

UNIVERSIDADE ABERTA



**IDENTIFICAÇÃO DE ALGUNS FACTORES DE RISCO PARA A
OBESIDADE E VISUALIZAÇÃO DA INFORMAÇÃO**

Janaína Maria Trinchão Silva Carvalho

Mestrado em Bioestatística e Biometria

2022

UNIVERSIDADE ABERTA



**IDENTIFICAÇÃO DE ALGUNS FACTORES DE RISCO PARA A
OBESIDADE E VISUALIZAÇÃO DA INFORMAÇÃO**

Janaína Maria Trinchão Silva Carvalho

Mestrado em Bioestatística e Biometria

Orientadora:

Professora Doutora Teresa Paula Costa Azinheira Oliveira

Coorientadora:

Professora Doutora Elisabete Carolino

2022

Resumo

A obesidade tem se apresentado como uma doença de preocupação mundial e que precisa de medidas preventivas por conta de um crescimento significativo em termos globais. A patologia obesidade potencializa o desenvolvimento de doenças futuras como a hipertensão, diabetes, cardiopatias, e o câncer. Para o estudo da obesidade é pertinente analisar não só as questões genéticas, mas também as clínicas e as sócios-demográficas.

O objetivo desse trabalho é identificar alguns fatores de riscos a obesidade e buscar maior visibilidade da informação, apoiando a comunidade médica e científica a buscar tratamentos preventivos, personalizados visando o combate, proporcionando melhor qualidade de vida a sociedade como um todo.

Nesta dissertação foi investigada o perfil de 714 pacientes relacionados a variáveis sociais, clínicas e genéticas usando técnicas univariadas, bivariadas e o modelo de Regressão Logística que foi fundamental para identificação dos fatores de risco para a obesidade.

A visualização da informação como uma ferramenta de apresentação também foi de grande importância para traçar o perfil do nosso paciente e dar apoio para melhores análises e tomada de decisões.

Com o uso do software estatístico (SPSS) e de softwares computacionais na área da visualização da informação (*Tableau Public e Power BI*), foi possível identificar e propor um melhor modelo que melhor demonstrasse as características e o conhecimento sobre a obesidade.

Palavras-chaves: Regressão Logística, Visualização da Informação, Obesidade.

Summary

Obesity has presented itself as a disease of global concern and that needs preventive measures due to significant growth in global terms. Obesity pathology enhances the development of future diseases such as hypertension, diabetes, heart disease, and cancer. For the study of obesity, it is pertinent to analyze not only genetic issues, but also clinical and socio-demographic ones.

The objective of this work is to identify some risk factors for obesity and seek greater visibility of information, supporting the medical and scientific community to seek preventive treatments, personalized aiming to combat, providing better quality of life to society as a whole.

In this dissertation, the profile of 714 patients related to social, clinical and genetic variables was investigated using univariate, bivariate techniques and the Logistic Regression model, which was fundamental for the identification of obesity risk factors.

The visualization of information as a presentation tool was also of great importance to outline the profile of our patient and provide support for better analysis and decision making.

With the use of statistical software (SPSS) and computer software in the area of information visualization (Tableau Public and Power BI), it was possible to identify and propose a better model that better demonstrated the characteristics and knowledge about obesity.

Keywords: Logistic Regression, Information Visualization, Obesity.

Dedico este trabalho a **Deus**. Sem ele nada seria possível.

À minha mãe **Nilda**, pela força, incentivo e carinho.

Ao meu esposo **Alexander** por acreditar em mim, me apoiando sempre nos estudos, e me cuidando durante toda a essa trajetória.

À minha amada filha **Ana Carolina** por todo amor e compreensão.

Nada disso seria uma realidade se não houvessem vocês na minha vida. Gratidão.

Agradecimentos

A Deus, que sempre esteve cuidando de mim e sendo a minha força para realização de cada sonho.

A Prof. Elisabeth Carvalho (in memoriam) por se colocar à minha disposição desde quando fui sua aluna na disciplina Visualização da Informação, propondo ser minha orientadora e me apoiando na construção de um projeto interessante e que de imediato tivesse a aprovação da UAB.PT, quando nem mesmo sabia que direção tomar.

A Professora Elisabete Carolino que é uma profissional incrível, dedicada e muito competente quando se fala em orientar trabalhos acadêmicos e de vasto conhecimento também na área de estatística e biologia. Prof. Elisabete Carolino foi um presente de Deus na minha vida que de maneira simples foi me conduzindo ao objetivo final de maneira brilhante.

A Professora Teresa Paula Oliveira por todos os ensinamentos e aprendizados.

Aos membros da banca examinadora pela generosidade de fazer parte desse momento tão importante de superação para mim.

Aos professores da UAB.PT que com grande dedicação e competência fizeram a sua parte deixando cada ensinamento necessário para que esta etapa fosse concluída.

Ao colega Geraldo que me incentivou durante todos os anos desse curso, acreditando que chegaríamos ao fim e esteve a minha disposição sempre que precisei.

A minha família que tenho de mais precioso, Nilda Pires, Alexander Carvalho e Ana Carolina Carvalho, que são a razão de todo meu empenho e dedicação.

Por fim a todos que direta ou indiretamente colaboraram para que meu sonho se realizasse.

Índice

CAPÍTULO 1	
1 INTRODUÇÃO	01
1.1 Motivação, Hipótese e contribuição.....	03
1.2 Objetivos.....	05
1.2.1 Objetivo Geral.....	05
1.2.2 Objetivos Específicos	05
1.3 Metodologia.....	05
1.4 Organização da Dissertação	07
CAPÍTULO 2	
2 OBESIDADE, GENÉTICA E FATORES CLÍNICOS	
2.1 Obesidade e seus conceitos.....	09
2.1.1 Obesidade mundial e suas estatísticas.....	11
2.2 Genética e obesidade.....	16
2.3 Fatores demográficos	19
2.4 Fatores clínicos ligados a obesidade.....	26
2.4.1 Hipertensão Arterial	26
2.4.2 Diabetes Mellitus tipo 1 e tipo 2.....	29
CAPÍTULO 3	
3 TÉCNICAS ESTATÍSTICAS MAIS UTILIZADAS NOS ESTUDOS DE OBESIDADE.....	33
CAPÍTULO 4	
4 VISUALIZAÇÃO DA INFORMAÇÃO	
4.1 Visualização da Informação e sua importância	36
4.2 Técnicas de Visualização da Informação.....	42
4.3 Ferramentas Computacionais da Visualização da Informação.....	45
CAPÍTULO 5	
5. MODELO ESTATÍSTICO E MODELO VISUAL PROPOSTO	
5.1 Modelo Estatístico	47
5.2 Regressão Logística Binária	51
5.2.1 Introdução	51
5.2.2 Estimação dos Parâmetros do Modelo de Regressão Logística Binária	53
5.2.3 Avaliação da qualidade do ajuste do modelo de Regressão Logística Binária	55
5.2.4 Odds Ratio	64
5.3 Visualização da Informação e Sistemas Computacionais	65
5.3.1 O uso do <i>Tableau Public</i>	66
5.3.2 O uso do <i>BI Power</i>	70
CAPÍTULO 6	
6. APLICAÇÃO SOBRE A BASE DE DADOS E RESULTADOS	75
6.1 Base de Dados e Variáveis	75
6.1.1. Variável resposta e variáveis preditoras	75

6.1.2. Definindo algumas variáveis preditoras	76
6.2 Construção do Modelo	77
6.3. Software Utilizado	78
6.4 Análises Descritiva	79
6.4.1 Análises Univariada.....	79
6.4.2 Análise Bivariada	83
6.4.3 Seleção de variáveis e ajuste do modelo	83
6.4.4 Escolha do modelo mais adequado	113
CAPÍTULO 7	
7. APLICANDO A VISUALIZAÇÃO DA INFORMAÇÃO NA BASE DE DADOS.	116
7.1 Uso do <i>Tableau Public</i> na apresentação dos dados	116
7.1.2 Representações gráficas <i>Tableau Public</i>	117
7.2 Uso do Power BI na apresentação dos dados	120
7.2.1 Representações gráficas <i>Power BI</i>	120
7.3 Modelo visual mais adequado ao estudo.....	123
CAPÍTULO 8	
8. CONCLUSÃO	125
CAPÍTULO 9	
9. CONSIDERAÇÕES FINAIS E FUTUROS TRABALHOS.	128
REFERÊNCIAS BIBLIOGRÁFICAS.....	129
ANEXOS	
Anexo I	138
Anexo II	143
Anexo III.....	148

Índice de Tabelas

Tabela 01: Classificação baseado no IMC pela OMS para adultos com idade superior ou igual a 18 anos	10
Tabela 02: Prevalência de excesso de peso e obesidade por região (1980 - 2015)	15
Tabela 03: Número de pessoas com diabetes no mundo por região 2017 e em 2045.....	29
Tabela 04: Registro de artigos, dissertações e teses que apresentam técnicas estatísticas voltadas a área de epidemiologia incluindo obesidade.....	33
Tabela 05: Exemplo de algumas técnicas de Visualização de Informação.....	43
Tabela 06: Base para cálculo da classificação do modelo.....	60
Tabela 07: Referência dos valores de Kappa e a avaliação de concordância com o modelo.....	62
Tabela 08: Resultado do Teste de Normalidade da variável idade	80
Tabela 09: Resultado do Teste de Normalidade da variável Atividade de Fosfatase	81
Tabela 10: Resumo de casos inclusos no estudo da análise do modelo.....	86
Tabela 11: Codificação da variável dependente e independentes do estudo.....	87
Tabela 12: Teste de <i>Omnibus</i> sob o critério de seleção <i>Enter</i>	88
Tabela 13: Razão de Verossimilhança, Pseudo R Cox e Snell, Pseudo de <i>Nagelkerke</i> pelo critério de seleção <i>Enter</i>	88
Tabela 14: Teste de <i>Hosmer e Lemeshow</i> sob o critério de seleção <i>Enter</i>	89
Tabela 15: Classificação do modelo segundo o critério de seleção <i>Enter</i>	89
Tabela 16: Área sobre a curva Roc segundo o critério de seleção <i>Enter</i>	90
Tabela 17: Tabulação cruzada e a concordância de K de Cohen segundo o critério de seleção <i>Enter</i>	91
Tabela 18: Variáveis da equação do modelo segundo o critério de seleção <i>Enter</i>	92

Tabela 19: Lista dos casos de outliers segundo o critério de seleção <i>Enter</i>	94
Tabela 20: Teste de <i>Omnibus</i> sob o critério de seleção <i>Forward</i> condicional	96
Tabela 21: Razão de Verossimilhança, Pseudo R Cox e Snell, Pseudo de Nagelkerke pelo critério de seleção <i>Forward</i> condicional	97
Tabela 22: Teste de <i>Hosmer e Lemeshow</i> sob o critério de seleção <i>Forward</i> condicional	97
Tabela 23: Classificação do modelo segundo o critério de seleção <i>Forward</i> condicional	98
Tabela 24: Área sobre a curva Roc segundo o critério de seleção <i>Forward</i> condicional	99
Tabela 25: Tabulação cruzada e a concordância de K de Cohen segundo o critério de seleção <i>Forward</i> Condicional	100
Tabela 26: Variáveis da equação do modelo segundo o critério de seleção <i>Forward</i> condicional	101
Tabela 27: Lista dos casos de outliers segundo o critério de seleção <i>Forward</i> Condicional	103
Tabela 28: Teste de <i>Omnibus</i> sob o critério de seleção <i>Backward</i> condicional	105
Tabela 29: Razão de Verossimilhança, Pseudo R Cox e Snell, Pseudo de Nagelkerke pelo critério de seleção <i>Backward</i> condicional	105
Tabela 30: Teste de <i>Hosmer e Lemeshow</i> sob o critério de seleção <i>Backward</i> condicional	106
Tabela 31: Classificação do modelo segundo o critério de seleção <i>Backward</i> condicional no SPSS	106
Tabela 32: Área sobre a curva Roc segundo o critério de seleção <i>Backward</i> condicional no SPSS	107
Tabela 33: Tabulação cruzada e a concordância de K de Cohen segundo o critério de seleção <i>Forward</i> Condicional	108
Tabela 34: Variáveis da equação do modelo segundo o critério de seleção <i>Backward</i> condicional	109

Tabela 35: Lista dos casos de outliers segundo o critério de seleção <i>Backward</i> Condicional	112
Tabela 36: Indicadores dos critérios de seleção de variáveis.....	114

Índice de Figuras

Figura 01: Prevalência de obesidade segundo gênero e idade.....	20
Figura 02: Variação temporal da frequência de excesso de peso e de obesidade (%) no conjunto da população adulta das capitais dos estados brasileiros e do Distrito Federal entre 2006 e 2013, segundo sexo.....	21
Figura 03: Estimativa da prevalência do IMC igual ou acima de 30 Kg/m ² em países de diferentes regiões do mundo no ano de 2005.....	22
Figura 04: Esquema de processo de visualização. Adaptado de Card, Mackinlay e Scheiderman (1999).....	39
Figura 05: O Diagrama esquemático do processo de visualização.....	40
Figura 06: Processo simplificado de visualização da informação auxiliada por computador.....	40
Figura 07: A curva ROC e o poder discriminante do modelo.....	62
Figura 08: Tela de abertura do <i>Tableau Public</i>	67
Figura 09: Tela de interação do <i>Tableau Public</i> depois de adicionado os dados.....	68
Figura 10: Produções gráficas no <i>Tableau Public</i> com estudos na área da saúde.....	69
Figura 11: Produções gráficas no <i>Tableau Public</i> em outras áreas de estudo.....	69
Figura 12: Interface do <i>Software Power BI</i>	71
Figura 13: Uso do <i>Power BI</i> pelo SUS relacionado ao avanço da COVID-19 em junho de 2020, por cidade.....	72
Figura 14: Uso do <i>Power BI</i> pelo SUS relacionado ao COVID-19 aplicado ao mapa de pontos por bairro para demonstrar as regiões mais críticas em junho de 2020.	73
Figura 15: Uso do <i>Power BI</i> pelo SUS relacionado ao COVID-19 na organização de banco de dados através de tabela em junho de 2020.	73
Figura 16: Gráfico da Curva Roc segundo o critério de seleção <i>Enter</i>	90
Figura 17: Gráficos da análise dos resíduos segundo o critério de seleção <i>Enter</i>	95
Figura 18: Gráficos da Distância de Cook segundo o critério de seleção <i>Enter</i>	95
Figura 19: Gráfico da Curva Roc segundo o critério de seleção <i>Forward</i> condicional	99
Figura 20: Gráficos da análise dos resíduos segundo o critério de seleção <i>Forward</i> Condicional	103

Figura 21: Gráficos da distância de Cook segundo o critério de seleção Forward	104
Figura 22: Gráfico da Curva Roc segundo o critério de seleção Backward condicional	107
Figura 23: Gráficos da análise dos resíduos segundo o critério de seleção Backward Condicional	112
Figura 24: Gráficos da distância de Cook segundo o critério de seleção Backward Condicional	113
Figura 25: Gráfico da variável obesidade segundo a idade.....	117
Figura 26: Gráfico da variável obesidade segundo o sexo.	117
Figura 27: Gráfico da variável obesidade segundo a atividade fosfatase ácida.	118
Figura 28: Gráfico da variável obesidade segundo o fenótipo ACP-1.....	118
Figura 29: Gráfico da variável obesidade segundo o grupo sanguíneo MN	119
Figura 30: Gráfico da variável obesidade segundo o grupo sanguíneo Lewis.....	119
Figura 31: Gráfico da variável obesidade segundo o fenótipo da haptoglobina.	119
Figura 32: Gráfico da variável obesidade.	120
Figura 33: Gráfico da variável obesidade segundo a idade	120
Figura 34: Gráfico da variável obesidade segundo fenótipo ACP-1.....	121
Figura 35: Gráfico da variável obesidade segundo o género.....	121
Figura 36: Gráfico da variável obesidade segundo a atividade fosfatase ácida	121
Figura 37: Gráfico da variável obesidade segundo o grupo sanguíneo Lewis.....	122
Figura 38: Gráfico da variável obesidade segundo o fenótipo de haptoglobina.	122
Figura 39: Gráfico da variável obesidade segundo o grupo sanguíneo MN.....	122
Figura 40: Perfil dos pacientes amostrais por variáveis e maiores frequências.	127
Figura 41: Percentual de obesos adultos a nível mundial, destaque para Portugal Brasil.....	148

Índice de Anexos

Anexo I: Output do SPSS na Análise Univariada.....	138
Anexo II: Output do SPSS na Análise Bivariada.....	143
Anexo III: Figuras complementares	148

Lista de Abreviaturas e Siglas

OMS – Organização Mundial da Saúde

VIS – Visualização da Informação

DCNT – Doença Crônica Não Transmissível

IMC – índice de Massa Corpórea

IOTF - International Obesity Task Force

IBGE – Instituto Brasileiro de Geografia e Estatística

VIGITEL – Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico)

WHO – *World Health Organization*

HTA – Hipertensão Arterial

PAS – Pressão Arterial Sistólica

PAD – Pressão Arterial Diastólica

DM – Diabete Mellitus

DM1 – Diabete Mellitus Tipo 1

DM 2– Diabete Mellitus Tipo 2

INFOVIS – Visualização da Informação

BI – *Business Intelligence*

SUS – Sistema Único de Saúde

HP – Fenótipo de Haptoglobina

ACT - Teste de Controle da asma

Capítulo 1

1 INTRODUÇÃO

A obesidade é considerada hoje um problema de saúde pública além de ser um tema em destaque para a comunidade científica. Grande parte da população mundial vem apresentando inúmeros problemas de saúde e também nutricionais, o que torna a obesidade uma epidemia de grande interesse de estudo a nível global.

A obesidade é definida como a acumulação anormal ou excessiva de gordura que pode prejudicar a saúde. O fato é que as estatísticas que envolvem o fator obesidade têm chamado a atenção das comunidades médica e científica a nível mundial, pois os índices indicam que o excesso de peso triplicou desde 1975.

Esta epidemia engloba tanto adultos quanto crianças de forma alarmante, trazendo a necessidade de estudos cada vez mais avançados, em busca de respostas adequadas que possibilitem a implementação de ações preventivas para as novas gerações. A necessidade de ações não é visível, e precisam de ser postas em prática de forma dinâmica, para suavizar um quadro de saúde que tende a ser complexo nas próximas décadas. (Organização Mundial da Saúde – OMS, 2020).

O problema da obesidade vem chamado atenção de maneira continua nos últimos 30 anos. As estatísticas referentes a 2016 já indicavam que crianças e adolescentes representavam cerca de 18% delas com sobrepeso e obesidade. Essas crianças e adolescentes serão nossos futuros adultos, e a existência desse sobrepeso e da obesidade representa um fator potencial para o surgimento de doenças coronarianas, hipertensão, diabetes méllitus, nestes nossos futuros adultos elevando a taxa de morte, tomando-se necessário mais estudos que identifiquem fatores de riscos apontando mais explicações que fomentem a busca de ações e medidas preventivas. (Wu, B., Jiang, Y., Jin, X. et al., 2020)

A implementação de medidas preventivas para a obesidade é fundamental, pois o excesso de peso ou a obesidade pode conduzir a tipos variados de enfermidades que podem dar origem à morte, tais como a diabetes, problemas cardiovasculares, e vários tipos de câncer (Martins, 2018).

Um aspecto de extrema importância no estudo da obesidade é a genética. Podemos herdar, tanto do pai como da mãe, genes que nos tornam mais propensos a sofrer de obesidade. Mas não se pode restringir a explicação da obesidade à herança genética. A obesidade resulta da interação de múltiplos genes, fatores ambientais e fatores comportamentais. Muitas vezes, as formas graves de obesidade verificam-se quando pessoas que têm uma suscetibilidade genética vivem num ambiente favorável ao aumento de peso, têm uma alimentação desequilibrada e uma vida sedentária, entre outros fatores.

A qualidade na alimentação dos indivíduos, é um tópico diretamente ligado ao fator obesidade, pois ao longo dos anos a população tem modificado seus padrões na alimentação por influência da indústria de laticínios que se torna cada vez mais agressiva e atraente na forma de vender seus produtos nas prateleiras.

Os produtos alimentícios que causam grande impacto na nossa vida e estão de maneira fácil ao nosso alcance, são produtos à base de embutidos, conservantes, com alto índice de açúcar e sódio, que contribuem para uma mudança no padrão da forma alimentar da população causando a elevação dos índices estatísticos de sobrepeso. A alimentação é um dos fatores de extrema importância para o controle da obesidade, contudo este não vai ser o tema deste trabalho (Martins, 2018).

A obesidade exemplifica um problema sindêmico crucial que exige uma política de nível internacional intervenções para reduzir o poder e influência de empresas multinacionais, como Big Sugar e Big Food, que visam implacavelmente a baixa renda.

(AE Ivanescu, P Li2, B George, et al., 2019, p. 189).

Existe um interesse amplo em trazer soluções e realizar estudos neste tipo de epidemiologia, visto que este é um problema que vem se agravando na sociedade e traz uma preocupação mundial nos sistemas públicos e sociais, uma vez que os registros numéricos têm chamado atenção da classe médica e da Organização Mundial da Saúde.

1.1 MOTIVAÇÃO, HIPÓTESE E CONTRIBUIÇÃO

Com o avanço das metodologias estatísticas em tão pouco tempo, avaliar os padrões atuais de saúde relacionados com o sobrepeso torna o assunto um tópico mais do que relevante para um estudo acadêmico científico, demonstrando onde este trabalho encontra a sua motivação.

O interesse é realizar um estudo para identificar alguns factores de risco para a obesidade, tais como determinados factores genéticos (fenótipo da Haptoglobina, fenótipo do ácido fosfatase, atividade do ácido fosfatase, dois sistemas de grupos sanguíneos, o sistema MN e o sistema Lewis), factores clínicos (tais como diabetes e, hipertensão), e factores sócio-demográficos (idade e género), recorrendo a metodologias estatísticas adequadas.

Os factores genéticos são de grande interesse, pois a medicina caminha para o tratamento personalizado, isto é, de acordo com as características genéticas.

São vastos os estudos que relacionam a obesidade com a diabetes *méllitus* tipo 2 (DM2), e doenças cardiovasculares, nomeadamente que a obesidade constitui um fator de risco para o surgimento dessas patologias. No presente estudo vamos tentar identificar se um paciente com estas patologias, apresentam alguma tendência para a obesidade.

Quando trabalhamos com questões relacionadas a saúde, uma grande massa de dados é gerada nas pesquisas, pelo que é necessário o uso de recursos computacionais capazes de auxiliar as análises, por forma a que os resultados sejam mais precisos para a obtenção de tratamentos eficazes.

A visualização de dados aliado ao trabalho de obesidade trará representações gráficas vinculadas ao banco de dados disponível para este estudo e permitirá uma maior compreensão das informações, apoiando os profissionais da área da saúde na sua tomada de decisão aprimorando o conhecimento.

Assim, através de um estudo transversal com dados coletados em instituições de saúde de norte a sul de Portugal, baseando-se em pacientes clínicos obesos e não obesos, procurar-se-á representar um estudo através de um modelo

matemático adequado que possa explicar quais os principais factores de risco para a obesidade, dentre os estudados.

Além do citado anteriormente, pretende-se ainda propor um modelo visual baseado em técnicas de Visualização de Informação, que contribua de forma a potencializar a análise dos resultados assim como abrir caminhos para futuras decisões e tratamentos que apresente de maneira mais clara, formas de tratamento para este público-alvo.

Em apoio a esta investigação serão utilizadas ferramentas computacionais para facilitar a obtenção de resultados tanto estatísticos quanto das representações gráficas através da Visualização da Informação (VIS).

O objetivo deste estudo é então identificar um método estatístico (da área da Bioestatística) mais apropriado para realizar o tratamento dos dados nesta área que permita obter resultados mais satisfatórios para a identificação de alguns fatores de risco para obesidade. Propor a utilização de um modelo visual apoiado em técnicas de Visualização de Informação que auxilie, facilite e melhore a análise e a interpretação dos resultados obtidos.

Assim é assumido neste trabalho que:

É possível através do uso da metodologia estatística Regressão Logística identificar a relação entre obesidade e a genética através do banco de dados proposto nesse estudo, sendo a Visualização da Informação uma contribuição para analisar esses resultados.

Em suma, este trabalho pretende contribuir para alertar a sociedade com relação aos fatores de maiores riscos para a obesidade, assim como potencializar as análises por parte os profissionais de saúde a encontrar respostas para os seus pacientes e tornar os tratamentos mais assertivos e personalizados.

1.2 OBJETIVOS

1.2.1 OBJETIVO GERAL

Este trabalho tem por objetivo identificar fatores de risco para obesidade aplicando uma metodologia estatística adequada que contribua para uma melhor análise do caso em estudo, propondo um modelo visual apropriado para melhor suporte à interpretação dos resultados apoiado em técnicas de visualização de Informação.

1.2.2 OBJETIVOS ESPECÍFICOS

Encontrar métodos estatísticos adequados que permitam responder ao objetivo geral deste trabalho, nomeadamente na identificação de fatores de riscos em obesos e não obesos.

Propor um modelo visual apoiado em técnicas de Visualização de Informação que seja capaz de auxiliar esse processo de análise dos resultados do método estatístico escolhido aplicado a dados de indivíduos obesos e não obesos.

1.3 METODOLOGIA

De acordo com Marcone (2003, p.83) “ método é um conjunto de atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo – conhecimentos válidos e verdadeiros, traçando o caminho a ser seguido, detectando erros e auxiliando as decisões do cientista. ”

A metodologia do trabalho é uma parte da escrita de suma importância, visto que nela contém o registro passo a passo de como a pesquisa foi realizada, de forma a deixar bem claro para o leitor assim como para profissionais da área.

Para o desenvolvimento deste estudo será utilizada uma pesquisa bibliográfica que será baseado em uma seleção de artigos, livros, publicações em revistas conceituadas, dissertações e monografias relacionadas com o assunto de interesse. Isso servirá de base para todo referencial teórico necessário que a pesquisa exigirá, visto que toda e qualquer pesquisa parte do pressuposto em qualquer área do conhecimento que se faz necessária e adequada.

O estudo para identificar os fatores de risco para a obesidade neste trabalho segue uma metodologia de estudo transversal. A amostra clínica para o desenvolvimento deste estudo foi cedida pela base de dados do Laboratório de Genética - Centro de Endocrinologia e Metabolismo da Universidade de Lisboa.

A base de dados compõe uma amostra de 714 sujeitos da população portuguesa de pacientes obesos e não obesos, contemplando dados demográficos como idade e sexo, dados genéticos tais como fenótipo de haptoglobina, do ácido fosfatase e atividade do ácido fosfatase, e fatores clínicos tais como a hipertensão e diabetes.

Essa pesquisa é de natureza quantitativa. Com o uso de recursos computacionais e ferramentas estatísticas, inicialmente os dados serão apresentados através de análise univariada, bivariada, com apresentação de tabelas com o objetivo de conhecer o perfil dos sujeitos em estudo, caracterizando uma análise descritiva dos dados.

Após traçar o perfil desses sujeitos, realizar-se-á uma análise minuciosa das variáveis pertencentes a esse banco de dados, onde serão selecionadas algumas delas, com o uso de testes estatísticos apropriados que auxiliará na construção de um modelo matemático que possa explicar a relação dessas variáveis com a obesidade, e que possibilite condições suficientes para demonstrar quais as situações e ou variáveis que são ou não um fator de risco na obesidade.

O modelo estatístico selecionado para este estudo será a Regressão Logística Binária, tendo como variável dependente a presença ou ausência de obesidade, onde faremos a construção do modelo geral, avaliando a importância da inclusão e retirada de cada variável, registrando o comportamento e ajustando o modelo até que sejam identificados os fatores de risco de acordo com as variáveis propostas no estudo.

Simultaneamente ao uso de métodos estatísticos proposto aqui nesta pesquisa, teremos o incremento da visualização da informação, ou seja, proporemos um modelo visual que seja eficaz e auxilie a interpretação dos resultados. O modelo será implementado com o auxílio da Visualização da

Informação, com a utilização das ferramentas de alto nível conhecidas como *Tableau Public e Power BI*.

O principal objetivo da Visualização da Informação, é trazer um despertar maior por parte dos profissionais, que neste caso aqui são relacionados com a saúde, ativando a inteligência humana, fazendo-a ter uma percepção mais apurada direcionando para novas respostas dentro de sua investigação.

(Bertolini, 2009, p. 18).

Uma vez feitos todos os processos relacionados a estatística e gráficos visuais serão descritos os resultados e análise de dados. Por fim fica a sugestão de futuros trabalhos a serem desenvolvidos ainda sobre esta temática.

1.4 ORGANIZAÇÃO DA DISSERTAÇÃO

O presente trabalho é composto pelas seguintes seções:

O capítulo 2, apresenta conteúdos referentes a obesidade, a genética e aos fatores clínicos, que são itens de sustento a essa dissertação apontando a motivação desse estudo. Através das pesquisas e de dados estatísticos no mundo, sobre o comportamento da obesidade é possível encontrar índices e apontar a sua relação com a genética e os fatores clínicos envolvidos num todo. Neste será demonstrado também, fatores demográficos com dados globais no que diz respeito a epidemia obesidade.

O capítulo 3, descreve técnicas estatísticas que permitam responder os objetivos propostos que sustentam esta dissertação, buscando as técnicas estatísticas mais utilizadas para identificação de fatores de risco para obesidade, apontando a melhor ferramenta a ser utilizada nos dados propostos.

O capítulo 4 apresenta uma descrição sobre a Visualização da Informação, sua importância, suas técnicas e as principais ferramentas computacionais. Isso nos dará um norte para a escolha das principais ferramentas que iremos usar no apoio a apresentação e interpretação dos dados definidos para esse estudo.

O capítulo 5 é descrito o modelo estatístico escolhido assim como o modelo visual que serão propostos e utilizados no conjunto de dados selecionados para o desenvolvimento deste trabalho. As técnicas selecionadas serão descritas, mostradas e justificadas, afim de obter resultados pertinentes para análise, tratamento e visualização de dados.

O capítulo 6, realiza a aplicação no bando de dados proposto, onde usar-se-á a técnica estatística apontada para identificação de fatores de risco a obesidade, observando os resultados e verificando o grau de sucesso obtido.

O capítulo 7 implementa os dados ao modelo visual proposto usando as ferramentas de Visualização da Informação através dos softwares *Tableau Public* e *Power BI* em busca de identificar e propor um modelo que apõe as mais-valias introduzidas na interpretação dos resultados.

O capítulo 8 apresentam-se as conclusões do trabalho, assim como a avaliação dos resultados finais.

O capítulo 9 contém as limitações encontradas e sugestões para trabalhos futuros.

Capítulo 2

2. OBESIDADE, GENÉTICA E FATORES CLÍNICOS

2.1 OBESIDADE E SEUS CONCEITOS

A obesidade apresenta-se no cenário mundial como uma patologia ou uma doença crônica não transmissível (DCNT) que se destaca no meio epidemiológico com a mesma importância das doenças cardíacas, hipertensão, diabetes, depressão e até mesmo o câncer.

Segundo (Dicio, 2020) a palavra obesidade tem origem no latim *obsesitas*, *atis* e descreve seu significado como a condição da pessoa obesa, de quem acumula uma quantidade elevada de gordura, representando o aumento do peso. Este peso representa um aumento de 20% do seu corpo total, o que conduz à constituição de novas doenças.

A OMS (2000), diz que obesidade é definida como uma doença que se apresenta quando existe o excesso de gordura no corpo, e esse traz malefícios para saúde.

A obesidade é um distúrbio da composição corporal definido pelo excesso absoluto ou relativo de gordura corporal, resultante de um estado de desequilíbrio entre as calorias ingeridas e as calorias gastas, que pode ter um efeito adverso sobre a saúde e levar a expectativa de vida reduzida e aumento dos problemas de saúde (Esteves P., 2011 apud Nammi S. et al., 2004; Breda J. et al., 2005; Francis S. et al., 2006; Uwaifo G. et al., 2011; Bessesen D.H. et al., 2008 & Anthony S. et al., 2008).

Existem várias formas de avaliar o estado nutricional, quer seja por métodos convencionais quer seja por métodos não convencionais. Dentro dos métodos convencionais, uma das formas para avaliar o estado nutricional é através do índice de massa corporal (IMC) que é obtido pela divisão do peso corporal (em quilogramas) e pela estatura (em metros) ao quadrado, sendo este um indicador

apropriado para avaliação do estado nutricional de adultos (Gigante D., Barros A & Olinto M., 1997, Apud Anjos, L.A,1992).

O IMC é classificado pela OMS, utilizando o critério apresentado na tabela 01:

Tabela 01: Classificação baseado no IMC pela OMS para adultos com idade superior ou igual a 18 anos.

IMC(kg/m)	Classificação
<18,5	Baixo Peso
18,5 - 24,99	Normal
25 - 29,99	Sobrepeso
30 - 34,99	Obesidade Grau I
35 - 39,99	Obesidade Grau II
>=40	Obesidade Grau III

Fonte: (OMS, 2000)

Amaral e Pereira (2008, p.314) descreve que a “obesidade é uma doença crônica e multifatorial. É um distúrbio metabólico, traduzido por um aumento persistente do balanço positivo entre o consumo e o gasto de energia”.

A obesidade é uma doença multifatorial, que abrange tanto fatores ambientais como fatores genéticos. Destaca-se também por ser um grande fator de risco para outras doenças (Pereira et al., 2019).

A simples definição de obesidade esconde na sua etiologia fenótipo algo muito mais complexo do que apenas o excesso de gordura, e esta se apresenta através do metabolismo superando a questão do aumento de peso (Chooi, Y., Ding C & Magkos F,2019).

Tendo em consideração o metabolismo, registra-se o aumento de morbidade por doença crônica mais conhecidas como Diabetes Méllitus, principalmente a do tipo 2, doenças cardiovasculares, câncer e hipertensão (Chooi, Y., Ding C & Magkos F,2019).

Marinna Mensorio (2016), com relação á etiologia vai mais além relacionando a obesidade não só a fatores genéticos e físicos, mas também a fatores sociais e psicológicos, avaliando essa doença como uma das maiores epidemias do mundo ocidental, atualmente sendo avaliada como um grave problema de saúde pública.

Diante do exposto por estudiosos, fica evidente que a obesidade não é uma doença de fácil tratamento, e devido a isso cada vez mais existe interesse em estudos na busca de tratamentos que por sua vez caminha para algo mais personalizado.

Um dos fatores que desperta interesse global sobre o estudo da obesidade é a observação de uma curva sempre crescente nas estatísticas. A seguir será descrito um histórico dessas estatísticas a nível global para um maior entendimento dos fatos sobre o excesso de peso e/ou obesidade.

2.1.1 OBESIDADE MUNDIAL E SUAS ESTATÍSTICAS

Estudos globais demonstram um aumento dos números da obesidade de forma assustadora sem distinção, envolvendo países desenvolvidos tanto quanto os subdesenvolvidos despertando atenção de órgãos governamentais da saúde.

O que chama a atenção dos órgãos governamentais é o fato de num período de vinte anos, ou seja, em duas décadas o volume de casos ter duplicado. A seguir vamos observar o comportamento dos números relacionados com a obesidade em alguns lugares tais como Brasil, Portugal, Estados Unidos, Austrália, Arábia Saudita, Reino Unido, Finlândia, Suíça, Holanda e Itália que são sempre destaque nos estudos realizados.

Amaral e Pereira (2008) descrevem que os dados do IOTF - International Obesity Task Force (2000), mostraram que a incidência de obesidade tem aumentado de forma significativa em várias regiões do mundo, o que eleva a mortalidade trazendo consequências à sociedade.

Dados mais atualizados sobre a situação mundial das taxas de obesidade por país referentes ao ano de 2022 pode ser visto no link <https://worldpopulationreview.com/country-rankings/obesity-rates-by-country>. Com base neste link foi possível obter as imagens (Figura 41 no anexo III) que destacam as populações Portuguesa e Brasileira, conforme consta maior incidência de pesquisa neste trabalho. Como se pode constatar de acordo com a figura 41, atualmente esta taxa é ligeiramente superior no Brasil. Graças à multiplicidade de problemas de saúde de que a obesidade pode ser responsável, a taxa desta pode ser, e é frequentemente, usada como indicador geral de saúde de uma população, uma vez que se constata que os países mais saudáveis do mundo têm em geral menores taxas de obesidade.

Chooi C, Ding C e Magkos F (2019) afirma que “Globalmente, foram estimados 1,9 bilhão e 609 milhões de adultos que estão acima do peso ou obesos em 2015, respectivamente, o que representa o equivalente a 39% de toda população mundial.”

Segundo Gigante, Barros e Olinto (1997) no Brasil, a prevalência de pessoas obesas subiu de 21% para 32% no período do ano de 1974 a 1989. Na pesquisa de orçamentos familiares 2002-2003. Análise da disponibilidade domiciliar de alimentos e do estado nutricional no Brasil realizada pelo Instituto Brasileiro de Geografia e Estatística- IBGE, aponta que em todas as regiões brasileiras a percentagem de adultos com sobrepeso está acima de 30%, sendo o maior destaque para a região sul com 45% dos indivíduos classificados com a utilização do IMC referenciada pela tabela 01 (Silva, 2010).

Marcio Mancini (2015) aponta dados mais recentes de uma pesquisa realizada pelo Vigitel (O sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico) realizada em 2013 em todas as capitais brasileiras. Essa pesquisa iniciou-se em 2006 e é aplicada até os dias de hoje para controle dos fatores de riscos ligados à obesidade.

Os resultados colhidos pelo Vigitel em 2013 apontam que no conjunto dos estados pesquisados, o excesso de peso de 50% foi confirmando mais uma vez

com relação ao sobrepeso dos indivíduos residentes nos estados do Sul destacando-se a cidade de Porto Alegre com 62,1% na frequência de adultos (Mancinni, 2015).

Na Ásia um estudo na Arábia Saudita com dados de 33 anos até 2013, mostrou o problema grave de obesidade colocando o país como um dos sete principais países com aumento nesse tipo de doença. As estatísticas descrevem que quase 35% dos adultos desse país são obesos sendo 41% mulheres e 31% para homens. (Abdulrahman, Alanazi Talal, and Dalia Kamal Alnagar, 2021)

Nos EUA o comportamento dos dados em relação à obesidade, considerando os últimos 50 anos, cresceu de forma acentuada, visto que nesta nação são poucas as restrições alimentares. Atualmente o índice dos obesos ou excesso de peso é o dobro do índice de quem tem peso normal com base no IMC compreendido entre 25 a 30 kg/m² (Hruby, Adela & Frank, 2014).

