

RAE-CEA-10P18

**RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO: “ESTUDO
COMPARATIVO ENTRE PROCEDIMENTOS DE AVALIAÇÃO DE OBESIDADE EM
CRIANÇAS”**

**MARCOS NASCIMENTO MAGALHÃES
ANTONIO CARLOS PEDROSO DE LIMA
GISELA TUNES DA SILVA
RACHEL CANDIDO UBRIACO DE OLIVEIRA**

CENTRO DE ESTATÍSTICA APLICADA – CEA – USP
RELATÓRIO DE ANÁLISE ESTATÍSTICA

TÍTULO: Estudo comparativo entre procedimentos de avaliação de obesidade em crianças

PESQUISADORAS: Heidi Lui Reinhardt

Clarissa Tamie Hiwatashi Fujiwara

ORIENTADORA: Maria Edna de Melo

INSTITUIÇÃO: Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo.

FINALIDADE: Publicação

RESPONSÁVEIS PELA ANÁLISE ESTATÍSTICA: Marcos Nascimento Magalhães, Antonio Carlos Pedroso de Lima, Gisela Tunes da Silva e Rachel Candido Ubriaco de Oliveira.

REFERÊNCIA DESTE TRABALHO: MAGALHÃES, M; LIMA, A. C. P; SILVA, G. T. e OLIVEIRA, R. C. U. **Relatório de análise estatística sobre o projeto: Estudo comparativo entre procedimentos de avaliação de obesidade em crianças.** São Paulo, IME-USP, 2010 (RAE-CEA-10P18).

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

- BUSSAB, W.O. e MORETTIN, P.A. (2010). **Estatística Básica**. 6.ed. São Paulo: Saraiva. 526p.
- CLEVELAND, W.S. (1979) Robust Locally Weighted Regression and Smoothing Scatterplots, **Journal of the American Statistical Association**, **74**, 829-836.
- CONDE, W.L. and MONTEIRO, C.A. (2006) Body mass index cutoff points for evaluation of nutritional status in Brazilian children and adolescents. **J Pediatr**, **82**, 266-272.
- LANDIS, J.R. and KOCH, G.G. (1977). The measurement of observer agreement for categorical data. **Biometrics**, 33, 159–174.
- MAGALHÃES, M. e LIMA, A. C. P. (2010). **Noções de Probabilidade e Estatística**. 7ªed. São Paulo: EDUSP. 432 p.
- STOKES, E. M., DAVIS, C. S., and KOCH, G.G. (2001). **Categorical Data Analysis Using The SAS System**. 2ªed. Cary, NC: SAS Institute Inc., 499 p.

PROGRAMAS COMPUTACIONAIS:

MS-Excel 2007

MS-Word 2007

Minitab (versão 16.0)

R (versão 2.2.1)

TÉCNICAS ESTATÍSTICAS UTILIZADAS:

Análise Descritiva Unidimensional (03:010).

Análise de Regressão Clássica (07:020).

Regressão Logística (07:090).

ÁREA DE APLICAÇÃO:

Bioestatística (14:030).

ÍNDICE

Resumo	6
1. Introdução	7
2. Objetivos	9
3. Descrição do estudo	9
4. Descrição das variáveis	10
5. Análise descritiva	11
6. Análise Inferencial	19
APÊNDICE A- TABELAS	26
APÊNDICE B- GRÁFICOS	31
APÊNDICE C- EXPRESSÕES	55
APÊNDICE D- METODOLOGIA	58
APÊNDICE E- SUPOSIÇÕES VERIFICADAS.....	60
APÊNDICE F- DISCRIMINAÇÃO DO MODELO	65

Resumo

Este relatório descreve as análises estatísticas desenvolvidas com o objetivo de avaliar a adequabilidade de três procedimentos na discriminação em obesos e super obesos, para crianças e adolescentes com peso elevado. Uma vez que a amostra é composta por indivíduos em fase de crescimento, a metodologia utilizada na classificação baseia-se na construção de valores padronizados do índice de massa corpórea (IMC) para diferentes idades e gênero. São utilizados gráficos com curvas de percentis de IMC produzidos pelo Center for Disease Control and Prevention (CDC), pela World Health Organization (WHO) e um instrumento de referência nacional, conhecido como gráfico Conde & Monteiro (C&M). A correspondência, também é avaliada, entre a classificação obtida por cada um dos três procedimentos e algumas morbidades comprovadamente associadas com a obesidade, tais como a intolerância à glicose e a dislipidemia (níveis elevados ou anormais de lipídios e/ou lipoproteínas no sangue).

As análises feitas indicaram que à medida que o número de doenças e o percentual de massa de gordura aumentam, a chance de classificar um paciente como super obeso também aumenta e a intensidade como isso ocorre é a mesma para os três métodos. Verificamos que, pela amostra apresentada, não é possível detectar, estatisticamente, a existência de um melhor método para classificar um paciente em obesos ou super obesos.

1. Introdução

A obesidade é uma doença caracterizada pelo acúmulo excessivo de gordura corporal, que implica em danos à saúde e reduções importantes na qualidade e expectativa de vida.

No Brasil, a proporção de pessoas com excesso de peso aumentou no período de 1974 e 1989, notadamente nas mulheres. Assim como nas populações adultas, constata-se um número crescente de obesidade infantil, que ocorre de maneira mais acentuada em países economicamente desenvolvidos e em populações urbanizadas. O sobrepeso na infância e na adolescência está associado com diversas alterações metabólicas e endócrinas, como a hipertensão arterial, dislipidemia, resistência à insulina, glicemia alterada, entre outras, que ocasionam aumento no risco de doenças cardiovasculares. Alguns estudos sugerem que a mortalidade por essas doenças está diretamente associada com o tempo de duração da obesidade.

Para a população com idade superior a 18 anos, a obesidade é caracterizada pelo o índice de massa corpórea (IMC), definido pela razão do peso com o quadrado da altura do indivíduo, ser superior a 30 kg/m^2 , e são classificados como super obesos indivíduos com IMC superior a 50 kg/m^2 . Esse critério de definição de obesidade deve ser empregado com cautela em populações mais jovens, pois nessa fase do desenvolvimento ocorrem variações na composição corporal, com expressivo ganho de altura que ocorre de forma distinta para cada gênero (sendo mais precoce no gênero feminino) e nas diferentes faixas etárias.

A classificação mais frequente do estado nutricional para crianças e adolescentes é realizada de acordo com gráficos contendo curvas baseadas em percentis do IMC, segundo idade e gênero. O sobrepeso e a obesidade infantil, são definidos pelo *Center for Disease Control dos Estados Unidos* (CDC) como um valor de IMC igual ou acima dos percentis 85 e 95, respectivamente.

As curvas desenvolvidas pelo CDC em 2000 consideram diferentes valores das características, variáveis antropométricas peso e altura, para crianças e adolescentes de 2 a 20 anos e consistem em uma versão revisada das curvas do *National Center for Health Statistics* (NCHS) de 1977. A maior parte dos dados utilizados para construir estas curvas é proveniente do *National Health and Nutrition Examination Survey*,

iniciado na década de 1960 e que periodicamente, coleta informações sobre a saúde da população norte-americana.

A utilização de curvas baseadas no IMC para idade na definição do estado nutricional de crianças e adolescentes possibilitou soluções convenientes para a prática clínica e epidemiológica; entretanto, abriu espaço para debates em torno da sua utilização para avaliar grupos em crescimento. As principais questões referem-se à adoção de uma curva de referência local ou internacional e se os parâmetros aqui construídos seriam comparáveis com aqueles de curvas internacionais. Os pontos de corte para sobrepeso (*percentil* 85) e para obesidade (*percentil* 95) das curvas de IMC podem ser considerados arbitrários, pois não representam necessariamente o ponto no qual o risco para a saúde relacionado à adiposidade corporal começa a se elevar. Embora estes sejam os pontos de corte mais utilizados e estejam relacionados ao risco para a saúde ocasionado pela adiposidade, consistem em números convenientemente redondos e podem não ser apropriados para a análise de todas as populações.

Além do CDC, a Organização Mundial da Saúde (World Health Organization – WHO) também possui gráficos e tabelas para avaliação de IMC baseados em uma amostra mais representativa e que busca representar crianças de qualquer país, independente de etnia, condição socioeconômica e tipo de alimentação. Conde & Monteiro (2006), delinearam um gráfico nacional de referência baseado no IMC a partir de dados de crianças e adolescentes entre 2 e 19 anos, extraídos da Pesquisa Nacional de Saúde e Nutrição de 1989. Neste trabalho denotaremos essa curva por C&M.

A partir de cada uma das três curvas (CDC, WHO e C&M), serão calculados valores de IMC padronizados para diferentes idades, que serão denotados por ZIMC, utilizando-se as expressões descritas no Apêndice C.

Para a população de pacientes de peso elevado, a curva mais adequada para utilização será aquela que classifica como super obesos os pacientes com maior prevalência de doenças (hipertensão arterial, glicemia alterada, dilipidemia e resistência à insulina).

2. Objetivos

Os principais objetivos do estudo são:

- Verificar as diferenças de classificação em obesos e super obesos, utilizando cada uma das curvas: CDC, WHO e C&M.
- Identificar quais curvas fornecem classificação de obesos e super obesos com maior coerência em relação aos indicadores de morbidade baseados no percentual de massa de gordura e na ocorrência ou não de hipertensão arterial, glicemia alterada, resistência à insulina e dislipidemia.

3. Descrição do estudo

O estudo utiliza uma amostra com 281 pacientes, de ambos os gêneros, selecionados de forma voluntária dentre os pacientes que se candidatam para participar da Liga de Obesidade Infantil da Disciplina de Endocrinologia e Metabologia do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo.

Estabeleceu-se como critério de inclusão a obesidade sem causa genética determinada ou aparente. Os critérios de exclusão considerados são o diagnóstico de síndrome genética estabelecido (por exemplo, Síndrome de Prader-Willi, Síndrome de Bardet-Biedl ou outras), sinais clínicos sugestivos de síndromes genéticas – tais como critérios maiores e menores para Síndrome de Prader-Willi, retinite pigmentosa associada à obesidade e retardo mental – e hipo/hipertireoidismo. Pais ou responsáveis foram orientados a respeito dos objetivos do projeto e assinaram o Termo de Consentimento Livre e Esclarecido.

Os pacientes são caracterizados clinicamente por: (i) histórico clínico; (ii) exame físico completo, com análise específica dos parâmetros peso, estatura, circunferência da cintura, estadiamento puberal, presença de acantose, pressão arterial e frequência cardíaca. Em termos metabólicos, eles são caracterizados por meio de avaliação laboratorial que consiste em: (i) avaliação através de exames de sangue feitos em jejum de glicemia, insulina, cálculo do Índice HOMA-IR (*Homeostasis Model Assessment for Insulin Resistance*), colesterol total, HDL-colesterol, LDL-colesterol, triglicérides,

enzimas hepáticas ALT e AST (alanina aminotransferase e aspartato aminotransferase) e ácido úrico;(ii) avaliação hormonal das funções tireoidiana e do eixo adiposo-hipotalâmico, através das dosagens de hormônio tireotrófico (TSH), T4 livre e leptina (basal em jejum).

A padronização do IMC (ZIMC) foi calculada para cada uma das curvas do CDC, WHO e C&M utilizando o método LMS (Cleveland,W.S , 1979). Maiores detalhes são apresentados no Apêndice C (Expressão C.1).

4. Descrição das variáveis

As informações de cada paciente foram coletadas através de um questionário de exames laboratoriais. As variáveis são:

- Gênero: masculino ou feminino;
- Cor: branco ou outra (mulato, negro, parda e amarelo);
- Idade: (em meses);
- Peso: (em kg);
- Estatura: (em metros);
- P Estatura: percentil de estatura (padronização CDC);
- IMC: índice de massa corpórea (em kg/m^2);
- ZIMC(CDC): índice de massa corpórea padronizado para a curva norte-americana (adimensional);
- ZIMC(WHO): índice de massa corpórea padronizado para a curva fornecida pela WHO (adimensional);
- ZIMC(C&M): índice de massa corpórea padronizado para a curva nacional (adimensional);
- CA PM: circunferência medida no ponto médio (em cm);
- P PA Sistólica: percentil de pressão arterial sistólica (padronização CDC);
- P PA Diastólica: percentil de pressão arterial diastólica (padronização CDC);
- FC: frequência cardíaca (batimentos por minuto);
- MG (%): percentual de massa gorda;

- CT: colesterol total (mg/dL);
- Glicose: (mg/dL);
- HDL-c: *high density lipoprotein cholesterol* (mg/dL);
- Insulina: (μ U/ml);
- LDL-c: *low density lipoprotein cholesterol* (mg/dL);
- TG: triglycerides (mg/dL);
- HOMA IR: resistência à insulina (adimensional);
- HAS: hipertensão arterial (ausência/presença);
- Glicemia alterada: (ausência/presença);
- DLP: dislipidemia (ausência/presença);
- IR: resistência à insulina (ausência/presença).

5. Análise descritiva

As tabelas utilizadas nas análises e representações gráficas, que não se encontram no corpo do texto, são apresentadas no Apêndice A e Apêndice B, respectivamente.

Analisando o perfil dos pacientes, podemos observar pela Tabela A.1 e Gráfico B.1 que 57% dos pacientes são do gênero feminino. A Tabela A.2 mostra que a maioria dos pacientes (210 indivíduos) era da cor branca, enquanto os restantes se dividiam nas categorias parda, negro, amarelo e mulato (Gráfico B.2). Devido a essa grande diversificação entre os grupos de cor os pacientes foram agrupados em brancos e outros; 75% são brancos (Gráfico B.3).

Na Tabela 5.1, temos medidas-resumo para as variáveis quantitativas. A idade média dos pacientes é de aproximadamente 12 anos (147 meses); o paciente mais novo tem 6 anos (74 meses) e o mais velho, aproximadamente 18 anos (214 meses). A distribuição dos pacientes segundo as idades é ilustrada no *box-plot* apresentado no Gráfico B.4, em que podemos ver uma distribuição razoavelmente homogênea. Para buscar uma relação entre idade e gênero dos pacientes, foi construído o Gráfico B.5, em que pode ser verificado que a variabilidade da idade para o gênero feminino é um pouco maior do que a do gênero masculino, apesar de que, em ambos, as medianas se

apresentam bem próximas. O Gráfico B.6 pode ser utilizado para verificarmos se existe alguma sugestão de diferença de idades para o grupo de brancos e outros. Podemos observar que a mediana da idade para outros é menor que a observada para os brancos, porém não parece existir grande diferença entre os dois grupos.

A quantidade total de pacientes na amostra é de 281 indivíduos; alguns, entretanto, não possuem informações sobre algumas das variáveis utilizadas, assim, por exemplo, um dos indivíduos não possui a informação do peso e estatura e vários não contam com a variável de percentual de massa de gordura, pois o aparelho utilizado para essa medição esteve indisponível por aproximadamente 2 meses.

Tabela 5.1: Resumo das variáveis quantitativas.

Variáveis Contínuas	Número	Faltantes	Média	Desvio Padrão	Q1	Mediana	Q3	Mínimo	Máximo
Idade (meses)	281	0	146,8	34,7	124,0	144,0	177,0	74,0	214,0
Peso (kg)	280	1	81,7	26,3	61,6	80,2	98,5	26,6	183,6
Estatura (m)	280	1	2,1	9,8	1,5	1,6	1,6	1,2	165,4
P Estatura	280	1	69,0	27,1	50,0	75,0	93,3	1,0	100,0
IMC (kg/m ²)	280	1	33,3	6,8	28,5	32,4	37,7	17,2	62,1
ZIMC (CDC)	280	1	2,4	0,4	2,2	2,4	2,6	-0,1	3,3
ZIMC (WHO)	280	1	3,2	0,7	2,8	3,3	3,6	0,0	5,6
ZIMC (Conde&Monteiro)	280	1	3,6	0,9	3,1	3,5	4,1	0,1	6,6
CA PM	275	6	101,8	17,0	91,2	100,4	114,0	34,9	155,0
FC	271	10	83,1	14,6	73,3	82,0	90,0	54,0	134,0
MG (%)	241	40	39,2	7,0	34,0	39,0	43,0	18,0	60,0
CT	252	29	165,1	38,2	143,0	162,0	186,0	73,0	300,0
LDL-c	252	29	99,5	30,8	80,0	95,0	118,0	39,0	225,0

O peso dos pacientes na amostra varia de 27kg a 184kg. A estatura média é de 1,55 metros. Como esperado em indivíduos em fase de crescimento, tanto a variância do peso quanto a da estatura são razoavelmente altas. Este padrão também é observado no Gráfico B.7 e no Gráfico B.8. A média do percentil da estatura de 69,0 indica que as crianças devem ser razoavelmente altas, (Gráfico B.9). O IMC médio na amostra é 33,3, variando de 17,2 a 62,1 (Gráfico B.10). O índice padronizado ZIMC (CDC) apresenta média igual a 2,4 com desvio padrão de 0,4 e tem sua distribuição ilustrada pelo histograma (Gráfico B.11), sendo que o *box-plot* (Gráfico B.12) também foi construído e ressalta a existência de 3 valores atípicos, potenciais *outliers*. O Gráfico B.13 mostra que, no índice ZIMC, a etnia branca possui maior variabilidade e mediana,

quando comparado com as outras etnias agrupadas. O índice calculado segundo o padrão WHO tem sua distribuição ilustrada no Gráfico B.14. Na Tabela 5.1, observa-se que o ZIMC (WHO) tem média e desvio padrão superior ao ZIMC (CDC). O *box-plot* apresentado no Gráfico B.15 mostra um número maior de valores atípicos; o Gráfico B.16 não sugere diferenças importantes para o ZIMC (WHO) entre a etnia branca e as outras. O índice calculado, segundo as curvas C&M tem sua distribuição ilustrada no Gráfico B.17 e da Tabela 5.1, observa-se que é o índice com maior média e desvio-padrão dos três. O Gráfico B.18 mostra a existência de um maior número de valores atípicos comparativamente aos índices construídos pelo método CDC, porém inferior ao que é observado para o método WHO. Analisando o Gráfico B.19, não parece existir diferença importante entre os valores de ZIMC (C&M) para as diferentes etnias. A circunferência abdominal calculada no ponto médio teve média de 101,8 cm, e tem sua distribuição dos valores apresentada no Gráfico B.20. O valor médio para a frequência cardíaca foi de 83,1 /bpm, com a mediana próxima à média verificamos a presença de *outliers* acima do limite superior (Gráfico B.21).

