoas residentes no Ceará.

O índice de massa corporal (IMC), uma medida bastante utilizada na classificação antropométrica, é obtido da seguinte forma:

$$IMC = \frac{\text{peso em quilogramas}}{(\text{altura em metros})^2} \quad (1)$$

A Tabela 1 mostra a classificação antropométrica segundo o IMC proposta pela OMS para a população adulta<sup>4</sup>.

**Tabela 1:** Classificação antropométrica segundo IMC para população adulta

Classificação	IMC (kg/m <sup>2</sup> )
Baixo peso	< 18,5
Eutrofia	18,5 – 24,9
Sobrepeso	25 – 29,9
Obesidade	≥ 30

Seguindo essa classificação, 2,67% das pessoas consideradas neste estudo estavam abaixo do peso e foram, portanto, excluídas das análises. Foram feitos tratamentos adicionais na base de dados: (i) excluiu-se pessoas com mais de 60 anos de idade; (ii) desconsiderou-se pessoas cuja condição no domicílio fosse pensionista, empregado doméstico ou parente de empregado doméstico; (iii) excluiu-se as observações que tinham alguma informação faltante nas variáveis relevantes. O número final de observações e sua distribuição por gênero é apresentado na Tabela 2.

<sup>3</sup> Foi selecionado um adulto com mais de 18 anos de idade em cada domicílio.

<sup>4</sup> <http://www.who.int/topics/obesity/en/>. Acesso em 27 de setembro de 2017.

**Tabela 2:** Total de observações e distribuição por gênero, Ceará, 2013

Gênero	Observações	Percentual
Masculino	1.971.217	58,70
Feminino	1.386.796	41,30
Total	3.358.013	100

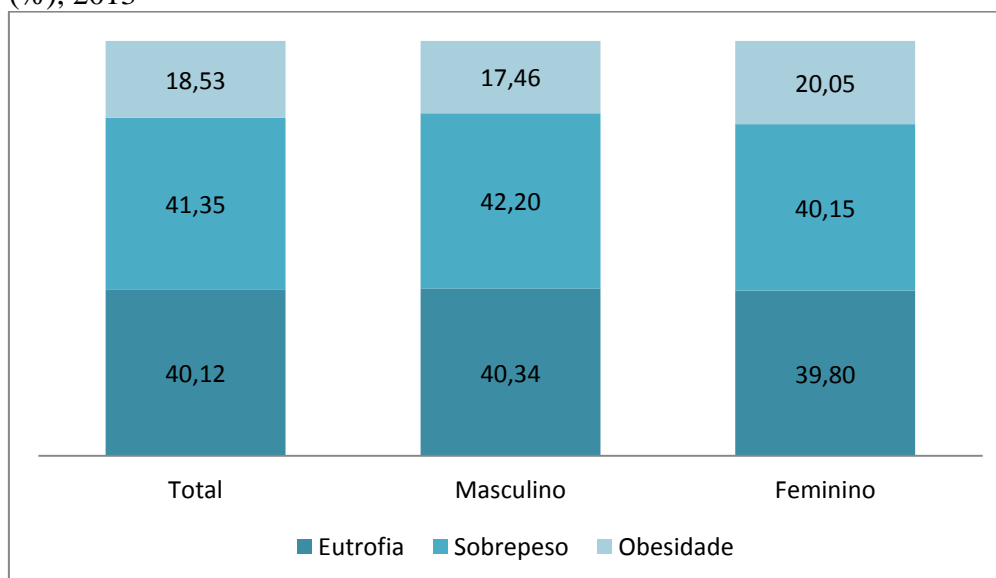
Fonte: Elaboração própria com dados da PNS (2013).

Uma vez excluídos os indivíduos com baixo peso, construiu-se uma variável categórica da seguinte forma:

$$y = \begin{cases} 0, & \text{eutrofia;} \\ 1, & \text{sobrepeso;} \\ 2, & \text{obesidade.} \end{cases} \quad (2)$$

Em que as categorias de  $y$  são ordenadas conforme o IMC dos indivíduos.

Mais da metade da população adulta cearense estava acima do peso, sendo 41,35% e 18,53% o percentual de pessoas com sobrepeso e obesidade, respectivamente (Figura 2). A Figura 2 também mostra que a incidência de obesidade, mensurada pelo IMC, se mostrou ligeiramente superior entre as mulheres do que entre os homens.

**Figura 2:** População adulta cearense por classificação antropométrica segundo IMC (%), 2013

Fonte: Elaboração própria com dados da PNS (2013).

O IMC, apesar de ser um bom indicador, não é totalmente correlacionado com a gordura corporal (ABESO, 2009). Conforme destacado em ABESO (2009), suas principais limitações incluem o fato de não distinguir massa gordurosa de massa magra; e não refletir, necessariamente, a distribuição da gordura corporal. A medida da circunferência abdominal, por sua vez, tem um alto grau de associação com a gordura corporal total. A OMS sugere pontos de corte para a classificação antropométrica segundo a circunferência abdominal, conforme mostra a Tabela 3.

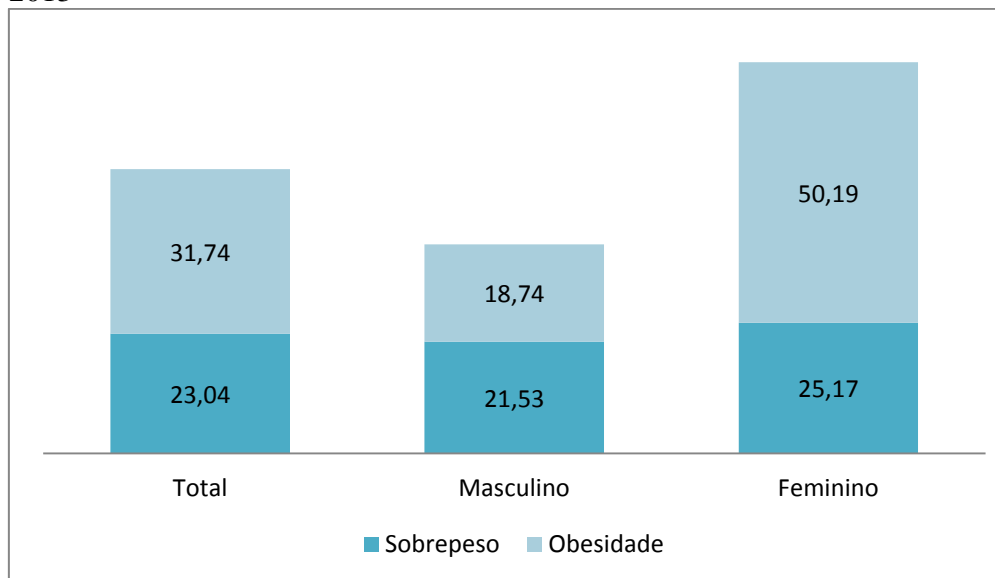
**Tabela 3:** Classificação antropométrica segundo a circunferência abdominal

Classificação	Circunferência abdominal (cm)	
	Homem	Mulher
Sobrepeso	$\geq 94$	$\geq 80$
Obesidade	$\geq 102$	$\geq 88$

Assim como em Siqueira et al. (2015), os resultados da Figura 3 indicam que a obesidade abdominal se mostrou bem mais alarmante entre as mulheres (50,19%),

sendo mais do que o dobro do percentual observado entre os homens (18,74%). Isto é um indicativo de que as mulheres estão muito mais sujeitas a complicações decorrentes da obesidade do que os homens, demandando, assim, uma atenção especial na formulação de políticas na área da saúde.