Segundo Esteves (2011) nos EUA entre 1960 e 1962 a porcentagem de adultos obesos era de 13,4% passando a registrar nesse intervalo 14,5%. Entre 1971 e 1974 esta porcentagem passou a registrar 15%, em 1980 assumiu 23,3% chegando a 30,9% nos anos 2000. Nos últimos 10 anos nos EUA, a prevalência aumentou para 35% nos adultos. Cerca de 100 milhões de adultos registram sobrepeso ou obesidade, dos quais 4,7% é de obesidade mórbida.

Com dados da 3ª Pesquisa Nacional de Exames de Saúde e Nutrição nos EUA considerando 5.799 homens e 6.429 mulheres entre 30 a 102 anos de idade, buscou identificar a relação da obesidade geral e adiposidade abdominal do indivíduo com a previsão de mortalidade. Demonstrou-se neste estudo que um adulto com IMC mais alto em ambos os sexos e a medida da circunferência da cintura em homens foram associados a sobrevida de indivíduos mais velhos, circunstância essa que não ocorre para indivíduos com meia idade que ao apresentarem uma alta medida na cintura-coxa para ambos os sexos e cintura-quadril em mulheres reforça a questão da associação positiva a mortalidade. (Reis JP, Macera CA, Araneta MR, Lindsay SP, Marshall SJ, Wingard D, 2009)

Em 2020 os autores Tim Aldair e Alan D. Lopez apontam em seu artigo que nos EUA e na Austrália o papel do sobrepeso e da obesidade nas tendências de mortalidade por doenças cardiovasculares resultou em mortalidade prematura que aumenta a uma taxa crescente anual atingindo índice de 3% para ambos os sexos, se comparado a outros tipos de doenças que não são cardiovasculares.

Ainda na Austrália um estudo realizado em 2012 destacou previsões futuras do índice de massa corpórea e da prevalência de peso, visto que este estudo apontou um aumento dessas variáveis em 0,4 e 0,8% ao ano respectivamente, levando em consideração que até o ano de 2025 teríamos estatísticas de 83% e 75% de homens e mulheres respectivos acima dos 20 anos com sobrepeso. (Haby M, Markwick A. et al.,2012)

Na Europa a prevalência da obesidade está na maior parte dos países. A variação nos países europeus está entre 5% e 40%. O destaque na obesidade está para países europeus como Jugoslávia, Grécia, com prevalência em torno de mais 35%, Roménia, República Checa, Inglaterra, com mais de 20% (Esteves, 2011).

Peter Scarborough e Linda J. Cobiac em 2021 realizaram um estudo com base em dados do Health Survey for England de período igual a 27 anos para modelar as trajetórias futuras e índice de massa corpórea na região da Inglaterra da qual mostrou-se que os adultos acima dos 18 anos terão uma prevalência de aumento no risco do excesso de peso até o ano de 2035.

Um outro estudo sobre uma projeção de modelagem da obesidade morbida para 2035 em todo Reino Unido confirma que a prevalencia de obesidade morbida aumente até 2035 com destaque para os países da Inglaterra, Escócia e o País de Gales, onde espera-se o registro dos maiores índices sobre a obediidade. (Keaver L, Xu B. et al, 2020)

Contraopondo totalmente aos índices anteriores, temos a Suíça, Holanda e Itália que registam uma taxa de obesidade em torno de 10% da população. Em

Portugal a prevalência de excesso de peso e obesidade na população adulta é acima de 10% (Esteves, 2011).

Para uma melhor percepção global dos dados da obesidade, na tabela 02 é possível observar os números relacionados por região continental. Para esta análise, considera-se a prevalência de IMC acima de 25 kg/m² representando o excesso de peso e IMC maior do que 30 kg/m² representando uma pessoa com obesidade (Chooi C, Ding C & Magkos F, 2019).

Tabela 02: Prevalência de excesso de peso e obesidade por região (1980 - 2015)

Regiões Continentais	Excesso de Peso (IMC > 25kg/m ²)		Obeso (IMC > 30kg/m ²)	
	1980	2015	1980	2015
	Oriente	37,90%	49,60%	11,80%
África	18,50%	34,50%	6,20%	12,70%
América	45,30%	64,20%	12,90%	28,30%
Europa	48,0%	59,60%	14,5%	22,90%
Ocidente	9,70%	28,20%	0,80%	4,90%
Ásia	10,90%	24,50%	1,70%	6,20%

Fonte: (Chooi, C. Et al.,2019)

Analisando os dados da tabela 02, é visível que de maneira global todas as regiões continentais sofreram um aumento preponderante nos últimos 35 anos (Chooi C, Ding C & Magkos F,2019). Este tipo de comportamento tem sido frequente independentemente dos países serem desenvolvidos ou subdesenvolvidos.

O aumento da prevalência da obesidade conforme foi evidenciado anteriormente, tem como consequência o desenvolvimento de doenças crônicas, que é uma das grandes preocupações governamentais. Neste estudo iremos entender um pouco sobre estas doenças e também verificar a questão genética que envolve a obesidade. É preciso obter um controle sobre esta situação global, entender o que está a acontecer em cada canto do mundo, comparando

tratamentos, considera-se um caminho para minimizar questões que envolvem obesidade e as outras doenças que se tornam consequência desse sobrepeso.

Sendo assim como muitos são os fatores relacionados ao estudo da obesidade, neste trabalho vamos ressaltar a questão da genética, assim como algumas variáveis de importância relacionada a questão sócio-demográfica (idade e gênero) do indivíduo e algumas doenças que estão correlacionadas com esse tema. Para isso a seguir procura-se conhecer quais as metodologias mais utilizadas no estudo da obesidade.

2.2 GENÉTICA E OBESIDADE

Com base em vários estudos, a obesidade é reconhecida como uma doença multifatorial, que envolve além de fatores genéticos também fatores ambientais.

Claudia Angeli (2008) também afirma que a obesidade é uma doença de herança multifatorial que embora seja influenciada por fatores genéticos na questão ambiental, essa proporção é em grau menor.

A OMS (2004) reforça afirmando que fatores genéticos contribuem para a susceptibilidade para a obesidade quando o indivíduo é exposto a fatores ambientais (Amaral e Pereira, 2008 apud OMS, 2004).

Porém, a obesidade também é o resultado de um conjunto de condições, que incluem fatores genéticos, físicos, sociais e comportamentais (Amaral & Pereira, 2008 Apud Koplan et al, 1999; Riebe et al 2002).

Devido ao aumento do número de indivíduos obesos, é de difícil compreensão o papel dos fatores genéticos na regulação do peso corporal. Um dos fatores que justifica o genoma humano conter variantes que conduzam à acumulação de gordura tem a sua explicação no chamado “genótipo económico”. Este termo “genótipo económico” caracteriza uma situação de momentos de escassez de alimentos, onde as pessoas registram um comportamento de gula, ingerindo grandes quantidades de alimentos, prevendo um período de escassez, crendo que nesse período a chance de sobrevivência era maior (Angeli, C. 2008).

Estudos voltados para a genética relacionados com a obesidade apontam para uma possível hereditariedade relativamente a esta epidemia. Segundo Vanessa Pereira et al, (2019) os estudos mostram que a relação entre obesidade e fatores familiares relacionados com a hereditariedade, comprometem de 20 a 25% dos humanos. Por isso a importância de avançar nos estudos sobre fatores genéticos em busca de um padrão traz contribuições importantes na construção de ferramentas capazes de realizar diagnósticos e tratamentos individualizados e ou personalizados.

As percentagens citadas acima passam a impressão de ser pouco representativa, mas não deixa de ser importante por dois fatores. O primeiro fator é que os números confirmam que a obesidade sim pode ser influenciada por fatores genéticos, e o segundo fator possibilita a realização de estudos que desenham o perfil do gene específico que contribui para o controle do excesso de peso nos indivíduos (Amaral & Pereira, 2008).

A transmissão familiar de obesidade é bem conhecida, acreditando-se que pode ser tanto por fatores genéticos, como por estilo de vida²². A ocorrência de obesidade dos pais levou a um risco de 1,3 entre os homens e 1,2 entre as mulheres para cada categoria da variável, chegando a um risco duas vezes maior para os entrevistados que informaram que tanto a mãe como o pai eram obesos (Gigante, Barros & Olinto, 1997: 244)

Um dos motivos de se estudar a relação da genética com a obesidade, é por compreender que este tipo de estudo científico pode potencializar propostas futuras de melhores tratamentos para combate da doença. Já foi citado anteriormente doenças crônicas que geram uma propensão e podem apresentar ligação à obesidade, das quais será de interesse do estudo analisar a relação que existe entre a obesidade e tais doenças.

A exemplo de uma doença, existem evidências fortes em vários estudos realizados que descreve a relação entre os indivíduos ter DM2 e excesso de peso apontando os fatores genéticos como um ponto importante para o desempenho desta relação (Mancinni, 2015).

Ainda segundo Mancinni (2015) é inequívoco que em sociedades em que as calorias são de acesso dispendioso e árduo, e muitas atividades físicas é desperdiçada para adquiri-las, a obesidade e a DM2 sejam incomuns.

A afirmação de Mancinni (2015) está muito associada a factores ambientais que ao associar-se a factores genéticos potencia a prevalência da obesidade a nível global. A mudança de comportamento no seu ambiente conjuntamente com pessoas que possuem o “gene da obesidade” contribui de forma intensa para o aumento das estatísticas.

Este aumento das estatísticas sugere de fato que existe uma predisposição da genética quando falamos da doença obesidade, não deixando de enfatizar a escolha do estilo de vida que é representado através do comportamento do apetite ou no trato com o gasto energético, sendo a maneira como a genética influencia na obesidade. Isto acentua-se principalmente quando nos referimos a determinadas famílias que têm um padrão de vida elevado em termos financeiros.

O fator genético também é determinante para doenças chamadas congénitas e um elemento de risco para doenças crónicas como diabetes, osteoporose, hipertensão, câncer e obesidade. Num outro momento vamos abordar sobre as que mais se destacam.

Uma diversidade de estudos na área da saúde, assim como metodologias, a citar o estudo em gémeos humanos monos e dizigóticos, estudo de modelos animais, associação e ligamento de genes candidatos em estudos epidemiológicos, estudos de caso-controle de indivíduos com ou sem obesidade têm comprovado a evidente participação da genética na obesidade.

Tendo isto sido registrado em vários estudos fica claro que sim existe a relação entre a genética e a obesidade. Em busca de compreender mais sobre estes indivíduos obesos que representam uma percentagem significativa na estatística mundial vamos em busca do perfil demográfico desses indivíduos em relação ao género e à idade.

2.3 FATORES DEMOGRÁFICOS

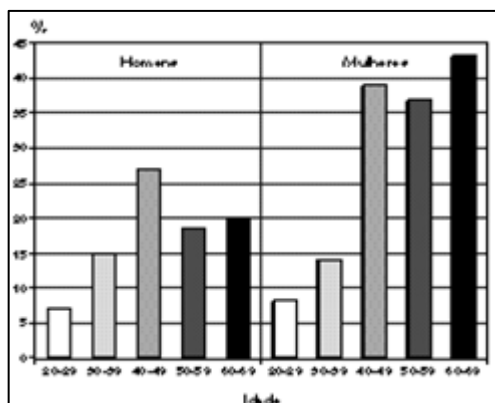
Tal como referido anteriormente, à obesidade é uma doença multifatorial, ou seja, resulta de vários fatores, pelo que se torna de grande importância traçar o perfil quanto ao género e à idade, visto que a busca por tratamentos personalizados é necessária.

Para determinar tratamentos individualizados e personalizados faz-se necessário conhecer como se comporta a obesidade na medida em que existe um aumento da idade (envelhecimento da população) e ou a mudança do género.

Considerando adultos, indivíduos com idade superior ou igual a 18 anos, a identificação do perfil dos indivíduos possibilitará uma maior preparação de ações e políticas públicas que possam retardar o crescimento dos índices, visto que a grande preocupação é que com o aumento da obesidade, tenhamos um número também crescente em outras enfermidades como cardiopatias, hipertensão, diabetes e câncer.

Diversos autores retratam as características mais marcantes com relação ao sexo e à idade dos indivíduos levando em consideração o excesso de peso e a obesidade (Gigante, Barros & Olinto, 1997, Angeli, 2008 & VIGITEL 2013).

Segundo a literatura, existem duas formas de se medir um indivíduo para avaliar se o mesmo é obeso ou não obeso. Uma das medições principais é o IMC, e outra forma bastante usada também é a medida da circunferência da cintura, sendo que neste estudo vamos nos fixar no IMC apenas. Os parâmetros do IMC já foram referidos anteriormente.



Fonte: (Gigante, Barros & Olinto,1997).

Figura 01: Prevalência de obesidade segundo gênero e idade.

O comportamento da obesidade através de um gráfico que relaciona a idade e o sexo, apresentado na figura 01, mostra que a obesidade é muito semelhante até aos 40 anos de ambos os sexos, mas que a partir dos 40 anos as coisas começam a ser representadas de forma distinta dando mais ênfase no sexo feminino no que diz respeito à obesidade, que duplica a sua condição inicial. Neste estudo a autora mediu os indivíduos usando o IMC considerando obesos os acima de 30 Kg/m², num estudo transversal com 1.035 pessoas com idade entre 20 e 69 anos de ambos os sexos na cidade de Pelotas-RS, Brasil (Gigante, Barros & Olinto,1997).

Gigante, Barros e Olinto (1997) em seus estudos também afirma que:

“As prevalências de obesidade foram maiores para mulheres e aumentaram de acordo com a idade, [...] chamou atenção que a prevalência entre os homens apresentou uma relativa estabilização a partir dos 40 anos, enquanto que para as mulheres as prevalências dobraram, a partir desta idade.”

(Gigante, Barros & Olinto,1997: 244)

Claudia Angeli (2008, p.1) em sua pesquisa escreve que “ Nossos resultados indicaram uma maior frequência de indivíduos com sobrepeso (IMC \geq 25 Kg/m²) e obesos (IMC \geq 30 Kg/m²) entre as mulheres (52% e 17,5%, respectivamente) do que entre homens (17,5% e 2,75%, respectivamente)”. A afirmação acima baseou-se no IMC, foi realizado com 790 indivíduos do Vale do Ribeira – SP, Brasil.

O Vigitel (Vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico) é um órgão responsável por monitorizar as frequências e as distribuições de fatores de risco e proteção para Doenças Crônicas Não Transmissíveis (DCNT) e iniciou o estudo no Brasil em 2006 realizando esta atividade até os dias atuais. A pesquisa do Vigitel é realizada em 26 capitais brasileiras e no Distrito Federal. O tamanho amostral estabelecido é de no mínimo 2000 mil indivíduos por capital. Esse trabalho é desenvolvido via telefone através de linhas fixas.

Tabela 4.8 Variação temporal da frequência de excesso de peso e de obesidade (%) no conjunto da população adulta das capitais dos estados brasileiros e do Distrito Federal entre 2006 e 2013, segundo sexo (Vigitel, 2006 a 2013).

	2006	2007	2008	2009	2010	2011	2012	2013	Variação anual média (pontos percentuais)*
Excesso de peso									
Homens	47,5	48,8	49,8	50,2	52,4	53,4	54,5	54,7	1,10
Mulheres	38,5	38,7	40,7	42,4	44,6	44,9	48,1	47,4	1,47
Obesidade									
Excesso de peso									
Ambos	42,6	43,4	44,9	46,0	48,2	48,8	51,0	50,8	1,30
Obesidade									
Homens	11,4	13,6	13,4	13,9	14,4	15,5	16,5	17,5	0,76
Mulheres	12,1	13,1	13,9	14,7	15,6	16,5	18,2	17,5	0,86
Ambos	11,8	13,3	13,7	14,3	15,1	16,0	17,4	17,5	0,82

*Corresponde ao coeficiente da regressão linear do valor do indicador sobre o ano do levantamento. Adaptada da referência Brasil, 2013.

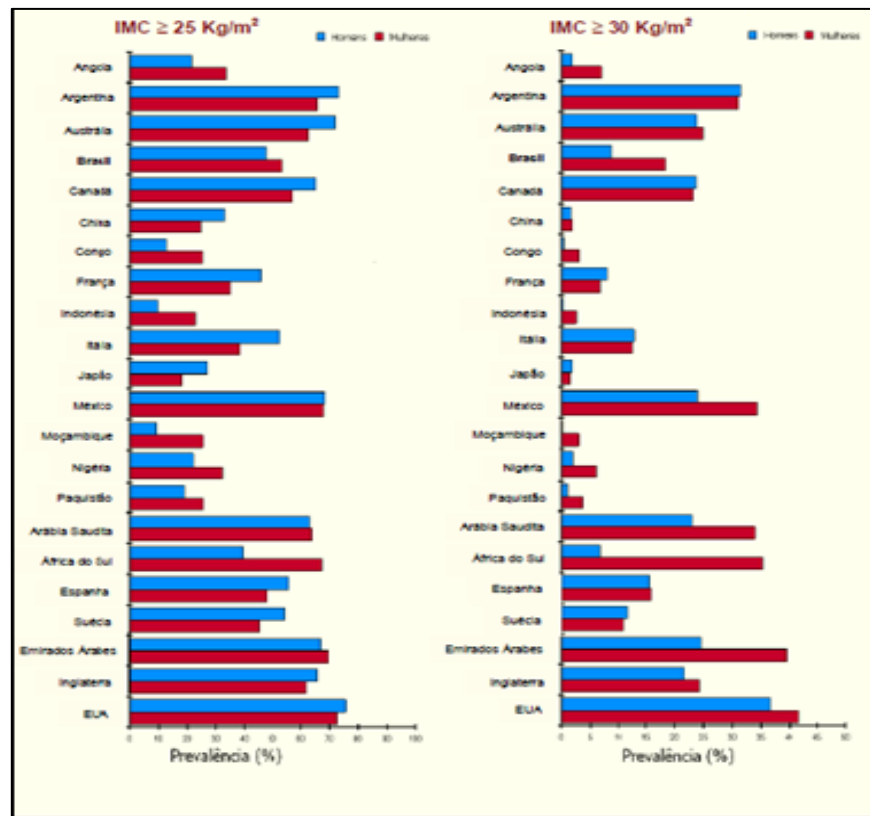
Fonte: (VIGITEL 2013, P.107).

Figura 02: Variação temporal da frequência de excesso de peso e de obesidade (%) no conjunto da população adulta das capitais dos estados brasileiros e do Distrito Federal entre 2006 e 2013, segundo sexo.

Na figura 02, o Vigitel (2013) relatou que em relação ao excesso de peso, ambos os gêneros tendem a aumentar a prevalência com o passar dos anos se referindo-se as mulheres até 54 anos e aos homens até 64 anos. Neste estudo o excesso de peso foi de 50,8% entre homens e 54,7 % entre as mulheres.

Relacionado com a obesidade, o Vigitel registou que existe um aumento do peso de três vezes mais em homens da faixa etária dos 18 aos 24 anos para a faixa

etária dos 55 aos 64 anos. As mulheres superam o resultado dos homens, com registos em mesmas faixas etárias confirmando aumentos maiores chegando até seis vezes mais em relação aos homens. (VIGITEL, 2013).



Fonte: (Angeli, 2008, p.8) apud WHO 2005)

Figura 03: Estimativa da prevalência do IMC igual ou acima de 30 Kg/m² em países de diferentes regiões do mundo no ano de 2005.

O estudo do VIGITEL em 2013 também registra que:

“Excesso de peso foi diagnosticado em cerca de metade da população e o de obesidade em 12,5% dos homens e em 16,9% das mulheres. Tanto o excesso de peso como a obesidade aumentam de frequência com a idade até a faixa etária de 45 a 54 anos em homens, e 55 a 64 anos em mulheres, declinando nas idades subsequentes. ” (VIGITEL 2013, p.102).

Numa relação comparativa de anos num estudo temporal da frequência percentual de excesso de peso e obesidade dos anos de 2006 a 2013, chega-se à conclusão que a média da frequência foi aumentando com o passar dos anos e pode ser comprovado na figura 02 (VIGITEL 2013, p.102).

Melo e Rissin (2015) realizaram uma pesquisa com indivíduos de 20 a 59 anos avaliando o sobrepeso e a obesidade de acordo com os padrões de IMC. Concluiu-se que as percentagens de sobrepeso e obesidade são respectivamente: na faixa etária de 20-29 (29,9% e 18,3%), de 30-39 (37,8% e 29,7%), de 40-49 (36,6% e 32,9%) e por fim de 50-59 (27,3% e 27,3%). Em relação ao género a autora aponta para os homens (31% e 24,3) enquanto que as mulheres (33,7% e 32,4%). A pesquisa foi baseada no IMC, realizada com 582 indivíduos no aglomerado urbano subnormal dos Coelhos – Recife, Brasil.

Os mesmos autores registram ainda que:

“A prevalência de excesso de peso foi de 62,5% (n=364), sendo maior nas mulheres (66,1%; n=251) do que nos homens (56,0%; n=113), com um aumento progressivo até a faixa etária de 40 a 49 anos, passando a declinar a partir de então. Nessa faixa, houve um risco de excesso de peso de 2,6 vezes.

(Melo & Rissin, 2015, p.1)

No Brasil ao descrever o perfil da obesidade com relação a faixa etária constatou-se que a obesidade aumenta gradativamente com a idade, sendo em maior prevalência em indivíduos acima dos 40 anos (Ferreira & Magalhaes 2006 apud Coitinho et al. 1991).

De forma ampla, a questão do género a nível mundial, é possível compreender com base na publicação da WHO (2005), que quantifica a prevalência do sobrepeso e obesidade em diferentes países do mundo associado ao género (Angeli, 2008, p.8 apud WHO 2005).

Através da figura 03 é possível visualizar como se comporta o género e também analisar as estimativas de prevalência com base no IMC de indivíduos com sobrepeso ($IMC \geq 25 \text{ Kg/m}^2$) e obesos ($IMC \geq 30 \text{ Kg/m}^2$) em 22 diferentes países.

Nestes países envolvidos nota-se no sobrepeso que a relação do género mulher e do género homem não chegam a ser tão relevantes quando comparadas as prevalências da obesidade. No sobrepeso podemos apontar destaque em poucos países citando a África do Sul, Moçambique, Congo, Indonésia e Angola.

No quesito obesidade, onde o IMC supera 30 Kg/m² observa-se uma diferença significativa entre os géneros homens e mulheres, sendo destaque para países como Angola, Brasil, Indonésia, Nigéria, África do sul e Emirados Árabes, onde as mulheres registam índices bem maiores se comparado com os homens.

Vale observar também que dentro dos países citados anteriormente, a África do Sul, Indonésia e Angola, apresentam índices diferenciados tanto para homens quanto para mulheres relacionados ao sobrepeso e também a obesidade.

No todo mundo no ano de 2013 observou-se que a taxa de obesidade vem crescendo e destacando uma prevalência maior em mulheres do que em homens. No período temporal de 1980 a 2013 afirma-se que a prevalência de excesso de peso aumentou de 28,8 para 36% em homens e de 29,8 para 38% em mulheres. Traduzindo para números absolutos tem-se que o crescimento foi de 857 milhões de indivíduos para 2,1 bilhões, o que confirma a necessidade de estudos para freiar esta epidemia (VIGITEL, 2013).

Quando relacionamos muitos países dentro do contexto da obesidade, estudos afirmam que existe uma diferença nos níveis de obesidade entre países desenvolvidos e subdesenvolvidos, demonstrando também que em lugares onde as condições de rendimentos são mais elevadas há uma maior incidência para a obesidade, e nos países menos favorecidos a incidência é menor.

Portugal é um dos países que não aparece na relação de países citados pela da World Health Organization - WHO (2005) baseado na figura 03, pois seus índices de obesidade estão próximos de 10 % da população (Angeli, 2008, apud WHO 2005).

Um estudo realizado pelo Instituto Nacional de Saúde em Lisboa – Portugal Continental, nos anos de 1955 a 1998 com uma amostra de 4328 indivíduos na faixa etária de 18 a 64 anos, constatou que 30% das mulheres e 41,1% dos homens registaram excesso de peso e 15,4% das mulheres e 12,9% dos homens eram obesos (Gaio, 2018).

O mesmo instituto citado anteriormente também realizou um estudo de âmbito nacional do período de 2003 a 2005, com indivíduos na faixa etária de 18 a 64 anos. A amostra de 8116 indivíduos apontou uma prevalência de 34,4% para as mulheres e 45,2% para os homens no excesso de peso e 13,4% nas mulheres e 15% dos homens para a obesidade (Gaio, 2018).

Em 2009 ainda em Portugal um estudo transversal “Hábitos alimentares e estilo de vida da população portuguesa” com uma amostra de 3529 indivíduos, considerando a faixa etária de 18 a 93 anos constatou que 27,8% das mulheres e 53,3% dos homens tinham excesso de peso e 10,4 % das mulheres e 11,2% dos homens eram obesos. Este mesmo estudo mais recente nos anos de 2015 a 2016 registrou a prevalência de 31% para mulheres, 45% para homens no excesso de peso e 23,7% para mulheres e 19,7% para homens com obesidade considerando 3435 indivíduos na idade de 18 a 64 anos (Gaio, 2018).

Vania Gaio (2018) através de dados recolhidos pelo Inquérito Nacional de Saúde com Exame Físico de 2015 realizou um estudo para descrever a prevalência de excesso de peso e obesidade da população Portuguesa em 2015. A autora trabalhou com uma amostra de 12.289 indivíduos dos quais incluía as variáveis IMC, sexo, idade, que é de interesse para este trabalho.

Relativamente à prevalência de excesso de peso e obesidade em Portugal, as taxas foram 38,9% e 28,7% respectivamente. No tópico excesso de peso os homens registraram 45,4% e na obesidade as mulheres destacaram-se com 32,1% superando os homens. Para o grupo etário a autora Vânia Gaio (2018) destaca o registo da obesidade sobretudo na faixa de 65 a 74 anos (41,8%) e de 45 a 54 anos o destaque é para o excesso de peso com 43,7%.

A abordagem do texto mostrou como se comporta o perfil, quanto ao género e a faixa etária dos indivíduos no mundo, no que diz respeito a excesso de peso e obesidade. A seguir deseja-se conhecer melhor o que a literatura fala sobre as doenças crónicas que se associam à obesidade, enfatizando a hipertensão, Diabetes Mellitus tipo 1 e Diabetes Mellitus tipo 2.

2.4 FATORES CLÍNICOS LIGADOS A OBESIDADE

2.4.1. HIPERTENSÃO ARTERIAL

A Hipertensão Arterial (HTA) é definida como uma condição onde os indivíduos apresenta níveis pressóricos altos continuamente propiciando lesão nas artérias e lesão no coração (Magalhães, 2018).

Renato Mello (2012) apresenta uma definição com base nas Organizações Nacionais e Internacionais que descreve que hipertensão arterial por pressão arterial sistólica (PAS) registrando valores superiores ou iguais a 140 mmHg e/ou pressão arterial diastólica (PAD) superior ou igual a 90 mmHg, aferidas em dois dias diferentes e em condições ideais.

Hipertensão Arterial é uma das doenças que se destaca quando relacionamos a obesidade. Isto dá-se pela condição onde o aumento da massa corporal tem sido reconhecido como um fator de risco para doenças hipertensivas (Alves et al., 2010 apud Mancinni, 2001).

“A obesidade é fator de risco independente para moléstias cardiovasculares e ocupa papel central na síndrome metabólica que envolve hipertensão arterial, resistência à insulina e dislipidemia” (GALVÃO, R. et al., 2002, p.1).

“A obesidade e a hipertensão Arterial determinam o aumento da massa cardíaca por mecanismos diferentes, e, quando ambas coexistem num mesmo paciente, ocasionam uma hipertrofia cardíaca mais acentuada e de instalação mais precoce”. (Robert, Ferreira & Zenella 2000 p. 69 apud Batlouni, M. et at, 1995 e Maza MP et at, 1998).

As citações acima confirmam que quando o paciente é obeso indica relação do paciente com a hipertensão arterial, o que potencia os problemas cardíacos através da hipertrofia cardíaca, tornando o paciente mais suscetível a índices de mortalidade.

Estudos relacionados à associação da obesidade com a hipertensão arterial afirmam que existe uma relação linear entre o IMC, ou seja, o peso corporal e a

pressão arterial de indivíduos em geral (Galvão e Kohlmann, 2002 apud Jones DW et al., 1994).

Outros estudos também afirmam o aumento de 3 a 8 vezes a frequência de hipertensão arterial quando o indivíduo é obeso. Contudo mesmo sabendo-se que existe uma relação entre a hipertensão arterial e o aumento de peso, os estudos ainda são superficiais no que diz respeito aos mecanismos que de fato fazem a hipertensão arterial crescer uma vez que o indivíduo aumenta o seu peso (Robert, Ferreira & Zenella 2000).

A autora Leila Alves, 2010 (apud Frohlich ED, 2002) descreve que os mecanismos fisiopatológicos que contribuem para o desenvolvimento da HTA associado a obesidade são complexos e multifatoriais. A citar alguns mecanismos temos a resistência à insulina, questões renais (retenção de sódio pelos rins), ativação do sistema nervoso e efeitos da leptina plasmática.

A leptina plasmática é definida como um hormônio peptídico que contém 167 aminoácidos e suas concentrações são diretamente proporcionais ao volume das células. Na história, a leptina foi nomeada como o hormônio das células adiposas (ALVES, L. et al., 2010 apud MAFFEI M et al., 1995, Coleman DL (1978), Harris RB et al., 1987).

A obesidade propicia o aumento do nível plasmático de leptina, alterando funções como controle do apetite e da massa corporal assim como a regulação do sistema cardiovascular. Os efeitos da leptina na modulação da pressão arterial trazem implicações em doenças cardiovasculares (ALVES, L. et al., 2010 apud HALL JE et al., 2001, Bravo PE et al., 2006).

Um outro mecanismo para o aumento da relação entre a HTA e a obesidade que é citado na literatura destaca que com o envelhecimento do indivíduo parece ocorrer um aumento da pressão arterial. Quando é feita uma análise à prevalência da hipertensão arterial o destaque é para o indivíduo obeso quando comparado ao não obeso (Robert, Ferreira & Zenella 2000).

Esta relação à idade, a pressão arterial é uma questão que fica muito próxima dos estudos relacionados com o aumento da obesidade e a idade. Ou seja, o indivíduo obeso, conforme ele envelhece, o peso tende a crescer assim como ele passa a ter um quadro mais propício a aumento da pressão arterial.

Estudos que contribuíram para demonstrar a relação do aumento da HTA e obesidade que foram destaques na literatura, se deu por meio da deterioração da sensibilidade à insulina. (Robert, Ferreira & Zenella 2000).

No texto abaixo desenvolvido no estudo realizado pelos autores Robert, Ferreira e Zenella (2000) descreve que:

“Com base em dados levantados, não há dúvidas quanto à existência da associação de obesidade à HA. Estudos experimentais e em humanos também comprovam ser a obesidade um fator de risco capaz de provocar aumento nos níveis pressóricos. Apesar de diversos mecanismos terem sido propostos para explicar elevação da pressão arterial nos indivíduos obesos, a fisiopatogênese da HA associada à obesidade não está completamente esclarecida. Há evidências sobre a participação da resistência à insulina e hiperinsulina”.

(Robert, Ferreira & Zenella, 2000, p.132)

A prevalência da hipertensão arterial no mundo considerando valores acima de 140/90mmHg mostra que algumas populações excedem 30% conforme a Organização Mundial da Saúde. No Brasil essa prevalência é de 32,5% atingindo 36 milhões de indivíduos na fase adulta sendo que 60% desse total são de idosos. (Magalhaes, 2018)

Robert, Ferreira e Zanella (2000) ainda citam que:

“Fatores tais como a hiperinsulinemia, aumentos de atividade simpática, aumento de atividade do sistema renina angiotensina e alterações de hemodinâmica renal que levam a um aumento na reabsorção tubular de sódio, parecem contribuir para o desenvolvimento da obesidade no paciente hipertenso”.

(Robert, Ferreira & Zenella, 2000, p.172)

A seguir vamos conhecer acerca dos tipos de Diabetes Mellitus e entender como essa doença também pode se associar a obesidade assim como a hipertensão.

2.4.2. DIABETES MELLITUS TIPO 1 e TIPO 2

Com base em dados divulgados pela OMS (2009) estimou-se que 180 milhões de indivíduos ao redor do mundo tem diabetes e que a estimativa para o ano de 2030 é que esse valor dobre (WHO, 2009 apud LUFT, V.,2010).

Os registros de pessoas com Diabetes Mellitus (DM) no mundo por regiões em 2017 podem retratar a atualidade em relação a esta doença. Na Europa a prevalência é de 58 milhões tendo uma estimativa de crescimento para o ano de 2045 de 16% chegando a atingir 67 milhões acometidos pela doença. Na tabela 03 contém informações para as outras regiões do mundo. Este estudo foi baseado em indivíduos da idade entre os 20 e 79 anos (Atlas IDF, 2017).

Tabela 03: Nº de pessoas com diabetes no mundo por região 2017 e em 2045.

Região	Prevalência de DM 2017 (em milhões)	Incremento (%)	Prevalência de DM 2045 (em milhões)
Europa	58	16	67
Norte da África e Oriente Médio	39	72	67
América do Norte e Caribe	46	35	62
América do Sul e Central	26	62	42
África	16	156	41
Sudeste Asiático	82	84	151
Pacífico	159	15	183

Fonte: Atlas IDF 2017 – Diabetes no Brasil

Dados mundiais registam 425 milhões de pessoas portadoras de DM em 2017 estimando um crescimento de 45% chegando à marca de 629 milhões de pessoas em 2045. A população mundial apresenta 8% de prevalência de diabetes de 20-79 anos sendo que deste grupo a prevalência de idosos acima de 65 anos é

de 19%. O Brasil representa o 4º lugar no mundo na lista dos 10 países com maior número de pessoas diagnosticadas com diabetes. (Atlas IDF, 2017)

A Federação de Diabetes Internacional (2019) afirma que:

- 1 em cada 11 adultos (20-79 anos) tem diabetes (463 milhões de pessoas);
- 1 em cada 2 adultos com diabetes não é diagnosticado (232 milhões de pessoas);
- 1 em cada 5 pessoas com diabetes tem mais de 65 anos (136 milhões de pessoas);
- 10% dos gastos globais de saúde são gastos em diabetes (US \$ 760 bilhões);
- Mais de 1,1 milhão de crianças e adolescentes com menos de 20 anos tem diabetes tipo 1;
- 1 em cada 13 adultos (20-79 anos) tem tolerância a glicose diminuída (374 milhões de pessoas);

A DM é descrito como um distúrbio metabólico caracterizado pelo aumento da glicemia de forma persistente, por deficiência na produção de insulina pelo indivíduo, o que ocasiona complicações em um longo período da sua vida. (SBD, 2017-2018).

A DM faz parte de um grupo de doenças metabólicas que apresenta hiperglicemia e isso está associado a várias complicações e insuficiências de vários órgãos humanos tais como coração, cérebro, rins, olhos e nervos (Ministério da Saúde Brasileiro, 2006)

Vários estudos pelo mundo sobre o DM afirmam que esta doença é um importante e crescente problema de saúde pública, estando associada a complicações que diminuem a qualidade de vida e aumenta a mortalidade. A DM pode ser classificada do tipo 1 (DM1) e do tipo 2 (DM2), tendo como seus fatores causadores a genética, a biologia e o ambiente em que se vive (SBD, 2017-2018).

Na DM1 ocorre a destruição de uma célula denominada como beta, que possui um processo auto-imune, levando à condição de diminuição de insulina, fazendo-se necessário o indivíduo repor a dose necessária de insulina para evitar futuras complicações. Este tipo de diabetes é mais evidente em crianças e adolescentes na faixa de 10 a 14 anos (Ministério da Saúde Brasileiro, 2006).

A DM2 representa de 90% a 95% dos casos de DM e é comum em pessoas na quarta década de vida. Ela é considerada uma doença de forte herança familiar o que relaciona com fatores genéticos. A DM2 apresenta-se quando é detectada uma hiperglicemia de maneira constante e existe uma resistência dos tecidos à ação da insulina, aumentando a produção da glicose e várias outras consequências como por exemplo o aumento da lipólise, o aumento da absorção renal de glicose, que estão relacionados à célula beta que por sua vez está ligado ao pâncreas (SBD, 2017-2018).

Um outro principal motivo da incidência de DM2 motivado pelo excesso de peso a nível nacional se deu por conta da escolaridade e gênero do indivíduo em um estudo desenvolvido na Finlândia. (Mehta, N., Stenholm, S. et al., 2020)

A grande preocupação das autoridades é o fato da DM ser a principal causa das amputações feitas em pacientes, visto que isso causa um alto custo no sistema da saúde (Papellbaum, Moreira & Coutinho, 2004, apud Musselman et al., 2003).

Esta preocupação ocorre pela existência de uma percentagem significativa de indivíduos que desenvolvem a DM 2, na maioria dos casos como consequência da obesidade uma vez que a DM2 representa de 90 a 95% dos casos de DM como já citado anteriormente.

Elaine Braga Faustino (2011) em seus estudos declara que:

“A obesidade tem adquirido grande proporção e, conseqüentemente, causado mais problemas para a saúde da população, levando à descompensações diabéticas e impactando a qualidade de vida dos indivíduos”.

Uma vez que os índices de obesidade se comportam de forma ascendente no mundo é fato que conseqüentemente doenças a ela associadas se tornem preocupação. Estudos afirmam que existe uma relação do excesso de peso e da DM2, sendo estes índices no ano de 2000 a nível global entre 50% nos homens e 66% nas mulheres que apresentam DM2 que veio como consequência do excesso de peso. (Flor, 2015)

Esta relação fica ainda mais existente quando analisamos os dados e nota-se que a maioria dos indivíduos com DM2 são obesos. Logo é necessário um controle simultâneo das duas doenças.

Capítulo 3

3. TÉCNICAS ESTATÍSTICAS MAIS UTILIZADAS NO ESTUDO DE OBESIDADE.

Com o propósito de identificar as técnicas estatísticas mais utilizadas nos estudos de obesidade, realizou-se a pesquisa de publicações como artigos científicos, teses, dissertações ligadas ao tema de interesse na busca de identificar fatores de risco para obesidade. A tabela 04 apresenta diversas pesquisas realizadas de publicações que permitem avaliar quais as técnicas estatísticas mais usadas em estudos de obesidade que serviram de apoio na decisão do processo de análise dos dados propostos nesse trabalho. Ainda na tabela 04 registra-se o título da publicação, o ano, autores, técnicas estatísticas, software estatístico, assim como o endereço onde se encontram disponíveis na internet.

Tabela 04: Registo de artigos, dissertações e teses que apresentam técnicas estatísticas voltadas a área de epidemiologia incluindo obesidade.