Para o percentual de massa de gordura obtivemos um valor médio de aproximadamente 39% com um desvio padrão de 7%. Pelo Gráfico B.22 vemos que o percentual de massa de gordura se distribui de forma homogênea ao redor do primeiro e terceiro quartil (34% e 43% respectivamente). O Gráfico B.23 mostra que o percentual de gordura para o grupo de brancos parece ter uma variabilidade menor, porém, as medianas dos dois grupos estão bem próximas.

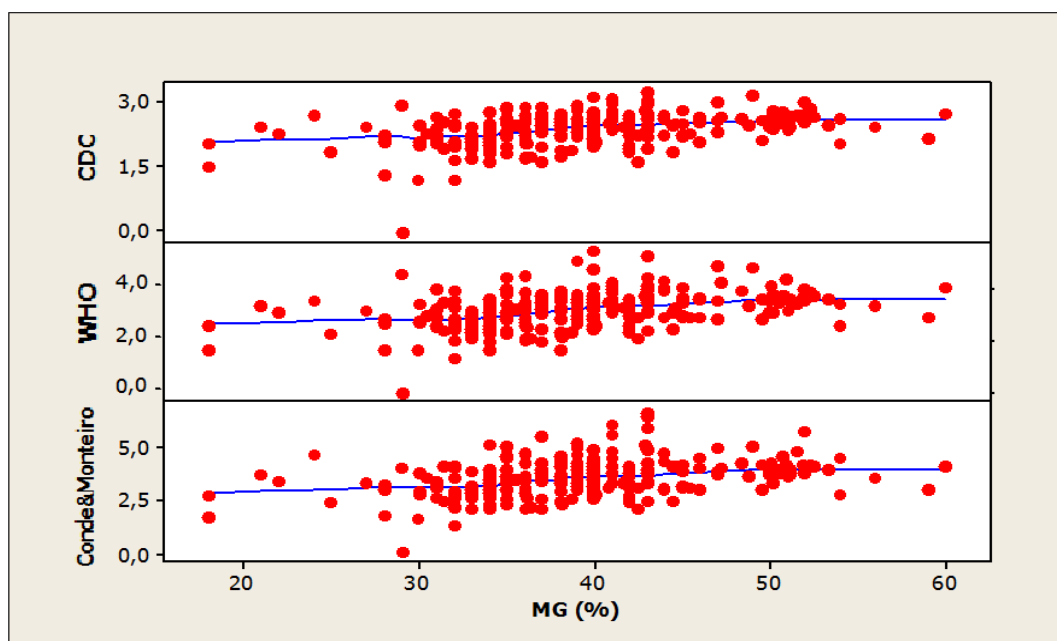
O valor médio de colesterol total foi de 165,1 mg/dL, analisando o Gráfico B.24 verificamos a presença de *outliers* e é possível notar que sua mediana se encontra abaixo da média. O LDL-c médio foi de 99,5 mg/dL e pelo Gráfico B.25 notamos a presença de três valores discrepantes acima do limite superior construído pelo *box-plot*.

Para avaliar a relação entre os 3 métodos, foram construídos diagramas de dispersão, que incluiu um curva suavizada construída pelo método *Lowess* (Cleveland, 1979), que busca mostrar o comportamento conjunto de dois índices. Se eles fossem exatamente iguais, deveriam estar assentados na bissetriz do gráfico. O Gráfico B.26 apresenta os valores índice calculado segundo o CDC em função do índice calculado pelo WHO. Verificamos que os valores obtidos pelo método WHO cresce a uma

velocidade maior do que os valores obtidos pelo método CDC. O Gráfico B.27 mostra a mesma relação entre o método americano e o método brasileiro, porém parece que o crescimento é um pouco mais acentuado para o método C&M do que o verificado no gráfico anterior. Comparando agora os valores do ZIMC calculados pelo método WHO em função do método C&M (Gráfico B.28), parece existir uma relação em que os valores calculados por C&M crescem a uma velocidade maior que o observado no método WHO; nota-se também uma maior dispersão entre os pontos no gráfico, o que sugere uma menor concordância entre os dois métodos.

Tendo em vista que a informação da variável MG (%) é uma importante forma de diagnosticar a obesidade, foram construídos gráficos de dispersão para avaliar o comportamento dos índices com relação à MG (%). A seguir é apresentado o Gráfico 5.1, em que os três métodos de obtenção do índice padronizado de IMC são representados em relação à porcentagem de massa de gordura, que incluiu um curva suavizada construída pelo método *Lowess*. Verifica-se que os pontos estão bastante dispersos, porém, parece existir uma tendência de que à medida que a massa de gordura aumenta, o mesmo parece ocorrer com o índice ZIMC. É possível verificar uma concentração da porcentagem massa de gordura no intervalo de 30% a 50%.

Gráfico 5.1: Dispersão dos índices ZIMC: CDC WHO e C&M segundo a massa de gordura.



Para quantificar as possíveis relações discutidas nos gráficos anteriores foram calculadas medidas de correlação de Spearman, (Bussab e Morettin, 2010), apresentadas na Tabela 5.2. Verificamos que, dentre os três métodos de cálculo do ZIMC, o que possui maior correlação com a porcentagem de massa de gordura é o método WHO (0,43), porém esse valor pode ser considerado relativamente baixo. Analisando a correlação entre os métodos de construção dos índices, observamos que todos possuem uma alta correlação, sendo a maior entre os métodos CDC e WHO, igual a 0,95.

Tabela 5.2: Correlação de Spearman para os diferentes índices ZIMC e a MG (%).

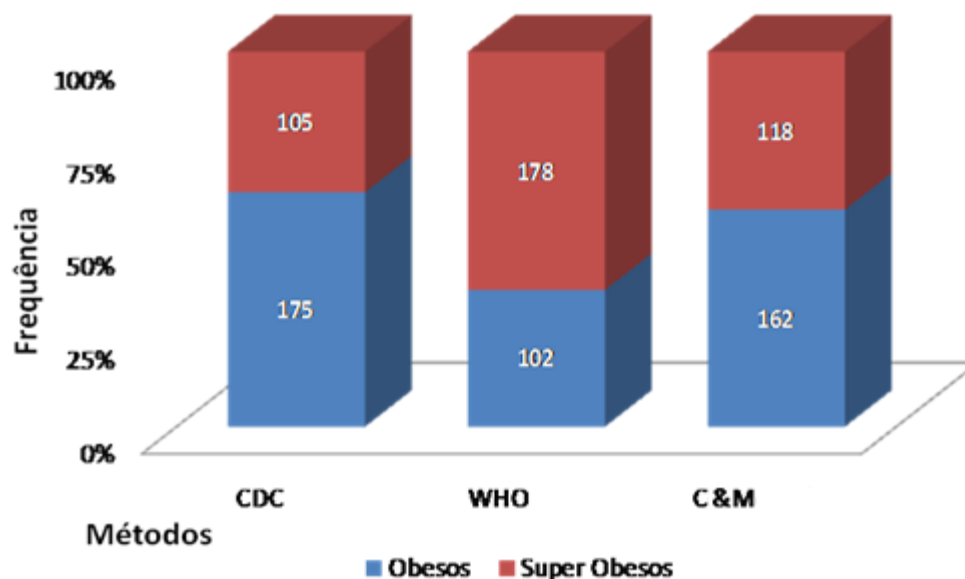
Correlação Spearman	MG %	CDC	WHO	C&M
MG (%)	1,00			
CDC	0,42	1,00		
WHO	0,43	0,95	1,00	
C&M	0,40	0,94	0,90	1,00

A amostra será dividida em dois grupos, obesos e super obesos. Seguindo metodologia proposta pelas pesquisadoras, um paciente foi considerado super obeso para o método CDC quando seu ZIMC era maior ou igual que 2,5; caso contrário, era considerado obeso. Para o método WHO, foi considerado super obeso quando o seu ZIMC era maior ou igual a 3,0, caso contrário era considerado obeso. Para o método C&M como não existe um valor de corte consagrado na literatura, foi feita uma análise de regressão (Stokes et. al. 2001), a fim de se construir um modelo que estabelecesse uma relação matemática entre o método CDC e o método C&M (detalhes são apresentados no Apêndice C). Após a obtenção dessa equação, foi obtido o valor de 3,71 para ponto de corte, que seria correspondente ao valor 2,5 na curva para o método CDC. Consequentemente, o paciente foi classificado como super obeso pelo método C&M se apresentava valor de ZIMC maior ou igual a 3,71; caso contrário, foi classificado como obeso.

Para avaliarmos os dados segundo a classificação obtida para os diferentes métodos foi feito um gráfico de barras sobrepostas (Gráfico 5.2) em que pode ser verificado que no método CDC, a maior parte dos pacientes (63%) foi classificada como

obesa; no método WHO, a maior parte (64%) foi classificada como super obesa; já pelo método C&M apenas 162 pacientes foram classificados como obesos, sendo que 42% foram classificados como super obesos.

Gráfico 5.2 : Classificação para as diferentes curvas de ZIMC.



A concordância entre índices calculados foi quantificada através de uma medida denominada índice Kappa, (Stokes et. al. 2001), para os métodos considerados dois a dois; expressões utilizadas são apresentadas no Apêndice C. Este índice é um método utilizado para avaliar a concordância entre escalas categóricas avaliadas por mais de um fator determinante. Os resultados possuem alternativas de interpretação devidas a diversos autores. Usaremos a de Landis e Koch (1977).

Tabela 5.3: Interpretação do índice Kappa sugerida por Landis e Koch.

Valores de Kappa	Interpretação
<0	sem concordância
0-0,19	concordância pobre
0,20-0,39	concordância fraca
0,40-0,59	concordância moderada
0,60-0,79	concordância substancial
0,80-1,00	concordância quase perfeita

Para avaliar a concordância entre o método CDC e WHO tem-se, pela Tabela A.3, que os métodos discordam em 75 pacientes; calculando o índice Kappa obtém-se um valor igual a 0,50 (com intervalo de confiança de 95% igual a [0,40;0,60]) que indica uma concordância moderada entre os métodos, pelo critério da Tabela 5.3. Para os métodos CDC e Conde & Monteiro observa-se que apenas 27 pacientes apresentam discordância (Tabela A.4), com o índice Kappa igual a 0,80 ([0,68; 0,92]). Tal valor mostra concordância quase perfeita. Avaliando a concordância para os métodos C&M e WHO tem-se que apenas 68 pacientes resultam discordantes (Tabela A.5), com o valor Kappa igual a 0,53 ([0,43; 0,64]), isto é, uma concordância moderada.

Dada a relação de obesidade com morbidade já estabelecida na literatura, uma forma de identificar a curva mais adequada para a classificação é avaliar se pacientes indicados como super obesos possuem um maior número de doenças.

As doenças do estudo são a hipertensão arterial, glicemia alterada, dislipidemia e resistência à insulina e são determinadas através das seguintes variáveis P PA Sistólica, P PA Diastólica, Glicose, HDL-c , TG e a HOMA-IR.

As informações sobre as doenças foram sumarizadas na Tabela 5.4, com as indicações da presença, ausência ou sem informação(verificadas também pelos gráficos B.29, B.30, B.31 e B.32). Após desconsiderar os dados faltantes, restaram 209 pacientes com informações de todas as doenças. Considerando o número de doenças apresentado por cada paciente, foi feito o Gráfico B.33.

Tabela 5.4: Número de pacientes segundo as ocorrências de doenças

Doenças	Presença	Ausência	Faltantes
Hipertensão Arterial	98 (35%)	172 (61%)	11 (4%)
Glicemia Alterada	18 (6%)	231 (82%)	32 (11%)
Dislipidemia	132 (47%)	114 (41%)	35 (12%)
Resistência à Insulina	145 (52%)	65 (23%)	71 (25%)

Elaboramos a Tabela 5.5 em que 14% dos pacientes não possuem doenças, 30% possuem uma doença, 39% possuem duas doenças, 16% dos pacientes possuem 3 doenças, e os restantes (10%) possuem todas as quatro doenças.

Tabela 5.5: Número de pacientes segundo o número de doenças

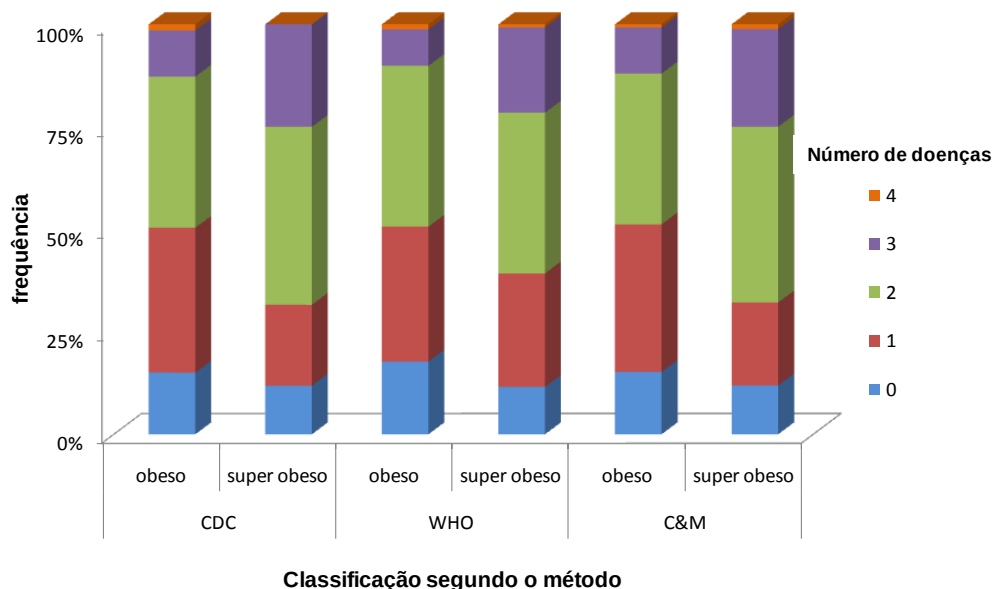
Número de doenças	Número	%
0	29	14%
1	62	30%
2	82	39%
3	34	16%
4	2	1%

Com o intuito de encontrar alguma relação entre os métodos de cálculos do ZIMC e a quantidade de doenças foram construídos *box-plots*. No Gráfico B.34 é apresentado o método CDC para as diferentes quantidades de doenças. Aparentemente os valores do ZIMC não se diferenciam muito para os diferentes números de doenças, mas é interessante notar que apesar do grupo com quatro doenças ser composto por apenas dois pacientes, para este método ambos são considerados pacientes obesos e não super obesos. Para o método WHO (Gráfico B.36), percebemos que o grupo que não apresenta doenças é o que apresenta a maior variabilidade. O grupo com três doenças apresenta menor variabilidade e maior mediana. No grupo com quatro doenças um dos pacientes é classificado como obeso e o outro como super obeso, o mesmo acontece para o método C&M (Gráfico B.36).

Para verificar a distribuição dos pacientes nos diferentes grupos de soma das doenças, para cada classificação (obeso e super obesos) nos diferentes métodos, foi construído um gráfico de barras sobrepostas (Gráfico 5.3), que mostra as quantidades e suas porcentagens dentro de cada classificação para os 209 pacientes. Para o método CDC, o grupo de obesos possui um total de 133 pacientes (Tabela A.6) e dentre estes observa-se que metade dos pacientes apresenta entre zero e uma doença e a outra metade apresenta duas doenças ou mais (Tabela A.7); o grupo dos super obesos possui um total de 76 pacientes e dentre estes 68% possuem duas doenças ou mais. Para o método WHO, verifica-se que 79 pacientes são classificados como obesos contra 130 que possui nenhuma ou uma doença, a classificação dos super obesos possui um total de 130 crianças e 61% delas apresenta duas doenças ou mais. Uma relação semelhante é feita para o método C&M: dentre 125 pacientes classificados

como obesos, 51% apresenta zero ou uma doença e dentre os 190 pacientes classificados como super obesos 68% possui duas doenças ou mais.

Gráfico 5.3: Número de doenças para obesos e super obesos segundo a classificação obtida para cada um dos 3 métodos.



6. Análise Inferencial

Um dos principais objetivos do projeto é avaliar qual dos três métodos (WHO, CDC ou C&M) classifica o paciente em obeso ou super obeso, coerentemente com a ocorrência das condições clínicas sabidamente relacionadas com a obesidade e avaliadas nos pacientes em questão. Em outras palavras, quanto maior o número de doenças apresentadas pelo paciente, maior deveria ser a chance dele ser classificado como super obeso, sendo que o mesmo pode ser dito com relação ao percentual de massa de gordura.

Para avaliar a relação entre a chance de classificação em super obeso e as condições clínicas mencionadas, foi utilizado o método de regressão logística (ver, por

exemplo, Stokes et al, 2001), para cada um dos 3 métodos de classificação considerados. Maiores detalhes são dados no Apêndice D.