**Figura 3:** População adulta cearense segundo sobrepeso e obesidade abdominal (%), 2013



Fonte: Elaboração própria com dados da PNS (2013).

Com o intuito de verificar a relação entre as classificações antropométricas baseadas no IMC e na circunferência abdominal, calculou-se o coeficiente V de Cramer, que mede a relação entre variáveis categóricas. Foi encontrado um valor de 0,4577, indicando que ambas as classificações apresentam alto grau de relação (WITTE & WITTE, 2009). Assim sendo, as análises realizadas neste estudo serão sobre a classificação segundo o IMC em detrimento da circunferência abdominal ou uma combinação de ambas.

O Quadro 1 define as variáveis independentes utilizadas neste estudo, que contêm informações referentes ao indivíduo, ao domicílio de residência e à área de moradia. Todas as variáveis foram categorizadas de forma a assumirem os valores zero e um.

**Quadro 1:** Definição das variáveis independentes

Variável	Definição
18 a 30 anos	1, se tem de 18 a 30 anos de idade; 0, caso contrário
31 a 40 anos	1, se tem de 31 a 40 anos de idade; 0, caso contrário
41 a 50 anos	1, se tem de 41 a 50 anos de idade; 0, caso contrário
51 a 60 anos	1, se tem de 51 a 60 anos de idade; 0, caso contrário
Branco	1, se branco(a); 0, caso contrário
Casado	1, se casado(a); 0, caso contrário
Direção*	1, se ocupação é classificada como “diretores e gerentes”, “profissionais das ciências e intelectuais”, “técnicos e profissionais de nível médio” ou “trabalhadores de apoio administrativo”; 0, caso contrário
Serviço*	1, se ocupação é classificada como “trabalhadores dos serviços, vendedores dos comércios e mercados” ou “ocupações elementares”; 0, caso contrário
Manual*	1, se ocupação é classificada como “trabalhadores qualificados, operários e artesãos da construção, das artes mecânicas e outros ofícios”, “operadores de instalações e máquinas e montadores” ou “membros das forças armadas, policiais e bombeiros militares”; 0, caso contrário
Ocupado	1, se ocupado(a); 0, caso contrário

continua...



## Continuação Quadro 1

Variável	Definição
Fundamental	1, se tem ensino fundamental completo ou médio incompleto; 0, caso contrário
Médio	1, se tem ensino médio completo ou superior incompleto; 0, caso contrário
Superior	1, se tem ensino superior completo; 0, caso contrário
Plano	1, se tem plano de saúde; 0, caso contrário
Rdpc1	1, se a renda domiciliar per capita está no primeiro quintil da distribuição; 0, caso contrário
Rdpc2	1, se a renda domiciliar per capita está no segundo quintil da distribuição; 0, caso contrário
Rdpc3	1, se a renda domiciliar per capita está no terceiro quintil da distribuição; 0, caso contrário
Rdpc4	1, se a renda domiciliar per capita está no quarto quintil da distribuição; 0, caso contrário
Rdpc5	1, se a renda domiciliar per capita está no quinto quintil da distribuição; 0, caso contrário
Feijão	1, se consome feijão ao menos 5 dias da semana; 0, caso contrário
Verdura e legume	1, se consome verduras e legumes ao menos 5 dias da semana; 0, caso contrário
Carne	1, se consome carne vermelha ao menos 4 dias da semana; 0, caso contrário
Frango	1, se consome frango ao menos 4 dias da semana; 0, caso contrário
Peixe	1, se consome peixe ao menos 4 dias da semana; 0, caso contrário
Fruta	1, se se consome fruta ao menos 5 dias da semana; 0, caso contrário
Refrigerante	1, se consome refrigerante ao menos 5 dias da semana; 0, caso contrário
Sal	1, se o se consumo de sal é alto ou muito alto; 0, caso contrário
Atividade	1, se praticou atividade física nos últimos 3 meses; 0, caso contrário
Tv	1, se assiste ao menos 4 horas de televisão por dia; 0, caso contrário
Bebida	1, se nunca bebeu; 0, caso contrário
Cigarro	1, se nunca fumou; 0, caso contrário
Urbano	1, se reside na zona urbana; 0, caso contrário
Fortaleza	1, se reside em Fortaleza; 0, caso contrário
Restormf	1, se reside na Região Metropolitana de Fortaleza exclusive Fortaleza; 0, caso contrário
Restouf	1, se reside nos demais municípios cearenses; 0, caso contrário

Fonte: Elaboração própria.

\*Variável criada com base na Classificação de Ocupações para Pesquisas Domiciliares - CBO Domiciliar.

Na Tabela 4 são apresentadas algumas estatísticas descritivas das variáveis independentes, a saber, média e desvio padrão. Os resultados são apresentados para a população adulta total e também desagregados por gênero.

Em geral, mais de 30% das pessoas consideradas têm entre 18 e 30 anos, aproximadamente 25% se autodeclarou branco(a); 37%, casado(a); e 95%, ocupado(a). O perfil de ocupação se difere entre homens e mulheres: quase metade das mulheres está alocada na classe denominada serviço (47,63%); enquanto entre os homens, os cargos classificados como serviço e manual contabilizam uma participação relativa de cerca de 40%, cada.

**Tabela 4:** Estatísticas descritivas das variáveis independentes

Variáveis	Total		Masculino		Feminino	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
18 a 30 anos	0,3567	0,4790	0,3760	0,4844	0,3292	0,4699
31 a 40 anos	0,2781	0,4481	0,2692	0,4435	0,2908	0,4541
41 a 50 anos	0,2255	0,4179	0,2281	0,4196	0,2218	0,4154
51 a 60 anos	0,1397	0,3467	0,1267	0,3326	0,1583	0,3650

continua...

Continuação Tabela 4

Variáveis	Total		Masculino		Feminino	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
Branco	0,2524	0,4344	0,2566	0,4367	0,2466	0,4310
Casado	0,3764	0,4845	0,3651	0,4814	0,3924	0,4883
Direção	0,2385	0,4262	0,2057	0,4042	0,2852	0,4515
Serviço	0,4185	0,4933	0,3778	0,4848	0,4763	0,4994
Manual	0,3013	0,4588	0,3733	0,4837	0,1988	0,3991
Ocupado	0,9583	0,1999	0,9569	0,2031	0,9603	0,1953
Sem instrução	0,3640	0,4811	0,4077	0,4914	0,3018	0,4590
Fundamental	0,1638	0,3701	0,1762	0,3810	0,1462	0,3534
Médio	0,3566	0,4790	0,3296	0,4701	0,3950	0,4889
Superior	0,1155	0,3197	0,0864	0,2810	0,1569	0,3637
Plano	0,1897	0,3920	0,1850	0,3883	0,1963	0,3972
Rdpc1	0,1851	0,3884	0,1994	0,3995	0,1648	0,3710
Rdpc2	0,1932	0,3948	0,1837	0,3873	0,2067	0,4050
Rdpc3	0,2051	0,4038	0,2013	0,4010	0,2106	0,4077
Rdpc4	0,1755	0,3804	0,1478	0,3549	0,2150	0,4108
Rdpc5	0,2410	0,4277	0,2678	0,4428	0,2029	0,4021
Feijão	0,7444	0,4362	0,8147	0,3885	0,6446	0,4786
Verdura e legume	0,1184	0,3231	0,0871	0,2819	0,1630	0,3694
Carne	0,3321	0,4710	0,3937	0,4886	0,2445	0,4298
Frango	0,3369	0,4726	0,3303	0,4703	0,3463	0,4758
Peixe	0,0666	0,2494	0,0693	0,2540	0,0628	0,2426
Fruta	0,4115	0,4921	0,3654	0,4816	0,4770	0,4995
Refrigerante	0,2174	0,4124	0,2257	0,4181	0,2055	0,4041
Sal	0,1302	0,3365	0,1564	0,3632	0,0928	0,2902
Atividade	0,4071	0,4913	0,4433	0,4968	0,3556	0,4787
Tv	0,1145	0,3184	0,1059	0,3077	0,1266	0,3325
Bebida	0,5598	0,4964	0,4385	0,4962	0,7323	0,4428
Cigarro	0,6766	0,4678	0,6281	0,4833	0,7455	0,4356
Urbano	0,8323	0,3736	0,8211	0,3832	0,8482	0,3588
Fortaleza	0,3308	0,4705	0,3157	0,4648	0,3523	0,4777
Restormf	0,2347	0,4238	0,2272	0,4190	0,2455	0,4304
Restouf	0,4345	0,4957	0,4572	0,4982	0,4022	0,4903