Autor/Ano	Título	Técnicas Estatísticas Utilizadas	Endereço Web
Mariath, Aline et al.2007	Obesidade e fatores de risco para o desenvolvimento de doenças crônicas não transmissíveis entre usuários de unidade de alimentação e nutrição	Teste qui-quadrado / Razão de Chances/ Regressão Logística / Statistic	https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2007000400017
Pontes, Luciano et al.2008	Prevalência de obesidade e análise do risco do excesso de peso na incidência de hipertensão em assistidos pelo programa de saúde da família de Picos (PI)	Estatística descritiva / Odds Ratio / T Student / R de Person / SPSS	https://www.efdeportes.com/efd123/prevalencia-de-obesidade-e-analise-do-risco-do-excesso-de-peso.htm
Moraes, Suzana et al.2004	Prevalência de sobrepeso e obesidade e fatores associados em escolares de área urbana de Chilpancingo, Guerrero, México, 2004	Regressão Logística Multivariada / Analise Univariada/ Odds Ratio / Teste de Wald/ Dummy	https://www.scielosp.org/article/csp/2006.v22n6/1289-1301/

Szwarcwald, Celia et al. 1992	Os caminhos da Estatística e suas incursões pela Epidemiologia	Modelo de Regressão e Estatística de Risco	https://www.scielo.org/articulo/csp/1992.v8n1/5-21/
Silva, Thais 2017	Obesidade geral e abdominal e seus determinantes em uma população adulta residente na zona rural do município de Pelotas, RS.	Stata 14.0 / Análise Descritiva / Intervalo de Confiança/ Teste qui-quadrado / Teste T / Análise Multivariada / Regressão Logística	http://epidemioufpel.org.br/uploads/teses/Thais.pdf
Carolino, Elisabete et al. 2010	Acid Phosphatase, Some Genetic Polymorphism and obesity risk factors in adult women	Modelo de Regressão Ordinal/ Odds Ratio / SPSS 17.0	https://www.researchgate.net/publication/233988326_Acid_Phosphatase_Some_Genetic_Polymorphism_and_obesity_risk_factors_in_adult_women
Gigante, Denise et al. 1997	Prevalência de obesidade em adultos e seus fatores de risco	Análise Multivariada / SPSS / Intervalo de Confiança/Regressão Logística / Odds Ratio	https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89101997000300004
Laranjeiras, Natércia et al. 2019	Efeitos Da Intervenção Alimentar Em Adultos Com Excesso De Peso Ou Obesidade	Statistic / Análise Descritiva / Teste qui-quadrado / Teste T-Student /	http://www.scielo.mec.pt/scielo.php?script=sci_abstract&pid=S2183-59852019000100005#:~:text=Verificou%2Dse%2C%20uma%20diminui%C3%A7%C3%A3o%20das,e%20obesidade%20classe%20%20respetivamente.
Silva, Wladimir 2010	Prevalência de sobrepeso/obesidade e fatores associados em adultos no Brasil	Análise Descritiva / Análise Univariada / Análise Multivariada / Regressão de Poisson/SPSS	https://nucidh.ufsc.br/files/2011/09/dissertacao_wladimir.pdf
Magalhaes, Elma et al. 2015	Prevalência e fatores associados ao ganho de peso gestacional excessivo em unidades de saúde do sudoeste da Bahia	Stata 9.1 / Teste Wald / Razão de Verossimilhança / Teste de Bondade / Regressão de Poisson /Análise Bivariada e Univariada	https://www.scielo.org/articulo/rbepid/2015.v18n4/858-869/pt/
Silva, Fernanda et al. 2019	Fatores ambientais associados à obesidade em população adulta de um município brasileiro de médio porte	Análise descritiva / Regressão Logística Binária / Odds Ratio / Teste de Wald/ Stata 13.1	https://www.scielo.br/pdf/csp/v35n5/1678-4464-csp-35-05-e00119618.pdf
Veloso, Helma et al. 2010	Prevalência e fatores associados à obesidade abdominal e ao excesso de peso em adultos maranhenses	Análise de Regressão Logística Multinomial / Análise Descritiva / Análise Univariada / Análise Multivariável	https://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-790X2010000300004
Monteiro, Paulo et al. 2004	Fatores de risco sociais, familiares e comportamentais para	Análise Bivariada/Análise Multivariada / Modelo	https://www.scielo.org/articulo/rpsp/2004.v16n4/250-258/

	obesidade em adolescentes	de Regressão Logística / Odds Ratio / Stata.	
--	---------------------------	--	--

Fonte: Pesquisas em Rede de Internet.

Diante da pesquisa apresentada na tabela 04, fica evidente que quando se trata de estudos que envolvem situações epidêmicas, a exemplo do tema de nosso interesse, a obesidade, observa-se o uso constante de Análises Multivariadas com o uso de Modelo de Regressão Logística, uso de Odds Ratio, Teste Wald, como sendo técnicas que podem apontar melhores resultados no estudo, o que nos certifica do uso das mesmas para aplicação no banco de dados proposto para o desenvolvimento desse trabalho. Em relação aos softwares estatísticos o *SPSS (Statistical Package for the Social Sciences)* é bem visto para aplicação das técnicas, assim será usado também para o nosso desenvolvimento.

Capítulo 4

4. VISUALIZAÇÃO DA INFORMAÇÃO

4.1 VISUALIZAÇÃO DA INFORMAÇÃO E SUA IMPORTÂNCIA

Com o avanço tecnológico e o uso cada vez maior dos computadores, usuários que utilizam sistemas de computadores para coletar dados e informações, a cada ano que passa têm-se deparado com um grande volume de dados. Neste caso é necessário o uso de técnicas, ferramentas, caminhos capazes de auxiliar na organização dessa quantidade de dados e de informações que contribuem para qualquer área do conhecimento.

Uma das técnicas que se destaca no meio computacional e que auxilia melhor a análise de dados e a tomada de decisão é a Visualização da Informação. O grande destaque desta técnica é a capacidade de compactar um grande volume de dados e obter análises mais claras, situação que não seria possível quando se usa uma simples visualização de dados científicos, ou seja, a Visualização da Informação se comporta de forma mais eficiente em relação ao volume de dados.

Ao escolher usar a técnica de Visualização da Informação que procura representar dados através de uma representação gráfica, o principal objetivo é assumir a necessidade de projetar e simplificar o entendimento destes dados, trazendo a maior compreensão do usuário. Para se obter sucesso nesse tipo de técnica é necessário seguir alguns passos tais como conhecer o tipo de dados utilizado e associar a melhor representação gráfica que otimize a melhor forma de análise (VAZ, 2004).

Este ramo da visualização da informação como um campo de pesquisa tem apenas 10 anos, mas de forma rápida se tornou ampla pela necessidade e versatilidade de ser uma técnica de abrangência interdisciplinar, visto que estreitou relações da visualização científica e da modelação de simulação.

A Visualização da Informação vem do inglês *Information Visualization* que é também abreviado usando a sigla InfoVis. Esta técnica que teve um maior interesse e desenvolvimento nos anos de 1990, possibilita as relações entre as variáveis, a forma gráfica, proporcionando um entendimento mais rápido se comparado com a análise utilizada com base em tabelas e listas (LOPES, 2018).

O significado da palavra visualizar, é dado por tornar visual, convertendo algo abstrato em imagem mental real, ou também tornar algo visível mediante determinado recurso (DICIO, 2020).

Com base no significado acima e na literatura sobre a visualização da informação, vamos ver abaixo definições de alguns autores.

Ana Lopes (2018) cita várias definições com base em vários autores. A seguir veremos algumas delas.

“ Freitas, Chubachi, Luzzardi e Cava (2001), Nascimento e Ferreira (2005) e Manovich (2011) definem a Visualização como “uma forma de representação visual de dados abstratos que facilita o entendimento ou permite que se descubram novas informações”. Contudo Manovich (2011, p. 150) prolonga um pouco esta definição, apontando que o objetivo da Visualização da Informação é “descobrir a estrutura de um conjunto de dados” que não é conhecida *a priori*. ”

(Lopes, 2018: 3)

A Visualização da Informação também poderia ser uma técnica que transforma os dados quantitativos que possui uma representação numeral numa apresentação visual, transformando códigos ou números em figuras visuais (Lopes, 2018 apud Manovich, 2011).

Segundo Pires (2001), InfoVis é a representação visual de qualquer tipo de informação. Plainsant (2004) define como uma representação gráfica que descreve um grande número de informações extraído de um conjunto de dados. Card, Makinlay e Scheiderman (1999) define como representações visuais de dados abstratos, que é possível desenvolver junto aos computadores e que permite ampliar a cognição do ser. Outros autores como Freitas, Chubachi, Luzzardi e Cava

(2001) descrevem como algo visual ou visível que é capaz de representar conceitos abstratos em figuras reais ou imagens mentalmente visíveis (Lopes, 2018).

Com o acumular de dados e com a velocidade necessária em obter respostas sobre análises e interpretação de dados, a Visualização da Informação passa a ter um papel importante na sociedade na construção de imagens visuais através de sistemas de símbolos que disponibilizam formas mais eficientes para a cognição da visão humana, visto que a relação computador e indivíduo é um caminho sem volta e intensa (Ware, 2004).

Devido à importância da técnica de Visualização da Informação, é possível observar as vantagens de se usar esta técnica que se destaca como um potencial recurso para interpretação e análise dos dados, sendo citada por autores como Ware (2004), Lopes (2018), Nascimento e Ferreira (2005).

De acordo com o autor Cole Ware (2004) a visualização apresenta uma série de vantagens tais como:

- Capacidade de interpretar enormes quantidades de dados, estando isso disponível a visão humana diante do uso da técnica;
- Percepção mais rápida de propriedades que não seriam percebidas sem o uso da técnica tornando marcas e linhas muito mais visíveis.
- Percepção mais rápida de um novo padrão associando locais a suas características.
- Permitir visibilidade quando existir problemas com os próprios dados como a forma de coleta, erros e artefatos.
- Permitir o controle de qualidade dos dados;
- Facilitar a compreensão de curta e larga escala dos dados.

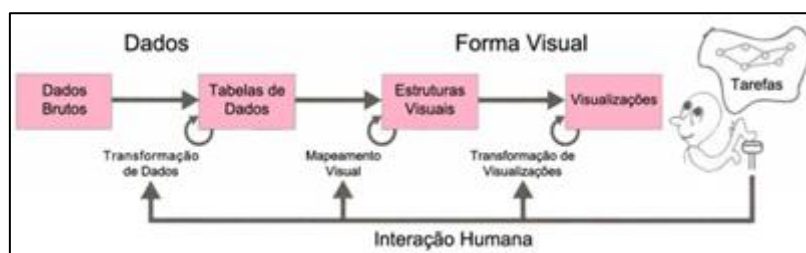
A autora Ana Lopes (2018) através da definição realizada por outros autores (Carde el at.,1999 e Fekete el at.,2008) lista várias características fornecidas pelo uso da representação visual que beneficia o poder na cognição humana.

- “ – Aumentar os recursos de memória e de processamento disponíveis;
- Reduzir a pesquisa por informação;
- Melhorar o reconhecimento de padrões através de representações visuais;
- Permitir a possibilidade de operações de inferência perceptiva;
- Utilizar mecanismos de atenção perceptiva para monitoramento de grandes eventos potenciais;
- Codificar informação num meio manipulável. ”

(Lopes, 2018: 5)

O processo de Visualização da Informação passa por etapas até que de fato a informação esteja pronta para gerar uma tomada de decisão. Na literatura há vários autores que demonstram de forma similar essas etapas.

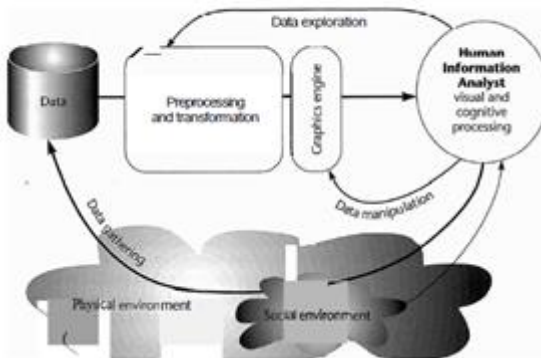
Ana Lopes (2018) através da proposta de Card, Mackinlay e Scheiderman (1999) prevê três etapas para visualização da informação, ou seja, efetivamente partindo dos dados brutos até a representação gráfica em si. A representação da autora pode ser vista na figura 04.



Fonte: (Lopes, 2018:84)

Figura 04: Esquema de processo de visualização. Adaptado de Card, Mackinlay e Scheiderman (1999)

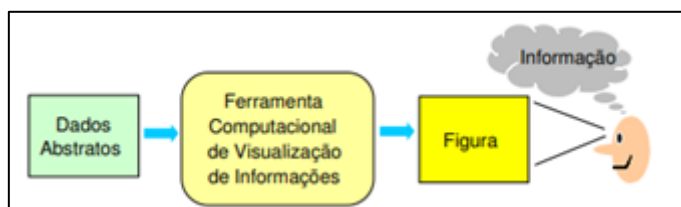
Para Colin Ware (2004) a visualização dos dados passa por quatro etapas sendo elas a coleta e armazenamento, a organização, o uso do hardware de exibição associado ao algoritmo gráfico e o sistema perceptivo de cognição do indivíduo. Esse processo pode ser visto na figura 05 abaixo.



Fonte:(Ware, 2004:4)

Figura 05: O Diagrama esquemático do processo de visualização.

Já os autores Nascimento e Ferreira (2005) descrevem as etapas na figura 06 de maneira bem compacta ou simplificada. Os mesmos justificam que essa representação poderia até ser chamada de visualização de dados, mas o foco é chamar de “Visualização de Informação” visto que a necessidade é obter informação acerca da relação entre os dados e não simplesmente os valores expostos em um gráfico propriamente dito.



Fonte: (Lopes, 2018:83 apud Nascimento e Ferreira, 2005).

Figura 06: Processo simplificado de visualização da informação auxiliada por computador.

A relação que existe entre a Visualização da Informação e a cognição visual humana, é utilizada para definir um modelo otimizado de modo a transformar as estruturas de dados em estruturas visuais, a esse processo chama-se de mapeamento visual. O grande desafio para os designers, é exatamente encontrar a visualização estratégica para o tipo de dados que se tem disponível.

O Mapeamento Visual é uma das etapas do processo de visualização posterior à transformação dos dados em tabelas. Esta fase da construção da estrutura virtual é definida em três partes sendo elas, o substrato visual, símbolos gráficos (marcas visuais) e propriedades gráficas. (Nascimento ,2011).

O fato é que mapear não é uma tarefa tão fácil para o design como já foi citado anteriormente, pois é possível que a escolha possa gerar resultados inesperados. Existem vários modelos propostos na literatura que são utilizados como estratégias para definir o mapeamento. Estes modelos são conhecidos como o *Blind Matching Procedure*, *Renaissance teams*, *Bottom-upGenerate-and-test*, *Top-down*, *Code based reasoning*. (Carvalho & Marcos, 2009)

Embora existam todos estes modelos que auxiliam na modelagem das formas gráficas, outras ferramentas computacionais têm sido desenvolvidas, com o intuito de realizar esta etapa do mapeamento de forma automática, analisando fatores como domínio, tipos de dados, dimensão e semântica auxiliando a escolha da melhor representação (Carvalho & Marcos, 2009).

O mapeamento visual dos dados nada mais é do que definir de que forma cada atributo será representado. No geral são considerados três tipos de atributos: nominais, ordinais e quantitativos. Os nominais são um conjunto de dados qualitativos sem uma ordem determinada, os ordinais são um conjunto de dados qualitativos que necessariamente implicam uma ordem e os quantitativos são elementos de um conjunto de dados numéricos que permitem fazer operações matemáticas (Silva, 2007).

As propriedades gráficas têm referências como sendo as marcas visíveis que compõem a marca visual e são representadas por propriedades muito utilizadas, tais como a posição da marca, a cor, o tamanho, a área ou o volume, entre outros fatores que apoiam a construção da representação gráfica de forma dinâmica necessária para chamar mais atenção do usuário dentro do substrato determinado. Essas propriedades incluem velocidade, animação, movimento, direção e sentido. As marcas visuais são figuras geométricas que servem para representar itens dos

dados ou uma posição sendo elas as linhas, figuras geométricas (área e volume) e figuras complexas chamadas de ícones (Nascimento,2011).

A seguir vamos conhecer as técnicas da Visualização da Informação e entender melhor de que maneira os dados através dessas técnicas se transformam em representações visíveis.

4.2 TÉCNICAS DE VISUALIZAÇÃO DA INFORMAÇÃO

A Visualização da Informação possui muitas técnicas e métodos que são aplicadas em ferramentas on-line e que representam muito bem um grande volume de dados. Esta técnica procura sempre uma forma de representação gráfica mais adequada e que seja eficiente no auxílio da interpretação de dados pelo leitor.

O autor André Grégio et al. 2009, descreve que:

“Existem muitas técnicas de visualização de dados, desde as mais simples e genéricas como gráficos simples de área, torta ou pizza, linhas, histogramas, pontos, etc. (que estão implementadas praticamente em todas planilhas eletrônicas) até as mais complexas e específicas como fatiamento (*slicing*) de volumes em três dimensões para apresentação de imagens bi-dimensionais, como as usadas para visualização de imagens tri-dimensionais como as de ressonância magnética nuclear”.

(Grégio et al. 2009:3)

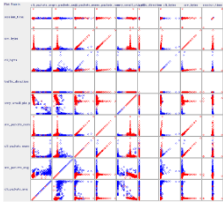
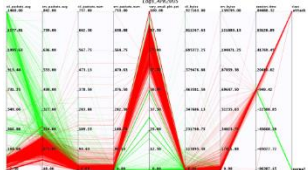
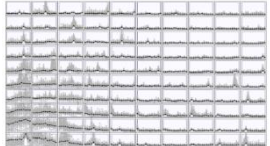
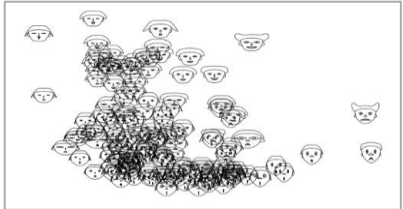
A Visualização da Informação aliada á computação gráfica permite e facilita a interação do indivíduo com a informação. Vários autores afirmam a sensibilidade na aplicação da técnica visto que é necessário direcionar determinada técnica a determinado tipo de dados para termos resultados satisfatórios, isso está ligado diretamente ao conjunto de dados. As técnicas podem ser de um simples gráfico que inclui histograma até mesmo representações complexas envolvendo acima de duas dimensões. (Freitas, Chubachi, Luzzardi & Cava, 2001 apud Lopes, 2018).

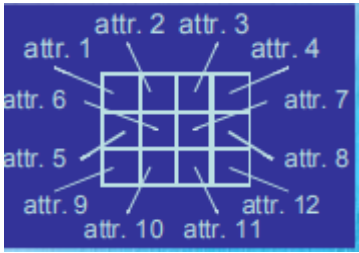

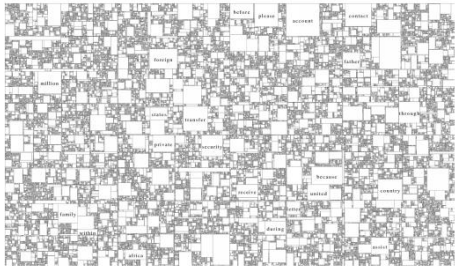
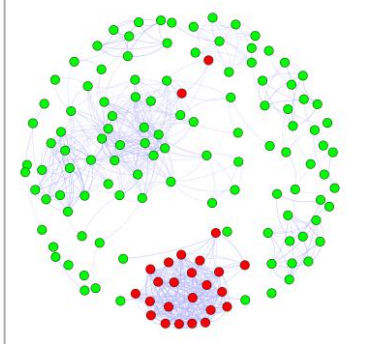
Atualmente a Visualização da Informação tornou-se uma ferramenta também ligada a várias áreas do saber, tais como Estatística, Medicina, Biologia,

Geografia, Bioquímica, Ambiental, com o objetivo de buscar soluções e apoiar a tomada de decisão.

Muitas são as técnicas que veem sendo utilizadas e incorporadas através de software de interfaces facilitadores para o usuário. Devido à quantidade extensa de técnicas de Visualização da Informação, e da sua taxonomia, apresenta-se na tabela 05 algumas delas, representando as mais utilizadas. O foco principal do uso destas técnicas é identificar a opção mais adequada para manipular os dados, atingir o objetivo e facilitar a compreensão dos resultados. No próximo item serão abordados os sistemas computacionais (software) utilizados para aplicar estas técnicas.

Tabela 05: Exemplo de algumas técnicas de Visualização de Informação.

Técnica	Mais utilizadas	Representação Gráfica
<p>Técnicas geométricas</p> <p>Permite a visualização dos dados através de transformações geométricas (ex. reorganizações, projeções) dos valores dos seus atributos</p>	<p>Matriz de espalhamento (<i>scatterplot matrix</i>)</p>	
	<p>Coordenadas Paralelas</p>	
	<p>Mapa Auto-Organizável de Kohonen (abreviadamente SOM, <i>Self-Organizing Map</i>)</p>	
<p>Técnicas baseadas em ícones</p> <p>Os dados multidimensionais a ser plotados como ícones cujas características</p>	<p><i>Chernoff faces</i></p>	

<p>correspondem a valores dos atributos dos dados</p>	<p>Codificação por formas (<i>shape coding</i>)</p>	
<p>Técnicas baseadas em pixels</p> <p>São similares às que usam ícones porque usam pequenos conjuntos em um arranjo espacial para representar os valores multidimensionais, e organizam estes conjuntos em arranjos maiores que podem representar as dimensões espaciais e/ou temporal do conjunto de dados</p>	<p><i>Pixel</i></p>	
<p>Técnicas hierárquicas</p> <p>Particionam as múltiplas dimensões dos dados de forma hierárquica em subespaços que podem ser visualizados (em 2 ou 3 dimensões)</p>	<p><i>Treemap</i></p>	
<p>Técnicas baseadas em grafos</p> <p>Servem para mostrar relações (arestas) entre objetos (vértices).</p>	<p><i>Clusters visíveis</i></p>	

<p>Técnicas tridimensionais</p> <p>Usam gráficos tridimensionais para visualização onde os dados são organizados em tipos de cenários, sendo muito mais efetivos na forma eletrônica (em <i>displays</i>) do que na forma impressa.</p>	<p>Paredes de perspectiva</p>	
--	-------------------------------	---

Fonte: Tabela criada pela autora (GRÉGIO et al., 2009)

4.3 FERRAMENTAS COMPUTACIONAIS DE VISUALIZAÇÃO DA INFORMAÇÃO.

É notório que o grande avanço na Visualização da Informação veio através do uso de computadores, em busca de soluções para uma grande massa de volume de dados. O uso da tecnologia através dos sistemas e software computacionais, contribuiu para atender à necessidade e o objetivo de seus usuários.

Diante da variedade de ferramentas torna-se uma tarefa difícil escolher a mais apropriada para seu banco de dados. A área de Visualização da Informação em relação a sistema divide-se em quatro categorias: i) Ferramentas de *Business Intelligence* (BI); ii) Ferramentas Analíticas; iii) Ferramentas de Visualização; iv) Ferramentas para trabalhos personalizados.

As ferramentas de Visualização de Informação são diversas, e por isso vamos relacionar neste trabalho as cinco mais utilizadas e citadas pelos autores em geral. As ferramentas *QlikView*, *Power BI*, *Google Data Studio*, *MicroStrategy* e *Tableau* são as mais utilizadas e encontram-se disponíveis no mercado atual.

QlikView é uma ferramenta de característica personalizável, tendo recursos que fazem a diferença na construção do visual, exigindo do usuário um tempo maior de dedicação para explorar todos os seus recursos. Geralmente é utilizada por usuários mais dedicados e que têm necessidades específicas ao ponto de personalizar para alcançar seus objetivos.

Power BI é um produto da *Microsoft Corporation*, que possui uma interface muito simples que atrai o público em geral. Este sistema traz soluções em análise de dados e visualizações em questão de minutos, o mesmo é usado pela área de negócios de organização do mundo inteiro. Um recurso muito usado no *Power BI* é o “arrastar e soltar” que permite criar, compartilhar e visualizar dados.

Google Data Studio é a mais nova ferramenta de análise do Google, de fácil uso, contendo gráficos apelativos visualmente que busca solucionar problemas e compartilhar com documentos com a mesma facilidade com é realizado. A versão dele é totalmente web, não tendo possibilidade de desktop como outros, mas permite ser editado e visualizado em ambiente online quando o documento criado é enviado a outro usuário.

MicroStrategy é um software considerado rápido, simples e fácil de baixar nos computadores. A permissão para conectar vários tipos distintos de banco de dados facilita muito para os usuários em geral, sejam eles online ou do excel podendo ainda combinar e associar as fontes diversas. As visualizações são simples, mas permitem também a criação de gráficos arrojados.

Tableau é a ferramenta mais utilizada atualmente, pois permite uma análise de dados em tempo real trazendo facilidade e flexibilidade para o usuário. Muito utilizado nos negócios, os gráficos são atrativos, os painéis interativos e de respostas rápidas. Ele é considerado o mais rápido sistema em relação às outras ferramentas por apresentar uma conectividade com inúmeras fontes de dados e ser muito intuitivo. O uso do *Tableau* não exige muito conhecimento prévio dos indivíduos, sendo um dos pontos fortes para de ser considerado um “queridinho” para análise e visualização de dados.

Neste trabalho a proposta para o desenvolvimento de representação gráfica com o banco de dados e através das ferramentas *Power BI* e *Tableau*, selecionadas por serem ferramentas muito amigáveis, mais utilizadas, além de fazerem parte do rol de plataformas de acesso gratuito. Sendo assim no capítulo 5 teremos de forma mais detalhada conhecimentos sobre as mesmas.

Capítulo 5

5. MODELO ESTATÍSTICO E MODELO VISUAL PROPOSTO

5.1 Modelo Estatístico

Muitos são os indicadores que interessa estudar na área da saúde tais como a mortalidade, morbidade, indicadores nutricionais, demográficos, sociais e ambientais. A estatística é uma ferramenta que apoia, traz resposta e auxilia na descrição de grandes volumes de dados que são gerados pelo mundo.

Atualmente o uso de métodos estatísticos pela comunidade dos profissionais da área da saúde, principalmente da área médica, vem crescendo e se destacando. A estatística aplicada em estudos na área de saúde, potencializa o conhecer e traçar o perfil de doenças que são de grande interesse e importância na sociedade. Este tipo de estudo permite realizar análise de dados, que nem sempre é tão simples para o público profissional da área médica, mas abre campo para melhores interpretações e decisões que potenciam as formas de tratamento.

No presente estudo interessa identificar alguns fatores de risco associados à obesidade. Para isso buscar-se-á nas referências quais os métodos mais utilizados pelos profissionais da saúde para encontrar a resposta desejada.

Segundo Ângela Paes (2010), “Os estudos da área da saúde, têm como objetivo identificar fatores associados a determinado desfecho (resposta) de interesse, frequentemente realizam “análises univariadas”, seguida por uma “multivariada” ”.

Uma análise univariada nos estudos estatísticos é utilizada quando temos uma variável resposta e uma variável independente, e é também chamada de multivariada quando a análise envolve mais do que uma variável independente.

Normalmente os estudos iniciam com uma análise univariada e progride para uma análise multivariada.

Ângela Paes (2010) ainda afirma que na área da saúde quando o interesse é investigar uma série de fatores, características ou eventos que influenciam as doenças, os métodos estatísticos mais utilizados são os modelos de regressão, que nada mais é que um modelo matemático que estabelece a relação entre a variável resposta e os fatores que interessa investigar.

Os modelos matemáticos são uma forma de interpretar o mundo e de fazer um recorte da realidade, auxiliando a interpretação de situações cotidianas e complexas. Um modelo é julgado bom se ele for o mais simples possível, tendo o cuidado para não ser inadequado (Montenegro, 2009).

Em vários estudos realizados na área de saúde que envolvem o interesse em buscar uma variável resposta para vários fatores envolvidos, observa-se o uso do chamado Modelo de Regressão, isso porque o objetivo é descobrir qual a relação existente entre essas variáveis independentes (fatores) com a variável resposta.

Na literatura temos uma ampla variedade de métodos estatísticos que são usados para realizar a modelagem de modelos matemáticos. Dentro dos métodos estatísticos diversos são os Modelos de Regressão cujo objetivo é analisar a relação da variável resposta com a variável independente. Na área de saúde depois de uma vasta pesquisa fica evidente o uso constante do modelo de Regressão Logística quando se trata de variável resposta com apenas duas alternativas.

Para uma maior compreensão, é possível citar, por exemplo quando nos referimos a modelos cuja variável resposta é dicotômica (0 – sem o diagnóstico, 1 – com o diagnóstico), pode-se usar o modelo normal linear, modelo *logit*, modelo normal censurado, modelo normal não-linear, modelo baseado na distribuição beta, na distribuição simplex ou um modelo de quase-verossimilhança.

Diante da variedade de modelos, por fim as literaturas recomendam para a situação de variável resposta dicotômica (0,1) o modelo de quase-verossimilhança, visto que é o mais adequado, ou seja, aquele que mais se aproxima de situações reais. (Montenegro, 2009)

Neste trabalho será desenvolvido exatamente um estudo transversal onde temos a variável resposta em duas alternativas (0 – Não obeso, 1 – Obeso) e será usado método da Regressão Logística Binária em busca de respostas.

Para avaliar a associação ou independência existente entre as variáveis propostas neste estudo foi utilizado o Teste de Qui-quadrado devido à sua grande aplicabilidade na área da saúde conforme pesquisa registada na tabela 04 apresentada no capítulo 3.

O Teste Qui-quadrado é utilizado na comparação de variáveis nominais, ou variáveis nominais versus ordinais que são possíveis separar por categorias. Segundo Jairo Fonseca e Gilberto Martins (2001), as frequências observadas são definidas através das tabelas de contingência de dupla entrada. As frequências esperadas são baseadas na definição de variáveis aleatórias independentes, ou seja, considera-se X e Y independentes se a distribuição conjunta de (X e Y) é igual ao produto das distribuições marginais de X e Y.

A estatística de teste Qui-quadrado é definida como:

$$x^2 = \sum_{i=1}^L \cdot \sum_{j=1}^C \frac{(Fo_{ij} - Fe_{ij})^2}{Fe_{ij}} \cap \chi^2_{(n^{\circ} \text{ linhas}-1)(n^{\circ} \text{ colunas}-1)} \quad (5.1)$$

Onde,

$$Fe_{ij} = \frac{(\text{soma da linha } i)(\text{soma da coluna } j)}{\text{total de observações}} \quad (5.2)$$

As hipóteses associadas ao teste Qui-quadrado são:

$$\left\{ \begin{array}{l} H_0: \text{As variáveis não estão associadas, ou são independentes} \\ H_1: \text{As variáveis estão associadas, ou são dependentes.} \end{array} \right. \quad (5.3)$$

Com base no cálculo do χ^2 na equação (5.1), na fixação do α (nível de significância) e com o auxílio da tabela χ^2 pode-se definir pela rejeição ou não da hipótese nula, ou seja , se χ^2 calculado na equação for menor que o quantil tabelado da distribuição χ^2 com $(n^\circ \text{ linhas}-1) \times (n^\circ \text{ colunas}-1)$ graus de liberdade e probabilidade $(1-\alpha)$, não se pode rejeitar H_0 , caso contrário rejeita-se H_0 . A tomada de decisão sobre H_0 , pode ser realizada também com reuso ao valor p , onde $p = P(\chi^2_{(n^\circ \text{ linhas}-1)(n^\circ \text{ colunas}-1)} \geq \chi^2)$ e se $p \leq \alpha$ então conclui-se pela rejeição de H_0 , caso contrário não se rejeita H_0 .

Ainda segundo os autores Jairo Fonseca e Gilberto Martins (2001), o Teste Qui-quadrado deve ser aplicado para amostras grandes e deve-se ter cuidado com as frequências esperadas menores que 5, pois isso traz impacto nos resultados. O indicado é juntar classes evitando uma frequência esperada menor que 5. O Teste Qui-quadrado é validado por esses pressupostos. A medida que os pressupostos não são atendidos uma das alternativas a opção é realizar o estudo dessa relação através do Teste Qui-quadrado por simulação de Monte Carlo.

As variáveis que serão submetidas nesse estudo em sua maioria são qualitativas, com exceção das variáveis idade e atividade da fosfatase ácida. Para avaliar a associação existente entre a variável resposta obesidade (qualitativa nominal e dicotômica), a idade e a atividade fosfatase ácida (quantitativas contínuas medidas em escala métrica) optou-se pelo uso do coeficiente de associação ETA através do software estatístico SPSS 22.0.

O coeficiente de correlação ETA é aplicado para avaliar o grau de relação existente entre uma variável quantitativa e outra variável categórica, cuja a resposta do coeficiente é sempre um valor no intervalo de 0 a 1 (Lira, 2004 apud Silveira, 1999).

Segundo a autora Lira (2004) o estimador do coeficiente de correlação ETA ($\eta_{Y,X}^2$) é a raiz quadrada da razão da soma de quadrados entre os grupos e a soma de quadrados total. Assim sendo a correlação ETA ao quadrado ($\eta_{Y,X}^2$)² é a razão entre a soma dos quadrados entre os grupos e a soma de quadrados total.

Assumindo valores no intervalo ($0 \leq \eta \leq 1$), interpreta-se uma maior correlação entre as variáveis qualitativa e quantitativa a medida que o valor de ($\eta_{Y,X}^2$)² se aproxima de 1.

Em busca do objetivo e conclusão desse estudo será explorado o uso da Regressão Logística Binária com todas as suas análises propostas através do SPSS, após realização de análise univariada e multivariada por apresentarem mais consistência em seus resultados em estudos semelhantes apresentados na literatura.

5.2 REGRESSÃO LOGÍSTICA BINÁRIA

5.2.1 INTRODUÇÃO

A Regressão Logística Binária é usada em casos em que a variável resposta é dicotômica e pretende-se estimar a presença ou ausência de determinada característica com base num conjunto de variáveis preditoras. Neste trabalho, pretende-se determinar a presença ou ausência de patologia (obesidade). Uma das funções do modelo de regressão logística binária é a determinação do *odds ratio* para cada fator estudado, possibilitando identificar os fatores que mais contribuem para ausência ou presença da patologia em estudo.

O modelo de regressão logística binária é um caso particular dos modelos lineares generalizados, neste caso ligado ao modelo *logit*. Para analisar uma função $\pi(x)$, considerando as variáveis independentes ($x_1, x_2, x_3, \dots, x_n$), e considerando $\pi(x)$ uma função de distribuição de probabilidade, da qual assume valores

respostas como 0 e 1, quando x varia na reta real é necessário realizar uma transformação logística, também chamada de *logit* para que seja possível uma representação gráfica (Liska, 2012).

$$\text{logit} = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1x_1 + \dots + \beta_nx_n \quad (5.4)$$

À razão $\left(\frac{\pi(x)}{1-\pi(x)}\right)$ chamamos de chance (*odds*). Se considerar A como um evento de interesse, define-se a razão de chance como o quociente entre a probabilidade de ocorrência de A e a probabilidade de ocorrência de não A.

Aplicando a função exponencial na expressão (5.4), temos que

$$e^{\text{logit}} = e^{\beta_0 + \beta_1x_1 + \dots + \beta_nx_n} \quad (5.5)$$

$$\frac{\pi(x)}{1-\pi(x)} = e^{\beta_0 + \beta_1x_1 + \dots + \beta_nx_n} \quad (5.6)$$

Por fim temos que o inverso desta função *logit* (5.6), é a função logística, o que nos resulta em,

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1x_1 + \dots + \beta_nx_n)}{1 + \exp(\beta_0 + \beta_1x_1 + \dots + \beta_nx_n)} \quad (5.7)$$

Onde $\pi(x)$ assume valores de (0,1), considerando uma variável resposta dicotômica, podendo ter o número de variáveis independentes necessárias para avaliar um diagnóstico. (Liska, 2012)

Para determinar a equação do modelo de regressão logística é necessário estimar os seus parâmetros. Um dos métodos usados para estimar os parâmetros de um modelo estatístico é chamado de Máxima Verossimilhança.

5.2.2 ESTIMAÇÃO DOS PARÂMETROS DO MODELO DE REGRESSÃO LOGÍSTICA BINÁRIA

Considerando o caso mais simples, isto é, o caso em que se tem uma só variável independente, para realizar o ajuste do modelo de regressão se faz necessário encontrar o vetor de parâmetros $\beta = [\beta_0, \beta_1]$. O método de Máxima Verossimilhança através do uso dos dados amostrais irá em busca de estimadores para esse modelo, ou seja, $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1]$, valores esses que maximizam o logaritmo da função de máxima verossimilhança. Esse tipo de aplicação permite encontrar estimadores para parâmetros que apresentam a maior probabilidade de repetir o padrão nas observações. (Queiroz, 2004)

A função de máxima verossimilhança é dada por

$$\lambda(\beta) = \prod_{i=1}^n f\left(\frac{Y_i}{X_i}\right) \quad (5.8)$$

Aplicando o Logaritmo na função (4.5), a expressão fica

$$L(\beta) = \ln[\lambda(\beta)] \quad (5.9)$$

$$L(\beta) = \ln \prod_{i=1}^n f\left(\frac{Y_i}{X_i}\right) \quad (5.10)$$

$$L(\beta) = \sum_{i=1}^n \{Y_i \ln[P(X_i)] + (1 - Y_i) \ln[1 - P(X_i)]\} \quad (5.11)$$

$$L(\beta) = \sum_{i=1}^n \left\{ Y_i \ln \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} + (1 - Y_i) \ln \left[1 - \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right] \right\} \quad (5.12)$$

$$L(\beta) = \sum_{i=1}^n \{Y_i(\beta_0 + \beta_1 X_i) - \ln(1 + e^{\beta_0 + \beta_1 X_i})\} \quad (5.13)$$

Fonte: (QUEIROZ, 2004)

Afim de encontrar β que maximize a função $L(\beta)$ deve-se derivar em relação a β igualando a zero. Assim chegamos às seguintes expressões,

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left(Y_i X_i \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right) \quad (5.14)$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left(Y_i X_i - \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right) \quad (5.15)$$

Fonte: (QUEIROZ, 2004)

Segundo Leandro Gonzales (2018), esses estimadores medem a taxa de variação do *logit* para uma unidade de variação na variável independente, representando a inclinação da linha de regressão entre a variável dependente e a sua variável independente.