Para o nosso estudo serão construídos 3 modelos logísticos, um para cada método, onde a variável resposta será a classificação em obeso (0) e super obeso (1), e os fatores cuja influência na chance de classificação em super obeso será avaliada serão o número de doenças presente nos pacientes (0 para nenhuma doença presente; 1 para uma doença presente; 2 para duas doenças presentes e 3 para três ou quatro doenças presentes) e percentual de massa de gordura.

A amostra foi reduzida para 192 crianças obesas após excluirmos os pacientes que não possuíam informação sobre alguma doença ou sobre o percentual de massa de gordura.

6.1 Modelo proposto

A partir da classificação pelos métodos, foi proposto a construção de três modelos que procura associar estes resultados, com o número de doenças e percentual de massa de gordura. (Detalhes no Apêndice D)

Com os resultados dos ajustes dos modelos foram construídas as Tabelas A.8, A.9 e A.10, para os métodos CDC, WHO e C&M respectivamente. Considerando um nível de significância igual a 5%, verificamos que todos os fatores explicativos são marginalmente significativos. Desta forma, podemos concluir que à medida que o número de doenças aumenta, a *chance* de classificar o paciente como super obeso aumenta (todas as estimativas apresentam sinal positivo nas tabelas). Além disso, analisando a estimativa associada com o percentual de massa de gordura, concluímos, como esperado, que à medida que esta aumenta, maior é a chance de classificação do paciente como super obeso. Notamos que essas conclusões estão de acordo como o esperado pela pesquisadora e confirmam os achados da análise descritiva.

Para avaliar o modelo, foi testado se existe interação entre as variáveis e concluímos que não foi possível verificar efeito de interação entre as variáveis. Isto pode ser visto pelos Gráficos B.37, B.38 e B.39, em que é possível verificar que, quanto maior o número de doenças e o percentual de massa de gordura, maior a probabilidade estimada.

6.2 Razão de chances

Para interpretar os parâmetros que se mostram estatisticamente significantes, foi calculada a razão de chances, para cada um dos fatores: número de doenças e percentual de massa de gordura. A interpretação da razão de chances se dá em termos relativos. Por exemplo, para o fator número de doenças, a razão de chances representa a chance de um paciente com uma doença ser classificado como super obeso comparativamente a um paciente sem doença ser classificado como super obeso, considerando que ambos os pacientes tenham o mesmo percentual de massa de gordura. Já para a massa de gordura, a razão de chances reflete o quanto muda a chance de classificação como super obeso ao aumentarmos o percentual de massa de gordura em uma unidade para pacientes com o mesmo número de doenças. A seguir discutimos as razões de chances calculadas, através dos modelos ajustados, para cada um dos três métodos utilizados.

6.2.1 Método CDC

Para o número de doenças, obtemos uma estimativa da razão de chances igual a $e^{0,490} = 1,63$, com um intervalo de confiança (IC) com coeficiente de confiança de 95% dado por [1,14; 2,34]. Desta forma, mantendo-se o percentual de massa de gordura constante, a ocorrência de uma doença aumenta a chance de classificação em super obeso em 60%. Alternativamente, podemos interpretar esse valor da seguinte forma: à medida em se aumenta o número de doenças em uma unidade, a chance de classificação em super obeso é multiplicada em 1,63, mantendo-se constante o percentual de massa de gordura. Podemos dizer ainda que a chance de um paciente que possui três doenças ser classificado como super obeso, é aproximadamente, cinco vezes a chance de um paciente que não possui doenças ser diagnosticado como super obeso.

A mesma relação foi utilizada para se calcular a razão de chances segundo o percentual de massa de gordura, em que à medida que se aumenta em uma unidade o percentual de massa de gordura, para pacientes com mesmo número de doenças, a

chance do diagnóstico do paciente ser super obeso é multiplicada por 1,12 vezes (ou seja, é 12% maior), com IC(95%)= [1,06; 1,18].

6.2.2 Método WHO

Obtemos uma estimativa da razão de chances, para o número de doenças, igual a $e^{0,346} = 1,41$, com IC(95%) dado por [1,00; 1,99]. Desta forma, mantendo-se o percentual de massa de gordura fixado, a presença de uma doença aumenta a chance de classificação em super obeso em 40%. De maneira análoga, podemos interpretar esse valor da seguinte maneira: à medida que aumenta o número de doenças em uma unidade, a chance de classificação em super obeso é multiplicada em 1,41, mantendo-se constante o percentual de massa de gordura. Ainda é possível concluir que a chance de um paciente que possui três doenças ser classificado como super obeso, é aproximadamente, quatro vezes a chance de um paciente que não possui doenças ser diagnosticado como super obeso.

Para calcular a razão de chances segundo o percentual de massa de gordura, foi utilizada a mesma relação, em que à medida que se aumenta em uma unidade o percentual de massa de gordura, para pacientes com mesmo número de doenças, a chance de o diagnóstico do paciente ser super obeso aumenta em 1,11 vezes (ou seja, é 11% maior), com IC(95%) igual a [1,05; 1,17].

6.2.3 Método Conde & Monteiro

Para o número de doenças, obtemos uma estimativa da razão de chances dada por $e^{0,497} = 1,64$, com IC(95%) igual a [1,16; 2,34]. Assim, fixando o percentual de massa de gordura, a ocorrência de uma doença aumenta a chance de classificação em super obeso em 60%. De maneira similar, é possível interpretar esse valor da seguinte forma: à medida em se aumenta o número de doenças em uma unidade, a chance de classificação em super obeso é multiplicada em 1,64, mantendo-se constante o percentual de massa de gordura. Podemos ainda dizer que a chance de um paciente que possui três doenças ser classificado como super obeso é aproximadamente cinco vezes a chance de um paciente que não possui doenças ser diagnosticado como super obeso.

A mesma relação foi utilizada para calcular a razão de chances segundo o percentual de massa de gordura, onde à medida que se aumenta em uma unidade o percentual de massa de gordura, para pacientes com mesmo número de doenças, a chance de o diagnóstico do paciente ser super obeso aumenta em 1,12 vezes (ou seja, é 12% maior), com IC(95%) igual a [1,06; 1,18]

Todos os modelos ajustados tiveram suas suposições verificadas. Maiores detalhes são apresentados no Apêndice E.

7. Outras Análises

Foram feitas outras análises para buscar diferenças entre os métodos, como por exemplo: (i) ao invés de considerarmos o número de doenças apresentado pelo paciente, utilizamos a presença (ou não) de cada uma das 4 doenças e ajustamos uma regressão logística para cada método. Assim, foram incluídos como fatores explicativos a ocorrência ou não de cada uma das doenças (dislipidemia, Glicemia Alterada, Hipertensão arterial, resistência à insulina) além do percentual de massa de gordura. O modelo ajustado mostrou que a resistência à insulina e a variável MG(%) são as únicas estatisticamente significantes para os 3 métodos, considerando um nível de significância igual a 5%; (ii) uma regressão logística para verificar se alguma doença ou o percentual de massa de gordura influenciam na discordância entre os métodos, foram feitas comparações dois a dois, ou seja, quando um método classificou o paciente como obeso e outro método como super obeso (CDC/WHO, CDC/C&M e WHO/C&M) e também comparações para os 3 métodos, ou seja, quando um método classificou o paciente como obeso e os outros dois como super obeso e vice e versa (CDC/WHO/C&M), em que zero era a resposta para a concordância e um para o caso de existir discordância. No ajuste desse modelo, não foi encontrada nenhuma variável estatisticamente significativa, ou seja, nenhuma doença ou percentual de massa de gordura influenciam na discordância entre os métodos; (iii) construímos um modelo logístico em que a variável resposta era obeso (0) e super obeso (1), e como variáveis explicativas o número de doenças excluindo a doença resistência à insulina (presente em 70% dos pacientes da amostra) e o percentual de massa de gordura, para a amostra toda (192 crianças, excluindo os valores faltante), como resposta obtivemos

que o número de doenças se torna significativo apenas para o método C&M e a variável MG (%) continua significativa para todos os métodos, isto poderia nos sugerir que o método do C&M é o único método que considera o número de doenças para a classificação em obesos e super obesos, com isso poderíamos falar que ele é o método que melhor faz o diagnóstico das crianças, mas isso apenas se fosse possível desconsiderar a doença resistência à insulina, o que não foi uma opção dada pelas pesquisadoras. (iv) foi feita uma análise discriminante para os 3 métodos, porém a proporção de acerto é muito próxima para estes 3 métodos e suas funções discriminantes também foram muito parecidas, o que não nos mostra diferença nas classificações dos métodos (v) uma análise de agrupamento foi feita com a intenção de separar em dois grupos segundo as doenças e o percentual de massa de gordura, e o critério de separação dos grupos foi a doença glicemia alterada (presente em 12 pacientes dos 192) o que não nos induziu a nenhuma conclusão.

8. Conclusões

O principal objetivo da análise foi identificar qual dos três métodos, CDC WHO e C&M, fornecem classificação de obesos e super obesos com maior coerência em relação ao percentual de massa de gordura e aos indicadores de morbidade baseados na ocorrência ou não de hipertensão arterial, glicemia alterada, resistência à insulina e dislipidemia.

As análises feitas indicaram que à medida que o número de doenças e o percentual de massa de gordura aumenta a chance de classificar um paciente como super obeso também aumenta. A intensidade como isso ocorre foi a mesma para os três métodos.

Através da razão de chances vemos que não existe diferença significativa entre os métodos tanto para o número de doenças quanto para o percentual de massa de gordura, desta forma concluímos que pela amostra apresentada, não é possível verificar, estatisticamente, a existência de um método melhor para classificar um paciente em obeso ou super obeso. Entretanto, avaliando exploratoriamente os resultados obtidos, notamos que a razão de chances para o método C&M, apesar de muito próxima do método CDC, foi ligeiramente maior, para o número de doenças e

igual para o percentual de massa de gordura, para esta amostra. O método WHO foi o que obteve os menores valores para as razões de chances calculadas, tanto para o número de doenças quanto para a porcentagem de massa de gordura.

APÊNDICE A
TABELAS

Tabela A.1: Proporção de pacientes por gênero.

Gênero	Número	%
Masculino	121	57%
Feminino	160	43%

Tabela A.2: Proporção de pacientes para os diferentes tipos de cor.

COR	Número	%
Branco	210	75%
Parda	2	1%
Negro	19	7%
Amarelo	4	1%
Mulato	46	16%

Tabela A.3: Concordância entre o método CDC e WHO.

WHO	CDC		Total
	obeso	super obeso	
obeso	101	1	102
super obeso	74	104	178
total	175	105	280

Tabela A.4: Concordância entre os métodos CDC e C&M.

C&M	CDC		Total
	obeso	super obeso	
obeso	155	7	162
super obeso	20	98	118
total	175	105	280

Tabela A.5: Concordância entre o método WHO e C&M.

C&M	WHO		Total
	obeso	super obeso	
obeso	98	64	162
super obeso	4	114	118
total	102	178	280

Tabela A.6: Soma das doenças para as diferentes classificações segundo os 3 métodos

Número de doenças	CDC		WHO		C&M	
	obesos	super obesos	obesos	super obesos	obesos	super obesos
0	20	9	14	15	19	10
1	47	15	26	36	45	17
2	49	33	31	51	46	36
3	15	19	7	27	14	20
4	2	0	1	1	1	1
Total	133	76	79	130	125	84

Tabela A.7: Porcentagem para soma das doenças para as diferentes classificações segundo os 3 métodos

Número de doenças	CDC		WHO		C&M	
	obesos	super obesos	obesos	super obesos	obesos	super obesos
0	15	12	18	12	15	12
1	35	20	33	28	36	20
2	37	43	39	39	37	43
3	11	25	9	21	11	24
4	2	0	1	1	1	1

Tabela A.8: Resultados do ajuste do modelo logístico proposto para o método CDC com todas as observações.

	Estimativa	Erro Padrão	Valor p
Constante	-5,903	1,142	<0,001
nº de doenças	0,490	0,184	0,008
MG(%)	0,115	0,027	<0,001

Tabela A.9: Resultados do ajuste do modelo logístico proposto para o método WHO com todas as observações.

	Estimativa	Erro Padrão	Valor p
Constante	-3,948	1,056	<0,001
nº de doenças	0,346	0,176	0,049
MG(%)	0,102	0,027	<0,001

Tabela A.10: Resultados do ajuste do modelo logístico proposto para o método C&M com todas as observações.

	Estimativa	Erro Padrão	Valor p
Constante	-5,568	1,111	<0,001
nº de doenças	0,497	0,180	0,006
MG(%)	0,112	0,027	<0,001

Tabela A.11: Discriminação do modelo logístico proposto para o método CDC.

Predito	Observado		
	Obeso	super obeso	Total
Obeso	83	20	103
super obeso	37	52	89
Total	120	72	192

Tabela A.12: Discriminação do modelo logístico proposto para o método WHO.

Predito	Observado		
	obeso	super obeso	Total
obeso	47	44	91
super obeso	24	77	101
total	71	121	192

Tabela A.13: Discriminação do modelo logístico proposto para o método C&M.

Predito	Observado		
	Obeso	super obeso	Total
obeso	86	27	113
super obeso	26	53	79
total	112	80	192

Tabela A.14: Resultados do ajuste do modelo logístico proposto para o método CDC sem as cinco observações discrepantes.

	Estimativa	Erro Padrão	Valor p
Constante	-7,207	1,283	<0,001
nº de doenças	0,516	0,194	0,008
MG(%)	0,145	0,030	<0,001

Tabela A.15: Resultados do ajuste do modelo logístico proposto para o método WHO sem as cinco observações discrepantes.

	Estimativa	Erro Padrão	Valor p
Constante	-5,038	1,174	<0,001
nº de doenças	0,413	0,184	0,025
MG(%)	0,128	0,030	<0,001

Tabela A.16: Resultados do ajuste do modelo logístico proposto para o método C&M sem as cinco observações discrepantes.

	Estimativa	Erro Padrão	Valor p
Constante	-7,587	1,322	<0,001
nº de doenças	0,546	0,194	0,005
MG(%)	0,159	0,032	<0,001

APÊNDICE B

GRÁFICOS

Gráfico B.1: Distribuição dos pacientes segundo gênero.

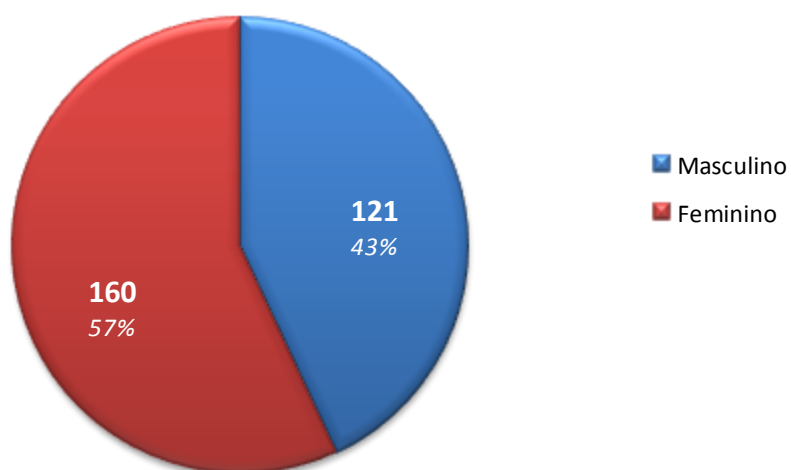


Gráfico B.2: Distribuição dos pacientes segundo grupos étnicos.

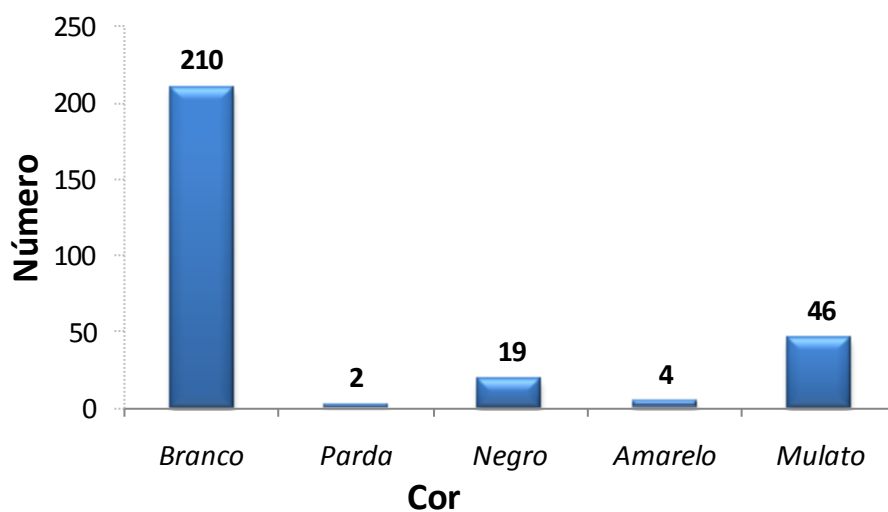


Gráfico B.3: Distribuição dos pacientes segundo grupos étnicos (agrupados)

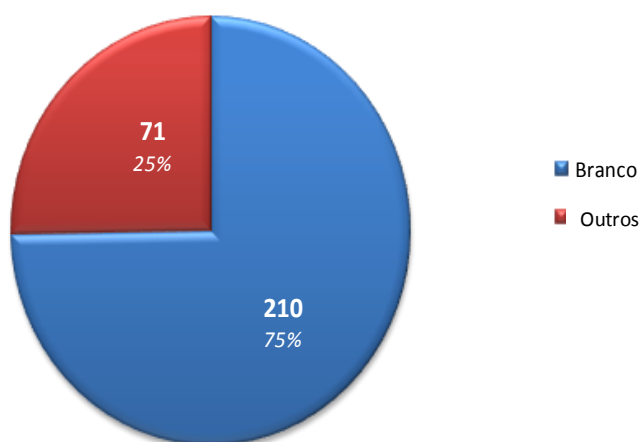


Gráfico B.4: *Box-plot* para a Idade.