Fonte: Elaboração própria com dados da PNS 2013.

A maioria dos homens não tem instrução ou possui o ensino fundamental incompleto (40,77%). O grupo educacional de maior destaque entre as mulheres é o que corresponde ao ensino médio completo ou superior incompleto. (39,50%). O percentual de mulheres que consumia verduras e legumes em pelo menos 5 dias da semana (16,30%) foi praticamente o dobro do percentual de homens (8,71%).

#### 4.2 Modelo Logit ordenado

A primeira etapa deste trabalho consiste em analisar os determinantes do sobrepeso e da obesidade na população adulta cearense em 2013. A variável dependente é constituída das seguintes categorias ordenadas segundo o IMC: eutrofia, sobrepeso e obesidade. Desta forma, foi estimado um modelo Logit ordenado, que não faz nenhuma suposição sobre a distribuição de probabilidade das variáveis independentes (GREENE, 2012). Tal modelo, em sua versão padrão, é construído a partir de uma variável latente,  $y^*$ , definida como segue:

$$y^* = x' \beta + \varepsilon \quad (3)$$

Em que,  $x$  é o vetor de variáveis explicativas;  $\beta$ , o vetor de parâmetros a ser estimado;  $\varepsilon$ , os fatores não observados; e  $y^*$ , a variável latente, constituída por  $J$  categorias, é não observada. Neste modelo, a variável  $y$  é observada:

$$\begin{aligned} y &= 0 \text{ se } y^* \leq 0 \\ y &= 1 \text{ se } 0 \leq y^* \leq \mu_1 \\ y &= 2 \text{ se } \mu_1 \leq y^* \leq \mu_2 \\ &\vdots \\ y &= J \text{ se } \mu_{J-1} \leq y^* \end{aligned} \quad (4)$$

Em que, os  $\mu$ 's são parâmetros desconhecidos satisfazendo:

$$0 < \mu_1 < \mu_2 < \dots < \mu_{J-1} \quad (5)$$

No caso em que  $y$  apresenta três categorias, como no presente estudo, a função de distribuição de probabilidade pode ser escrita como:

$$\begin{aligned} \text{Prob}(y = 0 | x) &= F(-x' \beta) \\ \text{Prob}(y = 1 | x) &= F(\mu_1 - x' \beta) - F(-x' \beta) \\ \text{Prob}(y = 2 | x) &= 1 - F(\mu_1 - x' \beta) \end{aligned} \quad (6)$$

Assim como nos modelos Logit binários, os efeitos marginais das variáveis explicativas  $x$  não são exatamente os coeficientes estimados. Para as três categorias da variável dependente, os efeitos marginais são dados por:

$$\begin{aligned} \frac{\partial \text{Prob}(y = 0 | x)}{\partial x} &= -f(-x' \beta) \beta \\ \frac{\partial \text{Prob}(y = 1 | x)}{\partial x} &= [f(-x' \beta) - f(\mu_1 - x' \beta)] \beta \\ \frac{\partial \text{Prob}(y = 2 | x)}{\partial x} &= f(\mu_1 - x' \beta) \beta \end{aligned} \quad (7)$$

Generalizando (7), somente os sinais das mudanças em  $\text{Prob}(y = 0 | x)$  e  $\text{Prob}(y = J | x)$  não são ambíguos em relação ao sinal dos coeficientes.

Em sua versão padrão, o modelo ordenado pressupõe que os coeficientes estimados,  $\beta$ , se mantêm constantes ao longo de todas as categorias da variável dependente (LONG & FREESE, 2006; ABREU et al., 2009). No entanto, essa hipótese pode ser violada em todas ou em algumas das categorias de  $y$ , dando origem, respectivamente, ao modelo ordinal generalizado e ao modelo ordenado de chances proporcionais parciais. No modelo ordinal generalizado, os  $\beta$ 's são diferentes para todas as categorias da variável dependente. Por sua vez, o modelo ordenado de chances proporcionais parciais permite que os  $\beta$ 's coincidam em algumas categorias e diverjam em outras.

A escolha do modelo adequado se deu pela aplicação do teste de Brant (1990), cuja hipótese nula é a versão padrão do modelo ordenado. Desta forma, o modelo ordinal generalizado é o mais indicado quando a hipótese nula do teste é rejeitada para todos os coeficientes. Quanto é rejeitada em alguns coeficientes e em outros não, o modelo de chances proporcionais parciais é o mais adequado.

### 4.3 Algoritmos de mineração de dados para classificação

O segundo objetivo deste trabalho consiste em comparar o desempenho do modelo Logit binário com diferentes algoritmos de mineração de dados na classificação

antropométrica da população adulta cearense. Para tal, a variável dependente foi recodificada gerando a variável a seguir:

$$z = \begin{cases} 0, & \text{eutrofia;} \\ 1, & \text{sobrepeso ou obesidade.} \end{cases} \quad (8)$$

O desempenho dos diferentes algoritmos de classificação utilizados neste trabalho será comparado pela acurácia global de cada modelo:

$$\text{Acurácia}_{\text{ global}} = \frac{\text{número de casos classificados corretamente}}{\text{número total de casos}} \quad (9)$$

Para a implementação dos algoritmos, a base de dados é dividida em dois conjuntos: treinamento e teste. Os modelos são desenvolvidos no primeiro conjunto e testados no segundo. Como os dados utilizados neste estudo são categóricos, os classificadores do tipo Árvore de Decisão e Bayesianos são os modelos mais indicados (Dugan et al., 2015).

Seja  $S$  uma base de dados composta por  $s$  observações<sup>5</sup>. Assuma que a variável a ser classificada é composta por  $m$  valores distintos definindo  $m$  classes  $C_i$ ,  $i = 1, \dots, m$ . Defina  $s_i$  como o número de observações que pertence à classe  $C_i$ . Cada amostra de dados  $X$  tem  $n$  atributos<sup>6</sup>, que podem ser expressos vetorialmente como  $X = (x_1, x_2, \dots, x_n)$ . Cada elemento  $x_i$  é um valor nominal ou contínuo que corresponde com o atributo  $A_i = (i = 1, \dots, n)$ . Suponha que o atributo  $A_i$  tem  $q$  valores distintos  $\{a_{i1}, a_{i2}, \dots, a_{iq}\}$ , e  $A_i$  pode particionar a base de dados original  $S$  em  $q$  subconjuntos,  $\{S_1, S_2, \dots, S_q\}$ , em que  $S_j$  contém dados com o valor  $a_{ij}$  em  $A_i$ . Então, seja  $s_{ij}$  o número de observações da classe  $C_i$  no subconjunto  $S_j$ .