Uma generalização de um modelo de regressão logística quando temos mais de uma variável independente, descreve-se com base na equação,

$$g(X) = \text{Logit}(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \ln\left[\frac{P(Y=1|X)}{P(Y=0|X)}\right] = X'\beta \quad (5.16)$$

$$X'\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (5.17)$$

Neste caso a função de verossimilhança, é semelhante à utilizada anteriormente em (5.8), e os seus parâmetros são estimados também através das derivadas parciais e igualando a zero. Assim obtemos os resultados,

$$L(\beta) = \sum_{i=1}^n [y_i \ln \pi_i + (1 - y_i) \ln(1 - \pi_i)] \quad (5.18)$$

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\pi}_i = 0 \quad (5.19)$$

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^n x_{ij} y_i - \sum_{i=1}^n x_{ij} \hat{\pi}_i = 0, \quad \text{para } j \in \{1, \dots, t\} \quad (5.20)$$

$\hat{\pi}_i$ é o estimador de π_i , usando a máxima verossimilhança (Gonzales, 2018 Apud Figueira, 2006).

5.2.3 AVALIAÇÃO DA QUALIDADE DO AJUSTE DO MODELO DE REGRESSÃO LOGÍSTICA.

Uma vez definido o modelo de regressão logística é necessário avaliar estes parâmetros e entender o quando eles são significativos ou não no modelo com o objetivo de termos diagnósticos mais precisos em relação à amostra em estudo.

Diversos autores citam como sendo as principais ferramentas de avaliação de um modelo, o teste da Razão de Verossimilhança, o Teste *Wald*, Pseudo R^2 de Cox e Snell, Pseudo R^2 *Nagelkerke*, Teste de *Hosmer-Lemeshow*, Teste de *Omnibus*, Curva ROC, Especificidade e Sensibilidade, Teste de concordância de *Kappa* e o diagnóstico dos outliers e casos influentes (Análise de Resíduos). Veremos a seguir cada um deles (Gonzales, 2018).

TESTE DE RAZÃO DE VEROSSIMILHANÇA

Este teste tem como objetivo testar simultaneamente se os coeficientes do modelo de regressão relacionados a β são todos nulos com exceção de β_0 .

O objetivo aplicando o teste e baseando-se na regressão logística é comparar os valores observados da variável resposta com os valores preditos usando o log da verossimilhança.

A função de verossimilhança que é utilizada para avaliar isto, e é realizada pela seguinte expressão,

$$D = -2 \ln \left[\frac{\text{Função de máxima verossimilhança do modelo corrente}}{\text{unção de máxima verossimilhança do modelo saturado}} \right] \quad (5.21)$$

$$D = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \quad (5.22)$$

Através dessa expressão deseja-se testar a hipótese,

$$H_0: \beta_1 = \dots = \beta_p = 0 \text{ V. s } H_1: \exists_{j=1, \dots, p}: \beta_j \neq 0 \quad (5.23)$$

Onde a estatística de teste é determinada por:

$G = D$ (modelo sem as p variáveis) – D (modelo com as p variáveis),

$$D = -2 \ln \left[\frac{\text{modelo corrente}}{\text{modelo saturado}} \right] \cap_{\text{sob } H_0} \chi^2_{(p-1)} \quad (5.24)$$

Com o uso da função de máxima verossimilhança surge a existência de modelos saturados e modelos correntes. Visto que inicia-se a identificação de variáveis que apresentam identificação com a variável resposta e outras não.

O modelo saturado é também conhecido como modelo completo, o mesmo tem p parâmetros que se igualam ao número n de observações ($p=n$). Este modelo também consegue atribuir toda a variação dos dados, tornando assim um modelo robusto, reproduzindo os próprios dados. Ele serve de base para validar o uso de um modelo corrente devido à sua robustez. (Marciano, 2009).

Por sua vez o modelo corrente é qualquer modelo sob investigação que apresenta $(n-1)$ parâmetros que será comparado ao modelo saturado.

Ao rejeitar-se essa hipótese nula, afirma-se que pelo menos um dos coeficientes é diferente de zero. Este teste ainda não é suficiente para avaliar cada coeficiente individualmente, sendo necessário aplicar o Teste de *Wald* que veremos em seguida.

TESTE DE *WALD*

O Teste de *Wald* será útil para avaliar a significância de forma individual de cada coeficiente estimado pelo modelo de regressão logística binária e avaliar se cada coeficiente é diferente de zero. Este teste analisa se a variável dependente tem relação significativa com uma variável independente de interesse (Cabral, 2013).

Neste caso o interesse é testar,

$$H_0: \beta_j = 0 \quad V.S. \quad H_1: \beta_j \neq 0, j = 0, \dots, p \quad (5.25)$$

A estatística de teste, sob as condições da hipótese nula, é,

$$W_j = \frac{\hat{\beta}_j}{\text{var}(\hat{\beta}_j)} \cap_{\text{sob } H_0} \chi^2_{(1)} \quad (5.26)$$

Ao se rejeitar a hipótese nula, afirma-se que existe uma relação entre a variável dependente e independente, pois o coeficiente da variável submetida ao teste é significativamente diferente de zero.

Um intervalo de confiança também pode ser definido para os parâmetros baseado na estatística de Wald de $100 \times (1-\alpha) \%$ sendo para o β_1 ,

$$IC(\beta_1, 1 - \alpha) = [\hat{\beta}_1 - z_{1-\alpha/2} DP(\hat{\beta}_1); \hat{\beta}_1 + z_{1-\alpha/2} DP(\hat{\beta}_1)] \quad (5.27)$$

E para a constante (intercepto) do modelo β_0 ,

$$IC(\beta_0, 1 - \alpha) = [\hat{\beta}_0 - z_{1-\alpha/2} DP(\hat{\beta}_0); \hat{\beta}_0 + z_{1-\alpha/2} DP(\hat{\beta}_0)] \quad (5.28)$$

Considerando a variabilidade respectiva a cada variável envolvida no modelo em estudo e o $Z_{1-\alpha/2}$ sendo o quantil Distribuição Normal padrão cuja probabilidade à sua esquerda é $1 - \alpha / 2$.

PSUDO R² DE COX E SNELL E R² NAGELKERKE

O Pseudo R² de Cox e Snell é assim chamado, visto que não existe em regressão logística uma estatística semelhante ao R² utilizado na regressão linear. O R² de Cox e Snell está numa escala similar ao modelo de regressão linear, variando de 0 a 1. Em relação à interpretação não podemos usar a mesma que usamos para a regressão linear, cujo valores próximos de 1 indicam um melhor ajuste do modelo. (Gonzales, 2018).

Vários autores apresentam diversas definições do Pseudo R², vamos usar a definição de Cox e Snell (1989) por ser bastante usada no meio computacional.

$$R_{CS}^2 = 1 - \left(\frac{L(\beta)_0}{L(\beta)_M} \right)^{\frac{2}{n}} \quad (5.29)$$

Sendo $L(\beta)_0$ e $L(\beta)_M$, os desvios para o modelo nulo (só contém o intercepto) e o modelo corrente (proposto) respectivamente.

Este valor de R^2 assume valores entre $[0,1]$ e quando o valor é menor que 1 indica a existência de uma relação entre as variáveis. Esta análise é interessante na comparação de modelos distintos, sendo possível avaliar qual das variáveis independentes contribuem mais para essa relação (Gonzales, 2018).

Segundo João Marôco (2011), o Pseudo R^2 de Nagelkerke é calculado como:

$$R_N^2 = \frac{R_{CS}^2}{1 - e^{-\frac{2LL_0}{n}}} \quad (5.30)$$

Este valor de R_N^2 pode assumir o valor 1, mas esses valores podem não ser de interpretação tão fácil como é na regressão linear.

TESTE DE *OMNIBUS*

O teste de *Omnibus* verifica se os coeficientes envolvidos no modelo de regressão são nulos. O teste baseia-se na estatística Qui-quadrada e no p-valor para analisar se pelo menos um dos coeficientes pode contribuir para qualidade desse modelo quando é rejeitada a hipótese nula. (Nardi, 2019)

As hipóteses subjacentes ao teste de *Omnibus* são:

$$\left\{ \begin{array}{l} H_0: \text{Todos os coeficientes do modelo são nulos.} \\ H_1: \text{Pelo menos um coeficiente do modelo não é nulo.} \end{array} \right. \quad (5.31)$$

TESTE DE *HOSMER E LEMESHOW*

Hosmer e Lemeshow (1980, 1982) propuseram um teste de dois tipos de agrupamentos baseado nas probabilidades estimadas. Estes grupos são descritos nos decis das probabilidades estimadas e na pré-definição de pontos de corte.

Para o cálculo da estatística de teste é necessário estimar as frequências esperadas, onde para $Y=1$ é realizada a soma de todas as estimativas dos indivíduos do grupo e para $Y=0$ é realizada a soma de uma unidade menos a soma de todas as estimativas dos indivíduos do grupo.

A estatística *Hosmer e Lemeshow* é calculada como \hat{C} , e tem distribuição Qui-quadrado com $g-2$ graus de liberdade.

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}, \quad (5.32)$$

Em que: n'_k é o número de indivíduos no k -ésimo grupo; $\bar{\pi}_k = \sum_{j=1}^{C_k} \frac{\pi_j}{n_k}$; C_k é o número total de combinações de níveis dentro do k -ésimo decil; $O_k = \sum_{j=1}^{C_k} y_j$ é número total de respostas dentro do grupo k .

As hipóteses subjacentes ao Teste *Hosmer e Lemeshow* são:

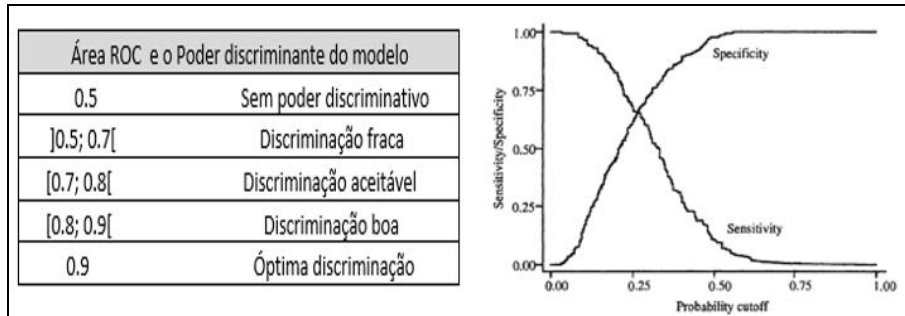
$$\left\{ \begin{array}{l} H_0: \text{O modelo ajusta-se aos dados} \\ H_1: \text{O modelo não se ajusta aos dados.} \end{array} \right. \quad (5.33)$$

CURVA ROC

Outro processo para avaliar a qualidade do modelo de regressão logística é a área da curva ROC (*Receiver Operating Characteristic*). A sensibilidade e a especificidade contam como um ponto de corte para determinar a positividade deste teste. Esta detecção é dada pela área abaixo da curva ROC. Este receptor capta a existência de um sinal na presença de ruído, ou seja, busca a probabilidade de se encontrar o sinal verdadeiro e sinal falso para um volume de dados para um ponto de corte determinado. A área tem uma variação de $[0,1]$ e auxilia a determinar os indivíduos que representam a característica de interesse.

Hosmer e Lemeshow (2000) e João Marôco (2011) trazem indicativos que podem servir de apoio para classificação da curva Roc dentro da variação da mesma. Na figura 07, segue descrito uma classificação do poder da área da curva Roc, assim como permite uma compreensão de um ponto de corte para análise,

verificando aproximadamente onde a sensibilidade e a especificidade se cruzam representada através da figura do gráfico.



Fontes: (Marôco, 2011: 829 & Hosmer e Lemeshow, 2000: 162)

Figura 07: A curva ROC e o poder discriminante do modelo.

ESPECIFICIDADE E SENSIBILIDADE

A sensibilidade e a especificidade permitem avaliar a eficiência classificatória do modelo em análise. Esses índices estão diretamente ligados ao percentual de classificados corretamente pelo modelo.

A classificação do modelo é calculada de duas maneiras: a classificação correta obtida pelo modelo e a classificação correta por acaso. Essas classificações podem ser definidas com base na tabela 06 a seguir (Marôco, 2011).

Tabela 06: Base para cálculo da classificação do modelo.

Característica Observada (Y)	Característica Estimada (\hat{Y})		
	0	1	Totais em linha
0	a	b	a+b
1	c	d	c+d
Totais em coluna	a+c	b+d	N

Fonte: Adaptado pela autora (Output SPSS).

A = % do total de casos corretamente classificado.

$$A = \frac{a+d}{N} \times 100 \quad (5.34)$$

B = % do total de casos corretamente classificado por acaso.

$$B = \left[\left(\frac{a+c}{N} \right)^2 + \left(\frac{b+d}{N} \right)^2 \right] \times 100 \quad (5.35)$$

Se a diferença entre estas medidas for superior a 25% considera-se que o modelo tem boas características classificatórias (Marôco, 2011).

A sensibilidade é calculada com base na classificação correta do modelo, da variável dependente, ou seja “1-sucesso” e a especificidade é a percentagem de classificação correta na característica “ 0-insucesso”. O que também pode ser definido abaixo como:

$$\text{Sensibilidade} = P[\hat{Y} = 1|Y = 1] \quad (5.36)$$

$$\text{Especificidade} = P[\hat{Y} = 0|Y = 0] \quad (5.37)$$

Fonte: (Marôco, 2011).

TESTE DE CONCORDÂNCIA KAPPA

O teste de concordância tem a finalidade de medir o grau de concordância entre as proporções derivadas de amostras dependentes. Definindo o cálculo do coeficiente Kappa o por:

$$\text{Kappa} = \frac{P(O) - P(E)}{1 - P(E)} \quad (5.38)$$

Considerando, P(O): proporção observada de concordâncias (soma das respostas concordantes dividida pelo total); P(E): proporção esperada de concordâncias (soma dos valores esperados das respostas concordantes dividida pelo total).

A interpretação para o coeficiente de Kappa segundo as autoras pode ser dada sob a visão de Landis e Koch apresentada na tabela 07. A medida pode variar de zero que representa nenhuma concordância e seu valor máximo 1, que significa total concordância

Tabela 07: Referência dos valores de Kappa e a avaliação de concordância com o modelo.

Valores de Kappa	Interpretação
<0	Ausência de concordância
0-0,19	Concordância pobre
0,20-0,39	Concordância leve
0,40-0,59	Concordância moderada
0,60-0,79	Concordância substantiva
0,80-1,00	Concordância quase perfeita

Fonte: (PAES & SILVA, 2012: 166)

Ainda segundo as autoras Rebeca Silva e Ângela Paes (2012) é possível realizar o teste de significância de Kappa, considerando as seguintes hipóteses para um teste monocaudal.

$$H_0: K = 0 \text{ vs } H_1: K > 0 \quad (5.39)$$

Sendo $K=0$ uma concordância nula, ou se acima de zero uma concordância acima do esperado pelo simples acaso.

ANÁLISE DE RESÍDUOS

Essa análise é importante para avaliar os *outlier* que podem comprometer o desempenho do nosso modelo. *Outlier* são casos influentes na estimação do modelo e devem ser cuidados para um melhor ajuste. Existem várias medidas aplicadas dentro da Regressão Logística para avaliar esses casos influentes que são os resíduos, *Leverage*, Distância de Cook e *DFbetas*. Neste estudo dar-se-á foco nos Resíduos de Person ou também chamado de resíduos estandardizados, e na Distância de Cook para avaliar o comportamento dos *outliers*. (Marôco, 2011).

Segundo o autor João Marôco, (2011), a definição de resíduo não estandardizado é a diferença entre o número de sucesso observados e o número de sucessos estimados pelo modelo conforme pode ser visto em (5.40).

$$e_j = y_j - \hat{y}_j = y_j - n_j \hat{\pi}_j \quad (5.40)$$

y_j = número de sucesso observado

\hat{y}_j = probabilidade de sucesso da observação j

n_j = número de observações j

$\hat{\pi}_j$ = probabilidade do sucesso estimado da observação j

Ainda segundo o autor João Marôco, (2011) o Resíduo de Person, também chamado de resíduo estandardizado pode ser calculado conforme a expressão (5.40) e com base no resíduo não estandardizado calculado abaixo:

$$e_j = \frac{e_j}{\sqrt{n_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \quad (5.41)$$

Esses resíduos para grandes amostras apresentam distribuição normal com média 0 e desvio padrão aproximadamente 1.

Para tornar esse desvio padrão constante e igual a 1, os resíduos estandardizados podem ser obtidos pela expressão (Marôco, 2011):

$$r_j = \frac{e'_j}{\sqrt{1 - h_j}} \quad (5.42)$$

r_j = Resíduo estandardizados com média 0 e desvio-padrão constante 1.

e'_j = Resíduo de Person

h_j = Valor da *Leverage*.

O valor de *Leverage* é utilizado para tornar os dados estandardizados uma vez que a variância sofre muita influência das observações que por sua vez também influenciam na estimação dos coeficientes do modelo.

$$e_j = n_j \hat{\pi}_j(x_j) [1 - \hat{\pi}_j(x_j)] x'_j (X' V X)^{-1} x'_j \quad (5.43)$$

Sendo $x'_j = [1, x_{1j}, x_{2j}, x_{3j}, \dots, x_{pj}]$ e X e V são matrizes das observações.

Outra medida que avalia a influência das observações na estimação dos coeficientes durante a construção do modelo é a Distância de Cook. Esta distância usa como base as medidas do Resíduo de Person e o valor de Leverage que pode ser calculado como apresentado na equação (Marôco, 2011).

$$DC_j = r_j^2 \frac{h_j}{(1 - h_j)} \quad (5.44)$$

5.2.4 ODDS RATIO

A Razão de chances (*Odds Ratio*) é muito utilizado na regressão logística pelo simples fato de existir uma relação entre a chance e o parâmetro estimado pelo modelo. Esta medida é muito útil nas análises relacionadas a epidemia, a mesma consegue retratar o quanto uma variável de interesse tem a possibilidade de estar ou não relacionada com a epidemia em estudo. A razão de chances é definida pela razão entre a chance de o acontecimento de interesse ocorrer ($Y=1$) nos indivíduos com $X=1$ e a chance de não ocorrer nos indivíduos com $X=0$.

A chance para quando o acontecimento de interesse não ocorre nos indivíduos $X=0$ é definida por,

$$\frac{\pi(0)}{1 - \pi(0)} \quad (5.45)$$

E para quando ocorre nos indivíduos $X=1$ é definida por,

$$\frac{\pi(1)}{1 - \pi(1)} \quad (5.46)$$

Assim, é possível analisar através desta estatística se a probabilidade de acontecimentos entre os dois tipos de indivíduos é a mesma.

As expressões que representam o acontecimento de interesse para as duas categorias distintas, são dadas por,

$$\pi(1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} \text{ e } \pi(0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \quad (5.47)$$

Logo, conseqüentemente a expressão que define o *Odds Ratio* é dada por,

$$OR = \frac{\pi(1)[1 - \pi(0)]}{\pi(0)[1 - \pi(1)]} = \exp(\beta_1), \quad (5.48)$$

tornando-se evidente a relação entre a estatística em questão e o modelo de regressão logística.

Na prática é possível determinar a razão de chances aplicando a função logarítmica obtendo uma estimativa de um parâmetro, sendo possível até determinar um intervalo de confiança para esta estimativa. (Gonçalves, 2013)

O intervalo de confiança exibe uma visão melhor da variabilidade amostral produzindo 100 x (1- α) % de confiança para taxa de chance.

Os limites inferior e superior do intervalo de confiança a 100 x (1 - α) % do *Odds Ratio* são definidos como,

$$IC(Odds Ratio, 1 - \alpha) = [e^{\beta_t}; e^{\beta_s}] \quad (5.49)$$

Denotado por β_t e β_s os limites do intervalo de confiança para β_j .

5.3 VISUALIZAÇÃO DA INFORMAÇÃO E SISTEMAS COMPUTACIONAIS

Com o crescimento dos *softwares* na área da Visualização da Informação, os mesmos podem garantir resultados com imagens mais bem definidas resultando melhores decisões em qualquer área do conhecimento. Por isso a proposta deste trabalho é usar esse *software* para visualização de dados para uma melhor interpretação.

Dando continuidade ao uso dos métodos selecionados para o desenvolvimento deste estudo, foram selecionadas duas ferramentas na área da Visualização da Informação que apontam boas referências no meio científico e mundial. A seguir abordar-se-á sobre o uso do *Tableau Public* e do *Uso do Power BI*.

5.3.1 O USO DO TABLEAU PUBLIC

O uso do *Tableau Public* tornou-se o mais popular do mundo, pelo fato de ser uma ferramenta de serviços gratuitos, que permite a qualquer usuário realizar publicação de dados interagindo com a internet. As visualizações construídas pelo *Tableau Public* podem ser colocadas em páginas Web, blogs, em mídias sociais e também realizar *download* (Ganchev, 2017).

Atualmente com a quantidade de dados relacionados à área de saúde, e com a necessidade de racionaliza-los, fica impossível realizar sem uma visualização gráfica. Naturalmente e por serem conhecidos a maioria dos usuários recorrem a realizarem listas, construções de tabelas e gráficos usando o pacote *office* da *Microsoft*, explorando o *Word*, *Excel*, *PowerPoint* e algumas ferramentas estatísticas disponíveis, o que pode por si só, ser uma visão muito limitada para uma tomada de decisão (Ganchev, 2017).

O *Tableau Public* é usado em plataforma *desktop* e serve para infográficos que pode ser usado de forma interativa, com gráficos, mapas, tabelas e filtros, compactando e ampliando a informação útil auxiliando na tomada de decisão. O mesmo pode ser visto através do endereço on-line <https://public.tableau.com/>.

Este *software* permite a construção de mapas interativos de pontos, bolhas ou áreas, assim como a criação de gráficos de barras, linhas, bolhas, barras empilhadas, linhas empilhadas, pizza, barras de Gantt, tree-maps.

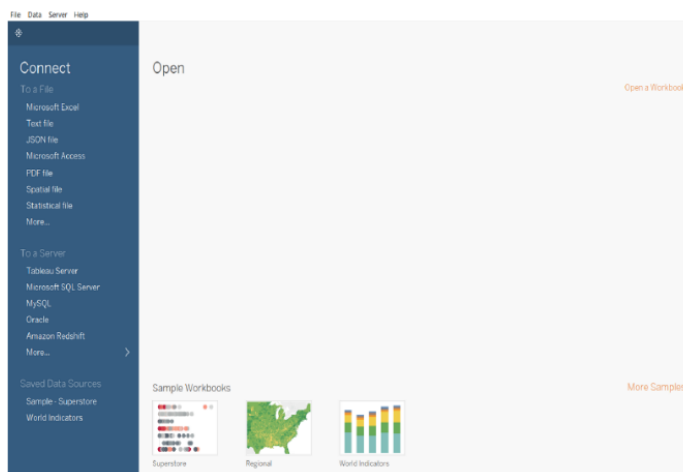
Conhecendo as vantagens desta ferramenta podemos citar a rapidez para criação de gráficos, facilidade na troca entre diversos tipos de visualizações, assim como, uma aplicação interativa, investigativa e na visualização de dados de grandes tabelas. O *Tableau Public* também suporta formato usado para armazenamento de dados geográficos assim como suporta conexão com o *Google*

Docs, realizando atualização em tempo real, caso seja feita uma atualização no banco de dados (Dados Finos, 2020).

As desvantagens partem do ponto de vista das limitações na gratuidade, na oferta de muitas opções demonstrando a impressão de um *software* complexo ao invés de simples, fazendo o usuário demorar mais na aprendizagem e tendo que personalizar bastante para ficar mais acessível. Com relação ao armazenamento a ferramenta se utiliza todo nas nuvens, dando acessibilidade ao público de todo mundo, sendo inadequado para dados empresariais que não podem tornar-se públicos (Dados Finos, 2020 & Sleeper, 2018).

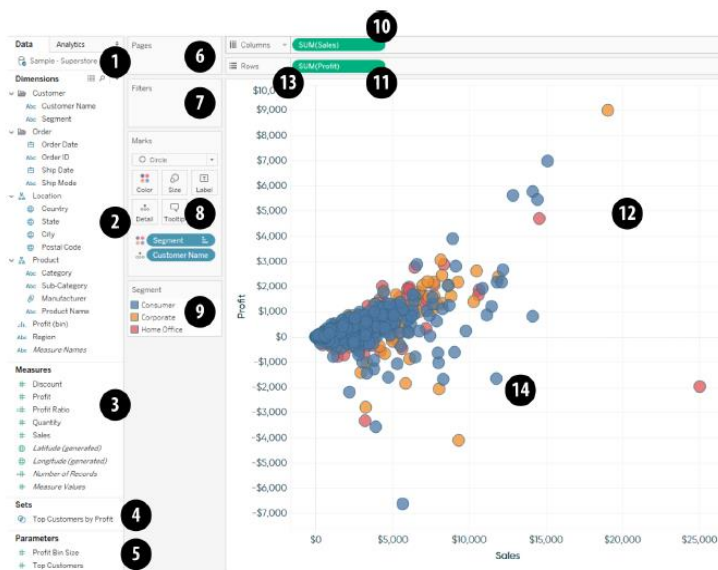
Seu sistema de entrada de dados se dá através de tabelas com extensões XLS, XLSX, CSV, arquivos de texto, TXT e JSON, *shapefiles*, SHP e PDF. No sistema de saída temos link para os gráficos interativos ou código de *embed*. (Dados Finos, 2020).

Ao abrir o sistema *on-line* vê-se uma tela com essa apresentação, onde deverá escolher a opção de conexão de dados apresentado na figura 08. Em seguida temos a tela depois de realizada a entrada de dados para iniciar as interações de acordo ao interesse do usuário na figura 09.



Fonte: *Tableau Software 2020*

Figura 08: Tela de abertura do *Tableau Public*.



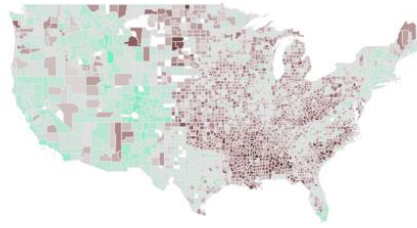
Fonte: *Tableau Software 2020*

Figura 09: Tela de interação do *Tableau Public* depois de adicionado os dados.

Para conhecimento das terminologias do *Tableau Public* vamos acompanhar a figura 9, com as chaves numeradas e descrever o que cada uma significa. Chave 1 – Fonte de dados, 2 – Área dimensões do painel de dados, 3 – Área de medidas do painel de dados, 4 – Define a área do painel de dados, 5 – Área de parâmetros do painel de dados, 6 – Prateleira de páginas, 7 – Prateleira de filtros, 8 – Cartão de Marcas, 9 – Legenda, 10 – Prateleira de colunas, 11 – Prateleira de Linhas, 12 – Planilha/Exibição, 13 – Pílula, 14 – Marca. (Sleeper, 2018)

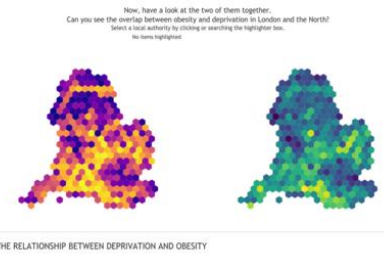
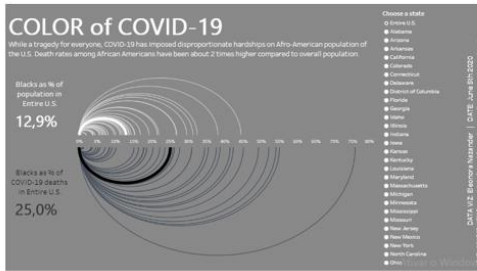
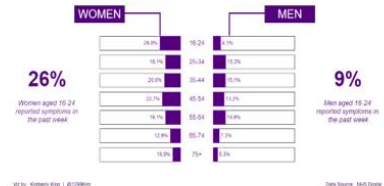
Vejam-se agora alguns gráficos produzidos no *Tableau Public* utilizados para solucionar problemas diversos em todo mundo inclusive na área da saúde tema de interesse deste estudo. Na figura 10 têm-se aplicações da produção visual pública na área da saúde e na figura 11 representações de produções gráficas em outras áreas do conhecimento.

The Geography of Social Capital in America: Obesity and Other Socioeconomic Factors



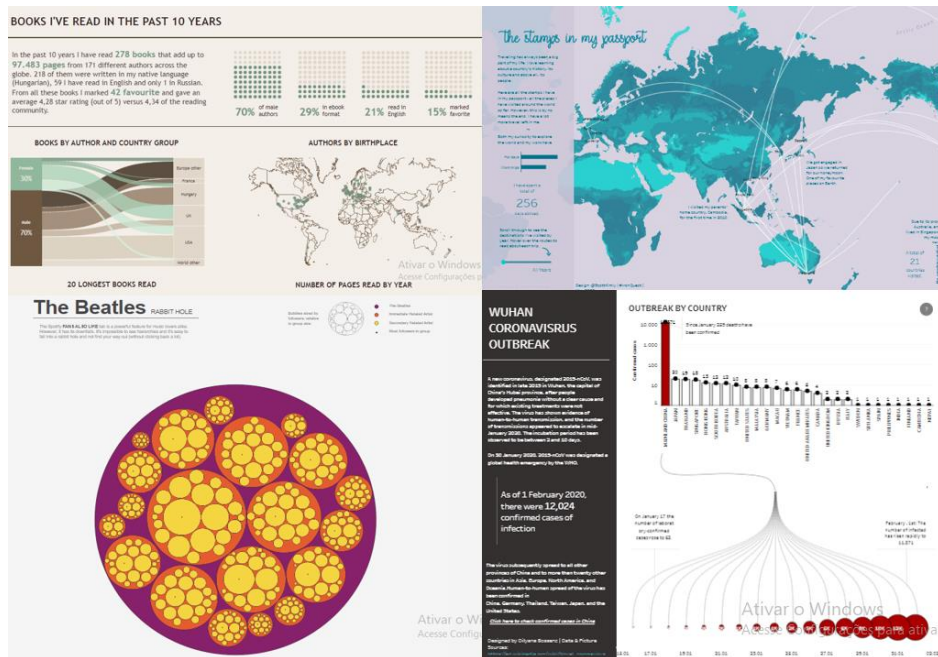
Mental Health in the UK

According to statistics body NHS Digital, at any one time, a sixth of the population in England aged 16 to 64 have a mental health problem. Within this survey, Women are more likely to report symptoms of a common mental illness over men.



Fonte: Tableau Software 2020

Figura 10: Produções gráficas no *Tableau Public* com estudos na área da saúde.



Fonte: Tableau Software 2020

Figura 11: Produções gráficas no *Tableau Public* em outras áreas de estudo.

5.3.2 USO DO POWER BI

O *Power BI* é uma ferramenta tão popular e perspicaz quanto a *Tableau Public*, esta ferramenta foi desenvolvida pela Microsoft em 24 de julho de 2015. A Microsoft disponibiliza essa plataforma on-line ou em desktop ficando à escolha do usuário. A mesma se tornou a plataforma líder de mercado na análise de *Business Intelligent* por anos consecutivos classificada pela *Gartner Magic Quadrant*.

O *Power BI* tem uma versão gratuita da qual se pode efetuar o *download* da plataforma no endereço <https://powerbi.microsoft.com/pt-br/desktop/> permitindo a um usuário trabalhar com grande volume de dados, criar visualizações atraentes, obter relatórios ricos, interativos, assim como criar diferentes formas que valorizam sua apresentação, facilitando a Visualização da Informação.

Segundo a Microsoft (2015), o *Power BI* surge das 3 opções (*Power Query*, *Power Pivot* e *Power View*) já existentes no *software Excel*. Além disso, traz diversas funcionalidades que resultam como gerador de relatório e apoio da tomada de decisão. A seguir indicam-se componentes que compõem a plataforma.

“*Power Query* (recurso desenvolvido para combinar e transformar dados), *Power View* (ferramenta de visualização da Microsoft), *Power Map* (o sistema 3D para visualizar dados espaciais), *Power Pivot* (um recurso de modelagem de dados), *Power Q&A* (a parte do sistema que responde a queries feitas em linguagem natural) e *Power BI Desktop* (que permite desenvolver soluções que utilizam todas essas tecnologias”.

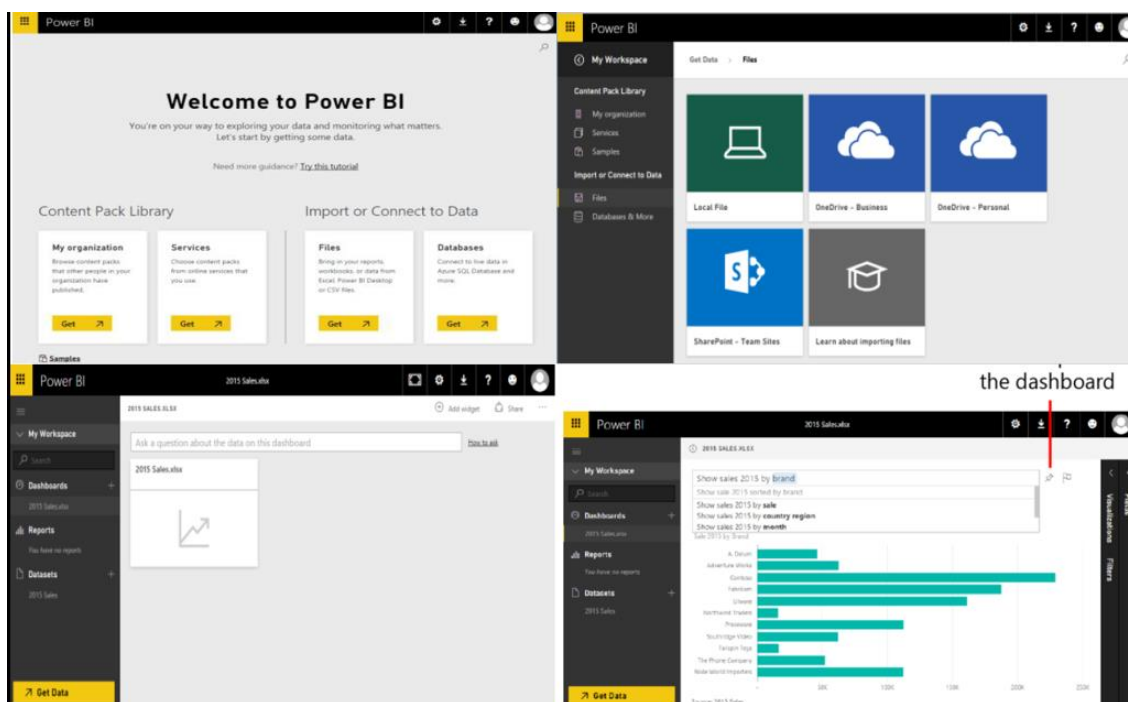
(ARBIT, 2020)

As vantagens oferecidas pelo *Power BI* tornam-na muito utilizada pelas empresas para atender às suas necessidades de BI (*Business Intelligence*) sendo elas: i) a possibilidade de compartilhar visualização de dados nas nuvens de forma global; ii) acesso a análise de dados corporativo; iii) acelera a preparação de *Big Data* através do Azure; iv) não é necessário grande domínio para usar a ferramenta; v) produz relatórios e modelos cognitivos usando dados estruturados e não

estruturados incluindo textos e imagens; vi) mantem uma integração com o *EXCEL* incomparável. (Microsoft, 2020)

O maior interesse de quem usa as ferramentas de BI em relação aos recursos oferecidos são os *dashboards*, seguido da visualização, *Drill-Down*, relatórios, análises preditivas, mineração de dados, ETL (*Extract, Transform and Load*) e OLAP (processamento analítico on-line). Esta pesquisa foi realizada pela SelectHub e contou com a participação de 600 empresas de diversos setores. (Silva, 2019 apud Conrad, 2018)

A interface do *Power BI* pode ser vista na figura 12, que consta a tela de abertura, tela de entrada de dados, tela após a entrada dos dados e a tela de uma exibição natural de dados solicitada.



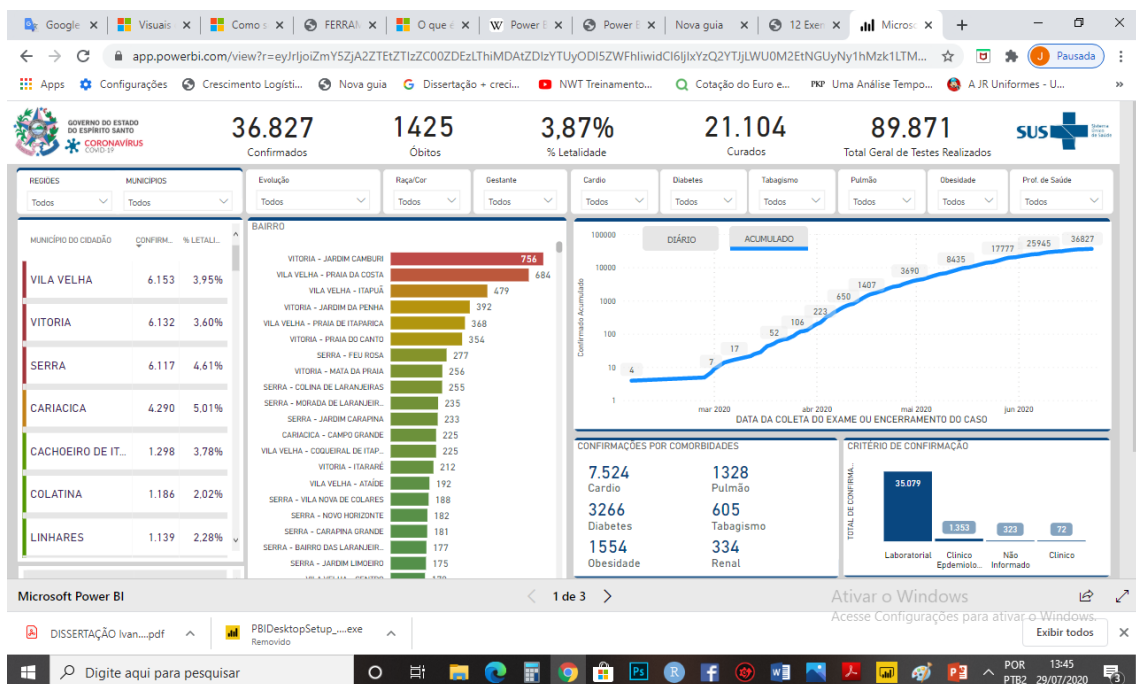
Fonte: Power BI Software 2020

Figura 12: Interface do Software Power BI

No setor da saúde a ferramenta é utilizada no mapeamento da ocorrência de doenças e na prevenção de surtos, como por exemplo ocorrência entre as doenças, hipertensão, diabetes, obesidade, assim como cruzando desses dados com

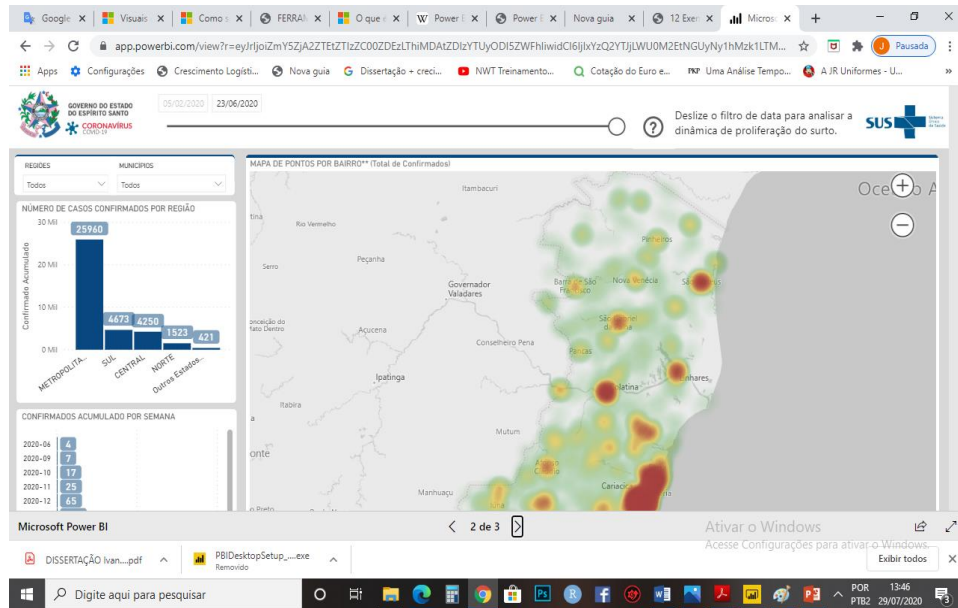
vacinação e saneamento. Uma outra situação bem interessante é a prevenção de riscos de novas doenças e o custo que as mesmas possam gerar ou seja, é possível conhecer a situação atual de um sistema de saúde e identificar quais as necessidades mais urgentes na sua população.