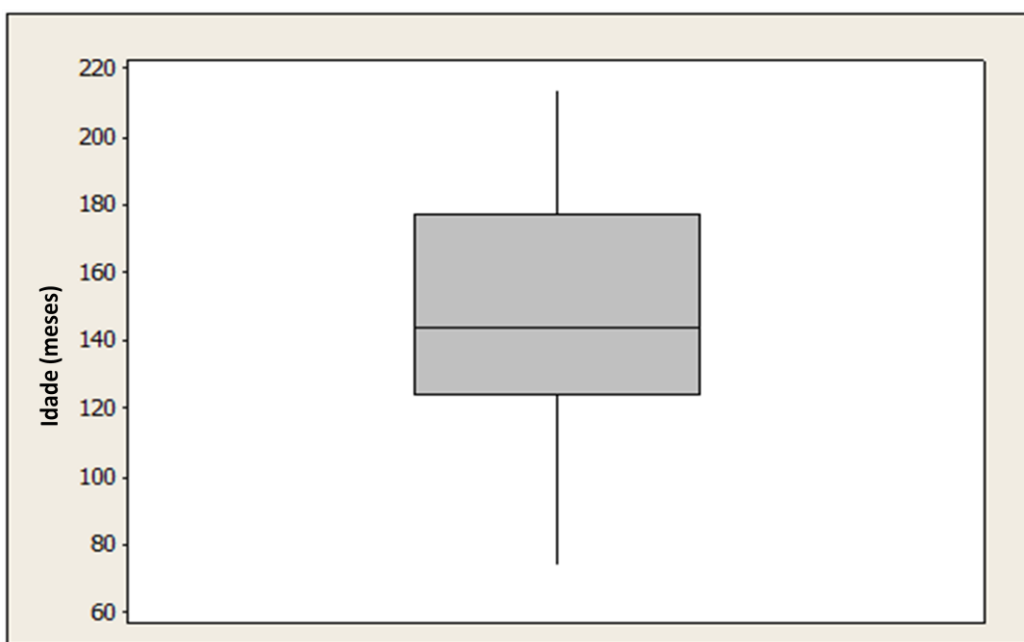


Gráfico B.5: *Box-plot* para a Idade segundo gênero.

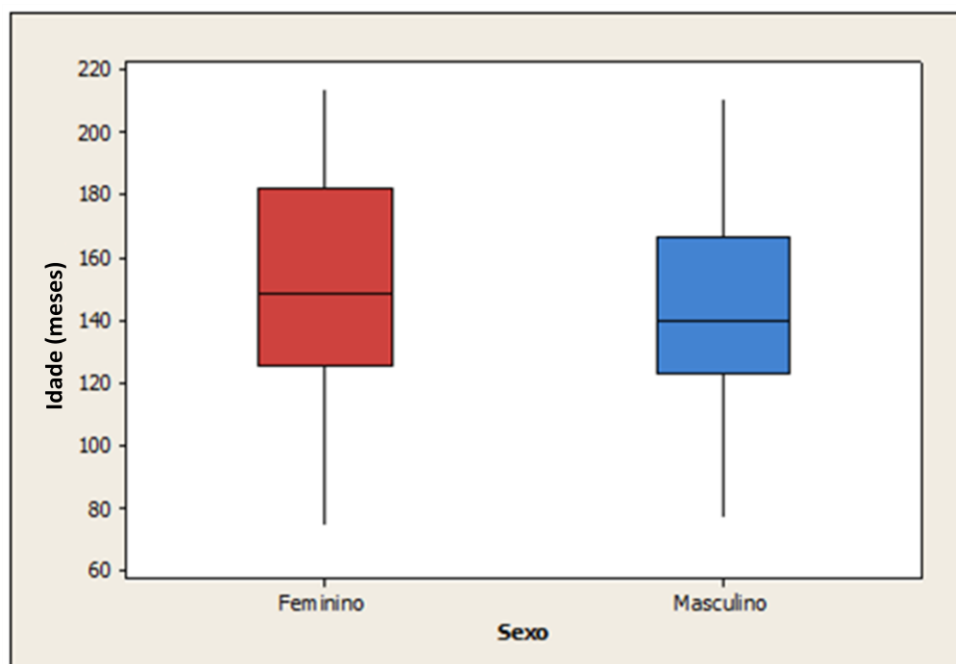


Gráfico B.6: *Box-plot* para a Idade segundo Cor.

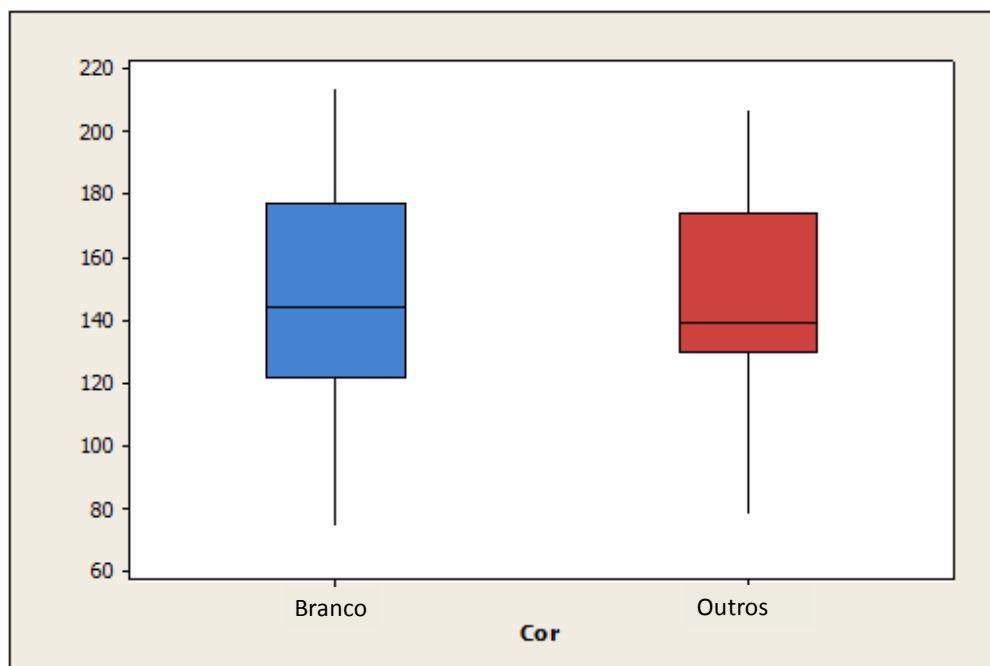


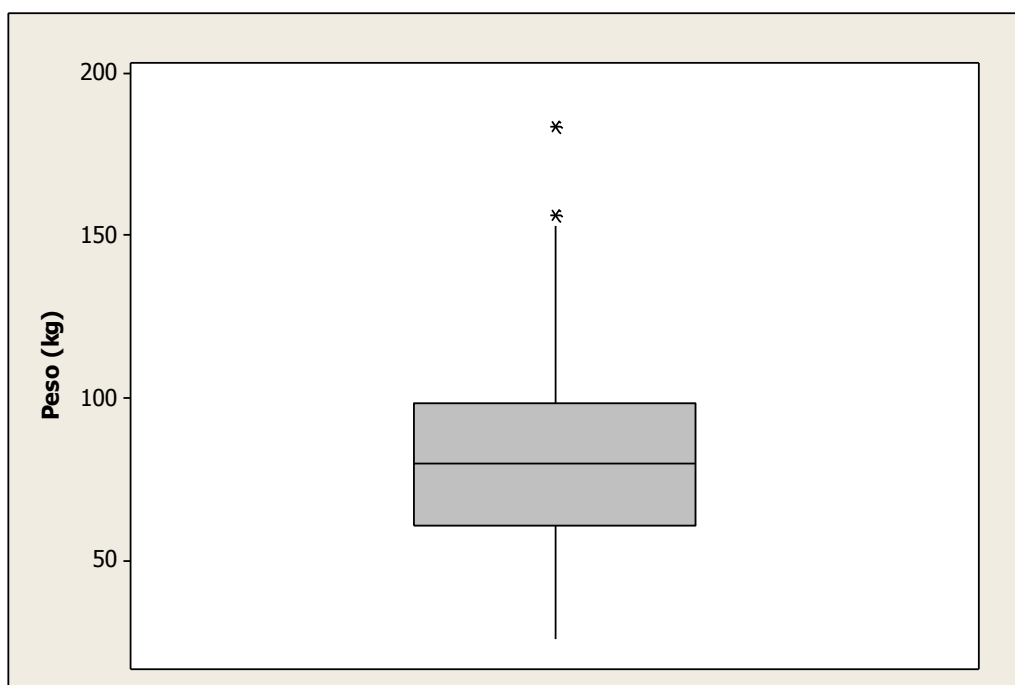
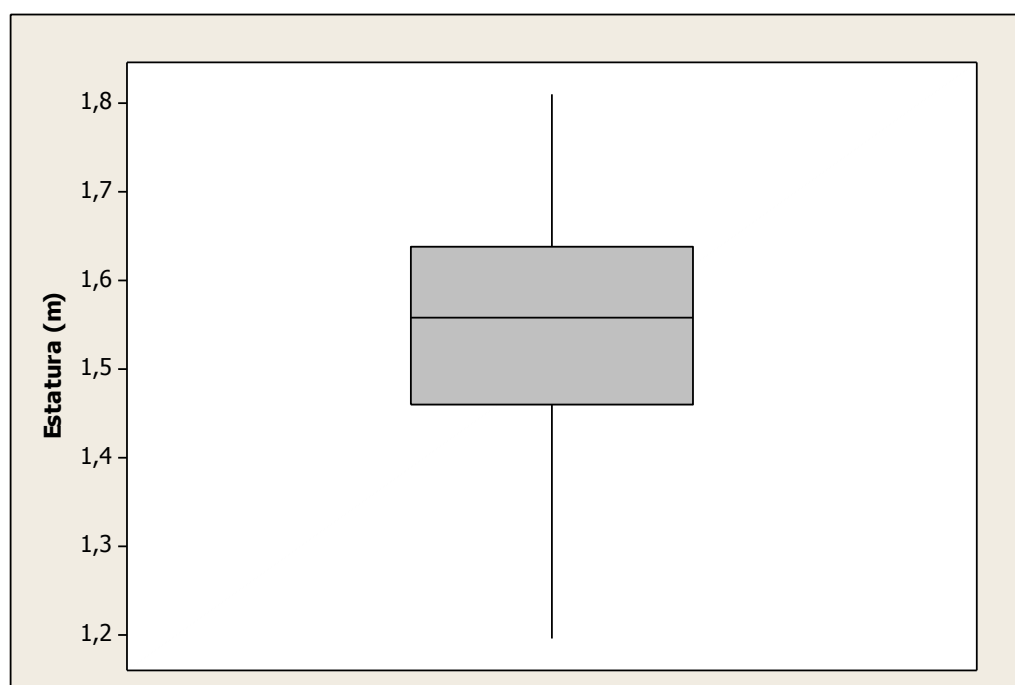
Gráfico B.7: *Box-plot* para o Peso.**Gráfico B.8:** *Box-plot* para a estatura.

Gráfico B.9: *Box-plot* para o percentil da estatura.

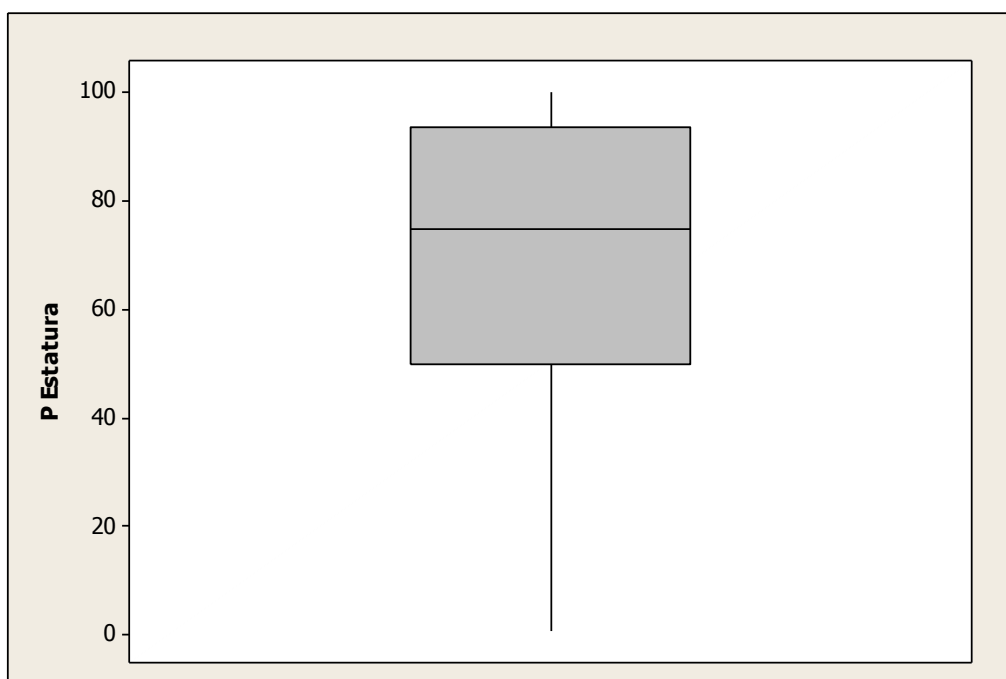


Gráfico B.10: *Box-plot* para o IMC.

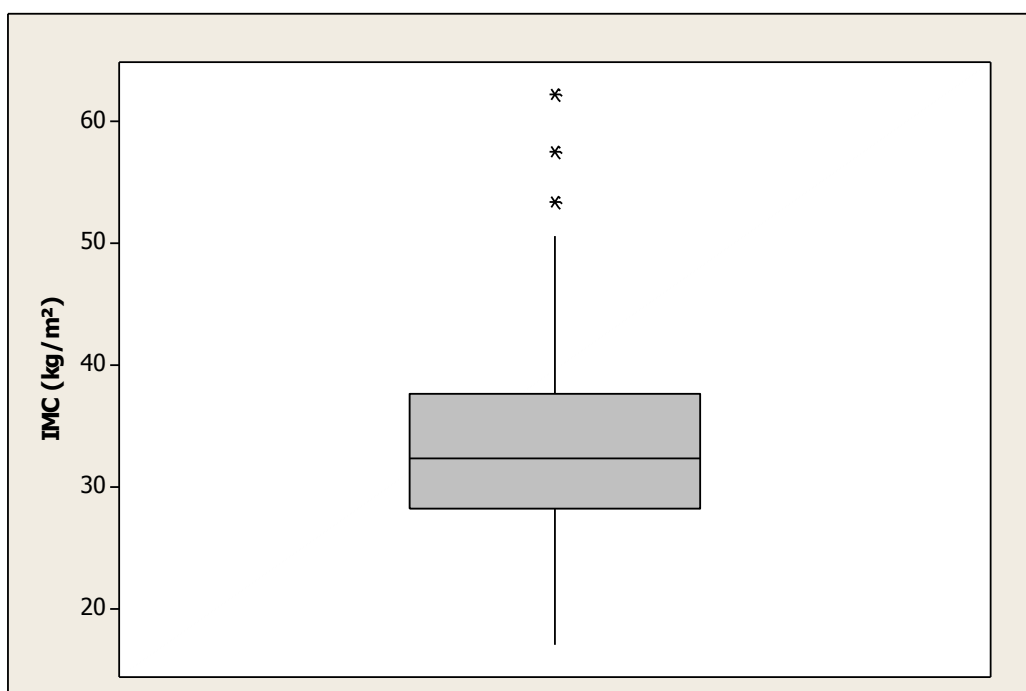


Gráfico B.11: Histograma para o os valores de ZIMC pelo método CDC.

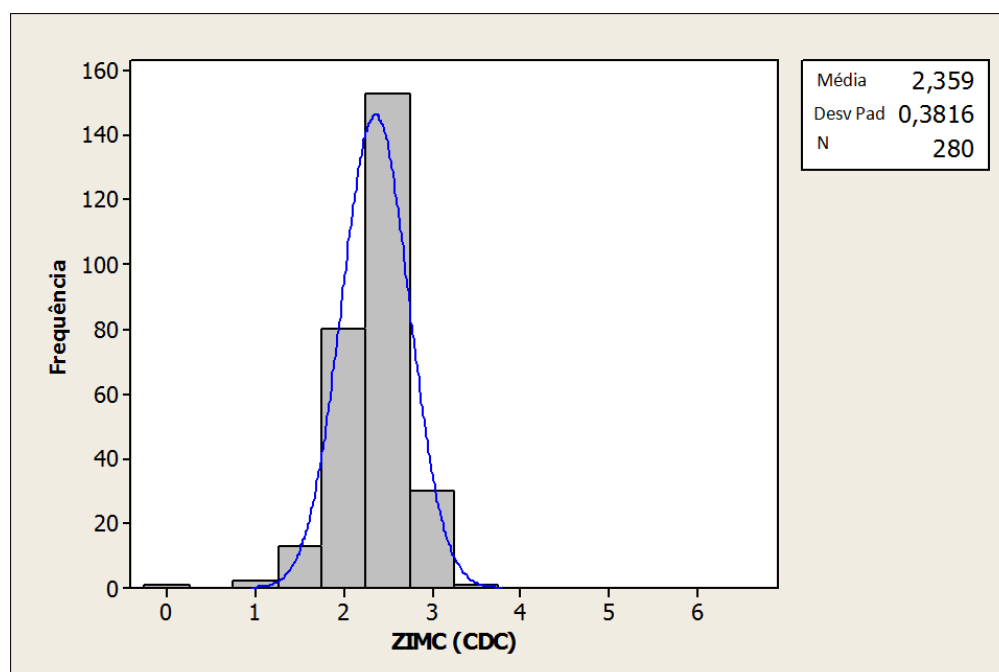


Gráfico B.12: *Box-plot* para o IMC padronizado pelo método CDC.