#### 4.3.1 Árvores de decisão

Este algoritmo gera um modelo em estrutura de árvore, que é uma estrutura hierárquica, composta por arestas consistindo de diferentes atributos e possui três tipos de nós: nó raiz, nós intermediários e nós finais representando uma classe. A árvore é construída com base no ganho de informação. O procedimento é descrito de forma breve a seguir. Em primeiro lugar, deve-se calcular a seguinte medida de informação:

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (10)$$

$$p_i = \frac{s_i}{s} \quad (11)$$

Em que,  $p_i$  é a probabilidade de uma observação aleatória pertencer à classe  $C_i$ .

Em segundo lugar, calcula-se a entropia  $E(A_i)$ , que é a informação esperada baseada na partição pelo atributo  $A_i$ :

$$E(A_i) = \sum_{j=1}^q \frac{s_{1j} + s_{2j} + \dots + s_{mj}}{s} I(s_{1j}, \dots, s_{mj}) \quad (12)$$

$$I(s_{1j}, \dots, s_{mj}) = \sum_{i=1}^m p_{ij} \log_2(p_{ij}) \quad (13)$$

<sup>5</sup> Esta seção baseia-se em Zhang et al. (2009).

<sup>6</sup> Na linguagem de mineração de dados, atributos correspondem às variáveis independentes.

$$p_{ij} = \frac{s_{ij}}{|S_j|} \quad (14)$$

Em que,  $|S_j|$  é o número de observações no subconjunto  $S_j$ .

Logo, a informação obtida em ramificar a árvore em  $A_i$  é dada por:

$$Gain(A_i) = I(s_1, s_2, \dots, s_m) - E(A_i) \quad (15)$$

O atributo  $A_i$  com o maior ganho de informação é escolhido como nó raiz e suas arestas são formadas de acordo com os diferentes valores de  $a_{ij}$ ,  $j=1, \dots, q$ . Este processo é repetido até que todas as observações pertençam à mesma classe, que rotulará o nó final. Ou seja, quanto mais alto na árvore, mais informativo é o atributo. Neste trabalho será implementado o algoritmo C4.5 (Quinlan, 1993), que é um dos mais aplicados na classificação de obesidade (Abdullah et al., 2016).

### 4.3.2 Classificadores Bayesianos

Os classificadores Bayesianos fazem a classificação com base na probabilidade de dada observação pertencer à determinada classe. Redes bayesianas (*Bayesian networks*) e Naïve Bayes são os dois modelos fundamentais nesse grupo de classificadores:

“Naïve Bayes algorithms assume that the effect that an attribute plays on a given class is independent of the values of other attributes. In practice, dependencies often exist among attributes; hence Bayesian networks are graphical models, which, unlike naïve Bayesian classifiers, can describe joint conditional probability distributions.” (ZHANG et al., 2009)

Dada uma amostra de dados,  $X$ , cuja classe será determinada, o algoritmo Naïve Bayes classifica determinada observação como pertencente à classe  $C_i$  se, e somente se:

$$P(C_i \setminus X) > P(C_j \setminus X), \text{ para } 1 \leq j \leq m, \neq i \quad (16)$$

Em que a probabilidade a posteriori  $P(C_i \setminus X)$ ,  $1 \leq i \leq m$ , pode ser determinada pelo Teorema de Bayes:

$$P(C_i \setminus X) = \frac{P(X \setminus C_i)P(C_i)}{P(X)} \quad (17)$$

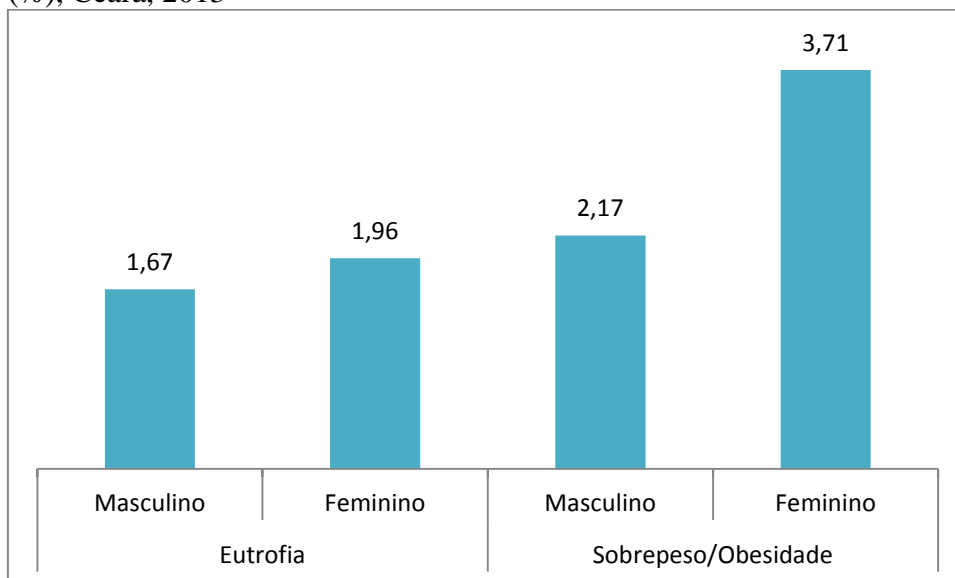
## 5. Análise de resultados e discussões

### 5.1 Prevalência de DCNT por classificação nutricional segundo IMC

Nesta seção, a população adulta cearense foi classificada em duas categorias segundo o IMC: eutrofia e sobrepeso/obesidade. Os resultados apresentados a seguir revelam que a prevalência do diagnóstico de algumas DCNT se mostrou maior entre a parcela da população adulta cearense acima do peso. Ademais, foi um problema que afetou as mulheres de forma mais acentuada.

A prevalência do diagnóstico de diabetes entre as mulheres acima do peso foi 3,71% (Figura 4) contra 1,96% entre as mulheres com peso normal. Já entre os homens, tais rubricas representaram, respectivamente, 2,17% e 1,67%.