Para entender ainda mais a importância do *Power BI*, segue dados reais sobre uma situação atual enfrentada no mundo que é o Corona Vírus, onde o sistema único de saúde (SUS) do Governo do Estado do Espírito Santo no Brasil em 2020 criou através da plataforma *dashboards* em tempo real para que toda a população do estado visualizasse informações sobre a situação da evolução da pandemia. Vejam-se nas figuras 13,14 e 15 como usaram de forma eficiente a plataforma com dados de fatos ocorridos até 23.06.20 no estado.



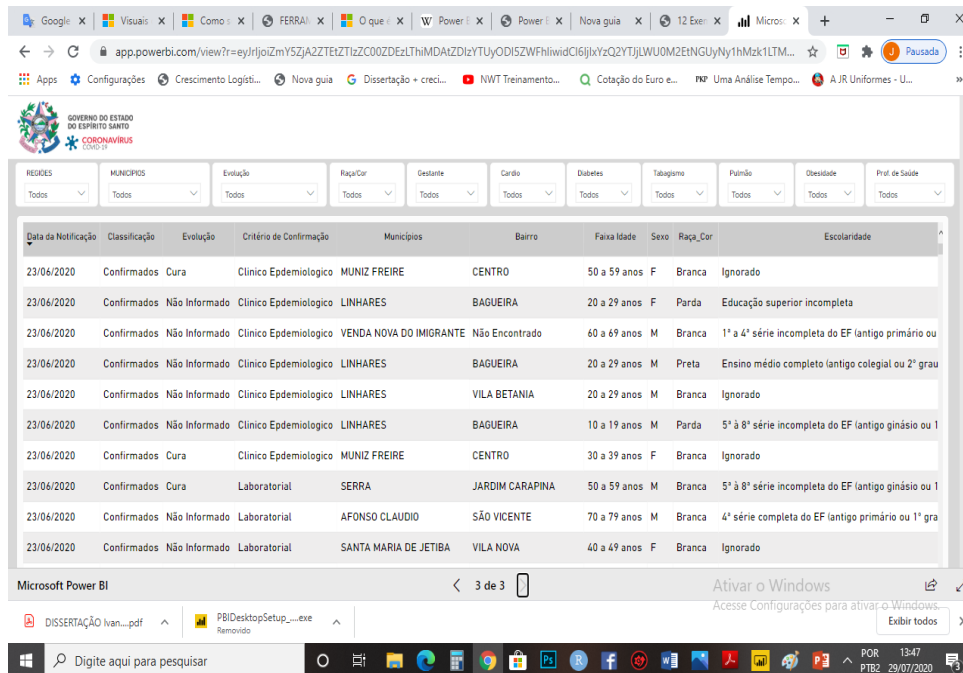
Fonte: SUS / ES-Brasil 2020

Figura 13: Uso do *Power BI* pelo SUS relacionado ao avanço da COVID-19 em junho de 2020, por cidade.



Fonte: SUS / ES-Brasil 2020

Figura 14: Uso do *Power BI* pelo SUS relacionado ao COVID-19 aplicado ao mapa de pontos por bairro para demonstrar as regiões mais críticas em junho de 2020.



Fonte: SUS / ES-Brasil 2020

Figura 15: Uso do *Power BI* pelo SUS relacionado ao COVID-19 na organização de banco de dados através de tabela em junho de 2020.

O *Power BI* e o *Tableau Public* serão ferramentas utilizadas no desenvolvimento do modelo visual proposto nesse estudo e o faremos usando suas características importante para criação de gráficos, relatórios e análises.

Capítulo 6

6. APLICAÇÃO SOBRE A BASE DE DADOS E RESULTADOS

Neste capítulo será utilizado junto a uma aplicação de dados reais, as opções metodológicas selecionadas anteriormente, que incluem métodos estatísticos, tais como a Estatística Descritiva e Regressão Logística Binária. Aqui ficará evidente o perfil dos participantes selecionados no estudo, assim como a relação das variáveis selecionadas em busca de respostas que facilite a identificação das possíveis variáveis que podem ser fatores de risco com a obesidade. Essas variáveis de influência para obesidade serão apontadas com base na construção de alguns modelos matemáticos do qual buscar-se-á um melhor ajuste em busca de chegar ao objetivo deste trabalho.

6.1 BASE DE DADOS E VARIÁVEIS

Para esse estudo transversal foi admitida uma amostra de 714 indivíduos de várias instituições de saúde do norte a sul de Portugal, que apresentou dados coletados através da ficha de atendimento de seus pacientes preenchidas pelos médicos. O histórico da coleta apresentava pacientes saudáveis ou com a patologia de obesidade, assim como dados demográficos (idade e género), dados genéticos (fenótipo de haptoglobina, fenótipo ACT, grupo sanguíneo Lewis e grupo sanguíneo MN), assim como fatores clínicos (hipertensão, diabetes, e atividade fosfatase ácida). Essas informações foram relevantes para uma conclusão do estudo.

6.1.1 VARIÁVEL RESPOSTA E VARIÁVEIS PREDITORAS

Para variável resposta foi utilizado o diagnóstico de obesidade em duas categorias (0 – Não obeso, 1 – Obeso, considerando a classificação do IMC segundo a OMS). As variáveis preditoras para o estudo foram: idade (anos), género (0 – feminino, 1 – masculino), sistema de grupo sanguíneo Lewis (1 - a+b-, 2 - a-b+, 3 - a-b-, 4 - a+b+), Sistema de grupo sanguíneo MN (1 – MM, 2 – MN, 3 - NN), fenótipo de haptoglobina (0 – 1.1, 1 – 2.1, 3 - 2.2), atividade de fosfatase ACP-1

(mmol/g Hb/H), tipo de diabetes (0 – não diabético, 1 – tipo 1, 2 – tipo 2), hipertensão (0 – normal, 1 – hipertenso), Fenótipo ACT(1 – AA , 2 – BB, 3 – CC , 4 – AB, 5 – AC , 6 – BC), IMC (0 – abaixo do peso, 1 – peso normal, 2 – sobrepeso, 3 – Obeso, de acordo com a classificação do IMC segundo a OMS).

A classificação do IMC neste estudo é realizada com base na Organização Mundial da Saúde, sendo assim temos classes divididas em abaixo do peso, normal, sobrepeso e obeso. A nossa variável resposta desse estudo é definida como uma variável dicotômica (0 – Não obeso, 1 – Obeso) gerada da junção das classes fornecidas pela OMS, referenciando (0 – abaixo do peso ou normal e 1 – sobrepeso e obeso).

6.1.2 DEFININDO ALGUMAS VARIÁVEIS PREDITORAS

SISTEMA DE GRUPO SANGUÍNEO LEWIS

Segundo Bonifácio e Novaretti (2009) o sistema de grupo sanguíneo Lewis são:

“Os antígenos Lewis são provenientes do plasma e adsorvidos nos eritrócitos. A sua expressão depende da interação de duas fucosiltransferases diferentes, produtos dos genes FUT2 e FUT3 localizados no cromossomo 19p13.3. Estão expressos em muitos tecidos humanos, incluindo células da mucosa da superfície gástrica. *Helicobacter pylori*, o maior agente causador de úlceras gástricas, liga-se às células da mucosa da superfície gástrica que contém fucose, especificamente antígenos Leb e H (cadeias tipo 1), sendo o antígeno Leb o receptor predominante para esse organismo. ”

(Bonifácio & Novaretti, 2009)

SISTEMA DE GRUPO SANGUÍNEO MN

O Sistema de grupo sanguíneo MN é semelhante ao sistema de grupo sanguíneo RH. No sistema do grupo MN são encontrados os anticorpos anti-M e anti-N, os antígenos MN são situados em duas glicoproteínas situadas na membrana do eritrócito. Além disso são codificados diretamente pelos genes correspondentes, ao antígeno M e o antígeno N. (Pschisky, 2003 apud Calhoun,1992.; Bello-Gonzales et al.,2001)

FENÓTIPO DE HAPTOGLOBINA

De acordo a autora Elisabete Carolino et al. (2005) a haptoglobina é descrita como:

“...uma proteína plasmática, alfa 2 globulina, codificada no cromossoma 16 e libertada pelo fígado em resposta à inflamação aguda estando envolvida na imunidade inata. São-lhe atribuídas funções antioxidante, neoangeogénica e moduladora de síntese de prostaglandinas entre outras. Apresenta três fenótipos distintos, 1.1, 2.1 e 2.2, de acordo com a proteína sintetizada. Esta proteína poderá estar associada à susceptibilidade para situações patológicas comuns, tais como diabetes e doenças cardiovasculares, assim como a outro tipo de patologias, nomeadamente neoplasias – carcinoma *in situ* (CIN) e carcinoma.”

(Elisabete Carolino et al. 2005: 144)

FENÓTIPO ACT

Fenótipo de uma doença descreve as possíveis características que podem ser observadas durante o processo da doença que são as reações químicas, morfológicas, bioquímicas, fisiológicas assim como a observação da resposta sobre o tratamento (Filho, 2020). O Fenótipo ACT - Teste de Controle da asma, trata-se de um controle sobre estas características nos pacientes que são asmáticos.

ATIVIDADE DE FOSFATASE

A avaliação da Atividade de Fosfatase é de grande importância para medicina visto que a mesma ajuda a diagnosticar doenças em seres humanos. A Fosfatase alcalina permite avaliar questões no fígado e na parte óssea enquanto que a Fosfatase ácida auxilia no diagnóstico de câncer prostático (Câmara, 2011).

6.2 CONSTRUÇÃO DO MODELO

Antecedendo a construção do modelo será realizada uma análise descritiva das variáveis, de um modo geral para maior conhecimento e visibilidade dos dados do estudo. Esta fase contemplará uma análise univariada, seguida de uma análise bivariada relacionando cada variável independente com a variável resposta obesidade.

Após a análise descritiva, será aplicada a regressão logística binomial com o apoio do Software SPSS onde através dessa técnica serão aplicados três métodos distintos para a inclusão das variáveis independentes no modelo. Esses

métodos são conhecidos como método *Enter*, método *Forward Condicional* e o método *Backward Condicional*. Desejamos com isso buscar o modelo mais apropriado para explicar a relação entre as variáveis e a obesidade.

Através da Regressão Logística Binária e dos três métodos citados acima, será avaliado, através de diversas técnicas, a qualidade do modelo apontado como um mais adequado que explique essa relação.

Os métodos citados acima se utilizam do Teste da Razão de Verossimilhança através da estatista G,

$$G = -2(\text{Loglikelihood}(\text{mod. reduz}) - \text{Loglikelihood}(\text{mod. compl.})) \quad (4.1)$$

que segue uma distribuição Qui-quadrado com p-q graus de liberdade, testando sob a hipótese nula que os coeficientes das variáveis excluídas são nulos. Juntamente com os métodos serão realizados, o Teste de Wald, Teste de Omnibus, Teste de Hosmer e Lemeshow, Curva de Roc, Coeficiente Kappa, assim como a análise de resíduos que permitirá uma comparação de cada modelo com p variáveis conforme critérios adotados por cada método.

Os resultados serão considerados significativos ao nível de significância de 5%. Para testar a normalidade dos dados das variáveis quantitativas, utilizou-se o teste *Kolmogorov-Smirnov*, dado que a amostra é dimensão superior a 50.

6.3. SOFTWARE UTILIZADO

O Software utilizado tanto para análise descritiva como para construção do modelo de regressão logística binária será o SPSS 22.0 para Windows. Na proposta do modelo para a Visualização da Informação estaremos usando os softwares *Tableau Public 2020.3* e *Power BI Desktop* atualização 2.91.383.0.

6.4 ANÁLISE DESCRITIVA

6.4.1 ANÁLISE UNIVARIADA

OBESIDADE

A obesidade é a nossa variável dependente ou resposta para o nosso estudo. Em termos de classificação, a variável obesidade foi determinada como “0” para pacientes que não possui obesidade e “1” para pacientes que possuem obesidade. Esses parâmetros foram criados a partir de dados obtidos pelo IMC do paciente e associados aos padrões mundiais determinados pela OMS. Dos 714 indivíduos da amostra, 417 apresentaram como dados válidos. Do total de 417 válidos registra-se 238 indivíduos considerados sem obesidade e 179 são considerados obesos, uma proporção de 57,07% e 42,92% respectivamente. Este resultado pode ser visto na tabela 01(anexo I) e mostra que temos quase metade da amostra com casos de obesidade.

IDADE

Dos 714 indivíduos da amostra, 644 responderam sobre a sua idade. A amostra apresenta indivíduos com idade que variam dos 13 aos 98 anos. A idade média dos participantes é de aproximadamente 38 anos com uma variabilidade de 19 anos. A variável idade apresenta uma curva muito concentrada na faixa etária dos 20 aos 40 anos tendo uma mediana de 36 anos, o que caracteriza uma população em sua maioria de jovens. Esses resultados podem ser apreciados na tabela 02 (anexo II).

Um teste de normalidade foi realizado para verificar se a distribuição da variável idade na amostra possui uma distribuição normal. Os testes utilizados foram os de Kolmogorov-Smirnov, dado que a dimensão da amostra é superior a 50, conforme pode ser visto na tabela 08. No teste verifica-se que o p-valor é igual a 0,00, ou seja, $p < 0,05$, significando que os dados relativos à idade a variável não se ajustam a distribuição normal, considerando um nível de significância de 5%.

Este fato é muito fácil de acontecer quando se trata de dados clínicos, visto que os dados apresentam uma grande variabilidade. É importante ressaltar que o testes kolmogorov-smirnov é bastante rigoroso, permitindo facilmente rejeitar a hipótese nula devido aos valores estarem mais concentrados no início da curva (Paes, 2012).

Tabela 08: Resultado do Teste de Normalidade da variável idade com o apoio do SPSS.

Testes de Normalidade			
Kolmogorov-Smirnov^a			
	Estatística	gl	Sig.
Idade	,143	644	,000
a. Correlação de Significância de Lilliefors			

Fonte: Output do Software SPSS.

GÊNERO

A variável gênero está dividida em duas categorias, sendo elas, masculina e feminina. No total da amostra, a maioria são mulheres totalizando 540, representado 81,1%. Os homens são em 126 representando 18,9%, considerando que 48 indivíduos omitiram a sua resposta. Esse diagnóstico pode ser visto na tabela 03 (anexo I).

ÍNDICE DE MASSA CORPOREA (IMC)

A variável IMC foi dividida em 4 classes de acordo com os critérios da OMS, a primeira corresponde aos indivíduos que estão abaixo do peso (0), a segunda faz referência aos indivíduos com peso normal (1), a terceira são os que apresentam sobrepeso (2) e a quarta representa os indivíduos obesos (3).

Analisando a distribuição dos participantes por classe de IMC, verificamos que dos 571 respondentes, 28,7 % deles apresentam sobrepeso, seguido de 26,9% que apresentam peso normal, 17,4 % de obesos e apenas 7% apresenta-se abaixo do peso. É possível observar na tabela 04 (anexo I), que mais de 50% da amostra apresenta peso normal e sobrepeso. Os que não responderam ao questionário representam 20% da amostra.

SISTEMA DE GRUPO SANGUINEO LEWIS

Com relação ao sistema sanguíneo Lewis dos pacientes, a variável também está dividida em 4 grupos, o primeiro grupo corresponde aos indivíduos que são classificados em “a-b-”, o segundo grupo faz referência aos indivíduos com classificação “a-b+”, o terceiro grupo são os que apresentam “a+b-” e o último grupo são indivíduos com “a+b+”.

De acordo com a tabela 05 (anexo1), verifica-se que 82,8% correspondem a indivíduos pertencentes aos grupos “a-b-“ e “a-b+”, sendo muito insignificante o número de indivíduos no grupo “a+b+”. O grupo “a-b+” foi excluído do estudo por não ter um número mínimo suficiente para a análise.

SISTEMA SANGUÍNEO MN

A variável MN foi dividida em 3 grupos, sendo o primeiro grupo indivíduos com característica “MM”, o segundo grupo os que apresentam “MN” e o terceiro os com características “NN”. Na tabela 06 (anexo I) é possível verificar que a maioria dos participantes apresentam característica “MM” sendo 48,2 %, respectivamente 27,1 % e 24,7% ficam por conta dos grupos “MN” e “NM”.

FENÓTIPO DE HAPTOGLOBINA (HP)

O Fenótipo de Haptoglobina está dividida em 3 grupos, sendo o primeiro grupo indivíduos com característica “1.1”, o segundo grupo os que apresentam “2.1” e o terceiro os com características “2.2”.

O resultado desta variável pode ser observado na tabela 07 (anexo I), onde 48,7 % tem característica do fenótipo de haptoglobina “2.1”, 35,2% com o fenótipo de haptoglobina “2.2” e 16,1% com o fenótipo de haptoglobina “1.1”. A maioria dos indivíduos se concentram nos grupos “2.1 e 2.2”.

TIPO DE DIABETES

A variável onde classifica o indivíduo sobre a presença de diabetes e o seu tipo ficou dividida em 3 grupos, sendo o primeiro grupo indivíduos com característica de ausência de diabetes, o segundo grupo os que apresentam a diabetes do tipo 1 e o terceiro grupo os que apresentam diabetes do tipo 2.

Na tabela 08 (anexo I) a distribuição dos respondentes baseado nos grupos citados acima, verifica-se que 97,5 % dos participantes se enquadram na ausência de diagnóstico de diabetes, enquanto que 1,9% apresentam diabetes do tipo 2 e apenas 0,6 % apresentam diagnóstico do tipo 1.

HIPERTENSÃO ARTERIAL (HTA)

A variável hipertensão foi dividida em apenas 2 grupos, os que apresentam diagnóstico e os que não apresentam, ou seja, o indivíduo é normal ou ele é hipertenso.

O observado na análise de acordo com a tabela 09 (anexo I), foi que 72,7% apresentam diagnóstico de ausência de hipertensão e 27,3% são considerados hipertensos.

ATIVIDADE DE FOSFATASE ÁCIDA – ACP-1 (MMOL/G HB/H)

Na amostra de 714 indivíduos, 400 registraram a informação sobre a sua atividade de fosfatase ACP-1. Os dados apresentaram indivíduos com atividade mínima de 19,02 (mmol/g Hb/ H) e máxima de 1.162,28 (mmol/g Hb/ H) retratando uma alta variabilidade. Em média os participantes da amostra apresentam uma atividade de fosfatase de 325,70 (mmol/g Hb/ H) com uma mediana de 293,18 (mmol/g Hb/ H), o que mostra um ponto importante visto que temos 1 participante que está elevando a média do grupo. Todas essas informações estão registradas na tabela 10 (anexo I).

Avaliando a normalidade dos dados observamos o mesmo ocorrido com a variável quantitativa idade. Os testes de normalidade de kolmogorov-smirnov conforme tabela 09, apresenta o p-valor igual a 0,00, o que significa que os dados da variável Atividade de Fosfatase ACP1 também não se ajustam a uma distribuição normal, considerando um nível de confiança de 5%.

Tabela 09: Resultado do Teste de Normalidade da variável Atividade Fosfatase ácida ACP-1.

Testes de Normalidade			
Kolmogorov-Smirnov^a			
	Estatística	gl	Sig.
Atividade Fosfatase ácida - ACP1 (mmol/g Hb/H))	,100	400	,000
a. Correlação de Significância de Lilliefors			

Fonte: Output do Software SPSS.

FENÓTIPO ACP-1

A variável Fenótipo ACP-1 foi dividida em 6 grupos (1 – AA, 2 – BB, 3 – CC, 4 – AB, 5 – AC, 6 – BC). Ao analisar a variável observou-se que os grupos que mais se destacam são “BB” com 38,3% seguindo do “AB” com 28,6%. O grupo que menos aparece entre os participantes é o de “CC” com 5,7%. Esses resultados podem ser vistos na tabela 11(anexo I).

6.4.2 ANÁLISE BIVARIADA

A análise bivariada aplicada aqui tem o objetivo de avaliar a associação ou o grau de dependência existente entre a variável resposta obesidade (0 – Não obeso, 1 – Obeso) e as demais variáveis envolvidas nesse estudo conforme já foram citadas anteriormente na análise univariada.

Para realizar esse estudo de associação recorreu-se ao Teste Qui-quadrado de independência. Vale ressaltar que quando os pressupostos do Teste Qui-quadrado não forem atendidos, a nova opção de tratamento para avaliar essa associação passou a ser através da simulação de Monte Carlo.

Com relação as variáveis quantitativas Idade e Atividade de Fosfatase Ácida (ACP-1) utilizou-se a medida ETA que permite avaliar o grau de dependência dessas variáveis de interesse por se tratar de uma variável resposta qualitativa dicotômica e uma independente quantitativa.

OBESIDADE versus GÉNERO

Com base nos resultados do Teste Qui-quadrado apresentado na tabela 01 (anexo II), não existe uma associação entre as variáveis obesidade e género considerando um p-valor de 0,245 que é maior igual a 0,05, não rejeitando a hipótese nula do teste de independência referenciado em (5.3).

OBESIDADE versus SISTEMA SANGUÍNEO LEWIS

Para análise bivariada da variável grupo sanguíneo Lewis os dados foram compactados em apenas 3 categorias (a+b-, a-b+, a-b-) devido a pequena representatividade da categoria a+b+ como pode ser visto na tabela 19 na análise univariada. Com respeito ao Teste Qui-quadrado e atendendo aos pressupostos do mesmo visto na tabela 02 (anexo II), não existe uma associação entre as variáveis obesidade e o grupo sanguíneo Lewis considerando um p-valor de 0,387 que é maior igual a 0,05, não rejeitando a hipótese nula do teste de independência.

OBESIDADE versus SISTEMA SANGUÍNEO MN

Com base nos resultados do teste Qui-quadrado e atendendo aos pressupostos do mesmo, na tabela 03 (anexo II) verifica-se que existe uma associação entre as variáveis obesidade e o grupo sanguíneo MN considerando um p-valor de 0,02 que é menor igual a 0,05, rejeitando a hipótese nula de independência do teste.

OBESIDADE versus FENOTÍPO DE HAPTOGLOBINA

Com base nos resultados do teste Qui-quadrado e atendendo aos pressupostos do mesmo visto na tabela 04 (anexo II), não existe associação entre as variáveis obesidade e a haptoglobina considerando um p-valor de 0,311 que é maior igual a 0,05, não rejeitando a hipótese nula do teste de independência.

OBESIDADE versus TIPO DE DIABETE

Com base nos resultados na simulação de Monte Carlo apresentado na tabela 05 (anexo II), podemos afirmar que existe associação entre as variáveis obesidade e tipo de diabetes considerando um p-valor de 0,03 que é menor igual a 0,05, rejeitando a hipótese nula do teste de independência.

OBESIDADE versus HIPERTENSÃO

Com base nos resultados do teste Qui-quadrado e atendendo aos pressupostos do mesmo, na tabela 06 (anexo II) verifica-se que existe uma associação entre as variáveis obesidade e hipertensão considerando um p-valor de 0,00 que é menor que 0,05, rejeitando a hipótese nula de independência do teste.

OBESIDADE versus FENÓTIPO ACP-1

Atendendo aos pressupostos do Teste Qui-quadrado, na tabela 07 (anexo II) apresenta-se a existência da associação entre as variáveis obesidade e fenótipo ACP-1 considerando um p-valor de 0,001 que é menor igual a 0,05, rejeitando a hipótese nula de independência do teste.

OBESIDADE versus IDADE

Para verificar o grau de relação existente entre a obesidade e a idade, sendo a idade uma variável quantitativa foi utilizada a medida ETA. O resultado pode ser visto na tabela 08 (anexo II), do qual apresenta-se um valor de 0,729 o que explica que existe sim uma significativa relação de dependência sobre essas variáveis.

OBESIDADE versus ATIVIDADE FOSFATASE ÁCIDA ACP-1

Semelhante a variável idade também foi aplicado a medida ETA conforme apresentado na tabela 09 (anexo II), do qual mostra um valor de 0,996, que representa uma relação ainda mais forte com relação a idade, indicando também uma relação de dependência existente entre a obesidade e a atividade de fosfatase ACP-1. Depois de toda análise bivariada realizada observou-se que as variáveis que apresentaram uma associação com a variável resposta obesidade foram o Sistema Sanguíneo MN, Tipo de diabetes, Hipertensão, Fenótipo ACP-1, Idade e Atividade Fosfatase Ácida ACP-1. Em contrapartida as variáveis gênero, Sistema Sanguíneo Lewis e Fenótipo de haptoglobina não apresentaram associação com a obesidade. A seguir busca-se avaliar de maneira mais significativa a importância de cada variável através da construção de um modelo matemático usando a Regressão Logística Binária. Assim é possível conquistar ainda mais informações da real relação entre estas variáveis e a obesidade.

6.4.3 SELEÇÃO DE VARIÁVEIS E AJUSTE DO MODELO

Para uma análise mais consistente sobre a inclusão das variáveis de interesse para um futuro ajustamento do modelo com a utilização do método de Regressão Logística Binária, do qual se objetiva identificar os fatores de risco para a patologia de obesidade serão aplicados três critérios de seleção de variáveis, a considerar Enter, o *Forward condicional* e o *Backward condicional*.

Estes critérios permitiram identificar as variáveis que devem ser incluídas no modelo, assim como a identificação de um melhor modelo ajustado, através de um comparativo entre eles.

Todos os métodos aplicados abaixo seguem o número de casos incluídos na análise conforme a tabela 10. Nesta análise foi incluído 85 casos equivalente a 24,9% da amostra total, visto que no banco de dados estes possuíam informações completas acerca das variáveis independentes e dependentes.

A maneira como as variáveis foram codificadas pode ser vista na tabela 11. Vale ressaltar que quanto utilizamos o modelo de regressão logística e as variáveis independentes de natureza qualitativa, medidas em escala nominal, possuem mais de 2 categorias, na regressão utilizam-se de variáveis *dummy* tantas quantas o número de categorias menos 1.

Tabela 10: Resumo de casos incluídos no estudo da análise do modelo.

Resumo de processamento do caso			
Casos não ponderados ^a		N	Porcentagem
Casos selecionados	Incluído na análise	85	24,9
	Casos ausentes	256	75,1
	Total	341	100,0
Casos não selecionados		0	,0
Total		341	100,0

a. Se a ponderação estiver em vigor, veja a tabela de classificação para o número total de casos.

Fonte: Output do Software SPSS.

Tabela 11: Codificação da variável dependente e independentes do estudo.

Codificação de variável dependente	
Valor original	Valor interno
Normal	0
Obeso	1

Codificações de variáveis categóricas							
	Frequência	Codificação de parâmetro					
		(1)	(2)	(3)	(4)	(5)	
Fenótipo ACP-1	AA – Categoria de referência	11	,000	,000	,000	,000	,000
	BB	31	1,000	,000	,000	,000	,000
	CC	7	,000	1,000	,000	,000	,000
	AB	22	,000	,000	1,000	,000	,000
	AC	11	,000	,000	,000	1,000	,000
	BC	3	,000	,000	,000	,000	1,000
Sist. Sang. Lewis	a+b- – Categoria de referência	20	,000	,000			
	a-b+	20	1,000	,000			
	a-b-	45	,000	1,000			
Sist.Sang. MN	MM – Categoria de referência	43	,000	,000			
	MN	21	1,000	,000			
	NN	21	,000	1,000			
F. Haptoglobina	1.1 – Categoria de referência	17	,000	,000			
	2.1	39	1,000	,000			
	2.2	29	,000	1,000			
Sexo	Feminino – Categoria de referência	80	,000				
	Masculino	5	1,000				

Fonte: Output do Software SPSS.

MÉTODO DE SELEÇÃO *ENTER*

O critério de seleção *Enter* realiza uma análise incluindo todas as variáveis independentes no modelo. Após várias análises, é possível identificar a

significância da inclusão de cada variável, assim como a qualidade do modelo ajustado, contendo todas as variáveis.

Com base na tabela 12 o teste de coeficiente de modelo Omnibus, que aponta a estatística Qui-quadrado $\chi^2 = 63,311$ e um p-valor = 0,000, sendo (p-valor <0,05) conclui-se que o modelo incluindo as variáveis independentes é significativo, conforme hipóteses do teste de Omnibus citada em (5.31) implicando que pelo menos uma das variáveis independentes contribui para explicar a patologia da obesidade.

Tabela 12: Teste de Omnibus sob o critério de seleção Enter no SPSS.

Testes de coeficientes de modelo Omnibus				
		Qui-quadrado	df	Sig.
Etapa 1	Etapa	63,311	14	,000
	Bloco	63,311	14	,000
	Modelo	63,311	14	,000

Fonte: Output do Software SPSS.

Na tabela 13 registra-se o valor da Razão de Verossimilhança (log -2 = 54,512), que ao nível de significância fixado previamente ($\alpha = 0,05$) e com base nas suas hipóteses citadas em (5.23) nos permite concluir que não se rejeita H_0 , ou seja o modelo se ajusta aos dados. Neste mesmo resumo encontram-se os valores de o $R^2_{CS} = 0,525$ e o $R^2_N = 0,700$ que são o Pseudo R^2 de Cox e Snell e o Pseudo R^2 de Nagelkerke anteriormente citados em (5.29) e (5.30). Esses resultados apontam a adequabilidade do modelo aos dados.

Tabela 13: Razão de Verossimilhança, Pseudo R Cox e Snell, Pseudo de Nagelkerke pelo critério de seleção Enter no SPSS.

Resumo do modelo			
Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	54,512 ^a	,525	,700

a. Estimación finalizada no número de iteração 20 porque o máximo de iterações foi atingido. Não é possível encontrar a solução final.

Fonte: Output do Software SPSS.

O teste de ajustamento de Hosmer e Lemeshow apresentado na tabela 14 cujas a hipótese podem ser vistas em (5.33), apresenta $\chi^2 = 6,123$ e p-valor = 0,525 considerando um $\alpha = 0,05$ não se rejeita H_0 , ou seja, isso nos mostra que os valores ajustados se aproximam dos valores esperados ajustados pelo modelo, o que mais uma vez demonstra que o modelo se ajusta aos dados.

Tabela 14: Teste de Hosmer e Lemeshow sob o critério de seleção Enter no SPSS.

Etapa	Qui-quadrado	df	Sig.
1	6,123	7	,525

Fonte: Output do Software SPSS.

Na tabela 15 através da classificação do modelo é possível avaliar a boa capacidade do mesmo, ou seja, o percentual de total de casos corretamente classificado é $A = 87,06\%$ apresentado na equação (5.34), e o percentual de casos classificados corretamente por acaso é $B = 50,06\%$ apresentado na equação (5.35). A diferença entre eles ($A-B$) é de $36,99\% > 25\%$ o que aponta uma boa capacidade do modelo.

Na sensibilidade o percentual é de $88,37\%$ com base na equação (5.35) o que representa a proporção de obesos classificados corretamente pelo modelo. A especificidade foi de $87,80\%$ com base na equação (5.36), indicando a proporção de indivíduos não obesos classificados corretamente. Ambos afirmam a boa capacidade do modelo.

Tabela 15: Classificação do modelo segundo o critério de seleção Enter no SPSS.

Etapa 1	Observado		Previsto		
			Obesidade		Porcentagem correta
	Normal	Obeso	Normal	Obeso	
	Normal		36	6	85,7
	Obeso		5	38	88,4
	Porcentagem global				87,1

a. O valor de recorte é ,500

Fonte: Output do Software SPSS.

Outra maneira de avaliar o ajuste de modelo é através da curva ROC. Na figura 16 temos a representação gráfica e a tabela 16 que define quantitativamente a área da curva indicando 0,931, um p-valor de 0.000. Considerando um $\alpha=0,05$, e significando conforme critério citado pelo autor Hosmer e Lemeshow (2000) na figura 07 anteriormente, temos uma ótima capacidade de ajuste do modelo. A área de 0,931 demonstra o quanto a curva de Roc está distante da reta diagonal do gráfico, o que demonstra também um poder classificatório do modelo.

Tabela 16: Área sobre a curva Roc segundo o critério de seleção Enter no SPSS.

Área sob a curva

Variável(eis) de resultado de teste: Probabilidade prevista

Área	Erro Padrão ^a	Sig. assintótico ^b	Intervalo de Confiança 95% Assintótico	
			Limite inferior	Limite superior
,931	,026	,000	,881	,981

a. Sob a suposição não paramétrica b. Hipótese nula: área real = 0,5

Fonte: Output do Software SPSS

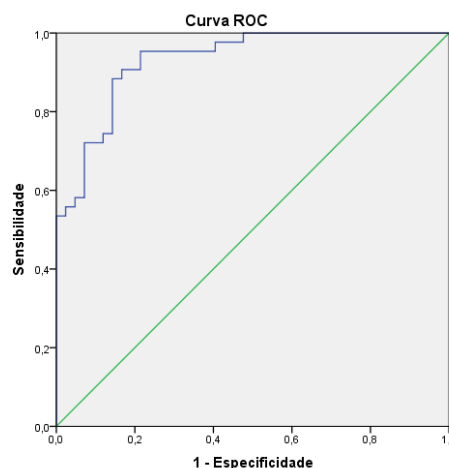


Figura 16: Gráfico da Curva Roc segundo o critério de seleção Enter

Outra forma, ainda, de avaliar o modelo, é através do coeficiente de concordância Kappa e seu teste de significância. De acordo com a tabela 17, o

teste de significância apresenta um p-valor < 0,05, ou seja, o Kappa é considerado diferente de zero segundo hipótese nula citada em (5.39). Uma vez que esse Kappa é diferente de zero e considerando o valor de Kappa = 0,741 conclui-se que existe uma concordância substantiva tomando como base a tabela 07.

Tabela 17: Tabulação cruzada e a concordância de K de Cohen segundo o critério de seleção Enter.

Obesidade * Grupo previsto Tabulação cruzada

Contagem

		Grupo previsto		Total
		Normal	Obeso	
Obesidade	Normal	36	6	42
	Obeso	5	38	43
Total		41	44	85

Medidas Simétricas

		Valor	Significância Erro Padrão ^a	Aprox. X ^b	Aprox. Sig.
Medida de concordância	Kappa	,741	,073	6,834	,000
N de Casos Válidos		85			

a. Não considerando a hipótese nula.

b. Uso de erro padrão assintótico considerando a hipótese nula.

Fonte: Output do Software SPSS

No estudo das variáveis na equação, cujo modelo *Enter* adiciona todas as variáveis na construção do modelo, é possível avaliar a significância de cada uma através da estatística de Wald e da Odds Ratio ou razão de chances que permitirá avaliar o grau de contribuição para a variável dependente.

Na tabela 18, contém o estudo das variáveis da equação, pelo critério de seleção Enter, através do teste de *Wald* cujas hipóteses estão referenciadas em (5.25) concluído-se que as variáveis Idade ($\chi^2_{\text{wald}}=15,589$; $p = 0,000$), MNgrupo(2) representado pelo Grupo NN ($\chi^2_{\text{wald}}=4,465$; $p = 0,035$ (a Haptoglobina (2) representando o fenótipo Hp2.2 ($\chi^2_{\text{wald}}= 4,795$; $p = 0,029$) e atividade fosfatase ACP1 ($\chi^2_{\text{wald}}=6,410$; $p = 0,011$), apresentam um efeito significativo sobre a probabilidade de contribuir para a patologia de obesidade. Diferentemente das variáveis gênero, Sistema Sanguíneo Lewis e Fenótipo ACP1 que apontam

coeficientes iguais a zero, não contribuindo de maneira significativa para a obesidade em conformidade com o critério, o que era expetável, dados os resultados da análise bivariada anteriormente realizada.

Tabela 18: Variáveis da equação do modelo segundo o critério de seleção Enter.

		Variáveis na equação						95% C.I. para EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Inferior	Superior
Etapa 1ª	Idade	,168	,043	15,589	1	,000	1,183	1,088	1,286
	Sexo(1)	,335	3,013	,012	1	,912	1,397	,004	512,477
	Sist. Sang. MM			4,831	2	,089			
	Sist. Sang. MN	,171	1,023	,028	1	,867	1,187	,160	8,807
	Sist. Sang. NN	-2,287	1,083	4,465	1	,035	,102	,012	,847
	Haptoglobina 1.1			7,735	2	,021			
	Haptoglobina 2.1	,579	,992	,341	1	,559	1,785	,255	12,480
	Haptoglobina 2.2	-2,672	1,220	4,795	1	,029	,069	,006	,755
	Atividade Fosfatase (mmolGhbH)	-,013	,005	6,410	1	,011	,987	,977	,997
	Fenótipo ACP1 (AA)			6,467	5	,263			
	Fenótipo ACP1 (BB)	-2,163	1,115	3,767	1	,052	,115	,013	1,021
	Fenótipo ACP1 (CC)	,382	1,577	,059	1	,808	1,465	,067	32,229
	Fenótipo ACP1 (AB)	,520	1,323	,154	1	,694	1,682	,126	22,486
	Fenótipo ACP1 (AC)	-,484	1,698	,081	1	,776	,617	,022	17,207
	Fenótipo ACP1 (BC)	-19,136	21196,8 86	,000	1	,999	,000	,000	.
	Sist. Lewis a+b-			2,417	2	,299			
	Sist. Lewis a-b+	-,852	1,142	,556	1	,456	,427	,045	4,004
	Sist. Lewis a-b-	,841	1,004	,702	1	,402	2,318	,324	16,574
	Constante	-1,805	1,701	1,126	1	,289	,165		

- a. Variável (is) inserida(s) na etapa 1: Idade, Sexo, MNGrupo, Haptogl, FosfacidACP1activitymmolgHbH, PhenotypeACP1, Lewis.