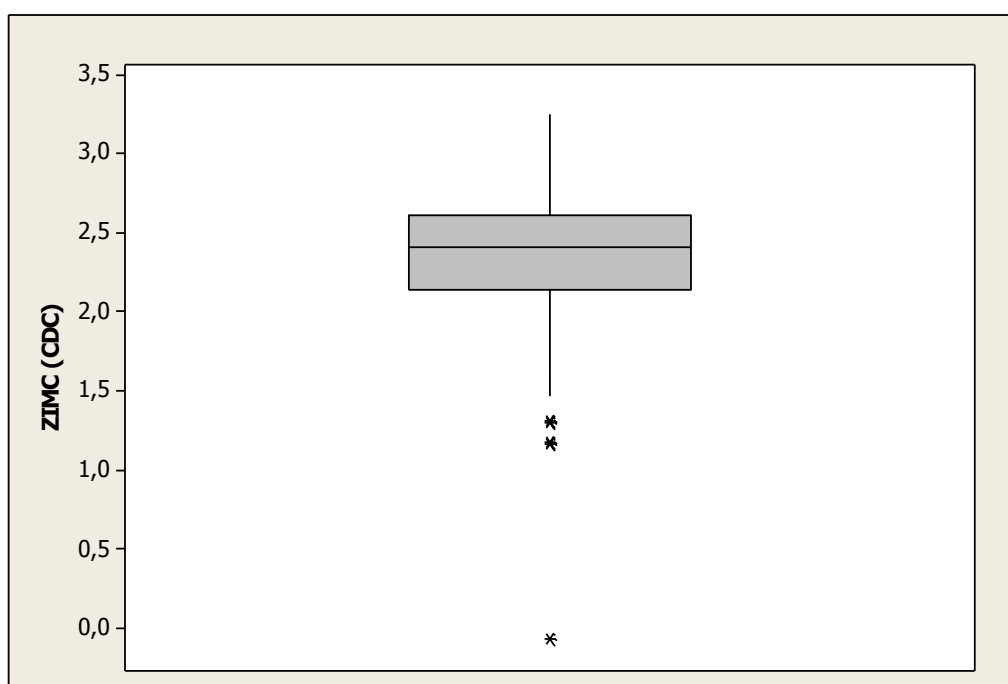


Gráfico B.13: *Box-plot* para o IMC padronizado pelo método CDC segundo cor.

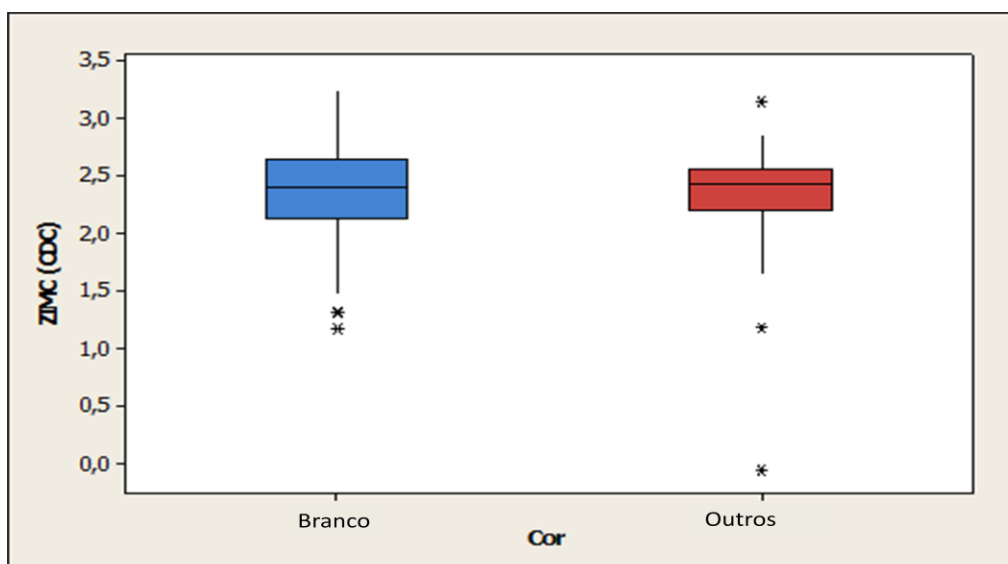


Gráfico B.14: Histograma para os valores de ZIMC pelo método WHO.

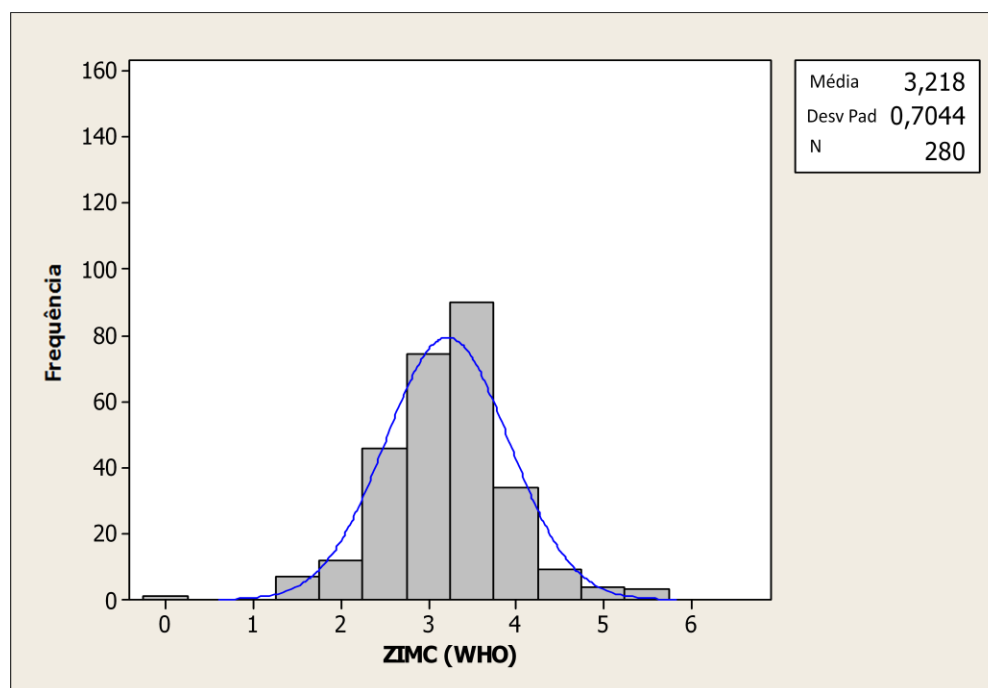


Gráfico B.15: *Box-plot* para o IMC padronizado pelo método WHO.

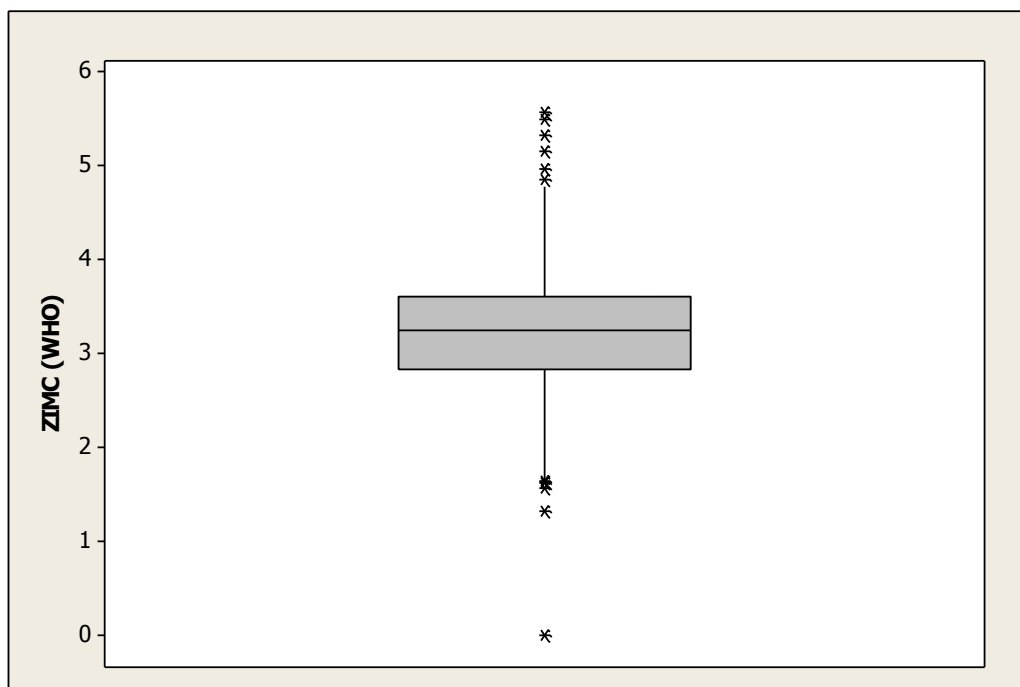


Gráfico B.16: *Box-plot* para o IMC padronizado pelo método WHO segundo cor.

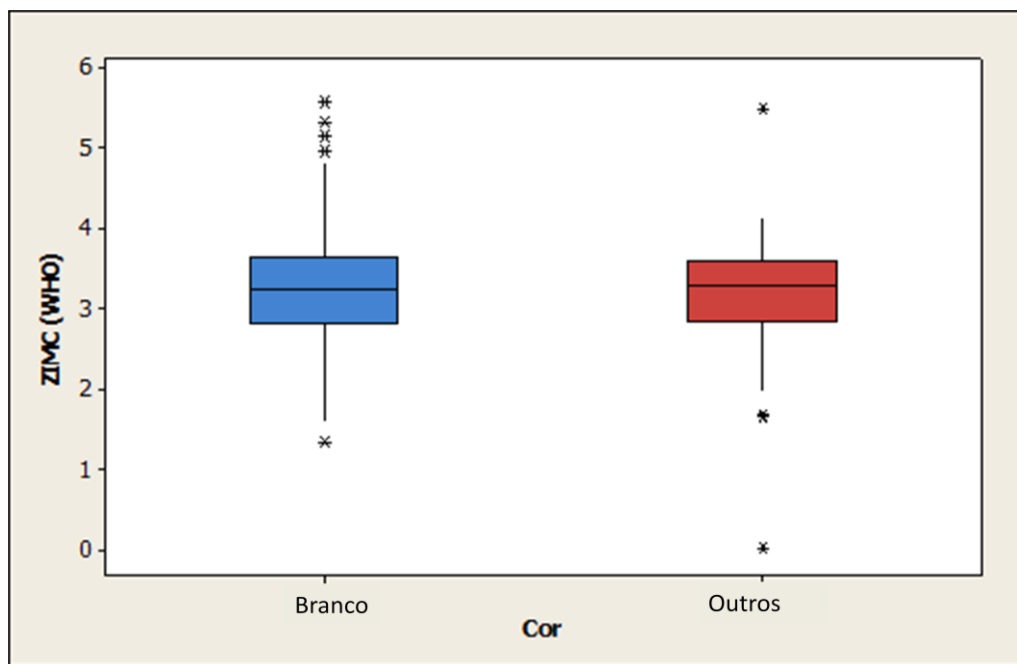


Gráfico B.17: Histograma para o os valores de ZIMC pelo método C&M.

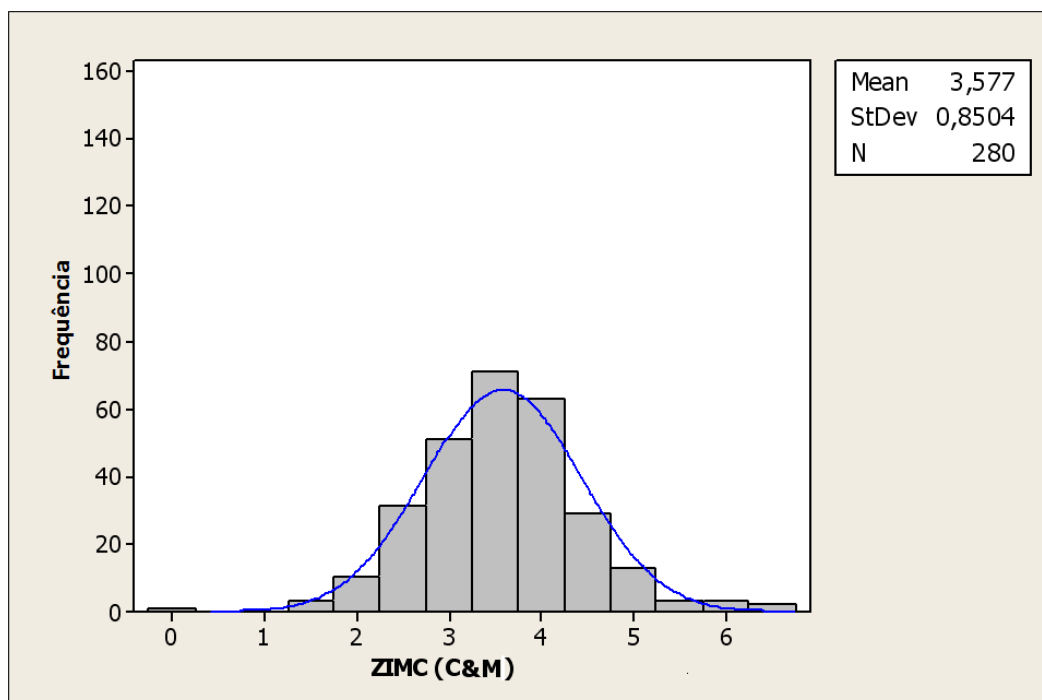


Gráfico B.18: *Box-plot* para o IMC padronizado pelo método C&M.

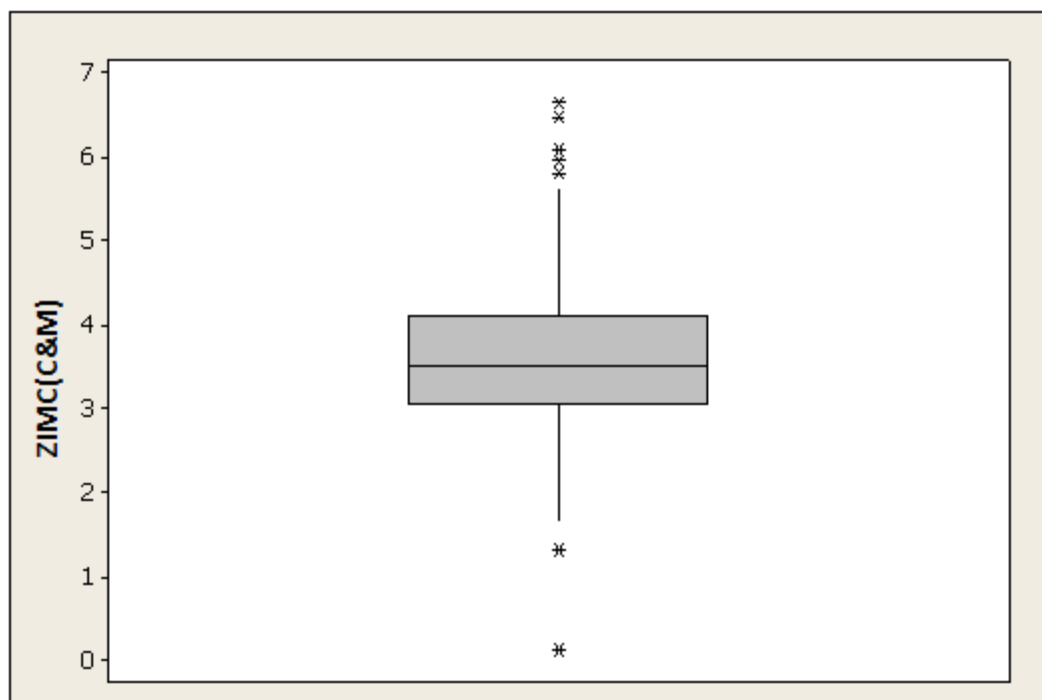


Gráfico B.19: *Box-plot* para o IMC padronizado pelo método C&M segundo cor.

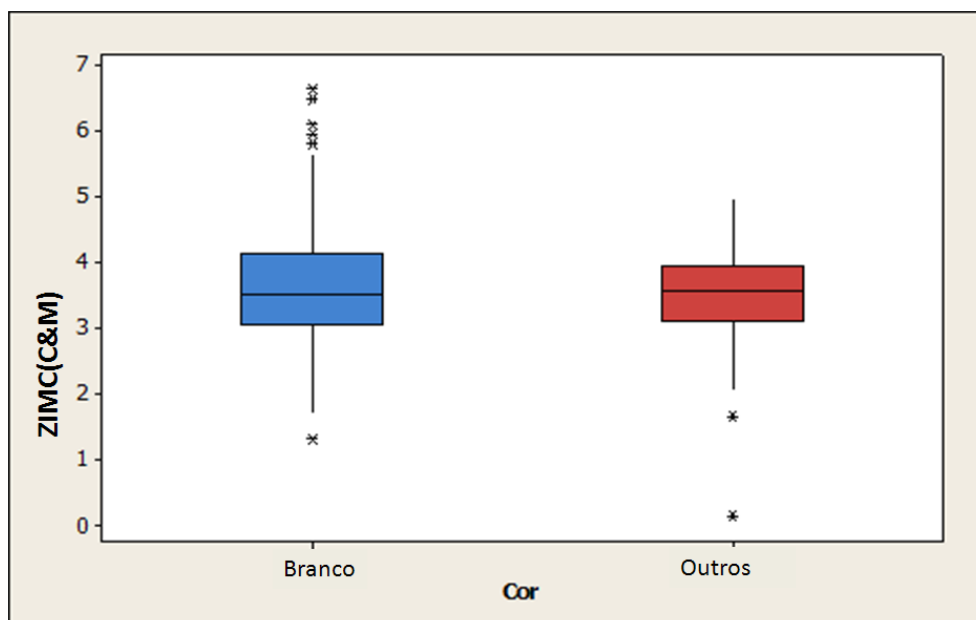


Gráfico B.20: *Box-plot* para a circunferência abdominal calculada no ponto médio.