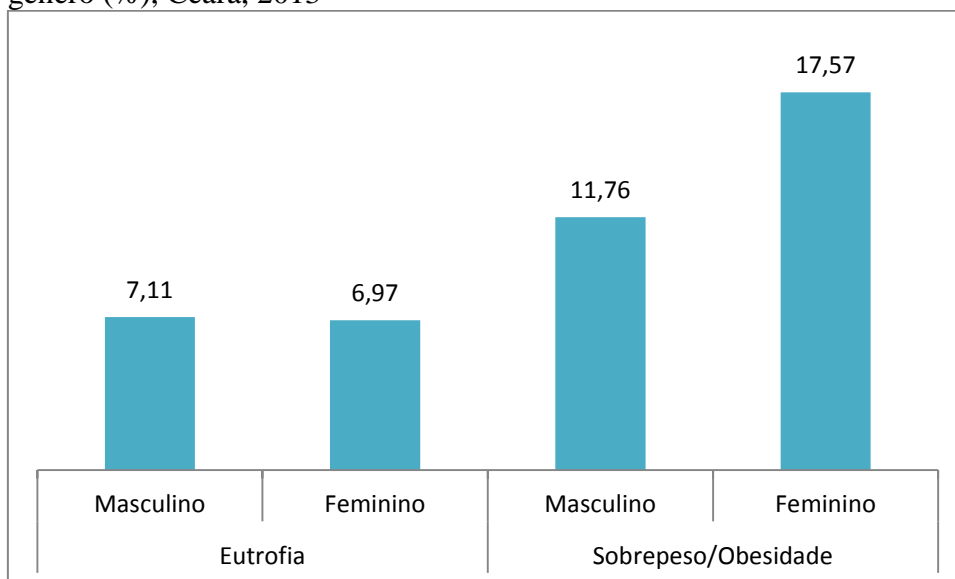
**Figura 4:** Prevalência do diagnóstico de diabetes por classificação nutricional e gênero (%), Ceará, 2013



Fonte: Elaboração própria com dados da PNS 2013.

A incidência de pessoas diagnosticadas com pressão alta também se mostrou superior entre as pessoas acima do peso (Figura 5). Neste grupo da população cearense, entre as mulheres 17,57% foram diagnosticados com pressão alta e entre os homens, 11,76%.

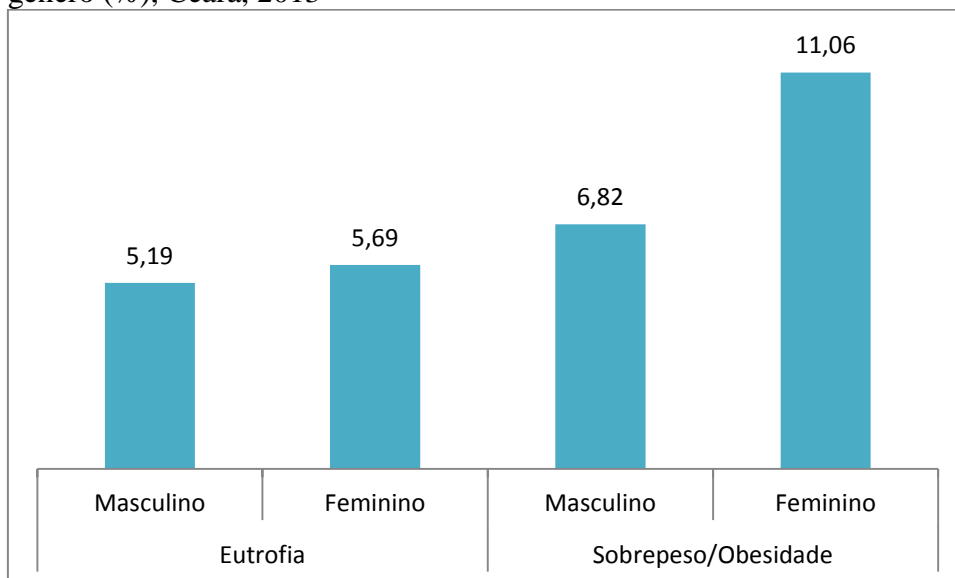
**Figura 5:** Prevalência do diagnóstico de pressão alta por classificação nutricional e gênero (%), Ceará, 2013



Fonte: Elaboração própria com dados da PNS 2013.

A Figura 6 indica que a prevalência do diagnóstico de colesterol alto foi semelhante entre homens e mulheres com peso normal, sendo muito mais elevada entre as mulheres acima do peso (11,06%) do que entre os homens (6,82%).

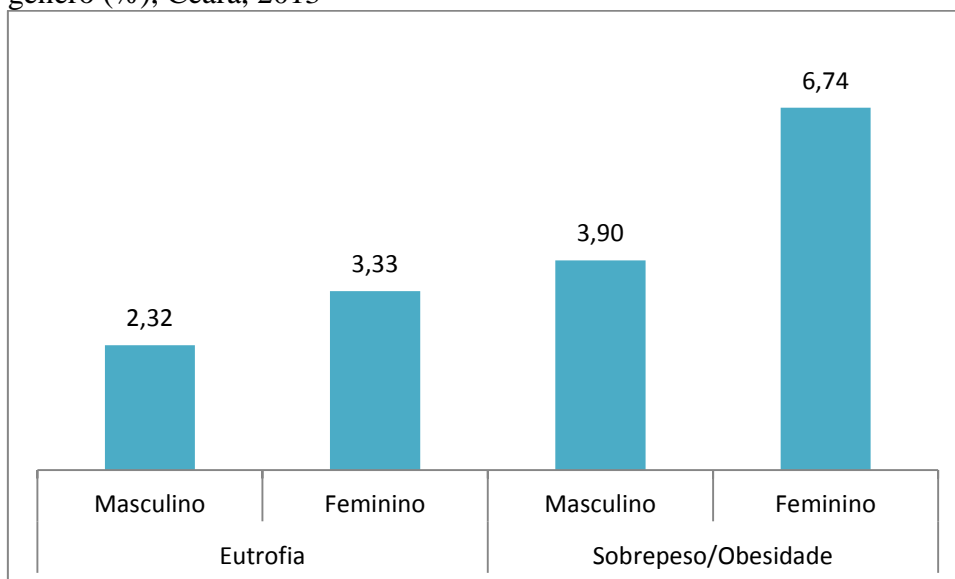
**Figura 6:** Prevalência do diagnóstico de colesterol alto por classificação nutricional e gênero (%), Ceará, 2013



Fonte: Elaboração própria com dados da PNS 2013.

A parcela de mulheres acima do peso diagnosticadas com reumatismo foi de 6,74% contra 3,33% das mulheres com peso normal. Já o percentual de homens eutróficos diagnosticados com reumatismo foi 2,32%, passando para 3,90% dos homens com sobrepeso ou obesos.

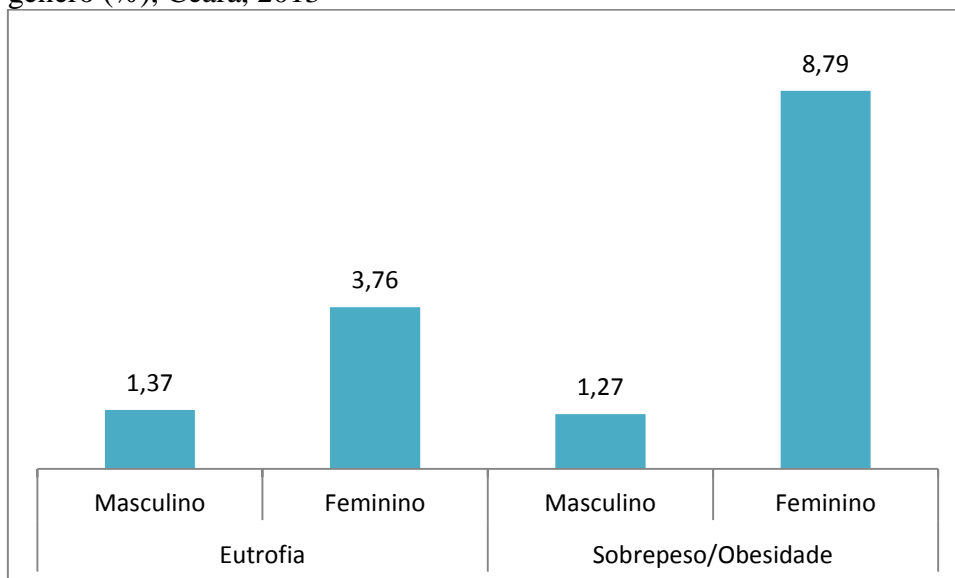
**Figura 7:** Prevalência do diagnóstico de reumatismo por classificação nutricional e gênero (%), Ceará, 2013



Fonte: Elaboração própria com dados da PNS 2013.