Fonte: Output do Software SPSS

Para avaliar o grau de importância dessas variáveis independentes e o quanto elas contribuem de maneira positiva ou negativa para o diagnóstico da obesidade, avaliou-se o Odds Ratio estimado. A avaliação será realizada sobre as variáveis idade, Sistema Sanguíneo NN, Fenótipo de Haptoglobina (2.2) e atividade fosfatase ACP1, devido a sua significância.

O Odds Ratio pode ser encontrado na tabela 18 como $\text{Exp}(\hat{\beta})$.

$\widehat{OR}_{\text{Idade}} = 1,167$, com I.C._{95%} = (1,088; 1,286) o que significa que a cada ano de idade a mais, o risco para a obesidade do paciente aumenta em 16,7%.

$\widehat{OR}_{\text{Sist. Sang. NN}} = 0,102$, com I.C._{95%} = (0,12; 0,847) representa indivíduos do grupo sanguíneo NN tendo como referência o grupo sanguíneo MM. Isso mostra que a chance de uma pessoa com o grupo sanguíneo NN desenvolver a patologia de obesidade é de -89,8% em relação ao grupo sanguíneo MM, tornando-o um fator protetor a doença com relação ao tipo sanguíneo de referência.

$\widehat{OR}_{\text{Haptoglobina 2.2}} = 0,069$, com I.C._{95%} = (0,06; 0,755) são indivíduos com o fenótipo Hp 2.2 tendo como referência o fenótipo Hp 1.1 que se apresenta como um fator de risco. Esta medida de 0,069 indica que a chance de um paciente com o fenótipo Hp 2.2 desenvolver a patologia de obesidade é de -93,1% relativamente a quem tem o fenótipo Hp1.1, ou seja, tem-se menos chance de apresentar a doença, tornando-se o fenótipo Hp2.2 um fator protetor.

$\widehat{OR}_{\text{Atividade Fosfatase ACP-1}} = 0,987$ com I.C._{95%} = (0,977; 0,997) indica que a cada unidade a mais de mmol/g Hb/ H, têm-se menos risco de diagnóstico de obesidade, sendo essa chance de - 0,013 %, o que também torna um fator protetor.

O modelo de Regressão Logística Binário com base no critério de seleção *Enter* é definido como:

$$\widehat{g}(X) = \text{Logit}(\hat{\pi}) = -1,805 + 0,168 \text{ idade} + 0,355 \text{ género} + 0,171 \text{ Sist. Sanguíneo MN} - 2,287 \text{ Sist. Sanguíneo NN} + 0,579 \text{ Haptoglobina 2.1} - 2,672$$

Haptoglobina 2.2 – 0,13 atividade fosfatase ácida - 2,163 Fenótipo ACP1(BB) + 0,382 Fenótipo ACP1(CC) + 0,520 Fenótipo ACP1(AB) – 0,484 Fenótipo ACP1(AC) – 19,136 Fenótipo ACP1(BC) - 0,852 Lewis (a-b+) + 0,841 Lewis (a-b-)

Conforme aplicado o critério *Enter* juntamente com o apoio da medida do *Odds Ratio* verifica-se que das variáveis significativas, as que apresentam o fator risco para a obesidade são a idade, a atividade fosfatase, o fenótipo de Haptoglobina e o sistema sanguíneo MN, destacando-se a variável independente idade com um índice significativo de 16,7% a cada acréscimo de ano na idade do paciente. As variáveis Fenótipo ACP-1 e o Sistema Sanguíneo Lewis não apresentaram significância para o modelo conforme medidas do Odds Ratio avaliada.

Na análise dos resíduos do modelo apresentado na figura 17, é visível que existem observações que apresentam candidaturas a *outlier*, pois as mesmas assumem valores do resíduo padrão acima do valor 2. Estas observações candidatas podem ser vista na tabela 19.

Tabela 19: Lista dos casos de outliers segundo o critério de seleção Enter.

Lista entre casos ^b						
Caso	Status selecionado ^a	Observado	Previsto	Grupo previsto	Variável temporária	
		Obesidade			Resid	ZResid
63	S	N**	,849	O	-,849	-2,374
75	S	N**	,835	O	-,835	-2,253
83	S	N**	,862	O	-,862	-2,503
194	S	O**	,091	N	,909	3,169

a. S = Selecionado, U = Casos não selecionados, e ** = Casos classificados incorretamente.

b. São listados os casos com resíduos estudentizados maiores que 2,000.

Fonte: Output do Software SPSS.

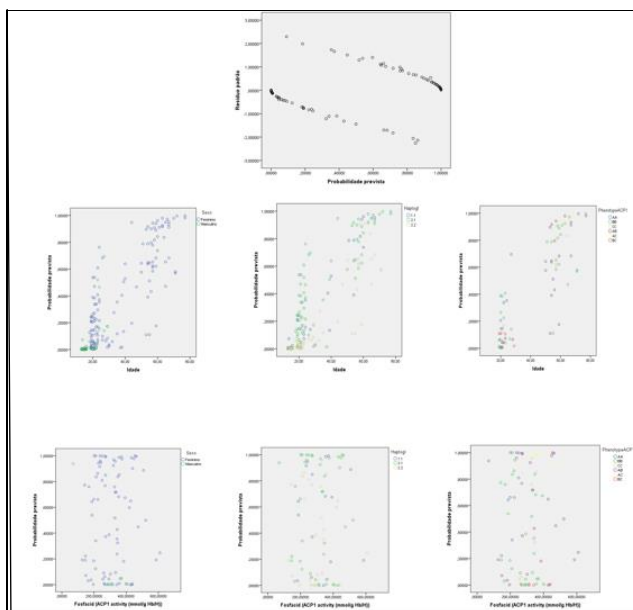


Figura 17: Gráficos da análise dos resíduos segundo o critério de seleção Enter.

Para avaliar melhor os outliers que influenciam na qualidade do ajuste do modelo é importante avaliar a Distância de Cook apresentada na figura 18. Nesta figura existem pontos onde a Distância de Cook é superior a 1, assumindo valores de 1,5 e 2, o que já aponta uma possível influência sobre os coeficientes do modelo.

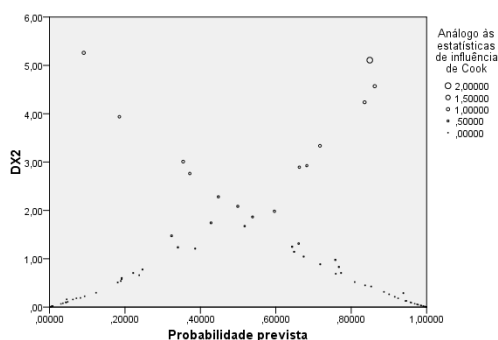


Figura 18: Gráficos da Distância de Cook segundo o critério de seleção Enter.

MÉTODO *FORWARD* CONDICIONAL

O método *Forward* Condicional define sua análise em etapas com base nas variáveis independentes de maior representatividade para construção do modelo, classificando o melhor modelo na sua última etapa de análise, finalizada com todas

as variáveis significativas. Na aplicação proposta por esse estudo, o critério atingiu seu melhor modelo na quarta etapa do processo. Assim a etapa 4 torna-se a referência de análise neste critério.

Na tabela 20 contém o teste de coeficiente de modelo Omnibus, e em todas as etapas apresentam um p-valor = 0,000, sendo um p-valor < 0,05, conforme teste de hipótese citado em (5.31), o que representa que a inclusão das variáveis uma a uma em cada etapa conclui-se um modelo significativo.

Tabela 20: Teste de Omnibus sob o critério de seleção *Forward* condicional.

Testes de coeficientes de modelo Omnibus				
		Qui-quadrado	df	Sig.
Etapa 1	Etapa	32,995	1	,000
	Bloco	32,995	1	,000
	Modelo	32,995	1	,000
Etapa 2	Etapa	4,074	1	,044
	Bloco	37,069	2	,000
	Modelo	37,069	2	,000
Etapa 3	Etapa	6,354	2	,042
	Bloco	43,423	4	,000
	Modelo	43,423	4	,000
Etapa 4	Etapa	6,682	2	,035
	Bloco	50,105	6	,000
	Modelo	50,105	6	,000

Fonte: Output do Software SPSS.

A estatística da razão de verossimilhança entre as etapas apresentadas na tabela 21, permite concluir que a etapa 4, com ($\log -2 = 67,719$) apresenta o melhor modelo, visto que quanto menor o valor da razão de verossimilhança melhor. Através do teste de hipótese citado em (5.23) e considerando $\alpha = 0,05$ conclui-se que não se rejeita H_0 , tendo um modelo ajustado aos dados. O Pseudo R Cox e Snell e no Pseudo de Nagelkerke ($R_{CS} = 0,445$ e o $R_N = 0,594$) conforme referência em (5.29) e (5.30) aponta boa adequabilidade do modelo aos dados.

Tabela 21: Razão de Verossimilhança, Pseudo R Cox e Snell, Pseudo de Nagelkerke pelo critério de seleção *Forward* condicional.

Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	84,829 ^a	,322	,429
2	80,755 ^b	,353	,471
3	74,400 ^b	,400	,533
4	67,719 ^c	,445	,594

- a. Estimaco finalizada no nmero de iterao 4 porque as estimativas de parmetro mudaram foram alteradas para menos de ,001.
- b. Estimaco finalizada no nmero de iterao 5 porque as estimativas de parmetro mudaram foram alteradas para menos de ,001.
- c. Estimaco finalizada no nmero de iterao 6 porque as estimativas de parmetro mudaram foram alteradas para menos de ,001.

Fonte: Output do Software SPSS.

No teste de ajustamento de Hosmer e Lemeshow apresentado na tabela 22, conforme hipteses citadas em (5.33), em todas as etapas o p-valor > 0,05, indicando a no rejeio da hiptese nula considerando um $\alpha = 0,05$, logo os dados se ajustam a cada modelo proposto sendo considerados bons.

Tabela 22: Teste de Hosmer e Lemeshow sob o critrio de seleo *Forward* condicional.

Etapa	Qui-quadrado	df	Sig.
1	6,174	7	,520
2	9,378	7	,227
3	11,917	7	,103
4	7,157	7	,413

Fonte: Output do Software SPSS.

Na tabela 23 a classificao do modelo aponta na etapa 4 um percentual global A=84,7% conforme (5.34) que so os casos corretamente classificados. Para B esse valor  de 50,34% que so os casos classificados corretamente por acaso, conforme citado em (5.35). A diferena entre eles (A-B)  de 34,36% > 25% e

permite confirmar uma boa capacidade do modelo. A sensibilidade e especificidade citados em (5.36) e (5.37) assumem valores de 88,37% e 87,18% respectivamente, ou seja, o modelo está prevendo corretamente a característica nos indivíduos que o possuem.

Tabela 23: Classificação do modelo segundo o critério de seleção *Forward* condicional.

Tabela de Classificação^a

Observado			Previsto		
			Obesidade		Porcentagem correta
			Normal	Obeso	
Etapa 1	Obesidade	Normal	29	13	69,0
		Obeso	5	38	88,4
	Porcentagem global				78,8
Etapa 2	Obesidade	Normal	31	11	73,8
		Obeso	6	37	86,0
	Porcentagem global				80,0
Etapa 3	Obesidade	Normal	35	7	83,3
		Obeso	6	37	86,0
	Porcentagem global				84,7
Etapa 4	Obesidade	Normal	34	8	81,0
		Obeso	5	38	88,4
	Porcentagem global				84,7

a. O valor de recorte é ,500

Fonte: Output do Software SPSS.

De acordo com avaliação da Curva Roc e pela tabela 24 a área da curva igual a 0,869 com p-valor de 0.000 considerando um $\alpha=0,05$, representa conforme critério do autor Hosmer e Lemeshow (2000) citado na figura 07, uma boa capacidade de ajuste do modelo. A área de 0,869 demonstra o quanto a curva de Roc está distante da reta diagonal do gráfico, o que aponta um poder classificatório do modelo. Essa representação pode ser vista na figura 17.

Tabela 24: Área sobre a curva Roc segundo o critério de seleção *Forward* condicional.

Área sob a curva

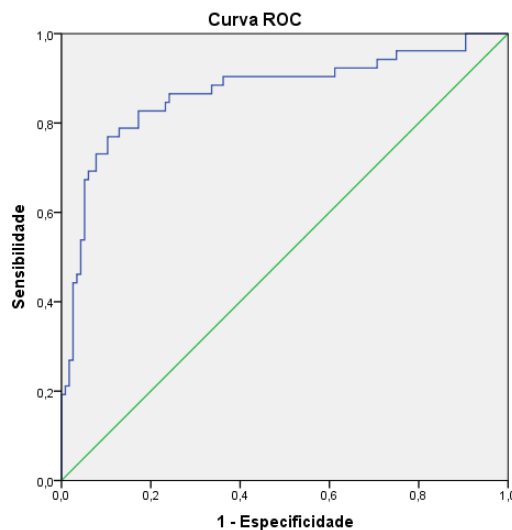
Variável(eis) de resultado de teste: Probabilidade prevista

Área	Erro Padrão ^a	Sig. assintótico ^b	Intervalo de Confiança 95% Assintótico	
			Limite inferior	Limite superior
,869	,034	,000	,802	,936

a. Sob a suposição não paramétrica

b. Hipótese nula: área real = 0,5

Fonte: Output do Software SPSS



Fonte: Output do Software SPSS.

Figura 19: Gráfico da Curva Roc segundo o critério de seleção *Forward* condicional.

O teste de significância de Kappa apresentou na tabela 25 um p-valor < 0,05, resultando em um $K > 0$ a um nível de 5%, segundo hipótese nula citada em (5.39). Sendo o $K > 0$, e o valor do coeficiente Kappa = 0,666 afirma-se que existe uma concordância substantiva tomando como base a tabela 07.

Tabela 25: Tabulação cruzada e a concordância de K de Cohen segundo o critério de seleção Forward Condicional no SPSS.

Obesidade * Grupo previsto Tabulação cruzada

Contagem

		Grupo previsto		Total
		Normal	Obeso	
Obesidade	Normal	104	12	116
	Obeso	12	40	52
Total		116	52	168

Medidas Simétricas

		Valor	Significância Erro Padrão ^a	Aprox. X ^b	Aprox. Sig.
Medida de concordância	Kappa	,666	,063	8,630	,000
N de Casos Válidos		168			

a. Não considerando a hipótese nula.

b. Uso de erro padrão assintótico considerando a hipótese nula.

Fonte: Output do Software SPSS

No estudo das variáveis da equação com base na etapa 4, descrito como o melhor modelo definido pelo critério, as variáveis independentes inclusas foram a Idade, Haptoglobina, Atividade Fosfatase e Levis conforme tabela 26.

Ainda de acordo com a tabela 26 na etapa 4, através do teste de *Wald* cujas hipóteses estão referenciadas em (5.25) conclui-se que as variáveis Idade ($\chi^2_{\text{wald}}=20,536$; $p = 0,000$), atividade fosfatase ACP1 ($\chi^2_{\text{wald}}= 6,546$; $p = 0,011$) e Levis (1) ($\chi^2_{\text{wald}}=4,065$; $p = 0,044$), apresentam um efeito significativo sobre a probabilidade de contribuir com a patologia de obesidade ao contrário da Haptoglobina que não contribui de forma significativa.

Tabela 26: Variáveis da equação do modelo segundo o critério de seleção *Forward* condicional.

		Variáveis na equação						95% C.I. para EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Inferior	Superior
Etapa 1 ^a	Idade	,074	,015	23,906	1	,000	1,077	1,046	1,110
	Constante	-3,359	,766	19,212	1	,000	,035		
Etapa 2 ^b	Idade	,087	,018	24,037	1	,000	1,091	1,054	1,130
	Atividade								
	Fosfatase (mmolGhbH)	-,005	,003	3,806	1	,051	,995	,989	1,000
	Constante	-2,266	,914	6,149	1	,013	,104		
Etapa 3 ^c	Idade	,109	,023	21,712	1	,000	1,115	1,065	1,167
	Atividade								
	Fosfatase (mmolGhbH)	-,008	,003	6,024	1	,014	,992	,986	,998
	Sist. Lewis a+b-			5,591	2	,061			
	Sist. Lewis a-b+	-1,874	,888	4,451	1	,035	,154	,027	,875
	Sist. Lewis a-b-	,000	,758	,000	1	1,000	1,000	,226	4,421
Etapa 4 ^d	Constante	-1,961	1,064	3,396	1	,065	,141		
	Idade	,127	,028	20,536	1	,000	1,135	1,075	1,199
	Haptoglobina 1.1			5,753	2	,056			
	Haptoglobina 2.1	,295	,863	,117	1	,732	1,343	,248	7,284
	Haptoglobina 2.2	-1,527	,919	2,761	1	,097	,217	,036	1,316
	Atividade								
	Fosfatase (mmolGhbH)	-,009	,004	6,546	1	,011	,991	,984	,998
	Sist. Lewis a+b-			5,780	2	,056			
	Sist. Lewis a-b+	-1,899	,942	4,065	1	,044	,150	,024	,948
	Sist. Lewis a-b-	,311	,827	,141	1	,707	1,365	,270	6,910
Constante	-2,259	1,478	2,336	1	,126	,104			

a. Variável(is) inserida(s) na etapa 1: Idade.

b. Variável(is) inserida(s) na etapa 2: FosfacidACP1activitymmolGhbH.

c. Variável(is) inserida(s) na etapa 3: Lewis.

d. Variável(is) inserida(s) na etapa 4: Haptogl.

Fonte: Output do Software SPSS.

Com base no Odds Ratio presente na tabela 26 como $\text{Exp}(\hat{\beta})$ conclui-se a seguir a relação das variáveis significativas com a obesidade.

$\widehat{OR}_{\text{Idade}} = 1,135$ e I.C._{95%} = (1,075; 1,199) o que significa que a cada ano de idade a mais, o risco para a obesidade do paciente aumenta em 13,5%.

$\widehat{OR}_{\text{Atividade Fosfatase}} = 0,991$ e I.C._{95%} = (0,984; 0,998) indica que a cada unidade a mais de mmol/g Hb/ H, têm-se menos risco de diagnóstico de obesidade, sendo essa chance de - 0,9 %, tornando-o um fator protetor.

$\widehat{OR}_{\text{Lewis a-b+}} = 0,150$, com I.C._{95%} = (0,24; 0,948) representando indivíduos do tipo sanguíneo a-b+ tendo como referência o tipo sanguíneo a+b-. Isso mostra que a chance de uma pessoa com o tipo sanguíneo a-b+ desenvolver a patologia de obesidade é de -85% em relação ao tipo sanguíneo a+b-, tornando-o um fator protetor a doença com relação ao tipo sanguíneo de referência a-b+.

O modelo de regressão logística com base no critério de seleção *Forward Condicional* é definido como:

$$\widehat{g(X)} = \text{Logit}(\hat{\pi}) = -2,259 + 0,127 \text{ idade} + 0,295 \text{ Haptoglobina 2.1} - 1,527 \text{ Haptoglobina 2.2} - 0,09 \text{ atividade fosfatase ácida} - 1,899 \text{ Lewis (a-b+)} + 0,311 \text{ Lewis (a-b-)}.$$

Conforme aplicado o critério *Forward Condicional* verifica-se que das variáveis significativas, as que apresentam o fator risco para a obesidade são a idade, a atividade fosfatase, e o sistema sanguíneo Lewis, destacando-se a variável independente idade que apresenta o risco de 13,5% a cada ano de acréscimo de uma unidade na idade do paciente. O tipo sanguíneo Lewis a-b+ destaca-se como um fator protetor em relação ao seu referencial tipo sanguíneo a+b-. A variável Haptoglobina não apresentou significância na etapa 4 desse critério.

Na análise dos resíduos do modelo apresentado na figura 20, é visível que existem várias observações que apresentam candidaturas a *outlier*, pois as mesmas assumem valores do resíduo padrão acima do valor 2 e podem ser vista na tabela 27.

Tabela 27: Lista dos casos de outliers segundo o critério de seleção Forward Condicional.

Lista entre casos^b

Caso	Status selecionado ^a	Observado	Previsto	Grupo previsto	Variável temporária	
		Obesidade			Resid	ZResid
63	S	N**	,940	O	-,940	-3,949
75	S	N**	,911	O	-,911	-3,194
83	S	N**	,923	O	-,923	-3,473
188	S	O**	,147	N	,853	2,413
208	S	O**	,132	N	,868	2,562
616	S	O**	,024	N	,976	6,423
632	S	O**	,005	N	,995	14,220
637	S	O**	,000	N	1,000	50,238
640	S	O**	,000	N	1,000	49,625
653	S	O**	,008	N	,992	11,420

a. S = Selecionado, U = Casos não selecionados, e ** = Casos classificados incorretamente.

b. São listados os casos com resíduos estudentizados maiores que 2,000.

Fonte: Output do Software SPSS.

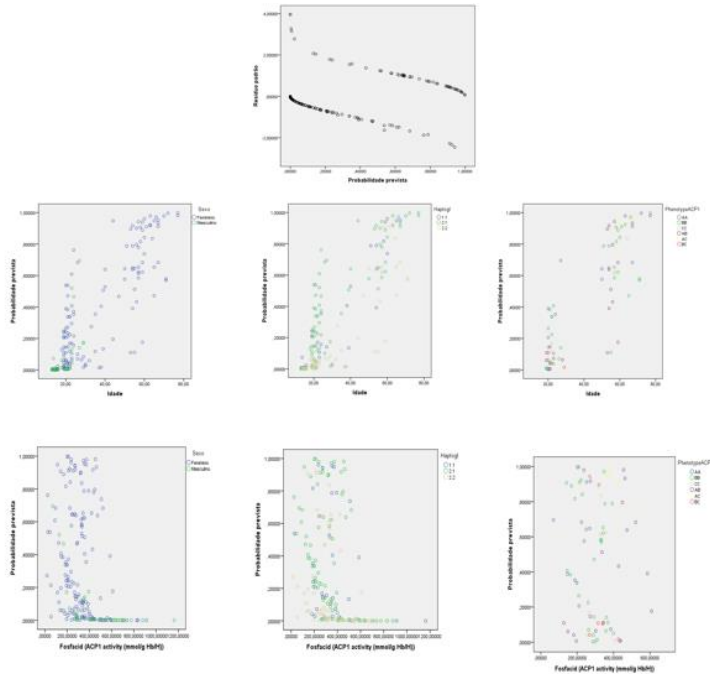


Figura 20: Gráficos da análise dos resíduos segundo o critério de seleção Forward Condicional.

Avaliando a Distância de Cook de acordo com a figura 21, alguns desses outlier assumem distância superior a 1, assumindo valores 2 atingindo até a distância de 5. Estes outliers podem apontar influência sobre os coeficientes do modelo e a qualidade do seu ajuste.

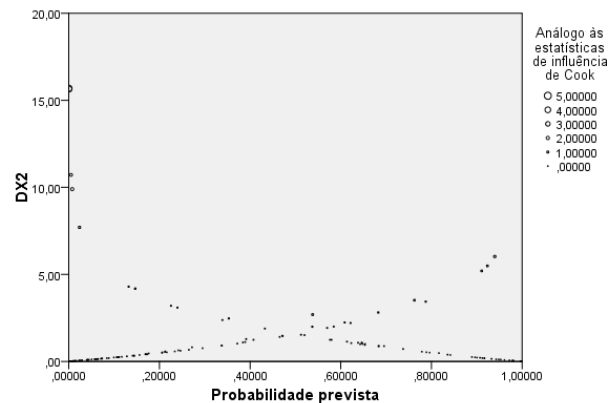


Figura 21: Gráficos da distância de Cook segundo o critério de seleção *Forward Condicional*.

MÉTODO *BACKWARD* CONDICIONAL

O método *Backward* Condicional define sua análise também em etapas, iniciando um modelo com todas as variáveis e retirando em seguida a variável de menor significância até definir o melhor modelo. O critério de seleção neste estudo, definiu seu melhor modelo depois de percorrer três etapas.

O teste de coeficiente de modelo Omnibus na tabela 28, em todas as suas etapas apresentaram um p-valor = 0,000, sendo um p-valor < 0,05, conforme teste de hipóteses citado em (5.31), o que representa que a inclusão das variáveis uma a uma em cada etapa conclui-se um modelo significativo.

Tabela 28: Teste de Omnibus sob o critério de seleção *Backward condicional*.

Testes de coeficientes de modelo Omnibus				
		Qui-quadrado	df	Sig.
Etapa 1	Etapa	63,311	14	,000
	Bloco	63,311	14	,000
	Modelo	63,311	14	,000
Etapa 2 ^a	Etapa	-,012	1	,912
	Bloco	63,299	13	,000
	Modelo	63,299	13	,000
Etapa 3 ^a	Etapa	-2,695	2	,260
	Bloco	60,604	11	,000
	Modelo	60,604	11	,000

a. Um valor de qui-quadrados negativos indica que o valor de chi-quadrado diminuiu a partir da etapa anterior.

Fonte: Output do Software SPSS.

A razão de verossimilhança entre as etapas apresentadas na tabela 29, permite concluir que a etapa 1, com ($\log -2 = 54,512$) apresenta o melhor modelo, visto que quanto menor o valor da razão de verossimilhança melhor. Através do teste de hipóteses citado em (5.23) é possível verificar que para $\alpha = 0,05$ que não se rejeita H_0 , tendo um modelo ajustado aos dados. O Pseudo R Cox e Snell e no Pseudo de Nagelkerke ($R_{CS} = 0,525$ e o $R_N = 0,700$) conforme referência em (5.29) e (5.30) apontando boa adequabilidade do modelo aos dados.

Tabela 29: Razão de Verossimilhança, Pseudo R Cox e Snell, Pseudo de Nagelkerke pelo critério de seleção *Backward condicional*.

Resumo do modelo			
Etapa	Verossimilhança de log -2	R quadrado Cox & Snell	R quadrado Nagelkerke
1	54,512 ^a	,525	,700
2	54,524 ^a	,525	,700
3	57,219 ^a	,510	,680

a. Estimação finalizada no número de iteração 20 porque o máximo de iterações foi atingido. Não é possível encontrar a solução final.

Fonte: Output do Software SPSS.

No teste de ajustamento de Hosmer e Lemeshow apresentado na tabela 30, conforme hipóteses citadas em (5.33), todas as etapas o p-valor > 0,05, o que indica que não se rejeita a hipótese nula considerando um $\alpha = 0,05$, logo os dados se ajustam a cada modelo proposto sendo considerando bons.

Tabela 30: Teste de Hosmer e Lemeshow sob o critério de seleção *Backward condicional*.

Etapa	Qui-quadrado	df	Sig.
1	6,123	7	,525
2	6,118	7	,526
3	2,988	7	,886

Fonte: Output do Software SPSS.

Na tabela 31 a classificação do modelo aponta na etapa 1 e 2 um percentual global A=87,1% conforme (5.34) que são os casos corretamente classificados. Para B esse valor é de 50,06% que são os casos classificados corretamente por acaso, conforme citado em (5.35). A diferença entre eles (A-B) é de 37,04% > 25% o que confirma uma boa capacidade do modelo. A sensibilidade e especificidade citados em (5.36) e (5.37) assumem valores de 88,37% e 87,80% respectivamente, ou seja, o modelo está prevendo corretamente a característica nos indivíduos que o possuem.

Tabela 31: Classificação do modelo segundo o critério de seleção *Backward condicional*.

Observado			Previsto		Porcentagem correta
			Normal	Obeso	
Etapa 1	Obesidade	Normal	36	6	85,7
		Obeso	5	38	88,4
Porcentagem global					87,1
Etapa 2	Obesidade	Normal	36	6	85,7
		Obeso	5	38	88,4
Porcentagem global					87,1
Etapa 3	Obesidade	Normal	32	10	76,2
		Obeso	5	38	88,4
Porcentagem global					82,4

a. O valor de recorte é ,500

Fonte: Output do Software SPSS.

De acordo com avaliação da curva Roc e pela tabela 32 a área da curva é igual a 0,931 com p-valor de 0.000 considerando um $\alpha=0,05$, significa conforme critério baseado no autor Hosmer e Lemeshow (2000) citado na figura 07, uma boa capacidade de ajuste do modelo. A área de 0,931 demonstra o quanto a curva de Roc está distante da reta diagonal do gráfico, o que demonstra um poder classificatório do modelo. Essa representação pode ser vista na figura 18.

Tabela 32: Área sobre a curva Roc segundo o critério de seleção *Backward condicional*.

Área sob a curva

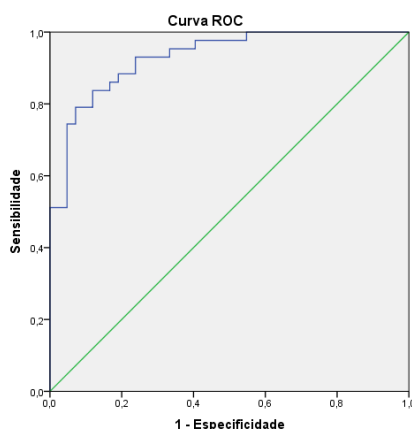
Variável(eis) de resultado de teste: Probabilidade prevista

Área	Erro Padrão ^a	Sig. assintótico ^b	Intervalo de Confiança 95% Assintótico	
			Limite inferior	Limite superior
,931	,026	,000	,880	,981

a. Sob a suposição não paramétrica

b. Hipótese nula: área real = 0,5

Fonte: Output do Software SPSS



Fonte: Output do Software SPSS.

Figura 22: Gráfico da Curva Roc segundo o critério de seleção *Backward condicional*.

O teste de significância de Kappa apresentou na tabela 33 um p-valor $< 0,05$, resultando em um $K > 0$ a um nível de 5%, segundo hipótese nula citada em (5.39).

Sendo o $K > 0$, e o valor do coeficiente Kappa = 0,647 conclui-se que existe uma concordância substantiva tomando como base a tabela 07.

Tabela 33: Tabulação cruzada e a concordância de K de Cohen segundo o critério de seleção *Backward Condicional*.

Obesidade * Grupo previsto Tabulação cruzada

Contagem

		Grupo previsto		Total
		Normal	Obeso	
Obesidade	Normal	32	10	42
	Obeso	5	38	43
Total		37	48	85

Medidas Simétricas

		Valor	Significância Erro Padrão ^a	Aprox. X ^b	Aprox. Sig.
Medida de concordância	Kappa	,647	,082	6,002	,000
N de Casos Válidos		85			

a. Não considerando a hipótese nula.

b. Uso de erro padrão assintótico considerando a hipótese nula.

Fonte: Output do Software SPSS

No estudo das variáveis da equação, com base a etapa 1, sendo o melhor modelo definido pelo critério observa-se a inclusão de todas as variáveis independentes conforme tabela 34.

Ainda na tabela 34 na etapa 1, através do teste de *Wald* cujas hipóteses estão referenciadas em (5.25) conclui-se que as variáveis Idade ($\chi^2_{\text{wald}}=15,589$; $p = 0,000$), Sistema Sanguíneo NN ($\chi^2_{\text{wald}}=4,465$; $p = 0,035$), Haptoglobina 2.2 ($\chi^2_{\text{wald}}=4,795$; $p = 0,029$), atividade fosfatase ACP1 ($\chi^2_{\text{wald}}= 6,410$; $p = 0,011$) apresentam significância sobre a probabilidade de contribuir com a patologia de obesidade.

Tabela 34: Variáveis da equação do modelo segundo o critério de seleção *Backward condicional*.

		Variáveis na equação						95% C.I. para EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Inferior	Superior
Etapa 1 ^a	Idade	,168	,043	15,589	1	,000	1,183	1,088	1,286
	Sexo(1)	,335	3,013	,012	1	,912	1,397	,004	512,477
	Sist. Sang. MM			4,831	2	,089			
	Sist. Sang. MN	,171	1,023	,028	1	,867	1,187	,160	8,807
	Sist. Sang. NN	-2,287	1,083	4,465	1	,035	,102	,012	,847
	Haptoglobina 1.1			7,735	2	,021			
	Haptoglobina 2.1	,579	,992	,341	1	,559	1,785	,255	12,480
	Haptoglobina 2.2	-2,672	1,220	4,795	1	,029	,069	,006	,755
	Atividade Fosfatase (mmolGhH)	-,013	,005	6,410	1	,011	,987	,977	,997
	Fenótipo ACP1 (AA)			6,467	5	,263			
	Fenótipo ACP1 (BB)	-2,163	1,115	3,767	1	,052	,115	,013	1,021
	Fenótipo ACP1 (CC)	,382	1,577	,059	1	,808	1,465	,067	32,229
	Fenótipo ACP1 (AB)	,520	1,323	,154	1	,694	1,682	,126	22,486
	Fenótipo ACP1 (AC)	-,484	1,698	,081	1	,776	,617	,022	17,207
	Fenótipo ACP1 (BC)	-19,136	21196,886	,000	1	,999	,000	,000	.
	Sist. Lewis a+b-			2,417	2	,299			
	Sist. Lewis a-b+	-,852	1,142	,556	1	,456	,427	,045	4,004
	Sist. Lewis a-b-	,841	1,004	,702	1	,402	2,318	,324	16,574
	Constante	-1,805	1,701	1,126	1	,289	,165		
	Etapa 2 ^a	Idade	,168	,043	15,519	1	,000	1,183	1,088
Sist. Sang. MM				4,873	2	,087			
Sist. Sang. MN		,168	1,023	,027	1	,870	1,183	,159	8,777
Sist. Sang. NN		-2,295	1,080	4,513	1	,034	,101	,012	,837
Haptoglobina 1.1				7,765	2	,021			
Haptoglobina 2.1		,575	,993	,336	1	,562	1,778	,254	12,444
Haptoglobina 2.2		-2,680	1,219	4,833	1	,028	,069	,006	,748
Atividade Fosfatase (mmolGhH)		-,013	,005	6,384	1	,012	,987	,977	,997
Fenótipo ACP1 (AA)				6,450	5	,265			
Fenótipo ACP1 (BB)		-2,169	1,113	3,796	1	,051	,114	,013	1,013
Fenótipo ACP1 (CC)		,372	1,575	,056	1	,813	1,451	,066	31,778

	Fenótipo ACP1 (AB)	,508	1,320	,148	1	,700	1,663	,125	22,080
	Fenótipo ACP1 (AC)	-,499	1,692	,087	1	,768	,607	,022	16,749
	Fenótipo ACP1 (BC)	-19,154	21181,854	,000	1	,999	,000	,000	.
	Lewis			2,400	2	,301			
	Lewis(1)	-,859	1,141	,566	1	,452	,424	,045	3,969
	Lewis(2)	,828	,994	,693	1	,405	2,288	,326	16,065
	Constante	-1,767	1,661	1,131	1	,288	,171		
Etapa	Idade	,145	,036	16,016	1	,000	1,156	1,077	1,241
3 ^a	Sist. Sang. MM			6,143	2	,046			
	Sist. Sang. MN	-,020	,991	,000	1	,984	,980	,141	6,836
	Sist. Sang. NN	-2,578	1,068	5,829	1	,016	,076	,009	,616
	Haptoglobina 1.1			7,624	2	,022			
	Haptoglobina 2.1	,286	,911	,098	1	,754	1,331	,223	7,938
	Haptoglobina 2.2	-2,767	1,195	5,366	1	,021	,063	,006	,653
	Atividade Fosfatase (mmolHgHbH)	-,011	,005	5,235	1	,022	,990	,981	,998
	Fenótipo ACP1 (AA)			7,756	5	,170			
	Fenótipo ACP1 (BB)	-2,261	1,082	4,370	1	,037	,104	,013	,868
	Fenótipo ACP1 (CC)	,477	1,542	,096	1	,757	1,612	,079	33,078
	Fenótipo ACP1 (AB)	,254	1,276	,040	1	,842	1,289	,106	15,725
	Fenótipo ACP1 (AC)	-,529	1,607	,108	1	,742	,589	,025	13,749
	Fenótipo ACP1 (BC)	-19,367	20894,094	,000	1	,999	,000	,000	.
	Constante	-,965	1,342	,517	1	,472	,381		

a. Variável(is) inserida(s) na etapa 1: Idade, Sexo, MNGrupo, Haptogl, FosfidACP1activitymmolHgHbH, PhenotypeACP1, Lewis.

Fonte: Output do Software SPSS.

Com base no Odds Ratio presente na tabela 34 como $\text{Exp}(\hat{\beta})$ conclui-se a seguir a relação das variáveis significativas com a obesidade baseado na etapa 1.

$\widehat{OR}_{\text{Idade}} = 1,183$ e $\text{I.C.}_{95\%} = (1,088; 1,286)$ o que significa que a cada ano acrescido na idade, o risco para a obesidade do paciente aumenta em 18,3%.

$\widehat{OR}_{\text{Sist. Sanguíneo NN}} = 0,102$, com $\text{I.C.}_{95\%} = (0,12; 0,847)$ representa indivíduos do tipo sanguíneo NN tendo como referência o tipo sanguíneo MM. Isso mostra que a chance de uma pessoa com o tipo sanguíneo NN desenvolver a patologia de obesidade é de -89,8% em relação ao tipo sanguíneo MM, tornando-o um fator protetor a doença com relação ao tipo sanguíneo de referência.

$\widehat{OR}_{\text{Haptoglobina 2.2}} = 0,069$, com I.C._{.95%} = (0,06; 0,75) são indivíduos com o fenótipo Hp 2.2 tendo como referência o fenótipo Hp 1.1 que se apresenta como um fator de risco. Essa medida de 0,069 indica que a chance de um paciente com o fenótipo Hp 2.2 desenvolver a patologia de obesidade é de -93,1% relativamente a quem tem o fenótipo Hp1.1, ou seja, tem-se menor chance de apresentar a doença, tornando-se o fenótipo Hp2.2 um fator protetor.

$\widehat{OR}_{\text{Atividade Fosfatase}} = 0,987$ com I.C._{.95%} = (0,977; 0,997) indica que a cada unidade a mais de mmol/g Hb/ H, têm-se menos risco de diagnóstico de obesidade, sendo essa chance de - 0,013 %, o que torna um fator protetor.