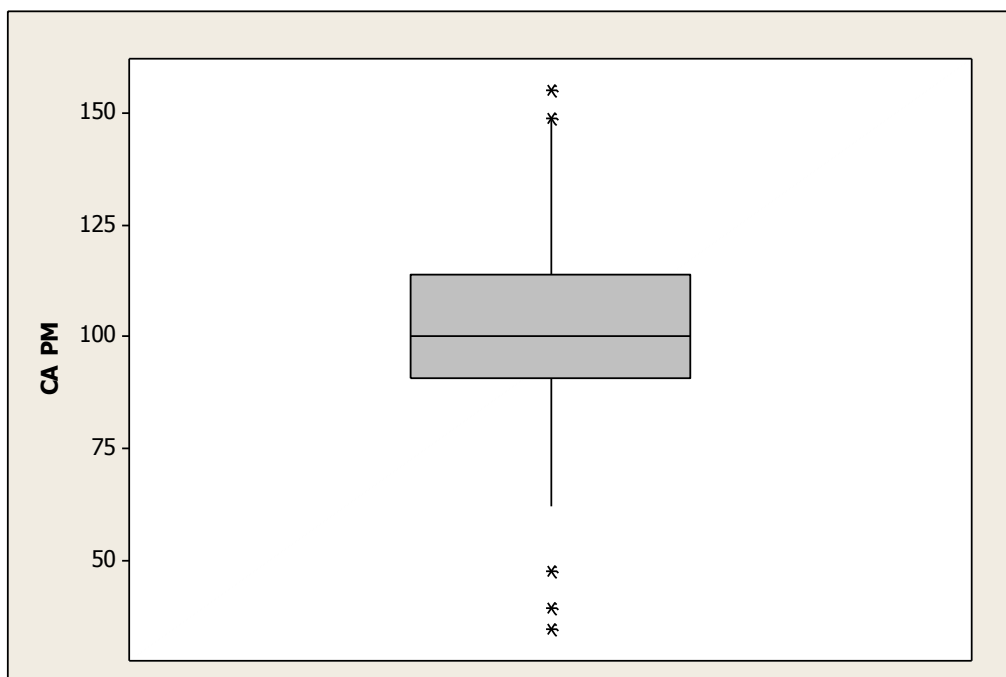


Gráfico B.21: *Box-plot* para a frequência cardíaca.

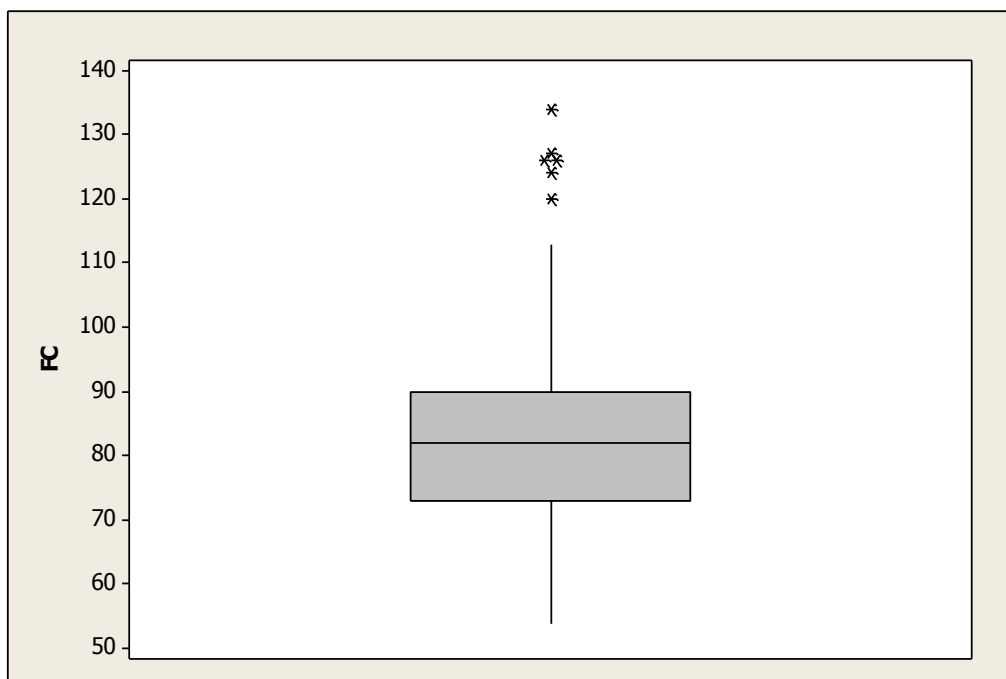


Gráfico B.22: *Box-plot* para a porcentagem de massa de gordura.

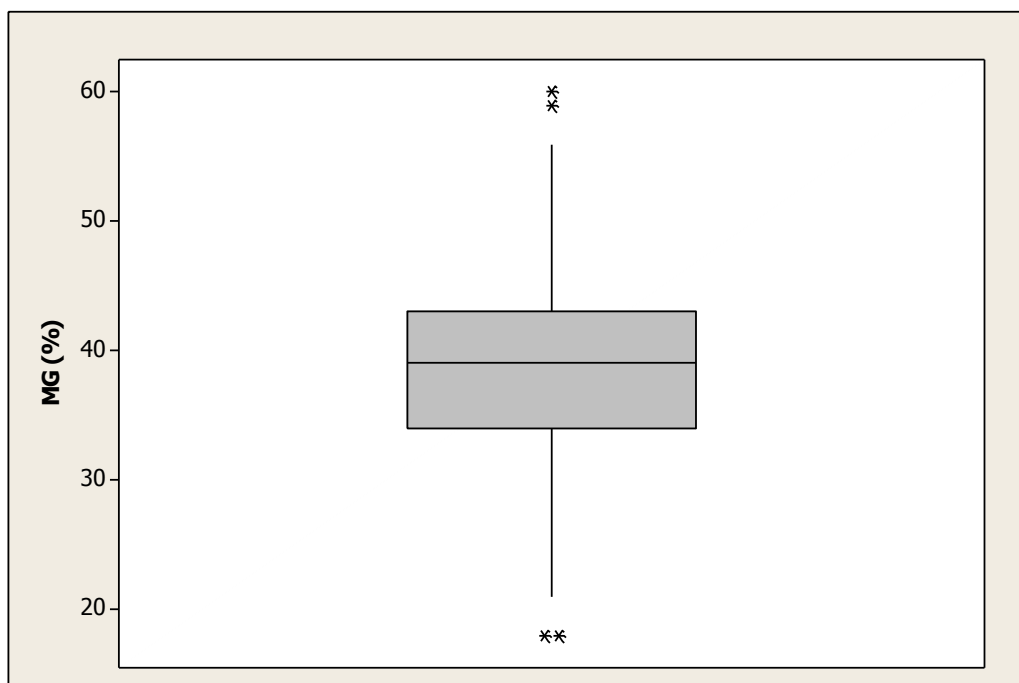


Gráfico B.23: *Box-plot* para a porcentagem de massa de gordura segundo cor.

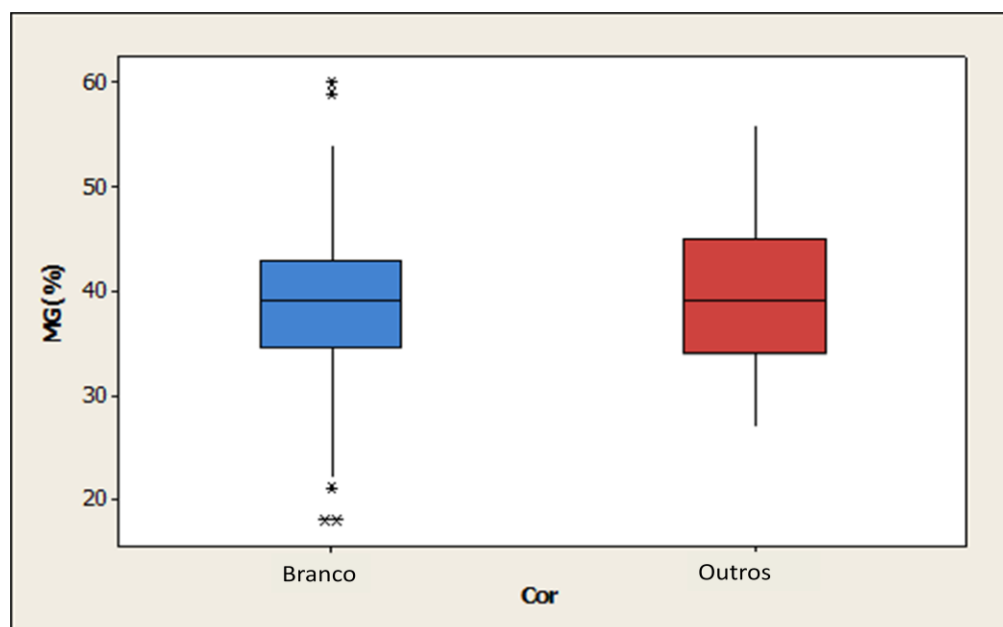


Gráfico B.24: *Box-plot* para o colesterol total.

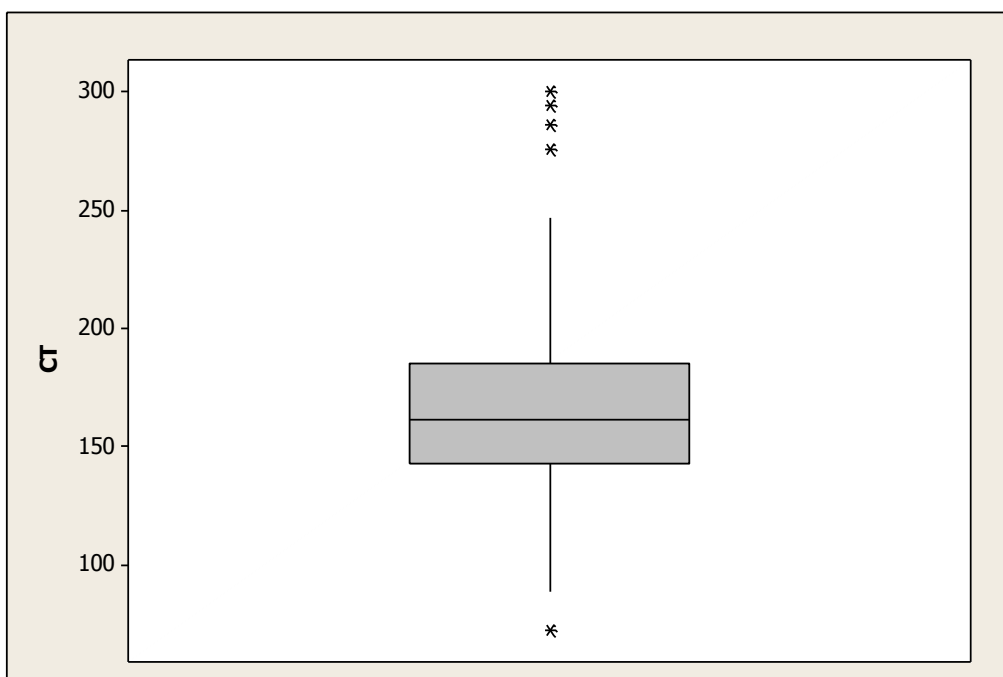


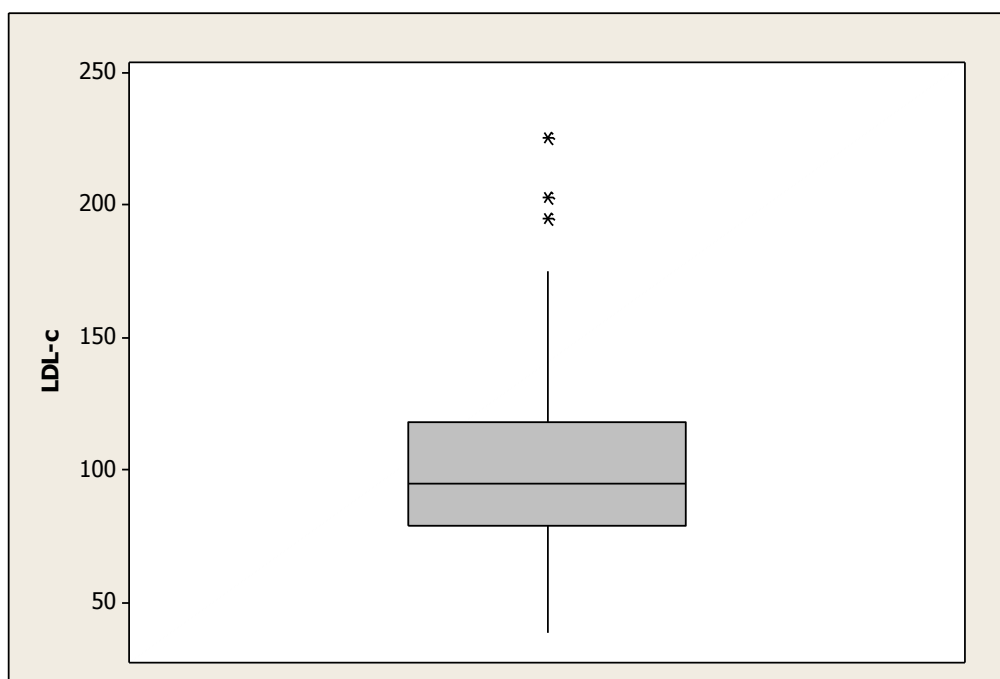
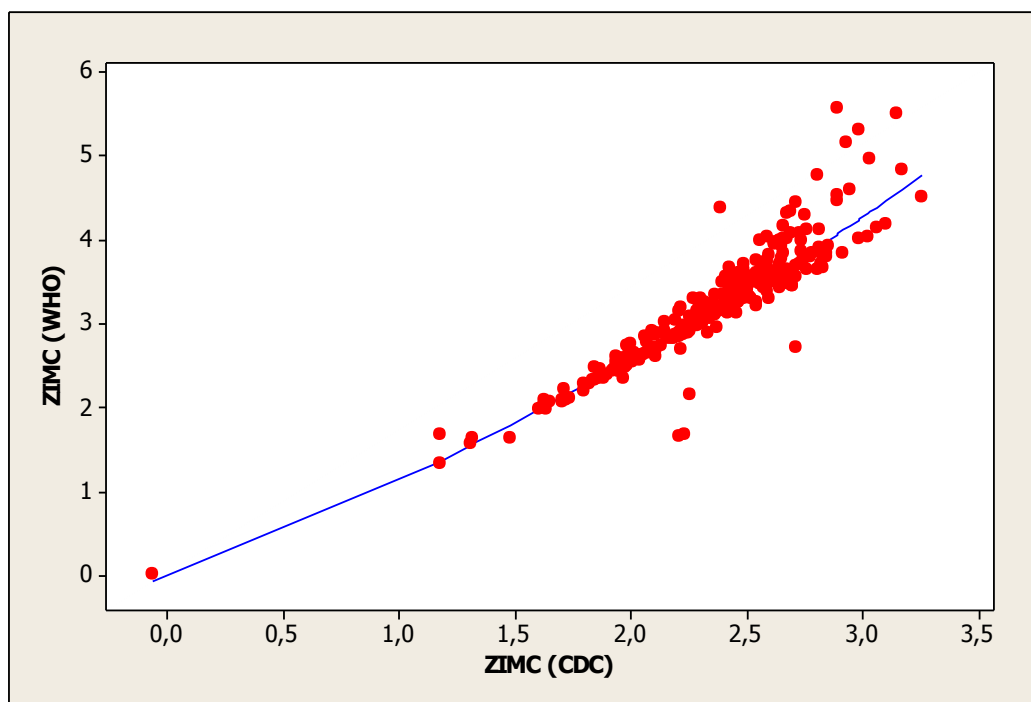
Gráfico B.25: *Box-plot* para o LDL-c.**Gráfico B.26:** Gráfico de dispersão para os valores de ZIMC pelo método de WHO segundo CDC.

Gráfico B.27: Gráfico de dispersão para os valores de ZIMC pelo método C&M segundo CDC.