Por fim, a prevalência do diagnóstico de depressão entre os homens foi semelhante entre aqueles com peso normal e acima do peso. No entanto, 8,79% das mulheres acima do peso foram diagnosticadas como sofrendo depressão, enquanto apenas 3,76% das mulheres de peso normal o foram.

**Figura 8:** Prevalência do diagnóstico de depressão por classificação nutricional e gênero (%), Ceará, 2013



Fonte: Elaboração própria com dados da PNS 2013.

Os resultados desta seção corroboram o fato de sobrepeso e obesidade serem fatores de risco para a ocorrência de DCNT. Assim sendo, e diante da atual epidemia da obesidade, a adoção de medidas que atuem na prevenção e no combate dessa doença é muito importante. Neste sentido, a análise dos fatores determinantes do sobrepeso e da obesidade se mostra relevante ao gerar informações úteis para os formuladores de políticas públicas.

## 5.2 Determinantes do sobrepeso e da obesidade

A análise do impacto dos determinantes do sobrepeso e da obesidade na população adulta cearense se deu por meio da estimação de um modelo Logit ordenado. Os dados foram desagregados por gênero para a realização das estimações. O teste de Brant indicou que o modelo ordinal generalizado foi o mais apropriado tanto entre os homens quanto entre as mulheres. Os efeitos marginais das variáveis explicativas são apresentados na Tabela 5.

Sobrepeso e obesidade aumentaram com a idade para ambos os sexos. Entre os homens, o fato de ser branco se mostrou fator de risco para sobrepeso e obesidade, enquanto que entre as mulheres observou-se resultado oposto (mulheres brancas têm probabilidade de terem peso normal 11,3 p.p. maior do que mulheres não brancas). Os homens casados têm uma chance de terem sobrepeso 11,1 p.p. maior do que homens não casados. Mulheres casadas também se mostraram mais suscetíveis a terem problemas de sobrepeso e obesidade.

Em relação ao tipo de ocupação, homens empregados nos cargos classificados como direção e serviços tem mais chance de serem obesos em comparação àqueles trabalhando na classe denominada manual. De fato, o crescente sedentarismo de muitas formas de trabalho é apontado pela OMS como um fator de risco para obesidade. Por sua vez, mulheres alocadas nas classes direção e serviço apresentaram uma maior probabilidade de terem peso normal quando comparadas com aquelas na classe manual. A condição na força de trabalho teve impacto divergente entre homens e mulheres. Homens (mulheres) ocupados(as) têm uma chance de terem peso normal 36p.p. maior (menor) do que àqueles(as) não ocupados(as).

A educação teve relação inversa com a obesidade entre as mulheres, enquanto que entre os homens não se observou uma relação clara. Este resultado corrobora o de



Ward et al. (2015), ao encontrarem relação negativa entre IMC e escolaridade para mulheres e padrão não estabelecido entre homens na população adulta de Belo Horizonte.

**Tabela 5:** Efeitos marginais estimados para classificação antropométrica segundo IMC no Ceará desagregado por gênero, 2013

Coeficientes	Masculino			Feminino		
	Eutrofia	Sobrepeso	Obesidade	Eutrofia	Sobrepeso	Obesidade
31 a 40 anos	-0,176*** (0,000938)	0,0792*** (0,00102)	0,0969*** (0,000874)	-0,164*** (0,00104)	0,135*** (0,00114)	0,0290*** (0,000997)
41 a 50 anos	-0,230*** (0,000949)	0,173*** (0,00110)	0,0573*** (0,000941)	-0,274*** (0,00104)	0,247*** (0,00124)	0,0277*** (0,00111)
51 a 60 anos	-0,158*** (0,00123)	0,0813*** (0,00141)	0,0766*** (0,00125)	-0,170*** (0,00134)	0,0927*** (0,00152)	0,0771*** (0,00141)
branco	-0,0516*** (0,000904)	0,0300*** (0,000897)	0,0216*** (0,000623)	0,113*** (0,00111)	-0,0618*** (0,00106)	-0,0515*** (0,000745)
casado	-0,102*** (0,000859)	0,111*** (0,000857)	-0,00902*** (0,000571)	-0,0541*** (0,000972)	0,0146*** (0,000981)	0,0395*** (0,000739)
direcao	-0,0639*** (0,00115)	-0,00172 (0,00123)	0,0656*** (0,000980)	0,142*** (0,00168)	-0,110*** (0,00164)	-0,0322*** (0,00109)
servico	-0,107*** (0,000834)	0,0210*** (0,000863)	0,0861*** (0,000676)	0,146*** (0,00129)	-0,112*** (0,00129)	-0,0334*** (0,000872)
ocupado	0,362*** (0,000639)	-0,322*** (0,00195)	-0,0394*** (0,00196)	-0,363*** (0,00220)	0,248*** (0,00207)	0,115*** (0,00114)
fundamental	-0,0590*** (0,00111)	0,0954*** (0,00109)	-0,0364*** (0,000730)	-0,0672*** (0,00152)	0,156*** (0,00146)	-0,0892*** (0,000801)
medio	-0,126*** (0,00106)	0,111*** (0,00106)	0,0148*** (0,000757)	0,0480*** (0,00128)	0,0639*** (0,00125)	-0,112*** (0,000825)
superior	0,182*** (0,00183)	-0,105*** (0,00174)	-0,0771*** (0,000785)	0,0794*** (0,00201)	-0,00769*** (0,00185)	-0,0718*** (0,00104)
rdpc2	-0,140*** (0,00106)	0,0380*** (0,00107)	0,102*** (0,00107)	-0,0143*** (0,00147)	-0,0624*** (0,00165)	0,0767*** (0,00143)
rdpc3	-0,0355*** (0,00115)	-0,0313*** (0,00105)	0,0668*** (0,00100)	-0,0270*** (0,00149)	-0,0677*** (0,00168)	0,0946*** (0,00143)
rdpc4	-0,190*** (0,00104)	0,119*** (0,00108)	0,0714*** (0,00108)	-0,0621*** (0,00150)	0,0644*** (0,00160)	-0,00235 (0,00123)
rdpc5	-0,206*** (0,000996)	0,0692*** (0,000298)	0,136*** (0,000832)	-0,0704*** (0,00154)	-0,0135*** (0,00170)	0,0839*** (0,00150)
plano	-0,0967*** (0,00115)	0,00216 (0,00118)	0,0945*** (0,000929)	-0,0263*** (0,00135)	0,0451*** (0,00129)	-0,0188*** (0,000891)
feijao	-0,0158*** (0,00101)	0,0777*** (0,000968)	-0,0619*** (0,000751)	0,0992*** (0,000989)	-0,0630*** (0,000986)	-0,0363*** (0,000762)
verdura e legume	-0,00142 (0,00147)	0,0258*** (0,00145)	-0,0244*** (0,000826)	0,0322*** (0,00118)	-0,0133*** (0,000517)	-0,0189*** (0,000667)
carne	-0,0150*** (0,000798)	0,0469*** (0,000786)	-0,0318*** (0,000512)	0,0748*** (0,00113)	-0,133*** (0,00118)	0,0579*** (0,000939)
frango	0,0169*** (0,000831)	0,0378*** (0,000819)	-0,0547*** (0,000527)	-0,0188*** (0,000987)	0,0389*** (0,00100)	-0,0201*** (0,000731)
peixe	0,0738*** (0,00155)	-0,0668*** (0,00150)	-0,00707*** (0,00101)	0,159*** (0,00217)	-0,102*** (0,00209)	-0,0574*** (0,00119)

continua...