O modelo de regressão logística com base no critério de seleção *Backward* Condicional é definido como:

$$\widehat{g(X)} = \text{Logit}(\hat{\pi}) = -1,805 + 0,168 \text{ idade} + 0,355 \text{ sexo} + 0,171 \text{ Sist. Sanguíneo MN} - 2,287 \text{ Sistema Sanguíneo NN} + 0,579 \text{ Haptoglobina 2.1} - 2,672 \text{ Haptoglobina 2.2} - 0,13 \text{ atividade fosfatase} - 2,163 \text{ Fenótipo ACP1(BB)} + 0,382 \text{ Fenótipo ACP1(CC)} + 0,520 \text{ Fenótipo ACP1(AB)} - 0,484 \text{ Fenótipo ACP1(AC)} - 19,136 \text{ Fenótipo ACP1(BC)} - 0,852 \text{ Lewis (a+b-)} + 0,841 \text{ Lewis(a-b-)}$$

No critério *Backward Condicional* a variável que apresenta o fator risco para a obesidade é a idade potenciando isso em 18,3% a cada ano a mais de vida. As variáveis atividade fosfatase e Lewis representam um fator protetor.

Conforme aplicado o critério *Forward Condicional* verifica-se que das variáveis significativas, as que apresentam o fator risco para a obesidade são a idade, a atividade fosfatase, e o sistema sanguíneo Lewis, destacando-se a variável independente idade com risco de 13,5% a cada ano de acréscimo de uma unidade na idade do paciente. O tipo sanguíneo Lewis a-b+ se destaca como um fator protetor em relação ao seu referencial tipo sanguíneo a+b-. A variável Haptoglobina não apresentou significância.

Na análise dos resíduos do modelo apresentado na figura 23, descreve observações que apresentam candidaturas a *outlier*, pois as mesmas assumem valores de resíduo acima do valor 2 que seria um referencial. Estas observações candidatas podem ser vista na tabela 35.

Tabela 35: Lista dos casos de outliers segundo o critério de seleção *Backward Condicional*.

Lista entre casos^b

Caso	Status selecionado ^a	Observado	Previsto	Grupo previsto	Variável temporária	
		Obesidade			Resid	ZResid
63	S	N**	,835	O	-,835	-2,253
90	S	N**	,862	O	-,862	-2,500
194	S	O**	,077	N	,923	3,473
201	S	O**	,138	N	,862	2,504

a. S = Selecionado, U = Casos não selecionados, e ** = Casos classificados incorretamente.

b. São listados os casos com resíduos estudentizados maiores que 2,000.

Fonte: Output do Software SPSS.

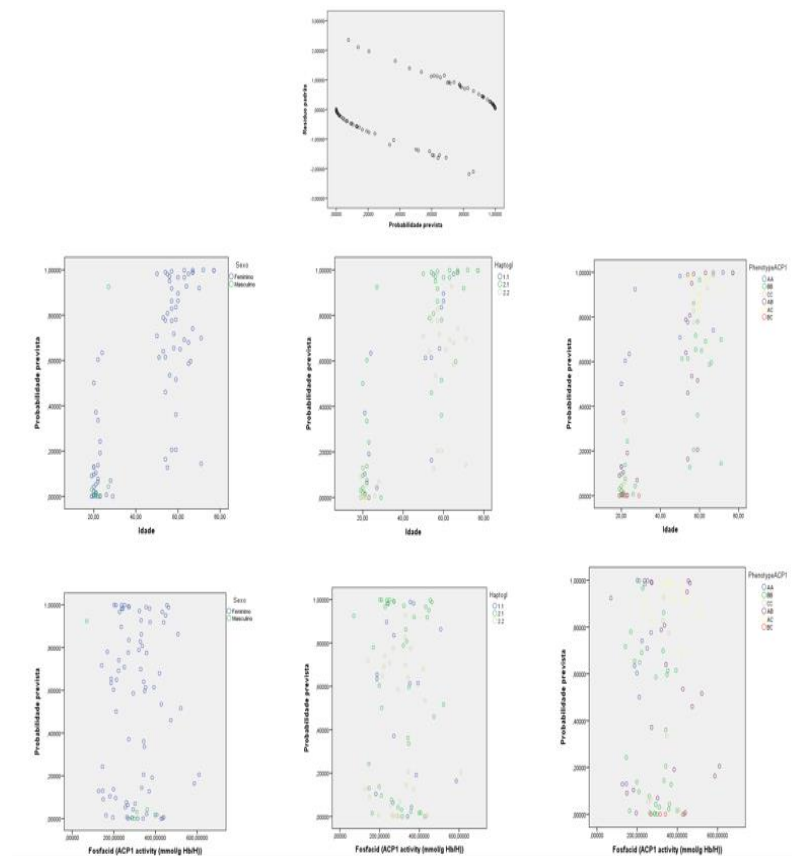


Figura 23: Gráficos da análise dos resíduos segundo o critério de seleção *Backward Condicional*.

Avaliando a Distância de Cook de acordo com a figura 24, alguns desses outliers assumem distância superior a 1, assumindo valores máximos de 2. Os outliers podem apontar influência sobre os coeficientes do modelo e a qualidade do seu ajuste.

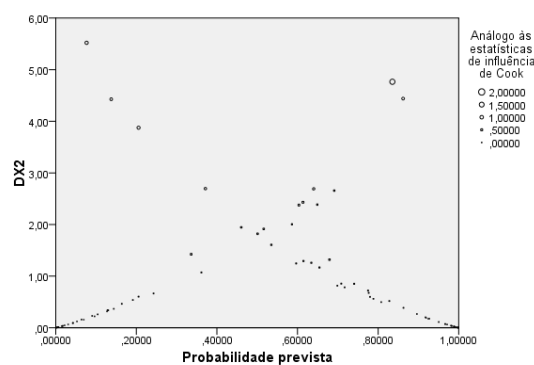


Figura 24: Gráficos da distância de Cook segundo o critério de seleção *Backward Condicional*.

A seguir será avaliado com base na análise dos critérios de seleção *Enter*, *Forward* condicional e *Backward Condicional* qual o melhor modelo indicado para identificar os fatores de risco para obesidade.

6.4.4 ESCOLHA DO MODELO MAIS ADEQUEDO

Para concluir sobre qual o modelo mais apropriado ou indicado para identificar alguns fatores de risco para obesidade, inicialmente realizaram-se três critérios de seleção distintos, a saber o *Enter*, *Forward Condicional* e o *Backward Condicional*. Uma vez realizado cada critério de seleção, foram avaliados indicadores importantes que apontam a qualidade do modelo e serve de base para uma melhor seleção. Os indicadores adotados como apoio para escolha foram a

Classificação do Modelo, Curva Roc, Coeficiente Kappa, Sensibilidade, Especificidade, e a Razão de Verossimilhança.

De acordo com os indicadores na tabela 36, observa-se que não existem diferenças significativas nos quesitos sensibilidade, especificidade e classificação do modelo. Ao avaliar a curva Roc e o coeficiente Kappa os modelos são considerados significativos, logo a maneira que cabe diante desse estudo para identificar alguns fatores de risco para obesidade, será aplicando os critérios da Regressão Logística que aponta a melhor seleção como um modelo significativo, todavia o modelo mais simples possível. Assim sendo, a escolha e sugestão será pelo critério de seleção *Forward Condicional* por apresentar significância e ou qualidade no ajuste e um modelo com 6 variáveis apenas.

Com relação aos resíduos, o modelo escolhido como o mais adequado, foi o que mais apresentou outliers como constatado na tabela 27. Esse critério também registrou os maiores resultados na distância de Cook apresentando na figura 21. É importante ressaltar que mesmo existindo estes outliers que implicam na construção os coeficientes do modelo, estes valores de outliers são reais, visto que os dados dos indivíduos que fazem parte desse estudo possuem fonte segura.

Tabela 36: Indicadores dos critérios de seleção de variáveis.

Critérios Indicadores	<i>Enter</i>	<i>Forward Condicional</i>	<i>Backward Condicional</i>
Razão de Verossimilhança	54,512	67,719	57,219
Sensibilidade	88,37%	88,37%	88,37%
Especificidade	87,80%	87,18%	87,80%
Classificação do modelo	87,10%	84,70%	82,40%
Curva Roc	0,931	0,869	0,931
Coeficiente Kappa	0,741	0,666	0,647
Variáveis no modelo	14	6	11

Fonte: Criada pela autora

O modelo de regressão logística selecionado e sugerido neste estudo como o mais adequado é dado por:

$$\widehat{g(X)} = \text{Logit}(\hat{\pi}) = -2,259 + 0,127 \text{ idade} + 0,295 \text{ Haptoglobina 2.1} - 1,527 \text{ Haptoglobina 2.2} - 0,009 \text{ atividade fosfatase ácida} - 1,899 \text{ Lewis (a-b+)} + 0,311 \text{ Lewis (a-b-)}.$$

No próximo capítulo serão aplicadas Técnicas da Visualização da Informação nas variáveis de interesse que também é parte desse estudo.

Capítulo 7

7. APLICANDO A VISUALIZAÇÃO DA INFORMAÇÃO NA BASE DE DADOS

Neste capítulo será realizada a Visualização da Informação através das ferramentas selecionadas *Tableau Public* e *Power BI*. Estas ferramentas foram escolhidas por serem relevantes a nível mundial assim como em estudos científicos. Aqui busca-se demonstrar representações visuais que apoiem e facilite nas interpretações dos dados estatísticos possibilitando uma tomada de decisão rápida, prática e objetiva, visto a necessidade real existente na área da saúde.

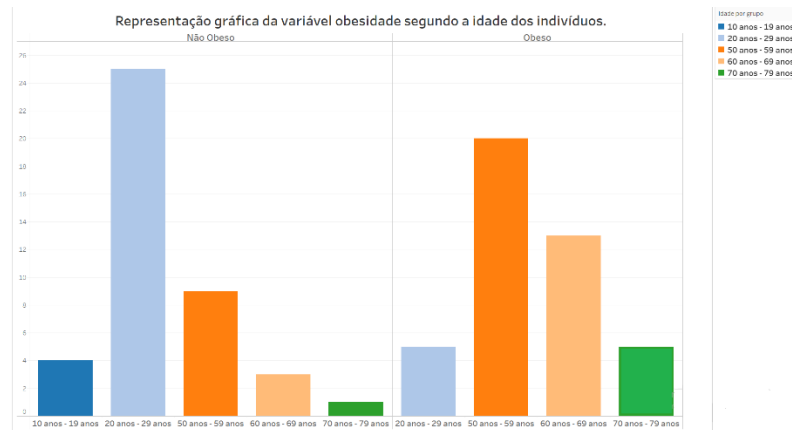
As aplicações realizadas a seguir serão da variável dependente obesidade relacionada a todos as variáveis independentes deste estudo citadas anteriormente.

7.1 USO DO *TABLEAU PUBLIC* NA APRESENTAÇÃO DOS DADOS

As representações gráficas com base no software *Tableau Public*, serão utilizadas para dar visibilidade as variáveis de importância neste estudo. O principal alvo é mostrar a relação existente entre a variável dependente obesidade com as demais variáveis. Será apresentado um gráfico da variável resposta cruzando com as variáveis independentes idade, género, grupo sanguíneo MN, grupo sanguíneo Lewis, atividade fosfatase ácida, Fenótipo ACP-1 e fenótipo de haptoglobina.

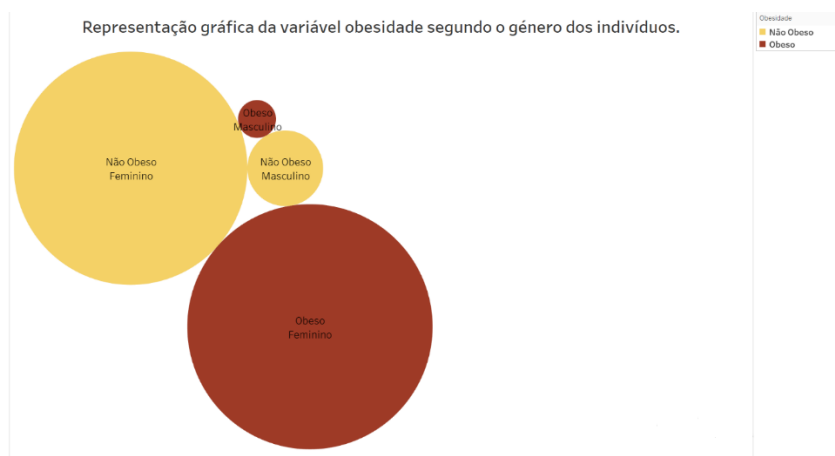
7.1.2 REPRESENTAÇÕES GRÁFICAS *TABLEAU PUBLIC*

Considerando o interesse em uma melhor interpretação dos dados através da Visualização da Informação serão desenvolvidos a seguir gráficos que permitam melhores análises sobre a relação existente entre a variáveis de interesse. A ferramenta aqui a ser utilizada é a *Tableau Public*.



Fonte: Output do *Software Tableau Public* versão 2020.1.

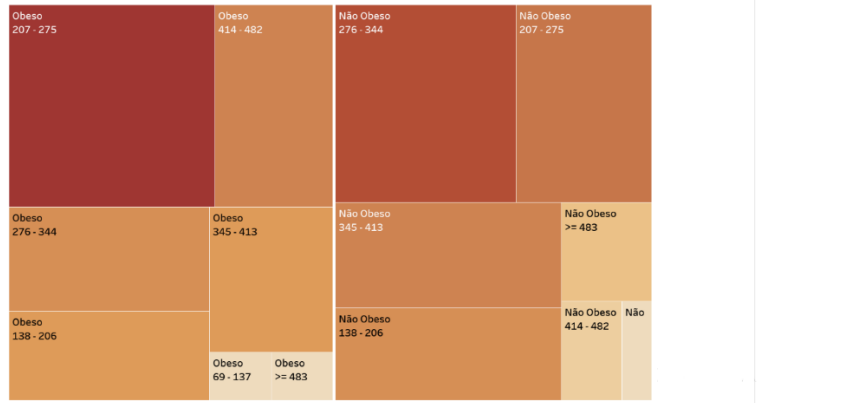
Figura 25: Gráfico da variável obesidade segundo a idade.



Fonte: Output do *Software Tableau Public* versão 2020.1.

Figura 26: Gráfico da variável obesidade segundo o sexo.

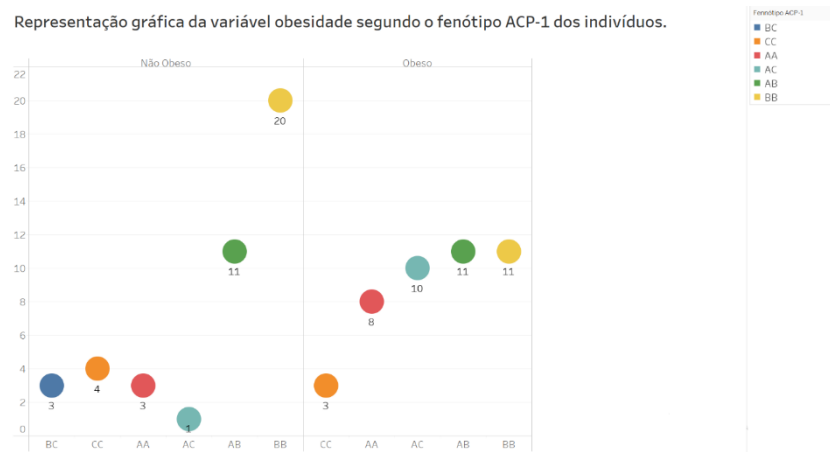
Representação gráfica da variável obesidade segundo grupo da atividade fosfatase ácida dos indivíduos.



Fonte: Output do *Software Tableau Public* versão 2020.1.

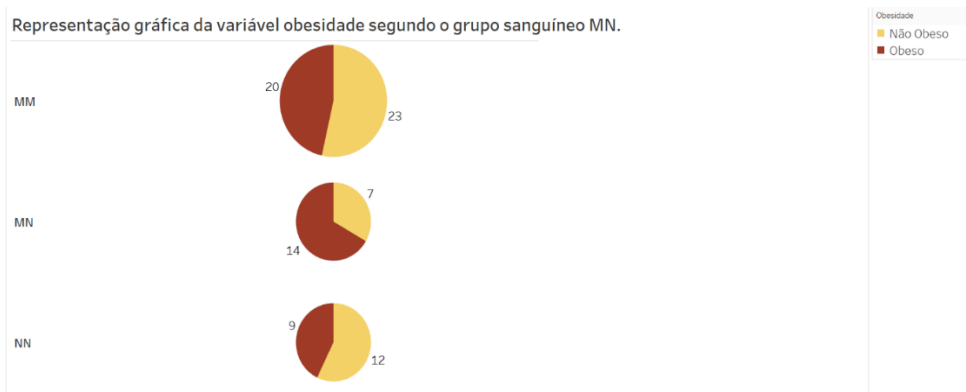
Figura 27: Gráfico da variável obesidade segundo a atividade fosfatase ácida.

Representação gráfica da variável obesidade segundo o fenótipo ACP-1 dos indivíduos.



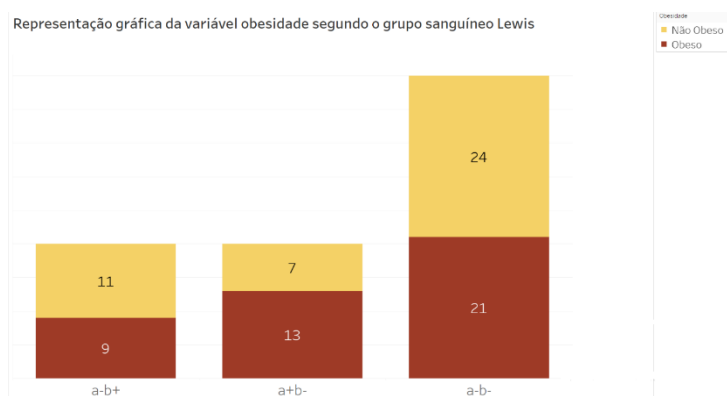
Fonte: Output do *Software Tableau Public* versão 2020.1.

Figura 28: Gráfico da variável obesidade segundo o fenótipo ACP-1.



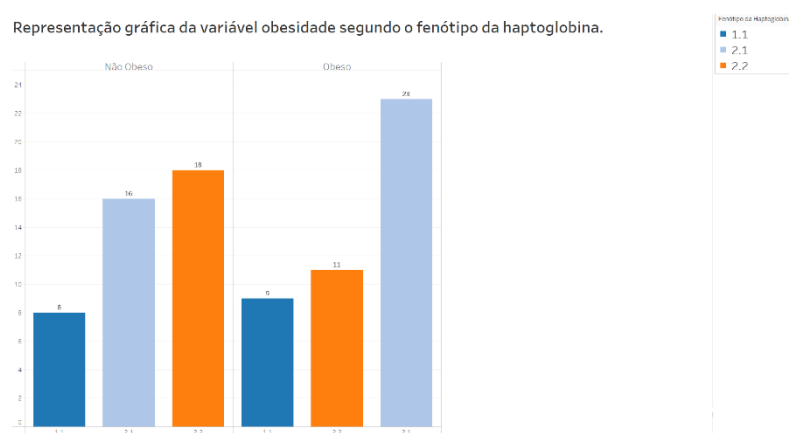
Fonte: Output do *Software Tableau Public* versão 2020.1.

Figura 29: Gráfico da variável obesidade segundo o grupo sanguíneo MN.



Fonte: Output do *Software Tableau Public* versão 2020.1.

Figura 30: Gráfico da variável obesidade segundo o grupo sanguíneo Lewis.



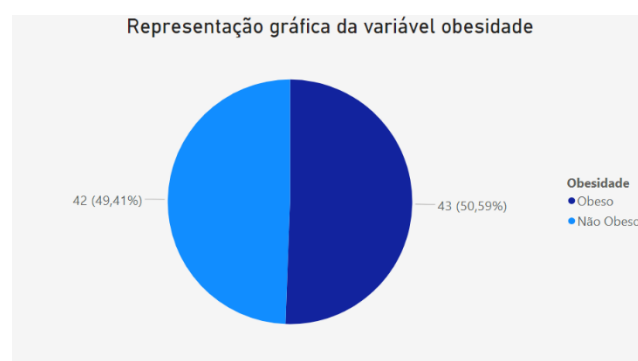
Fonte: Output do *Software Tableau Public* versão 2020.1.

Figura 31: Gráfico da variável obesidade segundo o fenótipo da haptoglobina.

7.2 USO DO *POWER BI* NA APRESENTAÇÃO DOS DADOS

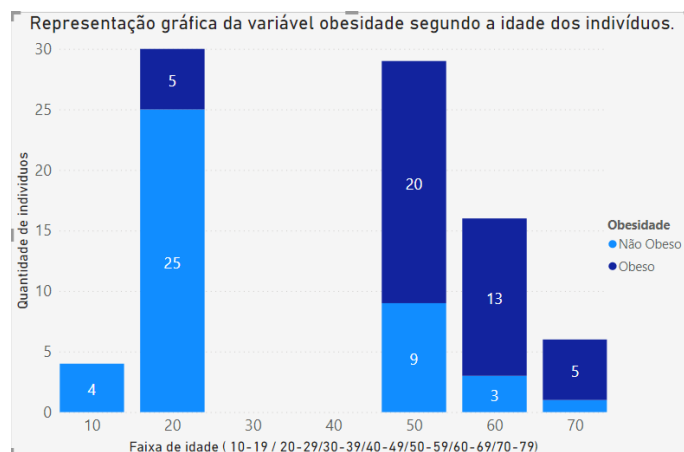
A representação gráfica apresentada nesta seção tem o mesmo foco da realizada anteriormente usando o *Tableau Public*, com o apoio agora do *Power BI*. Deseja-se apresentar gráficos da variável resposta obesidade em conjunto com as variáveis independentes idade, gênero, grupo sanguíneo MN, grupo sanguíneo Lewis, atividade fosfatase ácida, Fenótipo ACP-1 e fenótipo de haptoglobina.

7.2.1 REPRESENTAÇÕES GRÁFICAS *POWER BI*



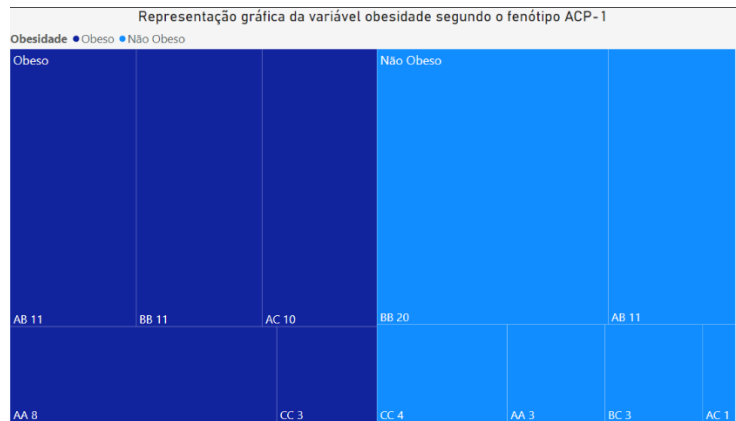
Fonte: Output do *Software Power BI* atualização 2.91.383.0.

Figura 32: Gráfico da variável obesidade.



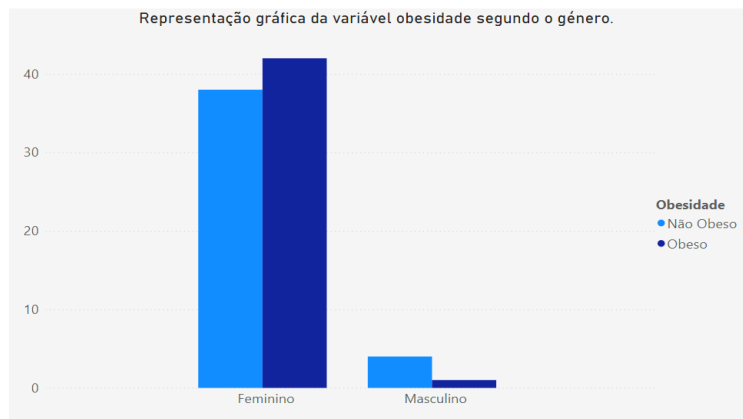
Fonte: Output do *Software Power BI* atualização 2.91.383.0.

Figura 33: Gráfico da variável obesidade segundo a idade.



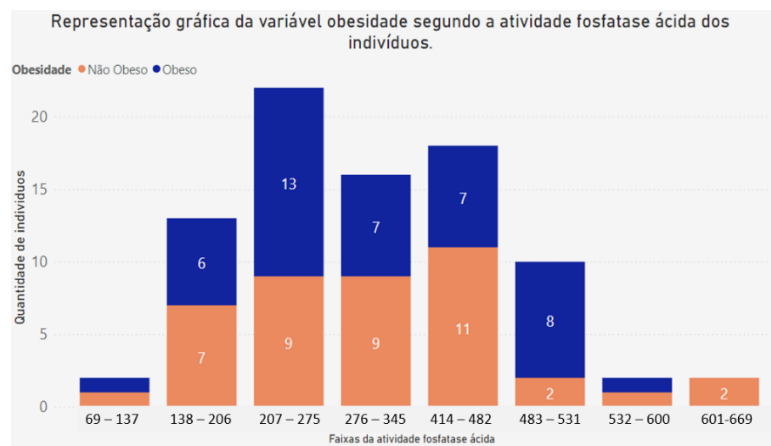
Fonte: Output do *Software Power BI* atualização 2.91.383.0.

Figura 34: Gráfico da variável obesidade segundo fenótipo ACP-1.



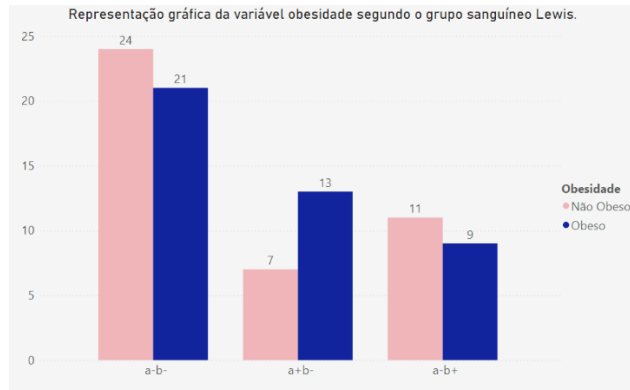
Fonte: Output do *Software Power BI* atualização 2.91.383.0.

Figura 35: Gráfico da variável obesidade segundo o gênero.



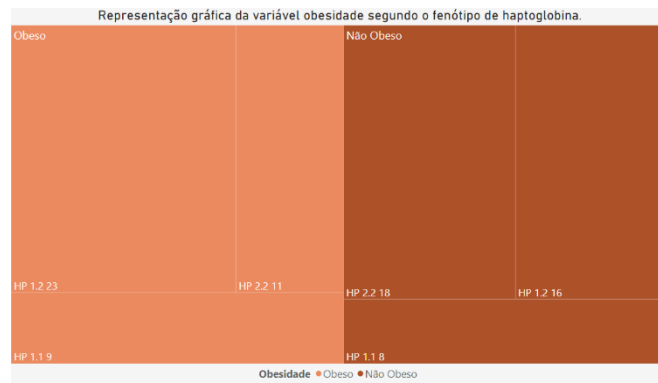
Fonte: Output do *Software Power BI* atualização 2.91.383.0.

Figura 36: Gráfico da variável obesidade segundo a atividade fosfatase ácida.



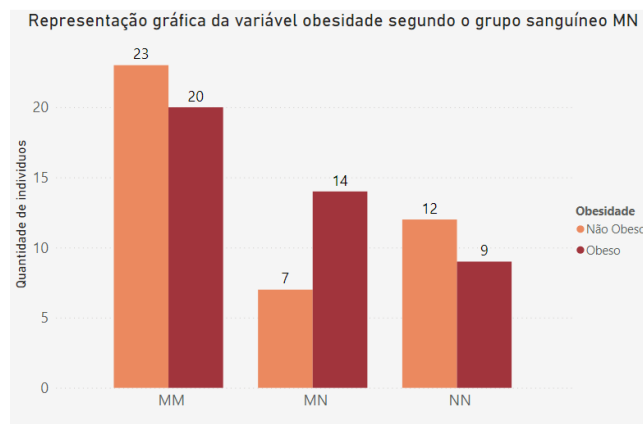
Fonte: Output do *Software Power BI* atualização 2.91.383.0.

Figura 37: Gráfico da variável obesidade segundo o grupo sanguíneo Lewis.



Fonte: Output do *Software Power BI* atualização 2.91.383.0.

Figura 38: Gráfico da variável obesidade segundo o fenótipo de haptoglobina.



Fonte: Output do *Software Power BI* atualização 2.91.383.0.

Figura 39: Gráfico da variável obesidade segundo o grupo sanguíneo MN.

7.3 MODELO VISUAL MAIS ADEQUADO AO ESTUDO.

O foco de propor um modelo visual mais adequado, objetiva uma melhor visualização da informação que potencie a capacidade de entendimento dos dados para quem vai analisar, trazendo maior dinâmica na tomada de decisões.

Para avaliar um bom modelo visual, foram selecionados dois softwares bastante utilizados atualmente que são o *Tableau Public* versão 2021.1 e *Power BI* atualização 2.91.383.0 conforme apresentação gráfica realizada anteriormente nos tópicos 7.1.2 e 7.2.1.

Com base no que foi construído anteriormente, o modelo visual do qual propõe e recomenda como o mais adequado quando se trata de dados voltados a área de saúde e/ou médica é o *software Tableau Public*.

A indicação do uso do *software Tableau Public* se dá pelas vantagens que o mesmo apresenta ao usuário quanto a sua construção gráfica e leitura do banco de dados. Estas vantagens foram possíveis avaliar durante a aplicação do banco de dados deste estudo e estão listadas a seguir:

- Facilidade e praticidade para o usuário manusear dados.
- Linguagem simplificada para cálculos.
- Permite uma interação fácil quando se deseja cruzar variáveis do estudo.
- Permite transformar com facilidade variáveis quantitativas em qualitativas.
- Apresenta variedade gráfica e sugestões visuais adequadas no foco de interesse possibilitando uma conclusão mais rápida.
- Simplicidade no visual e rapidez na resposta.
- Oferece muitos recursos quando se trata de detalhes na formatação.

Estas mesmas vantagens não foram possíveis observar na construção das representações gráficas sob as mesmas condições com o uso do *software Power BI*, visto que não foram encontradas as mesmas facilidades e simplicidades para conclusão gráfica e apresentação dos dados. A

plataforma *Power BI* não se apresenta de forma tão intuitiva para o usuário comum.

Algumas percepções com relação ao uso do *Power BI* durante o processo de aplicabilidade do banco de dados deste estudo estão relacionadas abaixo:

- Não disponibiliza uma relação entre as variáveis de forma tão simplificada.
- É preciso passar por muitas etapas até definir uma representação gráfica por completo.
- A reversão de uma variável qualitativa e quantitativa é limitada na definição das suas classes.
- Mais complexidade na realização de cálculos caso seja necessário.
- Ferramenta com foco em painéis, ou seja, permite criar várias visualizações simultâneas, mas o foco principal precisa ser o quanto eu consigo criar gráficos e interpretar com facilidade os dados.
- Mais difícil na interação com o usuário, não apresentando um destaque na relação sistema e indivíduo.

Pelos motivos relacionados acima proponha-se como um modelo adequado para representação gráfica, o uso do *Software Tableau Public* por apresentar facilidade de interação, rapidez e dinâmica que permite ao usuário sem muita complexidade chegar ao seu objetivo.

Capítulo 8

8. CONCLUSÃO

Com todo caminho desenvolvido neste trabalho, é possível perceber que a obesidade é uma patologia que precisa de uma atenção especial e que quanto mais estudos são realizados, mais próximos estamos de identificar tratamentos personalizados e proporcionar estrutura de apoio para construção de medidas preventivas aos futuros pacientes.

Atualmente identificar fatores que causam riscos à saúde dos indivíduos e aplicar medidas preventivas, pode significar uma redução de custos significativa nas esferas da saúde públicas, visto que a obesidade também pode levar a outras complicações e comorbidades.

Na investigação realizada neste trabalho, buscou-se identificar alguns fatores de risco para a obesidade com variáveis que envolvem a relação genética (Haptoglobina, grupo sanguíneo MN e grupo sanguíneo Lewis), clínica (Hipertensão, Diabetes e atividade fosfatase ácida) e sócio demográfico dos pacientes (Idade e gênero).

A identificação dos fatores de risco se deu através da construção do modelo matemático, com o uso da Regressão Logística com o critério *Forward Conditional*. Nesta proposta foi possível identificar as variáveis que foram relevantes em relação a variável resposta obesidade, ou seja, variáveis que podem apresentar ou não um fator de risco para obesidade.

Na tabela 45 na etapa 4 através do modelo *Forward Conditional* foi possível identificar as variáveis que apresentam risco a obesidade como a idade, Haptoglobina, atividade fosfatase ácida, e o sistema sanguíneo Lewis. Embora na etapa 4 tenha apontado o melhor modelo com a apresentação dessas variáveis, nem todas são significativas quando é realizado o teste de Wald cujos resultados se encontram apresentados na mesma tabela.

Com a avaliação do Teste de Wald e com o uso do *Odds Ratio* é possível determinar as variáveis que influenciam diretamente a variável dependente obesidade. Estas variáveis são a idade, atividade fosfatase, e o tipo sanguíneo Levis a-b+.

Com relação as variáveis identificadas acima podem-se observar que a idade faz parte do conjunto de variáveis sócio demográficas, e a atividade fosfatase ácida e o grupo sanguíneo Lewis fazem parte do grupo de variáveis genéticas.

Considerando as variáveis identificadas pelo modelo em estudo, destacou-se a variável independente Idade. Através do *Odds Ratio* (\widehat{OR})_{Idade} = 1,135 e I.C._{95%} = (1,075; 1,199) observado na mesma tabela 45, identificou-se que a variável idade é um fator de risco para obesidade aumentando a chance em 13,5% o risco a cada ano de acréscimo na idade do paciente.

Com relação as variáveis genéticas destacam-se pelo modelo *Forward Condicional* aplicado através de Regressão Logística Binária, a atividade fosfatase ácida e o grupo sanguíneo Levis a-b+, com base na medida do *Odds Ratio* as mesmas tornam-se um fator protetor.

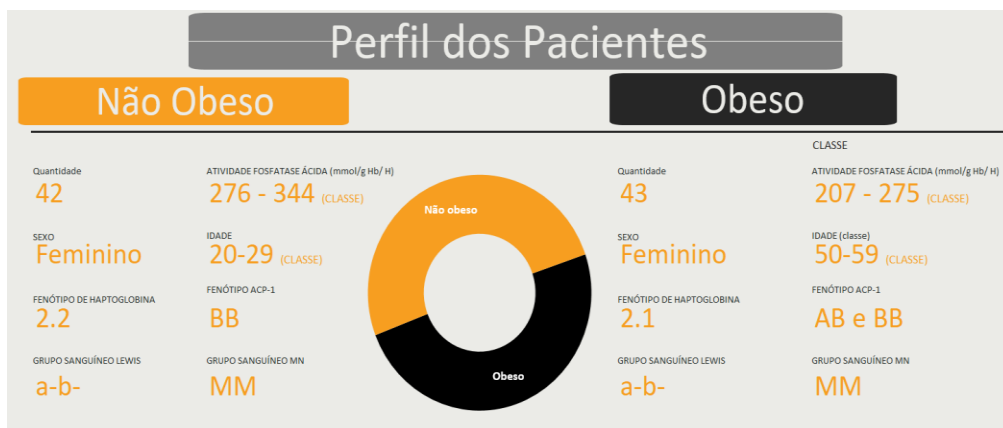
O *Odds Ratio* da atividade fosfatase ácida foi de 0,991 por cada (mmol/g Hb/ H) a mais, o que significa um menor risco para a obesidade, sendo esse índice de -0,009 % a cada mmol/g Hb/ H. O *Odds Ratio* do grupo sanguíneo Lewis tipo a-b+ que foi de 0,150, com I.C._{95%} = (0,24; 0,948) tendo como referência o tipo sanguíneo a+b-, ou seja, a chance de uma pessoa com o tipo sanguíneo a-b+ desenvolver a patologia de obesidade é de -85% em relação ao tipo sanguíneo a+b-, torna-o um fator protetor a doença com relação ao tipo sanguíneo de referência a-b+.

A conclusão desde trabalho aponta que a variável significativa como sendo um fator de risco para obesidade é a variável sócio demográfica idade, potenciando seu risco a cada ano a mais. A atividade fosfatase torna-se um fator protetor a cada unidade de aumento de (mmol/g Hb/ H) e o tipo sanguíneo a-b+ também contribui te forma protetora a obesidade em relação ao tipo sanguíneo de referência.

Considerando o aumento da idade um fator de risco para a obesidade abre-se um caminho na criação de tratamentos personalizados da classe médica para

seus pacientes, apresentando direcionamentos para controle da atividade fosfatase ácida e um despertar a quem tem um tipo sanguíneo a-b+ corroborando para um controle preventivo evitando o desenvolvimento de outras doenças que podem ser desencadeadas pelo simples fato de aumento de peso e/ou o não controle da atividade fosfatase ácida.

Uma técnica favorável que deu apoio a traçar o perfil dos participantes da pesquisa foi a visualização da informação. No capítulo 7 e através do modelo proposto com o uso do *software Tableau Public* foi possível identifica-lo. Na figura 40 contém de maneira visual informações colhidas através dos gráficos que foram apresentados nas figuras 25, 26, 27, 28, 29, 30 e 31 que permitiu visibilidade aos resultados.



Fonte: Criação do autor.

Figura 40: Perfil dos pacientes amostrais por variáveis e maiores frequências.

Em relação a Visualização da Informação neste estudo ficou claro que o software Tableau Public é uma ferramenta sugerida como modelo, devido a ser apontada como uma ferramenta facilitadora e amigável nas construções visuais e de apresentação para os usuários.

Capítulo 9

9. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Este estudo contou com um banco de dados amostral contendo 714 observações, que representa um número confortável para fazer um estudo. Ao passo que se avançou na investigação foi observado a existência de lacunas no banco de dados gerando uma limitação, ou seja, a continuidade do estudo só foi possível através de uma lapidação no banco de dados resultando 85 indivíduos que apresentavam todas as respostas das variáveis necessárias a este estudo. Diante dessa limitação assumimos uma nova amostra de referência no estudo.

Como limitação a conclusão da pesquisa faz referência para 85 indivíduos cujas informações se encontravam completas. No entanto, importa ressaltar que esse conjunto de dados retrata um perfil maioritariamente feminino, sendo efêmero o público masculino.

É importante ressaltar que na literatura e pesquisas científicas realizadas a nível mundial em estudos de obesidade, fica evidente a relação dessa patologia com a Diabete Millitus tipo 2. Sendo assim em posse de um banco de dados limitado neste estudo com apenas 85 indivíduos com informações completas, o perfil indicava apenas 1,9% dos indivíduos com Diabete Millitus tipo 2, motivo pelo qual a mesma nem se apresenta nos modelos de risco utilizados nas estimativas.