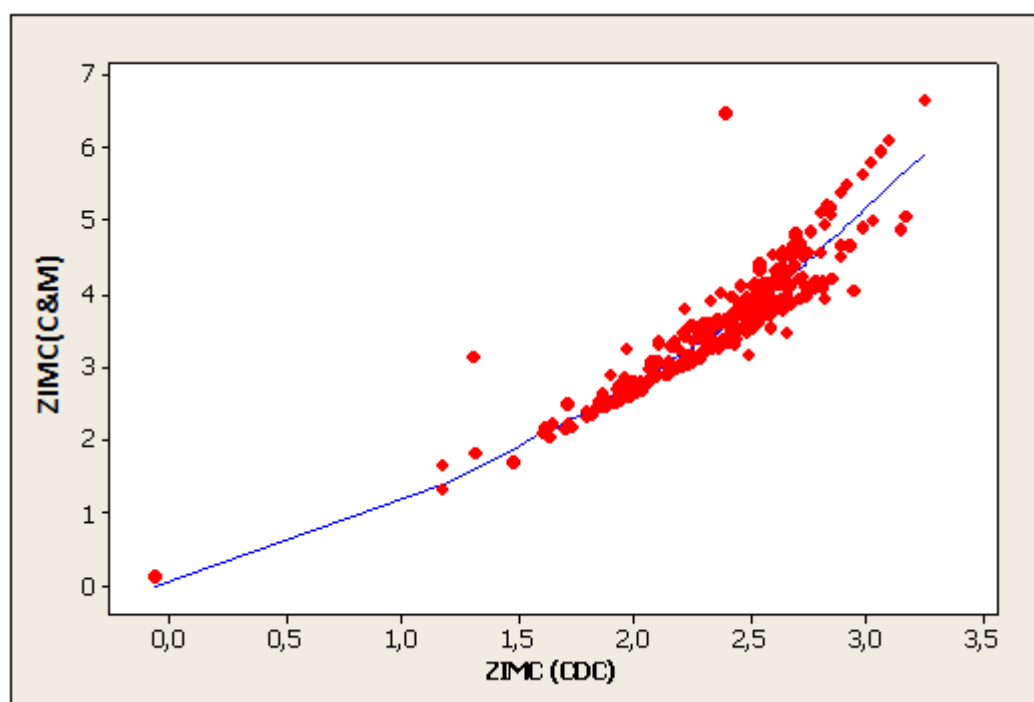


Gráfico B.28: Gráfico de dispersão para os valores de ZIMC pelo método C&M segundo WHO.

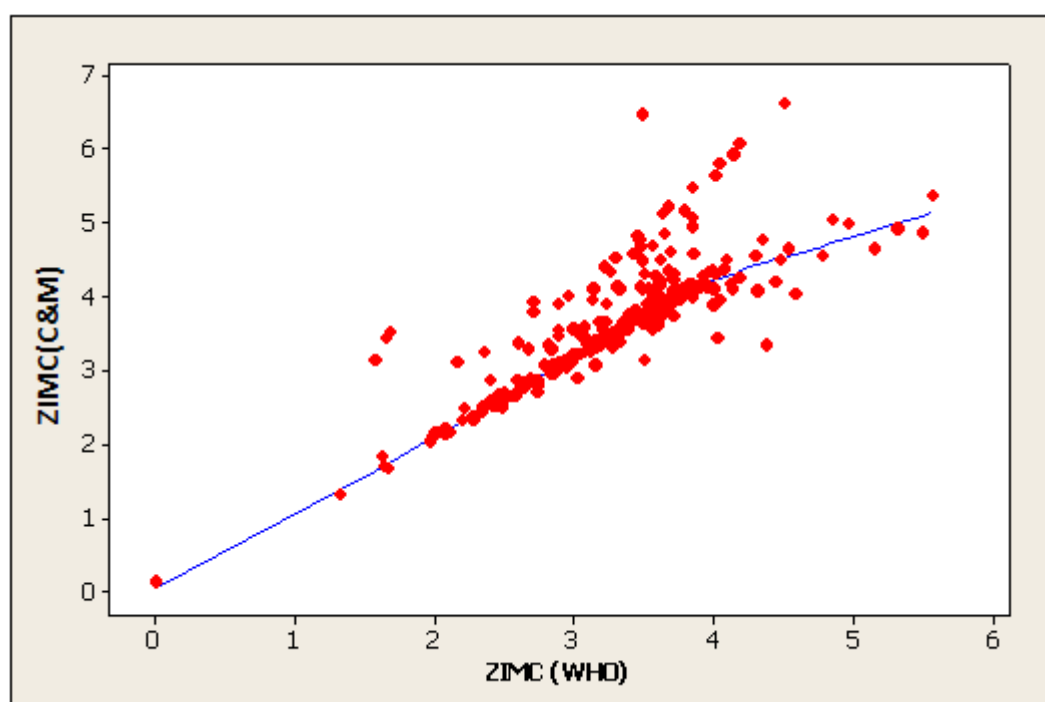


Gráfico B.29: Distribuição de pacientes segundo ocorrência de hipertensão arterial.

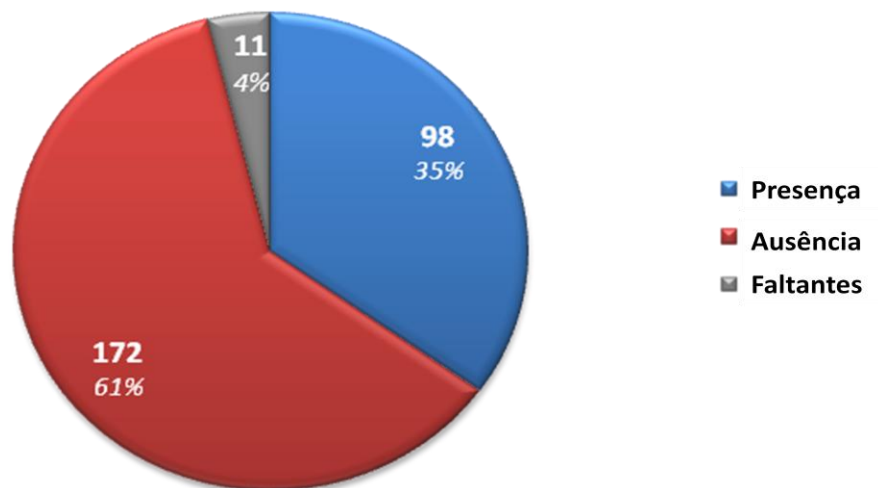


Gráfico B.30: Distribuição de pacientes segundo ocorrência de glicemia alterada.

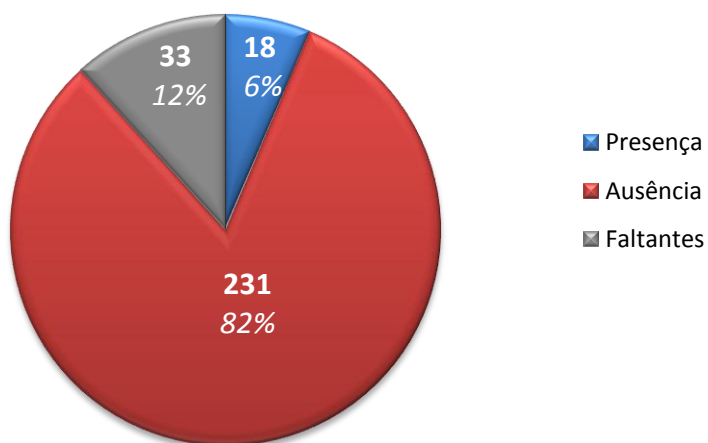


Gráfico B.31: Distribuição de pacientes segundo ocorrência de dislipidemia.

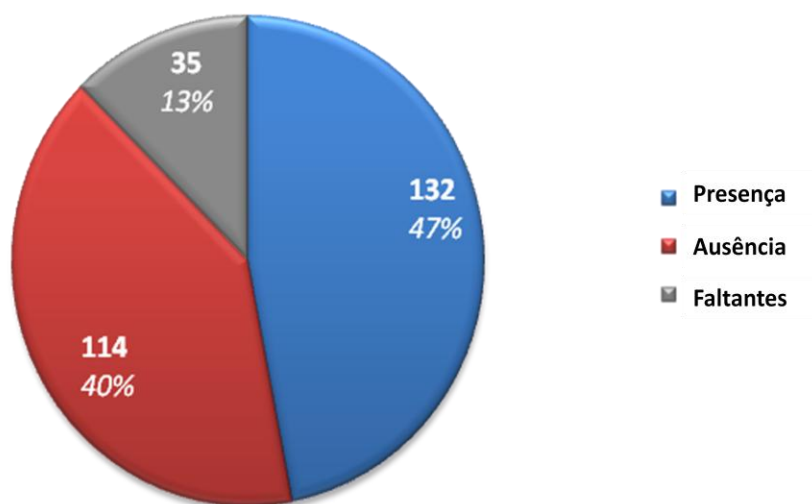


Gráfico B.32: Distribuição de pacientes segundo ocorrência de resistência à insulina.

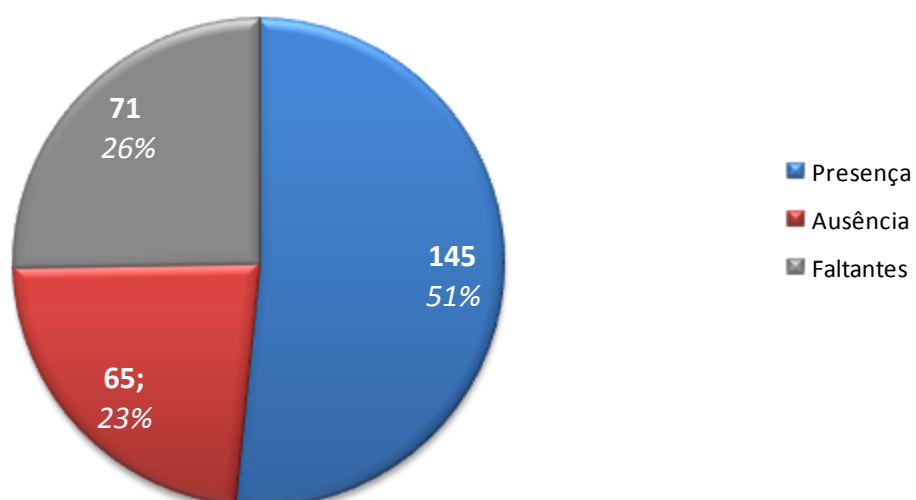


Gráfico B.33: Distribuição dos pacientes segundo número de doenças diagnosticadas

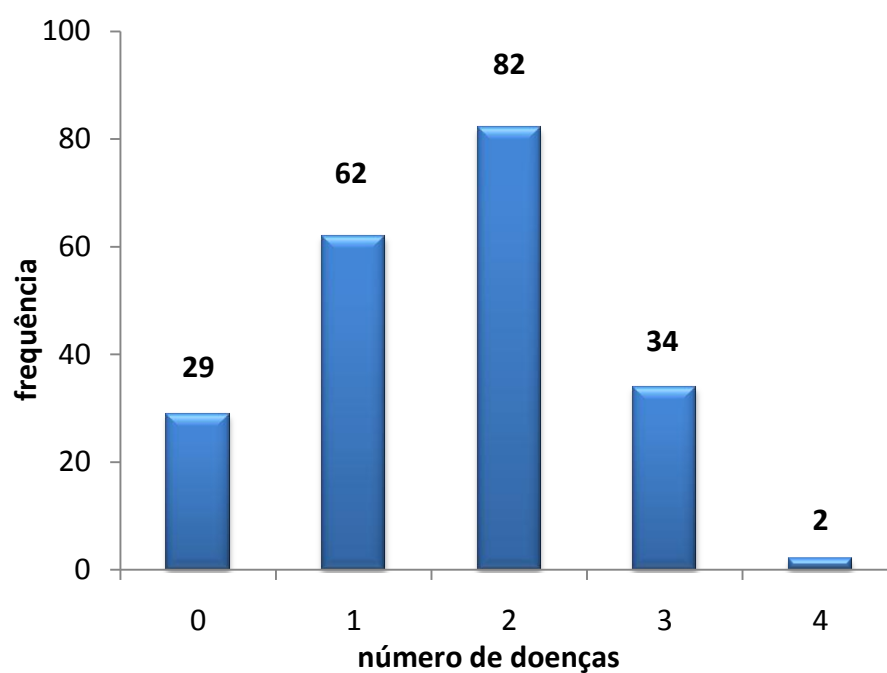


Gráfico B.34: Box-plot para ZIMC pelo método de CDC segundo número de doenças.

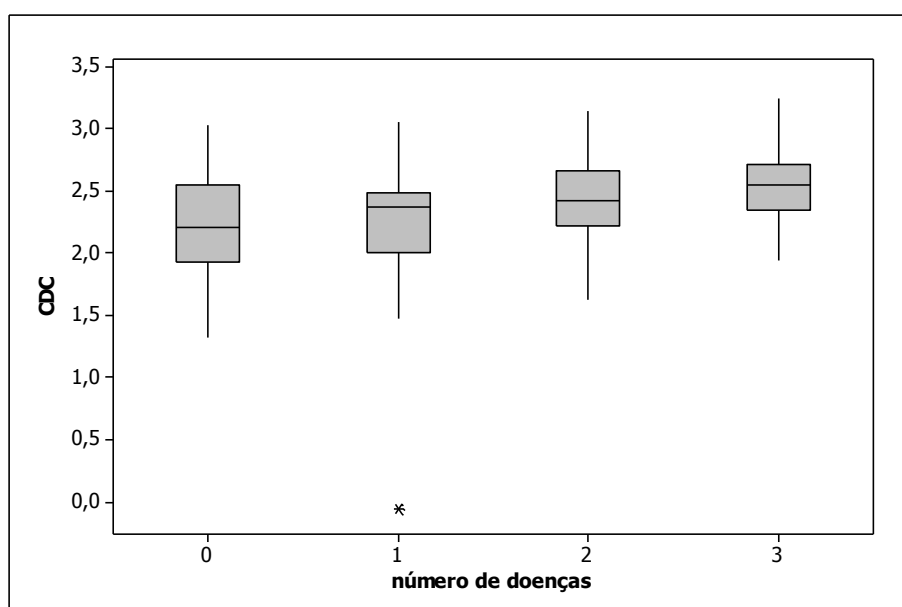


Gráfico B.35: *Box-plot* para ZIMC pelo método de WHO segundo número das doenças.

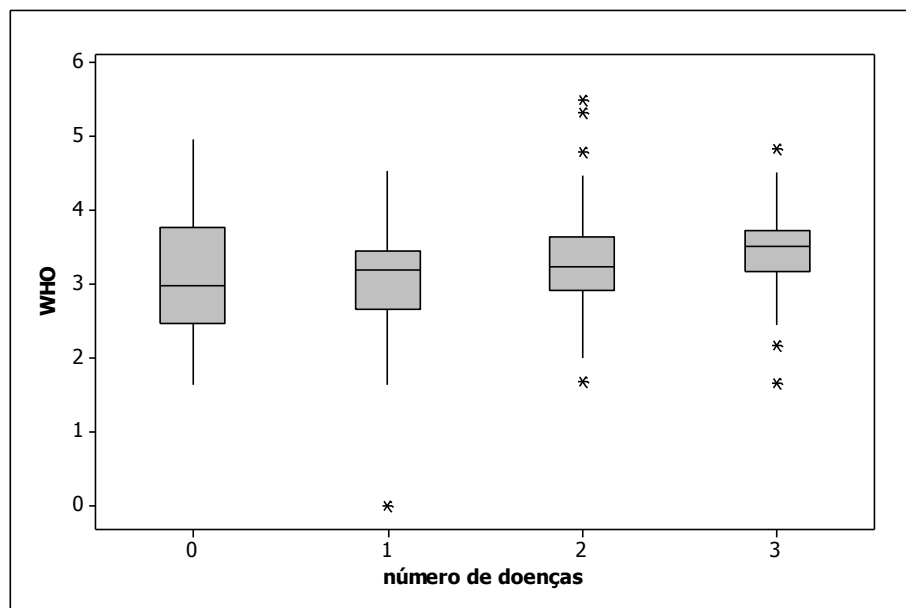


Gráfico B.36: *Box-plot* para ZIMC pelo método C&M segundo número das doenças.

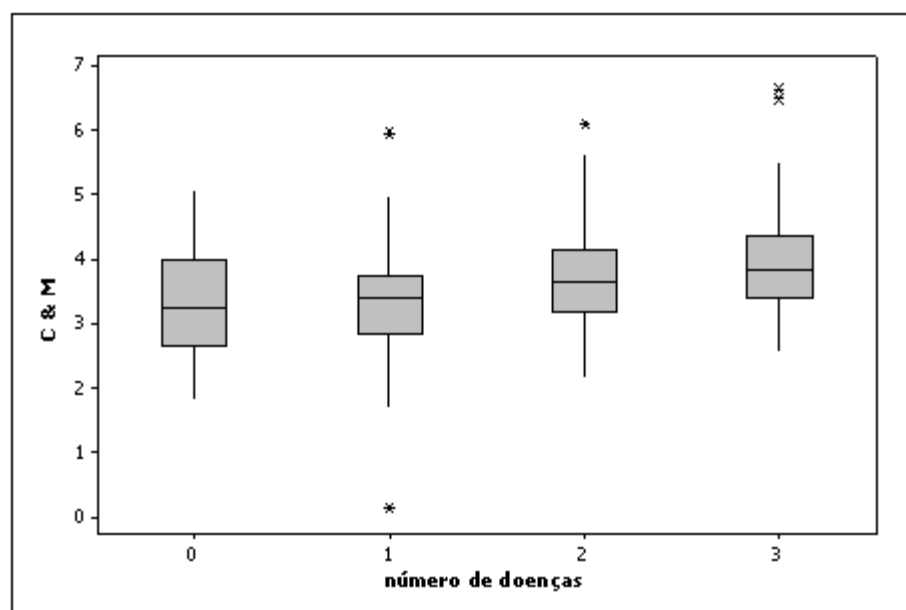


Gráfico B.37: Gráfico da probabilidade estimada segundo o número de doenças e percentual de massa de gordura para o método CDC.