Continuação da Tabela 5

Coeficientes	Masculino	Feminino
--------------	-----------	----------

	<b>Eutrofia</b>	<b>Sobrepeso</b>	<b>Obesidade</b>	<b>Eutrofia</b>	<b>Sobrepeso</b>	<b>Obesidade</b>
fruta	-0,0276*** (0,000807)	-0,0225*** (0,000806)	0,0501*** (0,000569)	-0,0910*** (0,000944)	0,0216*** (0,000966)	0,0694*** (0,000740)
refrigerante	0,0416*** (0,000941)	-0,0298*** (0,000909)	-0,0118*** (0,000620)	-0,0231*** (0,00117)	-0,0145*** (0,00118)	0,0376*** (0,000923)
sal	0,0232*** (0,00109)	-0,0936*** (0,00106)	0,0705*** (0,000863)	0,0198*** (0,00167)	-0,134*** (0,00166)	0,114*** (0,00146)
atividade	0,00261** (0,000833)	0,0517*** (0,000842)	-0,0543*** (0,000554)	-0,0602*** (0,000991)	0,0458*** (0,00100)	0,0144*** (0,000737)
tv	-0,0228*** (0,00123)	0,0269*** (0,00120)	-0,00414*** (0,000801)	-0,0977*** (0,00129)	0,163*** (0,00131)	-0,0649*** (0,000816)
bebida	0,104*** (0,000812)	-0,109*** (0,000817)	0,00560*** (0,000552)	0,142*** (0,00102)	-0,0691*** (0,00105)	-0,0725*** (0,000846)
cigarro	-0,0798*** (0,000913)	0,0737*** (0,000877)	0,00607*** (0,000591)	0,0812*** (0,00115)	-0,0528*** (0,00117)	-0,0284*** (0,000861)
urbano	0,0000273 (0,00107)	-0,0336*** (0,00103)	0,0336*** (0,000667)	0,0447*** (0,00133)	-0,0253*** (0,00141)	-0,0194*** (0,00111)
fortaleza	0,0536*** (0,000963)	-0,0459*** (0,000922)	-0,00770*** (0,000610)	-0,105*** (0,00113)	0,00746*** (0,00122)	0,0977*** (0,00103)
restormf	0,0125*** (0,00101)	0,0299*** (0,000995)	-0,0424*** (0,000592)	-0,0954*** (0,00113)	-0,0649*** (0,00132)	0,160*** (0,00119)

Fonte: Elaboração própria com dados da PNS 2013.

\*\*\*Significativo a 1%.

De um modo geral, maiores níveis de renda domiciliar per capita se mostraram como fator de risco para obesidade entre homens e mulheres. Por exemplo, homens no quintil mais rico da distribuição da renda domiciliar per capita têm uma chance de serem obesos 13,6p.p. maior do que aqueles no quintil mais pobre. Já entre as mulheres, esse aumento na chance de ser obesa corresponde a 8,39p.p. Ter plano de saúde foi fator de risco para obesidade (9,45p.p. a mais de chance em relação à categoria base) entre os homens, e foi fator de risco para sobrepeso (4,51p.p. a mais de chance em relação à categoria base) entre as mulheres.

Em relação à alimentação, o consumo regular de feijão, verdura e legume reduziu as chances dos homens serem obesos. No entanto, também reduziu a chance de terem peso normal, tendo efeito positivo sobre sobrepeso. Já entre as mulheres, tais variáveis foram fator de proteção contra sobrepeso e obesidade. O consumo de carne vermelha foi fator de risco para sobrepeso entre os homens (4,69p.p. a mais de chance em relação à categoria base). Já entre as mulheres foi o consumo de frango que aumentou as chances de terem sobrepeso em 3,89p.p. em comparação à categoria base. O consumo de peixe foi fator protetor nos dois sexos, sendo o efeito muito maior entre as mulheres na chance de ter peso normal (7,38p.p. para os homens e 15,9p.p. para as mulheres). O consumo de refrigerante foi fator de risco para obesidade entre as mulheres (3,76p.p. a mais de chance de serem obesas) e fator protetor de peso normal entre homens (4,16p.p. a mais de chance de terem peso normal).

A obesidade aumentou com o consumo de frutas e de sal em ambos os sexos. Além disso, passar mais de 4 horas por dia assistindo TV contribui para o aumento no sobrepeso em 2,69p.p. entre os homens e em 16,30p.p. entre as mulheres. O fato de não consumir bebidas alcoólicas aumentou a probabilidade de homens e mulheres terem peso normal. Por sua vez, nunca ter fumado foi fator de risco para sobrepeso e obesidade entre os homens, e contribuiu para o aumento da chance das mulheres terem peso normal.

Quanto ao local de moradia, homens residindo em Fortaleza e na Região Metropolitana de Fortaleza apresentaram menores chances de serem obesos em relação

aos homens residindo nos demais municípios cearenses. Resultado oposto se observou entre as mulheres. De um modo geral, os resultados dessa seção mostraram que o excesso de peso é afetado de maneira diferente entre homens e mulheres.

### 5.3 Acurácia dos algoritmos de classificação

Os algoritmos de classificação foram implementados nos dados desagregados por gênero. A divisão da base de dados entre os conjuntos de treinamento e teste seguiu a *cross*-validação de 10 passos. Ou seja, a amostra foi dividida em 10 partes iguais e em cada rodada, nove destas partes foi o conjunto de treinamento e uma, o conjunto de teste. Após as dez repetições deste exercício, foram obtidos a média e o desvio-padrão da acurácia de cada modelo. Os resultados são apresentados na Tabela 6 a seguir.

**Tabela 6:** Acurácia dos algoritmos na classificação antropométrica segundo gênero (%), 2013

Gênero	Logit	Árvore de decisão	Naïve Bayes	Rede bayesiana
Masculino	59,96	56,67	58,57	58,61
Feminino	66,85	62,99	62,49*	62,54*

Fonte: Elaboração própria com dados da PNS 2013.

\*Acurácia inferior ao modelo Logit (nível de significância de 5%).

Os modelos tiveram desempenho similar entre os homens, com aproximadamente 60% dos casos classificados corretamente (Tabela 6). Por sua vez, entre as mulheres, o modelo Logit teve desempenho superior ao dos classificadores bayesianos e similar ao da árvore de decisão.

A construção das árvores de decisão indicou que, entre os homens, a situação matrimonial foi o principal fator que contribuiu para o excesso de peso, já entre as mulheres foi a condição de ocupação no mercado de trabalho. Mais uma vez, sobrepeso e obesidade são afetados de maneiras distintas por seus determinantes entre homens e mulheres.