Em busca de mais respostas sobre o tema proposto neste trabalho e pelo fato da pertinência desse tema, sugere-se a continuação deste estudo como referência para trabalhos futuros buscando uma base de dados mais completa e consistente, com o objetivo de reafirmar o já identificado nesta obra, buscando-se identificar mais variáveis que apontem melhores indicações para tratamentos futuros, personalizados, e além disso, levar em consideração um aspecto extremamente importante que é **A PREVENÇÃO**.

REFERÊNCIAS BIBLIOGRÁFICAS

ABDULRAHMAN, ALANAZI TALAL, AND DALIA KAMAL ALNAGAR. 2021. “**Data Analysis and Computational Methods for Assessing Knowledge of Obesity Risk Factors among Saudi Citizens.**” *Computational and mathematical methods in medicine* vol. 2021 1371336. 26 Oct. 2021, <https://doi:10.1155/2021/1371336> [31 mar 2022].

AE IVANESCU, P LI2, B GEORGE, AW BROWN, SW KEITH, D RAJU AND DB ALLISON. 2016. **The importance of prediction model validation and assessment in obesity and nutrition research.** *International Journal of Obesity* 40, 887–894 © 2016 Macmillan Publishers Limited. [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(18\)32822-8/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)32822-8/fulltext) [31 mar 2022].

ADAIR, T., LOPEZ, AD. 2020. **O papel do sobrepeso e da obesidade nas tendências adversas de mortalidade por doenças cardiovasculares: uma análise de dados de múltiplas causas de morte da Austrália e dos EUA.** *BMC Med* 18, 199. <https://doi.org/10.1186/s12916-020-01666-y> [31 mar 2022]

ALLISON, Paul D. 2012. **Logistic Regression Using SAS®: Theory and Application**, Second Edition. Cary, NC: SAS Institute Inc.

ALVES, M. 2015. **Visualização de informação para simplificar o entendimento de indicadores sobre avaliação da ciência e tecnologia.** <https://repositorio.ufscar.br/bitstream/handle/ufscar/1151/6802.pdf?sequence=1&isAllowed=y> [17 jun 2020].

ALVES, L et al. 2010. **Leptina, hipertensão arterial e obesidade: importância das ações de enfermagem.** *Acta paul. enferm.*, São Paulo , v. 23, n. 2, p. 286-290 https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-21002010000200021 [04 jun 2020].

- AMARAL,O,PEREIRA,C. 2008. **Obesidade da genética ao ambiente.**
<https://repositorio.ipv.pt/handle/10400.19/382> [01 mai 2020].
- ANGELI, C.B. 2008.**Susceptibilidade Genética e outros Fatores de Risco Associados ao Sobrepeso e à Obesidade em População Afro-descendentes do Vale do Ribeira – SP.** https://teses.usp.br/teses/disponiveis/41/41131/tde-07072008-161259/publico/ClaudiaBAngeli_parcial.pdf [14 mai 2020].
- ATLAS IDF. 2017. **Número de pessoas com diabetes no mundo e por região em 2017 e em 2045 (20-79 anos).**
<https://www.diabetes.org.br/profissionais/images/2018/poster-atlas-idf-2017.pdf>
acesso 06 jun 2020. [06 jun 2020].
- ATLAS IDF. **Figuras Chaves.** 2019. <https://www.diabetesatlas.org/en/> [06 jun 2020].
- ARBIT, 2020. **Power BI: o guia completo da ferramenta da Microsoft .**
<https://blog.arbit.com.br/power-bi-guia-ferramenta-microsoft/> [29 de julho de 2020]
- BERTOLINI, R. 2009. **Estudo de caso sobre visualização de dados na área da saúde.**
<https://repositorio.ucs.br/xmlui/bitstream/handle/11338/1269/TCC%20Rogerio%20Bertolini.pdf?sequence=1&isAllowed=y> [12 mar 2020].
- BONIFÁCIO, S. L.; NOVARETTI, M. C. Z. 2009. **Funções biológicas dos antígenos eritrocitários.** *Rev. Bras. Hematol. Hemoter*, v. 31, n. 2, p. 104-11.
<https://www.scielo.br/pdf/rbhh/v31n2/aop1509.pdf> [07 dez 2020].
- BRAGA FAUSTINO, Elaine et al. 2011. **DIABETES MELITTUS: Busca Ativa em Portadores de Obesidade.** ISSN 2176-9133.
<https://revistas.ufpr.br/cogitare/article/view/21120> [07 jun 2020]
- CABRAL,C. 2013 - **Aplicação do Modelo de Regressão Logística num Estudo de Mercado .**
https://repositorio.ul.pt/bitstream/10451/10671/1/ulfc106455_tm_Cleidy_Cabral.pdf
[27 jul 2020].
- CÂMARA, B. 2011. **Enzimas de Diagnóstico.**
<https://www.biomedicinapadiao.com.br/2011/12/enzimas-de-diagnostico.html>
[07 dez 2020].
- CAROLINO,E., et al. 2005 - **Estatística jubilar : actas do XII Congresso Anual da Sociedade Portuguesa de Estatística**
https://www.researchgate.net/publication/289374303_Estatistica_Jubilar_Actas_d_o_XII_Congresso_Anual_da_Sociedade_Portuguesa_de_Estatistica [18 fev 2021].

- CARVALHO E., MARCOS A. 2009. **Visualização da Informação**. <https://core.ac.uk/download/pdf/55609941.pdf> [17 jun 2020].
- CHOOI YC, DING C, MAGKOS F. 2019. **The epidemiology of obesity. Metabolism**. [https://www.metabolismjournal.com/article/S0026-0495\(18\)30194-X/fulltext](https://www.metabolismjournal.com/article/S0026-0495(18)30194-X/fulltext) [04 mai 2020].
- COBIAC LJ, SCARBOROUGH P. 2021. **Modelling future trajectories of obesity and body mass index in England**. PLOS ONE 16(6): e0252072. <https://doi.org/10.1371/journal.pone.0252072> [31 mar 2022]
- DADOS FINOS. 2020. **Infográficos, visualização e jornalismo de dados**. Disponível em: <https://www.dadosfinos.info/> [27 jul 2020].
- DICIO, 2020. **Dicionário Online de Português**. <https://www.dicio.com.br/> [01 mai 2020].
- DICIO. **Dicionário Online de Português**. 2020. <https://www.dicio.com.br/> [17 jun 2020].
- ESTEVES, P. 2011. **Obesidade – Revisão Bibliográfica**. <https://ubibliorum.ubi.pt/bitstream/10400.6/924/1/Tese%20Final%20-%20Pedro%20Esteves.pdf> [04 mai 2020].
- FERREIRA, V. MAGALHÃES, R. 2006. **Obesidade no Brasil: Tendências atuais. 24** https://www.researchgate.net/publication/237470975_Obesidade_no_Brasil_Tendencias_atuais [06 jun 2020].
- FLOR, L., et al. 2015. **Diabetes burden in Brazil: fraction attributable to overweight, obesity, and excess weight**. Revista de Saúde Pública [online], v. 49, n. 00 <https://doi.org/10.1590/S0034-8910.2015049005571> [06 jun 2020].
- FILHO, P. 2020. **Asma Brônquica. Tipos de Asma. Classificação / Fenótipo / Endotipos**. http://www.asmabronquica.com.br/PDF/tipos_de_asma_classifica.pdf [07 dez 2020].
- GAIO, V., ANTUNES, L., BARRETO, M. GIL, A., KISLAIA, I, NAMORADO, S. RODRIGUES, A., SANTOS, A., NUNES, B., DIAS, C. 2018. **Prevalência de excesso de peso e de obesidade em Portugal: resultado do primeiro Inquérito Nacional de Saúde com Exame Físico (INSEF 2015)**. http://repositorio.insa.pt/bitstream/10400.18/5588/5/Boletim_Epidemiologico_Obse_rvacoes_N22_2018_artigo7.pdf [02 jun 2020].
- GALVÃO R, KOHLMANN JR. O. 2002. **Hipertensão arterial no paciente obeso**. Rev Bras Hipertens vol 9(3) https://www.researchgate.net/publication/237647272_Hipertensao_arterial_no_paciente_obeso [04 jun 2020].

- GANCHEV, D., 2017. **Application of Tableau Public Free Data Visualization Software in the Pesticide Science and Ecotoxicology Education.** *Mayfeb journal of agricultural science*.
https://www.researchgate.net/publication/319454203_Application_of_Tableau_Public_Free_Data_Visualization_Software_in_the_Pesticide_Science_and_Ecotoxicology_Education [26 jul de 2020].
- GIGANTE, D., BARROS A, OLINTO, M.T.1997. **Prevalência de obesidade em adultos e seus fatores de risco.** *Revista de Saúde Pública*.
https://www.researchgate.net/publication/26344318_Prevalencia_de_obesidade_em_adultos_e_seus_fatores_de_risco/citation/download [04 mai 2020].
- GONÇALVES, A. 2013. **Regressão Logística Aplicada à Pesquisa de Preditores de Morte.**
https://eg.uc.pt/bitstream/10316/33697/1/Regressao%20logistica%20aplicada%20a%20pesquisa%20de%20preditores%20de%20morte_AnaMargaridaGoncalves.pdf [26 jul 2020].
- GONZALES, L. 2018. **Regressão Logística e suas Aplicações.** Disponível em :
<https://monografias.ufma.br/jspui/bitstream/123456789/3572/1/LEANDRO-GONZALEZ.pdf> [26 jul 2020].
- GOVERNO DO ESTADO DO ESPIRITO SANTO. 2020. **Dados do Coronavírus (COVID-19)**
<https://app.powerbi.com/view?r=eyJrIjoiZmY5ZjA2ZTEtZTlzZC00ZDEzLThiMDAtZDIzYTUyODI5ZWFlwidiCl6jlxYzQ2YTJjLWU0M2EtNGUyNy1hMzk1LTM1ZWMyNWlxZmRhMSJ9> [29 jul 2020].
- GRÉGIO A., FILHO, B., MONTES A., SANTOS R. 2009. **Técnicas de Visualização de Dados aplicadas à Segurança da Informação.**
<http://www.lac.inpe.br/~rafael.santos/Docs/SBSEG/2009/sbseg2009.pdf> [17 jun 2020].
- HOSMER, D.W.; LEMESHOW, S. 2000. **Applied Logistic Regression.** 2nd ed. New York: John Wiley & Sons.
- HRUBY, ADELA & FRANK. 2014. **The Epidemiology of Obesity: A Big Picture.**
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4859313/> [04 mai 2020].
- KEAVER L, XU B, JACCARD A, WEBBER L. 2020. **Morbid obesity in the UK: A modelling projection study to 2035.** *Scand J Public Health*.48(4):422-427. doi: 10.1177/1403494818794814 . Epub 2018 Aug 30. PMID: 30160640. <https://pubmed.ncbi.nlm.nih.gov/30160640/> [31 mar 2022]
- LIRA, S. 2004. **Análise de Correlação: Abordagem teórica e de Construção dos coeficientes com aplicações .**
http://www.ipardes.pr.gov.br/sites/ipardes/arquivos_restritos/files/documento/2019-09/sachiko_dissertacao_2004.pdf [26 jul 2021].
- LISKA, G. 2012. **Classificação de Dados em Modelos com Resposta Binária via Algoritmo Boosting e Regressão Logística.**

<http://repositorio.ufla.br/bitstream/1/626/1/DISSERTA%C3%87%C3%83O%20Classifica%C3%A7%C3%A3o%20de%20dados%20em%20modelos%20com%20resposta%20bin%C3%A1ria%20via.pdf> [26 jul 2020].

LOPES, A. 2018 **Teoria e boas práticas da Visualização da Informação**. <https://repositorio.utad.pt/bitstream/10348/8630/1/Ana%20Daniela%20Lopes%20-%2038025%20-%20Teorias%20e%20boas%20pr%C3%A1ticas%20da%20Visualiza%C3%A7%C3%A3o%20da%20Informa%C3%A7%C3%A3o.pdf> [17 jun 2020].

LUFT, V. 2010. **Obesidade e diabetes : contribuição de processos inflamatórios e adipocitocinas, e a potencial importância de fatores nutricionais**. <https://www.lume.ufrgs.br/handle/10183/19022> [06 jun 2020].

MAGALHÃES, L., AMORIM, A., REZENDE E. 2018. **Conceito e aspectos epidemiológicos da hipertensão arterial**. Rev. Bras. de Hipertensão – Vol 25. <http://departamentos.cardiol.br/sbc-dha/profissional/revista/25-1.pdf> [06 jun 2020].

MARCIANO, F. 2009. **Principais tipos de resíduos utilizados na análise de diagnóstico em MLG com aplicações para os modelos: Poisson, ZIP e ZINB**. <file:///C:/Users/User/Downloads/Relat%C3%B3rio%20Final.pdf> [01 nov 2020].

MARÔCO, J. 2011 – **Análise Estatística com o Uso do SPSS. 5ª edição** .

MARTINS, A. 2018. **É preciso tratar a obesidade como um problema de saúde pública**. Rev. adm. empres. vol.58 no.3. São Paulo http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-75902018000300337#aff1 [12 mar 2020].

MARTINS, G. FONSECA, J. 1979. **Curso de estatística**. São Paulo: Atlas, 200 p. ou 6. ed. 1990.

MARTINS, G. A. 2008. **Estudo de caso: uma reflexão sobre a aplicabilidade em pesquisas no Brasil**. Revista de Contabilidade e Organizações, v. 2, n. 2, art. 1, p. 8-18. <http://www.spell.org.br/documentos/ver/4842/estudo-de-caso--uma-reflexao-sobre-a-aplicabili---> [09 abr 2020].

MANCINI, M.C. 2015. **Tratado de obesidade**. Segunda edição. Guanabara. Grupo Editorial Nacional.

MARCONI, M. LAKATOS E. 2003. **Fundamentos de metodologia científica**. 5 ed. São Paulo. 312 p. http://docente.ifrn.edu.br/olivianeta/disciplinas/copy_of_historia-i/historia-ii/china-e-india/view [04 jun 2020].

MELLO, R. 2012. **Doença cardiovascular, hipertensão arterial e obesidade em idosos: um estudo de base populacional**. <https://www.lume.ufrgs.br/handle/10183/143495> [04 jun 2020].

- MELO, S. RISSIN, M. 2015. **Excesso de Peso de Adultos Residentes em um Aglomerado Urbano Subnormal**. Rev Bras Promoç Saúde, Fortaleza, 28(2): 257-265 <https://periodicos.unifor.br/RBPS/article/view/3533/0> [02 jun 2020].
- MENSORIO M., COSTA JUNIOR, A. 2016. **Obesity and coping strategies: what is highlighted by literature?** http://www.scielo.mec.pt/scielo.php?script=sci_arttext&pid=S1645-00862016000300013&lang=pt [01 mai 2020]
- MEHTA, N., STENHOLM, S., MÄNNISTÖ, S., JOUSILAHTI, P., & ELO, I. (2020). **Excess body weight, cigarette smoking, and type II diabetes incidence in the national FINRISK studies**. *Annals of epidemiology*, 42, 12–18. <https://doi.org/10.1016/j.annepidem.2019.12.008> [31 mar 2022]
- MICHELLE M. HABY, ALISON MARKWICK, ANNA PEETERS, JONATHAN SHAW, THEO VOS. 2012. **Previsões futuras do índice de massa corporal e prevalência de excesso de peso na Austrália, 2005–2025**, *Health Promotion International*, Volume 27, Issue 2, June 2012, Pages 250–260, <https://doi.org/10.1093/heapro/dar036> [31 mar 2022]
- MICROSOFT. 2020. **Porque usar o Power BI?** <https://powerbi.microsoft.com/pt-br/what-is-power-bi/>. [27 jul 2020]
- MINISTÉRIO DA SAÚDE. 2006. **Caderno de Atenção Básica: Diabetes Mellitus**. https://bvsm.sau.gov.br/bvs/publicacoes/diabetes_mellitus.PDF [06 jun 2020].
- MONTENEGRO S.G., 2009. **Modelo de Regressão Logística Ordinal em Dados Categóricos na Área de Ergonomia Experimental**. <https://repositorio.ufpb.br/jspui/handle/tede/5273> [08 mai 2020].
- NASCIMENTO, H., & FERREIRA, C. 2005. **Visualização de informações – uma abordagem prática**. Artigo apresentado no XXV Congresso da Sociedade Brasileira de Computação, São Leopoldo. https://www.researchgate.net/publication/267403645_Visualizacao_de_Informacoes_-_Uma_Abordagem_Pratica [17 jun 2020].
- NASCIMENTO, H., & FERREIRA, C. 2011. **Uma introdução à visualização de informações**. https://www.researchgate.net/publication/307760992_An_introduction_to_information_visualization [17 jun 2020].
- NARDI, I. 2019. **O desenvolvimento de um modelo matemático para a previsão da aprovação da disciplina de cálculo 1 utilizando regressão logística**. <https://www.brazilianjournals.com/index.php/BRJD/article/view/4184> [18 fev 2021].

- OMS - Organização Mundial as Saúde. 2020. **Obesidade e excesso de peso.** <https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight> [12 mar 2020].
- PAES, A.; SILVA,R. 2012.**Por Dentro da Estatística: Teste de Concordância Kappa** <http://apps.einstein.br/revista/arquivos/PDF/2715-165-166.pdf> [27 jul2020].
- PAPELBAUM, M. et al. 2004. **Diabetes mellitus e transtornos alimentares: uma revisão sistemática.** https://www.researchgate.net/profile/Marcelo_Papelbaum/publication/288307337_Diabetes_mellitus_and_eating_disorders_A_systematic_review/links/56a60f1808aebf168e31faf4/Diabetes-mellitus-and-eating-disorders-A-systematic-review.pdf [06 jun 2020].
- PEREIRA, L.T.K.; GODOY,D.M.A TERCARIOL, D. 2009. **Estudo de caso como procedimento de pesquisa científica: reflexão a partir da clínica fonoaudiológica.** *Psicol. Reflex. Crit.* [online], vol.22, n.3, pp.422-429. ISSN 0102-7972. <https://doi.org/10.1590/S0102-79722009000300013> [09 abr 2020].
- PEREIRA V., RODRIGUES, C. E CORTEZ, FILIPA. 2019. **Fatores genéticos, epigenômicos, metagenômicos e cronobiológicos da obesidade.** n.17, pp.22-26. http://www.scielo.mec.pt/scielo.php?script=sci_abstract&pid=S2183-59852019000200004&lng=pt&nrm=iso [01 mai 2020].
- PSCHISKY, A. 2003. **Grupos Sanguíneos Humanos nos Livros Didáticos de Biologia – Análise de Conteúdo.** <https://repositorio.ufsc.br/bitstream/handle/123456789/85330/198249.pdf?sequence=1&isAllowed=y> [07 dez 2020].
- QUEIROZ,N. 2001. **Regressão Logística – Uma Estimativa Beysiana aplicada na identificação de fatores de risco para HIV, em doadores de sangue.** <http://www.tede2.ufrpe.br:8080/tede2/bitstream/tede2/5233/2/Niedja%20Maristone%20Oliveira%20Barreto%20Queiroz.pdf> [26 Jul 2020].
- RAMOS, M.R.; CAROLINO, E.; OLIVEIRA, T.; BICHO, M.. "Two-way MANCOVA: An application to public health". 2012.10.1063/1.4756503
- REIS, J. P., MACERA, C. A., ARANETA, M. R., LINDSAY, S. P., MARSHALL, S. J., & WINGARD, D. L. (2009). **Comparison of overall obesity and body fat distribution in predicting risk of mortality.** *Obesity (Silver Spring, Md.)*, 17(6), 1232–1239. <https://doi.org/10.1038/oby.2008.664> [27 mar 2022].
- RIBEIRO ARTUR BELTRAME, ZANELLA MARIA TERESA. 2000. **Tratamento da Hipertensão Arterial no Paciente Obeso.** 2000. *Rev Bras Hipertens* 2:172-8 <http://departamentos.cardiol.br/dha/revista/7-2/013.pdf> [07 out 2020].
- RIBEIRO FILHO, FERNANDO S.F. et al . 2000. **Obesidade, hipertensão arterial e suas influências sobre a massa e função do ventrículo esquerdo.** *Arq Bras Endocrinol Metab*, São Paulo , v. 44, n. 1, p. 64-71.

https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-27302000000100011 [04 jun 2020].

ROBERT, S., FERREIRA, G., ZANELLA, .2000. **Epidemiologia da Hipertensão associada à Obesidade.**

https://www.researchgate.net/publication/238692093_EPIDEMIOLOGIA_DA_HIPERTENSAO_ARTERIAL_ASSOCIADA_A_OBESIDADE [04 jun 2020]

RUSSO, A. 2016. **Introducing Microsoft Power BI?** Published by Microsoft Press. [29 jul 2020]

SBD. **Diretrizes da Sociedade Brasileira de Diabetes 2017-2018.**

<https://www.diabetes.org.br/profissionais/images/2017/diretrizes/diretrizes-sbd-2017-2018.pdf> [06 mai 2020].

SLEEPER, R. 2018. **Practical Tableau - 100 Tips, Tutorials, and Strategies from a Tableau Zen Master**, First Edition, Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

SILVA, C. 2007. **Considerações sobre o uso de Visualização de Informação no auxílio à gestão de informação.**

<https://cin.ufpe.br/~psgmn/Gestao%20da%20Informacao%20e%20do%20Conhecimento%20-%20GIC/Artigos/4.pdf> [17 jun 2020].

SILVA, I.2017. **Bussiness Intelligence com Microsoft Power BI.**

<https://docplayer.com.br/87479615-Business-intelligence-com-microsoft-power-bi.html> [28 jul 2020]

SILVA, L.2019. **Ferramentas de BI para inteligência Competitiva.**

<https://acervodigital.ufpr.br/bitstream/handle/1884/65124/LUCAS%20JOSE%20HARMATIUK%20DA%20SILVA.pdf?sequence=1&isAllowed=y> [29 jul 2020]

SILVA, V. 2010. **Prevalência de Sobrepeso / Obesidade e Fatores Associados em Adultos no Brasil.**

<https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/94481/276683.pdf?sequence=1&isAllowed=y> [06 mai 2020].

SILVA, C. 2007. **Considerações sobre o uso de Visualização de Informação no auxílio à gestão de informação.**

<https://cin.ufpe.br/~psgmn/Gestao%20da%20Informacao%20e%20do%20Conhecimento%20-%20GIC/Artigos/4.pdf> [17 jun 2020].

TABLEAU PUBLIC SOFTWARE. 2017 <https://public.tableau.com/> [27 jul 2020].

TISCHIER, A. 2013. **Caracterização do Perfil Corporal de Pacientes Obesos e Portadores de Hipertensão Arterial Sistêmica Admitidos em uma Clínica-Escola de Nutrição no Município de Lauro de Freitas - BA.**

<http://www.rbone.com.br/index.php/rbone/article/view/296/287> [06 mai 2020].

VAZ, F. CARVALHO L. 2004. **Visualização da Informação.**

http://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_003-04.pdf [17 jun 2020].

VIGITEL BRASIL 2013. **Vigilância de fatores de risco e proteção para**

doenças crônicas por inquérito telefônico / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância de Doenças e Agravos não Transmissíveis e Promoção da Saúde. – Brasília: Ministério da Saúde, 2014.

<https://www.saude.gov.br/images/pdf/2014/dezembro/09/Vigitel-2013.pdf> [02 mai 2020].

WARE, C. 2004. **Information Visualization: Perception for Design: Second Edition.**

https://www.researchgate.net/publication/224285723_Information_Visualization_Perception_for_Design_Second_Edition/citation/download [17 jun 2020].

WARE, C. 2004. **Information Visualization: Perception for Design: Second Edition.**

https://www.researchgate.net/publication/224285723_Information_Visualization_Perception_for_Design_Second_Edition/citation/download [17 jun 2020].

WHO (World Health Organization). 2000. **Obesity: Preventing and Managing**

the Global Epidemic. <https://apps.who.int/iris/handle/10665/42330> [04 mai 2020].

WORLD POPULATION REVIEW. 2022. **População Mundial por País.**

<https://worldpopulationreview.com/> [27 de março de 2022]

WU, B., JIANG, Y., JIN, X. *ET AL.* 2020. **Using three statistical methods to analyze the association between exposure to 9 compounds and obesity in children and adolescents: NHANES 2005-2010.** *Environ Health* 19, 94.

<https://doi.org/10.1186/s12940-020-00642-6> [23 fev 2022].

ANEXO I: Output do SPSS na Análise Univariada.

Tabela 01: Medidas Estatísticas da Variável Obesidade com o uso do SPSS.

Obesidade					
		Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido	Normal	238	33,3	57,1	57,1
	Obeso	179	25,1	42,9	100,0
	Total	417	58,4	100,0	
Ausente	Sistema	297	41,6		
Total		714	100,0		

Fonte: Output do Software SPSS

Tabela 02: Medidas Estatísticas da Variável Idade com o uso do SPSS.

Resumo de processamento de casos						
	Válido		Omisso		Total	
	N	Porcentagem	N	Porcentagem	N	Porcentagem
Idade	644	90,2%	70	9,8%	714	100,0%

Descritivas

		Estadística	Estadística do teste	
			Padrão	
Idade	Média	38,2973	,76042	
	95% de Intervalo de Confiança para Média	Limite inferior	36,8041	
		Limite superior	39,7905	
	5% da média aparada	37,5770		
	Mediana	36,0000		
	Variância	372,388		
	Erro Padrão	19,29737		
	Mínimo	13,00		
	Máximo	98,00		
	Amplitude	85,00		
	Amplitude interquartil	35,00		
	Assimetria	,425	,096	
	Curtose	-1,052	,192	

Fonte: Output do Software SPSS

Tabela 03: Medidas Estatísticas da Variável Género com o uso do SPSS

		Género		
		Frequência	Porcentagem	Porcentagem
			válida	acumulativa
Válido	Feminino	540	75,6	81,1
	Masculino	126	17,6	100,0
	Total	666	93,3	100,0
Omisso	Sistema	48	6,7	
Total		714	100,0	

Fonte: Output do Software SPSS.

Tabela 04: Medidas Estatísticas da Variável IMC com o uso do SPSS.

		Grupos de IMC		
		Frequência	Porcentagem	Porcentagem
			válida	acumulativa
Válido	Abaixo do peso	50	7,0	8,8
	Normal	192	26,9	42,4
	Sobrepeso	205	28,7	78,3
	Obeso	124	17,4	100,0
	Total	571	80,0	100,0

Ausente	Sistema	143	20,0	
Total		714	100,0	

Fonte: Output do Software SPSS.

Tabela 05: Medidas Estatísticas da Variável Sistema de Grupo Sanguíneo Lewis com o uso do SPSS.

Sistema Sanguíneo Lewis					
		Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido	a-b-	188	26,3	54,8	54,8
	a-b+	96	13,4	28,0	82,8
	a+b-	57	8,0	16,6	99,4
	a+b+	2	,3	,6	100,0
	Total	343	48,0	100,0	
Omisso	Sistema	371	52,0		
Total		714	100,0		

Fonte: Output do Software SPSS.

Tabela 06: Medidas Estatísticas da Variável Sistema de grupo sanguíneo MN com o uso do SPSS.

Sistema Sanguíneo MN					
		Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido	MM	185	25,9	48,2	48,2
	MN	104	14,6	27,1	75,3
	NN	95	13,3	24,7	100,0
	Total	384	53,8	100,0	
Omisso	Sistema	330	46,2		
Total		714	100,0		

Fonte: Output do Software SPSS.

Tabela 07: Medidas Estatísticas da Variável Fenótipo de Haptoglobina com o uso do SPSS.

Fenótipo de Haptoglobina					
		Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido	2.1	194	27,2	48,7	48,7
	2.2	140	19,6	35,2	83,9

	1.1	64	9,0	16,1	100,0
	Total	398	55,7	100,0	
Omisso	Sistema	316	44,3		
Total		714	100,0		

Fonte: Output do Software SPSS.

Tabela 08: Medidas Estatísticas da Variável Tipo de Diabete com o uso do SPSS.

		Tipo de Diabete			
		Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido	Ausente	310	43,4	97,5	97,5
	Tipo 2	6	,8	1,9	99,4
	Tipo 1	2	,3	,6	100,0
	Total	318	44,5	100,0	
Omisso	Sistema	396	55,5		
Total		714	100,0		

Fonte: Output do Software SPSS.

Tabela 09: Medidas Estatísticas da Variável Hipertensão com o uso do SPSS.

		HTA			
		Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido	Normal	412	57,7	72,7	72,7
	Hipertenso	155	21,7	27,3	100,0
	Total	567	79,4	100,0	
Omisso	Sistema	147	20,6		
Total		714	100,0		

Fonte: Output do Software SPSS.

Tabela 10: Medidas Estatísticas da Variável Atividade de Fosfatase ACP-1 com o uso do SPSS.

Resumo de processamento de casos							
		Casos					
		Válido		Omisso		Total	
N	Porcentagem	N	Porcentagem	N	Porcentagem	N	Porcentagem

Atividade	400	56,0%	314	44,0%	714	100,0%
Fosfatase ácida – ACP-1 (mmol/g Hb/H)						
Descritivas						Estatística do teste
				Estatística	Padrão	
Atividade	Média		325,7001300		8,05026816	
Fosfatase ácida – ACP-1 (mmol/g Hb/H)	95% de Intervalo de Confiança para Média		Limite inferior	309,8738881		
			Limite superior	341,5263719		
	5% da média aparada		314,7722528			
	Mediana		293,1810000			
	Variância		25922,727			
	Erro Padrão		161,00536322			
	Mínimo		19,02000			
	Máximo		1162,28000			
	Amplitude		1143,26000			
	Amplitude interquartil		181,71400			
	Assimetria		1,240		,122	
	Curtose		2,563		,243	

Fonte: Output do Software SPSS.

Tabela 11: Medidas Estatísticas da Variável Atividade de Fenótipo ACP-1 com o uso do SPSS.

Fenótipo ACP-1					
		Frequência	Porcentagem	Porcentagem válida	Porcentagem acumulativa
Válido	BB	87	12,2	38,3	38,3
	AB	65	9,1	28,6	67,0
	BC	26	3,6	11,5	78,4
	AA	19	2,7	8,4	86,8
	AC	17	2,4	7,5	94,3
	CC	13	1,8	5,7	100,0
	Total	227	31,8	100,0	
Omisso	Sistema	487	68,2		
Total		714	100,0		

Fonte: Output do Software SPSS.

ANEXO II: Output do SPSS na Análise Bivariada.

Tabela 01: Tabulação cruzadas com valores observados, esperados e Teste Qui-quadrado para as variáveis obesidade e gênero com o uso do software SPSS.

Obesidade * Gênero Tabulação cruzada					
			Gênero		Total
			Feminino	Masculino	
Obesidade	Normal	Contagem	179	51	230
		Contagem Esperada	183,7	46,3	230,0
	Obeso	Contagem	146	31	177
		Contagem Esperada	141,3	35,7	177,0
Total	Contagem		325	82	407
	Contagem Esperada		325,0	82,0	407,0

Testes Qui-quadrado			
	Valor	df	Significância Sig. (2 lados)
Qui-quadrado de Pearson	1,350 ^a	1	,245

a. 0 células (0,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 35,66.

Fonte: Output do Software SPSS.

Tabela 02: Tabulação cruzadas com valores observados, esperados e Teste Qui-quadrado para as variáveis obesidade e o Sistema Sanguíneo Lewis com o uso do software SPSS.

Obesidade * Sistema Sanguíneo Lewis - Tabulação cruzada

			Sistema Sanguíneo Lewis			Total
			a+b-	a-b+	a-b-	
Obesidade	Normal	Contagem	24	41	93	158
		Contagem Esperada	27,9	40,6	89,5	158,0
	Obeso	Contagem	20	23	48	91
		Contagem Esperada	16,1	23,4	51,5	91,0
Total		Contagem	44	64	141	249
		Contagem Esperada	44,0	64,0	141,0	249,0

Testes qui-quadrado

	Valor	df	Significância Sig. (2 lados)
Qui-quadrado de Pearson	1,897 ^a	2	,387

a. 0 células (0,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 16,08.

Fonte: Output do Software SPSS.

Tabela 03: Tabulação cruzadas com valores observados, esperados e Teste Qui-quadrado para as variáveis obesidade e o Sistema Sanguíneo MN com o uso do software SPSS.

Obesidade * Sistema Sanguíneo MN - Tabulação cruzada

			Sistema Sanguíneo MN			Total
			MM	MN	NN	
Obesidade	Normal	Contagem	112	31	31	174
		Contagem Esperada	97,5	40,3	36,1	174,0
	Obeso	Contagem	50	36	29	115
		Contagem Esperada	64,5	26,7	23,9	115,0
Total		Contagem	162	67	60	289
		Contagem Esperada	162,0	67,0	60,0	289,0

Testes qui-quadrado

	Valor	df	Significância Sig. (2 lados)
Qui-quadrado de Pearson	12,650 ^a	2	,002

a. 0 células (0,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 23,88.

Fonte: Output do Software SPSS.

Tabela 04: Tabulação cruzadas com valores observados, esperados e Teste Qui-quadrado para as variáveis obesidade e Fenótipo de Haptoglobina com o uso do software SPSS.

Obesidade * Fenótipo de Haptoglobina ACP-1 - Tabulação cruzada

			Fenótipo de Haptoglobina			Total
			1.1	2.1	2.2	
Obesidade	Normal	Contagem	27	88	77	192
		Contagem Esperada	29,9	91,0	71,1	192,0
	Obeso	Contagem	21	58	37	116
		Contagem Esperada	18,1	55,0	42,9	116,0
Total		Contagem	48	146	114	308
		Contagem Esperada	48,0	146,0	114,0	308,0

Testes qui-quadrado

	Valor	df	Significância Sig. (2 lados)
Qui-quadrado de Pearson	2,339 ^a	2	,311

a. 0 células (0,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 18,08.

Fonte: Output do Software SPSS.

Tabela 05: Tabulação cruzadas com valores observados, esperados e simulação de Monte Carlo para as variáveis obesidade e Tipo de Diabetes com o uso do software SPSS.

Obesidade * Tipo de diabetes Tabulação cruzada

			Tipo de diabetes			Total
			Ausente	Tipo 1	Tipo 2	
Obesidade	Normal	Contagem	176	1	0	177
		Contagem Esperada	172,2	1,2	3,6	177,0
	Obeso	Contagem	109	1	6	116
		Contagem Esperada	112,8	,8	2,4	116,0
Total		Contagem	285	2	6	293
		Contagem Esperada	285,0	2,0	6,0	293,0

Testes Qui-quadrado

	Valor	df	Significância Sig. (2 lados)	Sig. Monte Carlo (2 lados)		
				Sig.	Intervalo de Confiança 95%	
					Limite inferior	Limite superior
Qui-quadrado de Pearson	9,461 ^a	2	,009	,003	,002	,005

a. 4 células (66,7%) esperavam uma contagem menor que 5. A contagem mínima esperada é ,79.

Fonte: Output do Software SPSS.

Tabela 06: Tabulação cruzadas com valores observados, esperados e Teste Qui-quadrado para as variáveis obesidade e hipertensão com o uso do software SPSS.

Obesidade * HTA Tabulação cruzada

			HTA		Total
			Normal	Hipertenso	
Obesidade	Normal	Contagem	176	20	196
		Contagem Esperada	151,3	44,7	196,0
	Obeso	Contagem	108	64	172
		Contagem Esperada	132,7	39,3	172,0
Total		Contagem	284	84	368
		Contagem Esperada	284,0	84,0	368,0

Testes Qui-quadrado

	Valor	df	Significância Sig. (2 lados)
Qui-quadrado de Pearson	37,925 ^a	1	,000

a. 0 células (0,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 39,26.

Fonte: Output do Software SPSS.

Tabela 07: Tabulação cruzadas com valores observados, esperados e simulação de Monte Carlo para as variáveis obesidade e Fenótipo ACP-1 com o uso do software SPSS.

Obesidade * Fenótipo ACP-1 Tabulação cruzada

			Fenótipo ACP-1						Total
			AA	BB	CC	AB	AC	BC	
Obesidade	Normal	Contagem	5	38	10	23	3	16	95
		Contagem Esperada	9,2	34,9	6,7	25,7	7,7	10,8	95,0
	Obeso	Contagem	13	30	3	27	12	5	90
		Contagem Esperada	8,8	33,1	6,3	24,3	7,3	10,2	90,0
Total		Contagem	18	68	13	50	15	21	185
		Contagem Esperada	18,0	68,0	13,0	50,0	15,0	21,0	185,0

Testes qui-quadrado

	Valor	df	Significância Sig. (2 lados)
Qui-quadrado de Pearson	19,627 ^a	5	,001
a. 0 células (0,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 6,32.			

Fonte: Output do Software SPSS.

Tabela 08: Medida do ETA aplicado as variáveis obesidade e Idade com o uso do SPSS.

Medidas Direcionais

			Valor
Nominal por intervalo	Eta	Obesidade Dependente	,729
		Idade Dependente	,591

Fonte: Output do Software SPSS.

Tabela 09: Medida do ETA aplicado as variáveis obesidade e Atividade Fosfatase com o uso do SPSS.

Medidas Direcionais

			Valor
Nominal por intervalo	Eta	Obesidade Dependente	,996
		Atividade Fosfatase ácida – ACP-1 (mmol/g Hb/H)) Dependente	,033

Fonte: Output do Software SPSS.

ANEXO III

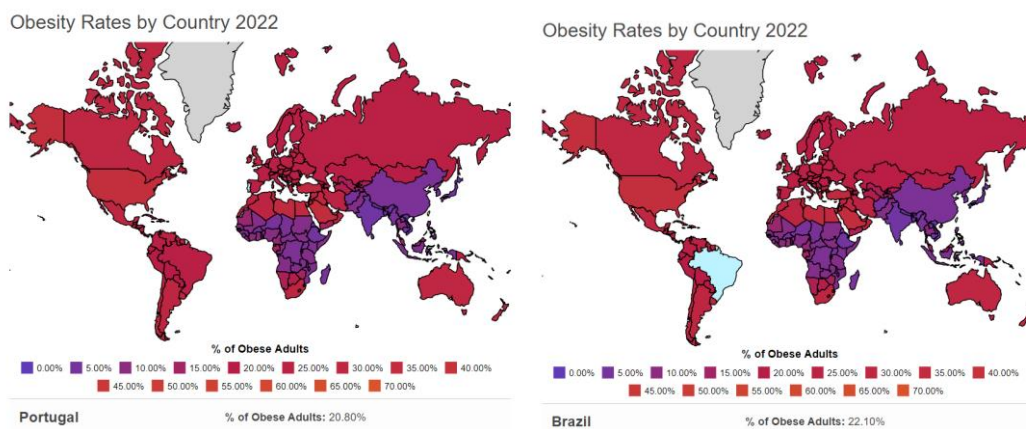


Figura 41: Percentual de obesos adultos a nível mundial, destaque para Portugal Brasil.