## 6. Considerações finais

Este estudo tem como objetivo contribuir para a literatura de Economia da Saúde no Ceará. Mais especificamente, visa favorecer o debate sobre sobrepeso e obesidade na população adulta cearense. Tal agravo, que representa fator de risco para as DCNT e gera outros efeitos perversos na qualidade de vida das pessoas, tem se tornado uma epidemia mundial ao longo dos últimos anos.

As informações utilizadas neste trabalho foram obtidas na PNS 2013 realizada pelo IBGE. A classificação antropométrica, feita com base no IMC, seguiu a proposição da OMS para a população adulta. As pessoas com baixo peso foram excluídas das análises, de forma que foram consideradas apenas aquelas com peso normal, sobrepeso e obesidade. Em vista da natureza da variável dependente, foram estimados modelos Logit ordenados com o intuito de analisar os determinantes do excesso de peso entre homens e mulheres residentes no Ceará. Dentre as variáveis explicativas, têm-se informações sobre o indivíduo, o domicílio e a região de moradia. Os resultados mostraram peculiaridades nos determinantes do excesso de peso entre os dois sexos. A prevalência de sobrepeso e obesidade aumentou com a idade e com a renda domiciliar per capita entre homens e mulheres. Assim como em Ward et al. (2015), a escolaridade apresentou relação negativa com a prevalência de excesso de peso entre as mulheres, não tendo padrão muito claro entre os homens. Alimentação saudável, representada pelo consumo regular de feijão, verduras, legumes e peixe, reduz a obesidade em ambos os sexos. Ademais, o consumo de frutas, o consumo excessivo de sal e número de horas

assistindo televisão foram fatores de risco para excesso de peso entre homens e mulheres.

Na segunda parte do trabalho, foram utilizados algoritmos de mineração de dados na classificação antropométrica da população adulta cearense. A construção das árvores de decisão mostrou que, novamente, a prevalência de excesso de peso é afetada de maneira diferente entre homens e mulheres. Entre os homens, a situação matrimonial foi o principal fator que contribuiu para sobrepeso e obesidade, já entre as mulheres foi a condição de ocupação no mercado de trabalho. A utilização de algoritmos de mineração de dados não apresentou ganhos em termos de acurácia quando comparados ao modelo Logit. Em trabalhos futuros, pode ser feita a comparação do modelo Logit com modelos de redes neurais artificiais e máquinas de vetores de suporte, por exemplo.

De um modo geral, os resultados indicam que a prevalência de excesso de peso é afetada de maneira diferente entre homens e mulheres, o que deve ser levado em consideração para uma melhor formulação de políticas para o enfrentamento deste agravo. Tais políticas devem ser compostas por medidas de diversas áreas, como saúde e educação, tendo em vista o perfil multifacetado dos determinantes do excesso de peso.

## **7. Referências bibliográficas**

ABDULLAH, F. S. et al. **Data Mining Techniques for Classification of Childhood Obesity Among Year 6 School Children**. In: International Conference on Soft Computing and Data Mining. Springer, Cham, 2016. p. 465-474.

ABESO - Associação Brasileira para o Estudo da Obesidade e da Síndrome Metabólica. **Diretrizes brasileiras de obesidade 2009/2010**. 3.ed. Itapevi, SP : AC Farmacêutica, 2009.

ABREU, M. N. S.; SIQUEIRA, A. L.; CAIAFFA, W. T. **Regressão logística ordinal em estudos epidemiológicos**. Rev Saúde Pública 2009; 43(1):183-194.

ANDRADE, R. G. et al. **Excesso de peso em homens e mulheres residentes em área urbana: fatores individuais e contexto socioeconômico**. Cad. Saúde Pública, Rio de Janeiro, 31 Sup:S1-S11, 2015.

BRANT, R. **Assessing proportionality in the proportional odds model for ordinal logistic regression**. Biometrics, v. 46, n. 4, p. 1171–1178, 1990.

CORREIA, L. L. et al. **Prevalência e determinantes de obesidade e sobrepeso em mulheres em idade reprodutiva residentes na região semiárida do Brasil**. Ciência & Saúde Coletiva, 16(1):133-145, 2011.

COSTA, E. C. et al. **Evolução do excesso de peso e fatores associados em mulheres de 10 a 49 anos em Pernambuco, Nordeste, Brasil**. Rev. Nutr., Campinas, 27(5):513-524, set./out., 2014.

DUGAN, T. M. et al. **Machine learning techniques for prediction of early childhood obesity**. Appl Clin Inform 2015; 6: 506–520.

FAYYAD, U; PIATETSKY-SHAPIRO, G; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases**. American Association for Artificial Intelligence, 1996.

- GREENE, W. H. **Econometric Analysis**. 7. ed. New Jersey: Prentice Hall, 2012.
- HAND, D.; MANNILA, H.; SMYTH, P. **Principles of Data Mining**. The MIT Press, 2001.
- LONG, J. S.; FREESE, J. **Regression Models for Categorical Dependent Variables Using Stata**, 2. ed. Stata Press, 2006.
- MCKINSEY GLOBAL INSTITUTE. **Overcoming obesity: An initial economic analysis**. 2014.
- MINISTÉRIO DA SAÚDE (BR). Secretaria de Vigilância em Saúde. Departamento de Vigilância de Doenças e Agravos não Transmissíveis e Promoção da Saúde. **Vigitel Brasil 2014: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico**. Brasília: Ministério da Saúde; 2015.
- Organização das Nações Unidas para Alimentação e Agricultura e Organização Pan-americana de Saúde. **Panorama de la seguridad alimentaria y nutricional**. Santiago, 2017.
- POCHINI, A.; WU, Y.; HU, G. **Data Mining for Lifestyle Risk Factors Associated with Overweight and Obesity among Adolescents**. In: Advanced Applied Informatics (IIAIAAD), 2014 IIAI 3rd International Conference on. IEEE, 2014. p. 883-888.
- QENDRAJ, D.; XHAF AJ, E. **Evaluating risk factors of being obese, by using id3 algorithm in Weka software**. European Scientific Journal, ESJ, v. 11, n. 24, 2015.
- ROEMLING, C.; QAIM, M. **Obesity trends and determinants in Indonesia**. Appetite, v. 58, n. 3, p. 1005-1013, 2012.
- SIQUEIRA, D. G. B. et al. **Diferenças entre sexos nos determinantes da obesidade abdominal em adultos de 40 anos ou mais: estudo de base populacional**. Rev. Nutr., Campinas, 28(5):485-496, set./out., 2015.
- WARD, J. et al. **Associação de fatores socioeconômicos com o índice de massa corporal, obesidade, atividade física e dieta em Belo Horizonte, Minas Gerais, Brasil: Estudo Saúde em Beagá**. Cad. Saúde Pública, Rio de Janeiro, 31 Sup:S182-S194, 2015.
- WITTEN, I. H.; FRANK, E. **Data mining : practical machine learning tools and techniques**. 2. Ed. (Morgan Kaufmann series in data management systems) 2005.
- WITTE, R. S.; WITTE, J. S. **Statistics**. 9. Ed. Wiley, 2009.
- ZHANG, L.; LIAO, T. F. **Multilevel Effects of Socioeconomic Status and Income Inequality on Obesity among Chinese Adults**. In: <http://paa2013.princeton.edu/papers/132519>. Acesso em 20 de julho de 2017.

ZHANG, S. et al. **Comparing data mining methods with logistic regression in childhood obesity prediction.** Information Systems Frontiers, v. 11, n. 4, p. 449-460, 2009.