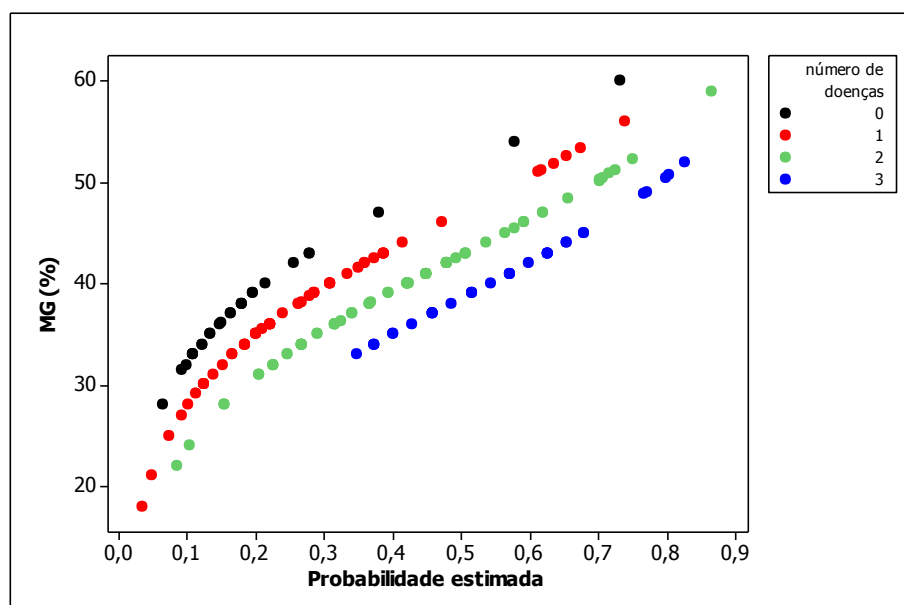


Gráfico B.38: Gráfico da probabilidade estimada segundo o número de doenças e percentual de massa de gordura para o método WHO.

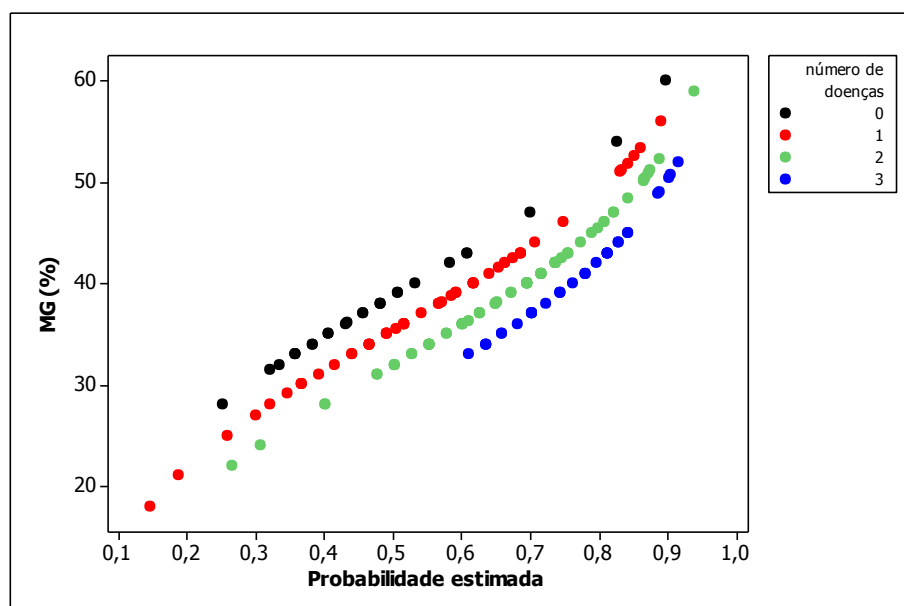


Gráfico B.39: Gráfico da probabilidade estimada segundo o número de doenças e percentual de massa de gordura para o método C&M.

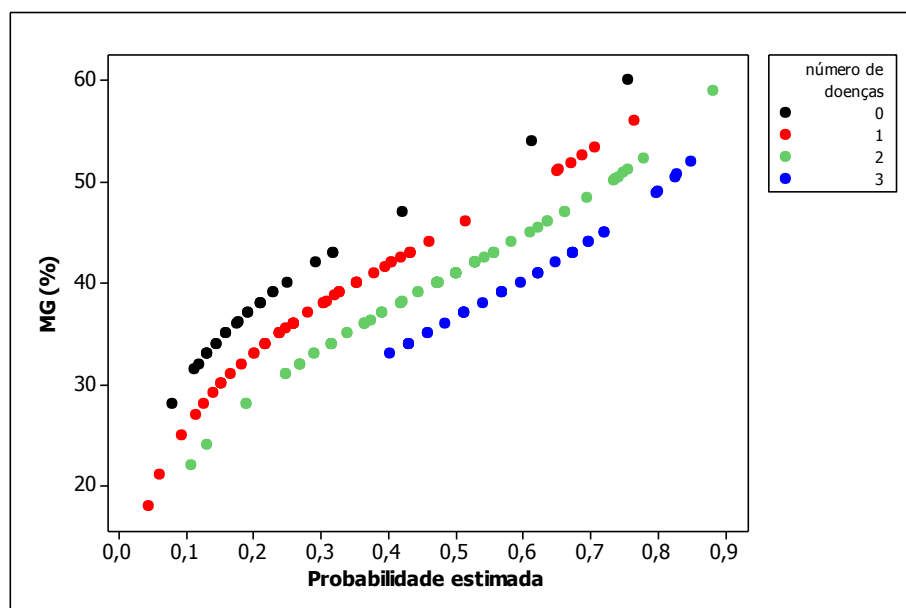


Gráfico B.40: Gráfico normal de probabilidades com banda empírica de 95% de confiança para os resíduos do modelo logístico proposto para o método CDC

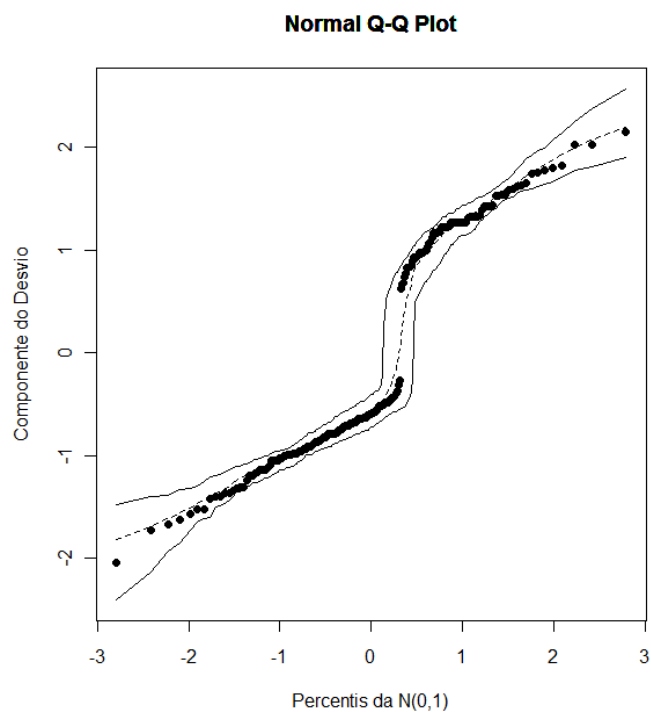


Gráfico B.41: Gráfico dos pontos influentes do modelo logístico proposto para o método CDC

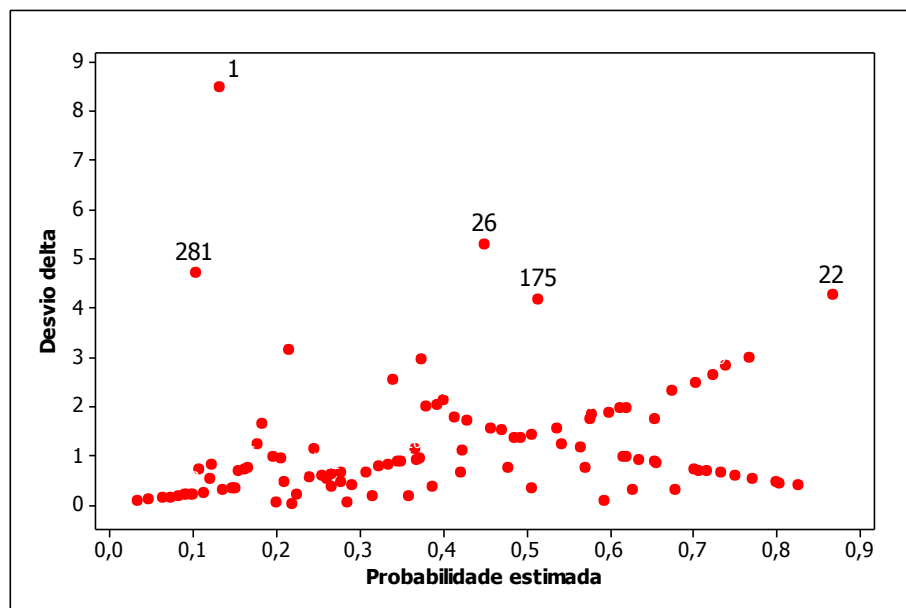


Gráfico B.42: Gráfico normal de probabilidades com banda empírica de 95% de confiança para os resíduos do modelo logístico proposto para o método WHO

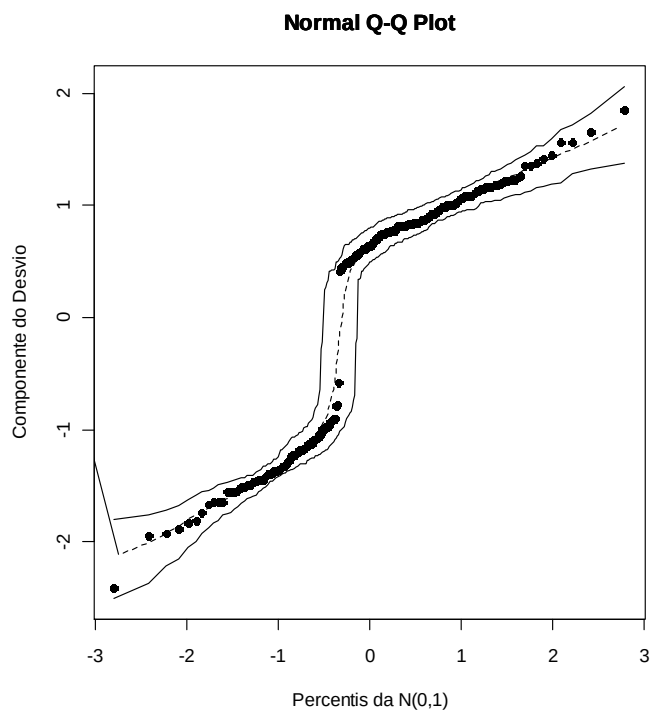


Gráfico B.43: Gráfico dos pontos influentes do modelo logístico proposto para o método WHO

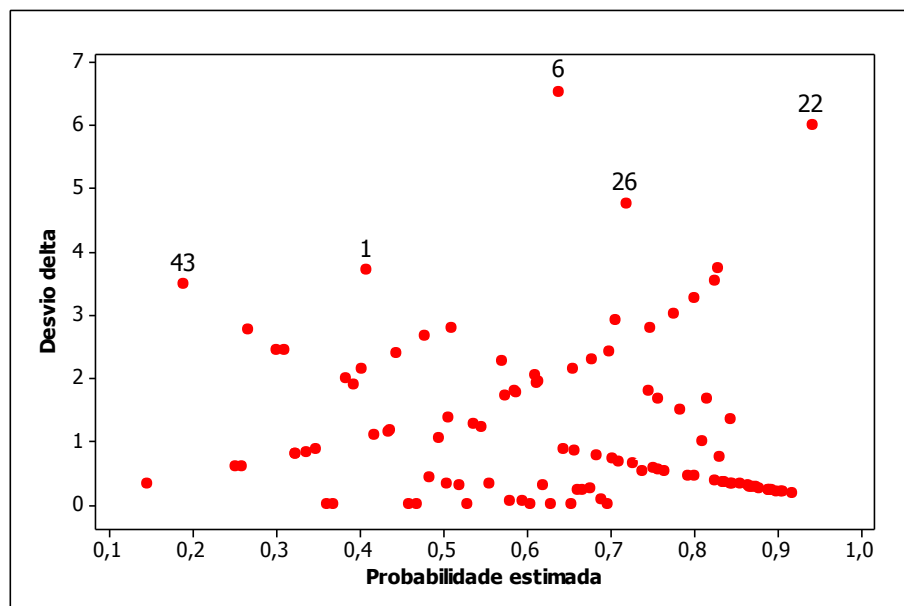


Gráfico B.44: Gráfico normal de probabilidades com banda empírica de 95% de confiança para os resíduos do modelo logístico proposto para o método C&M

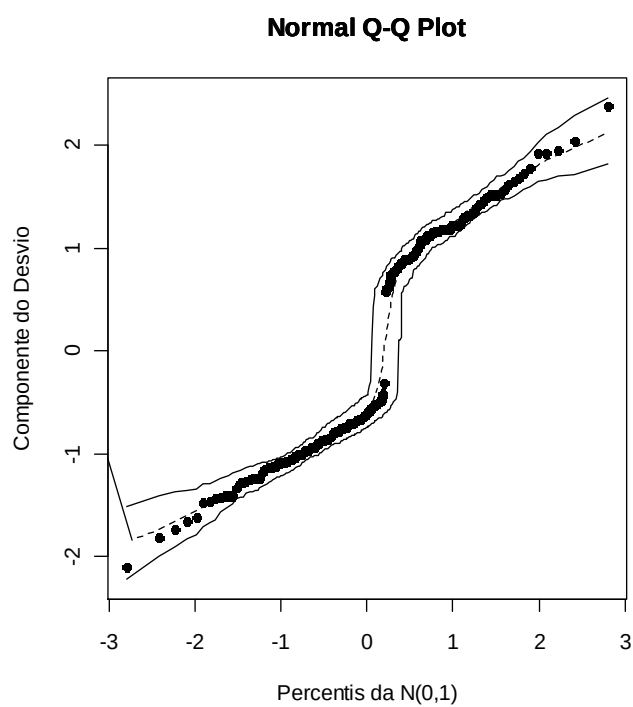
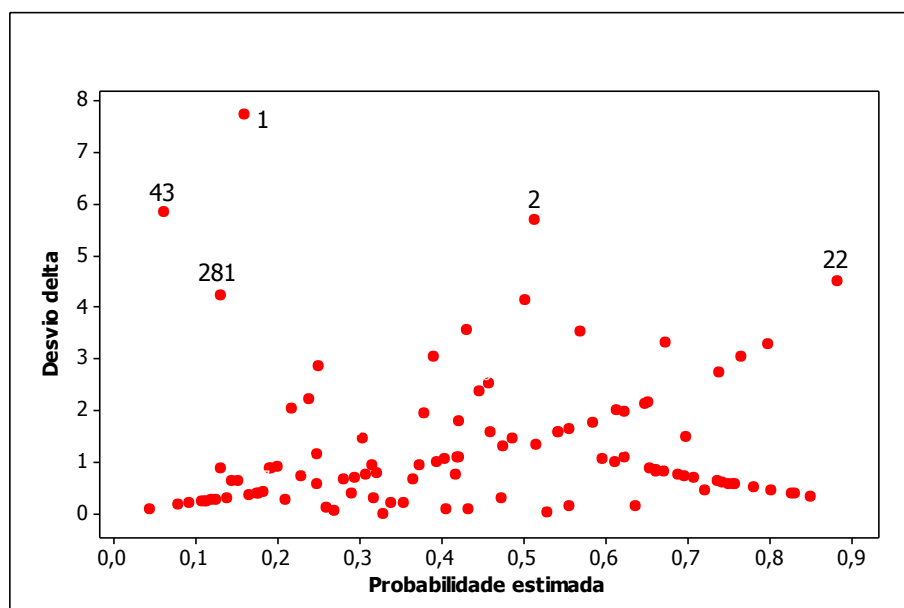


Gráfico B.44: Gráfico dos pontos influentes do modelo logístico proposto para o método C&M



APÊNDICE C
EXPRESSÕES

Expressão C.1: Cálculo do Z pelo método LMS:

O método LMS assume que, para dados independentes com valores positivos, a transformação Box-Cox idade-específica pode ser empregada para torná-los normalmente distribuídos; A amostra foi separada por gênero e para cada gênero foi separada em faixas etárias, cada uma de 3 meses, denominados estratos. Em cada estrato, foram calculados os parâmetros L, M e S. O parâmetro M expressa mediana do IMC (kg/m²) em cada estrato; S representa o coeficiente de variação de cada estrato e o parâmetro L, o coeficiente (Box-Cox) empregado para a transformação matemática dos valores do IMC com o objetivo de obter distribuição normal em cada estrato. O valor selecionado para o coeficiente L é aquele cuja transformação produza a menor soma dos quadrados dos desvios da variável.

$$Z = \frac{\left[\left(\frac{IMC}{M}\right)^L - 1\right]}{(LS)} \text{ para } L > 0$$

$$Z = \frac{1}{S} \ln\left(\frac{IMC}{M}\right) \text{ para } L = 0$$

Expressão C.2: Equação de regressão:

A regressão foi construída para se obter um valor de corte para classificação pelo método de Conde & Monteiro (C&M), já que este valor não existe na literatura.

$$ZIMC(C\&M) = 1,22 + 0,996 ZIMC(CDC)$$

Pela análise de resíduos, todas as suposições foram satisfeitas. Como o valor de corte do método CDC é 2,5, valor fornecido pela pesquisadora, teremos a seguinte equação:

$$ZIMC(C\&M) = 1,22 + 0,996 * 2,5 = 3,71$$

Portanto 3,71 será a nota de corte estabelecida para o método de C&M.

Expressão C.3: Cálculo do índice Kappa

$$\hat{k} = \frac{\hat{\theta}_1 - \hat{\theta}_2}{1 - \hat{\theta}_2}$$

$$\text{onde, } \hat{\theta}_1 = \frac{\sum x_{ii}}{n}, \hat{\theta}_2 = \frac{\sum x_{i.} + x_{.i}}{n^2}$$

Com x_{ii} , $x_{i.}$, $x_{.i}$, $x_{..}$ e n definidos na tabela abaixo:

Classificação B	Classificação A				
	A	B	...	z	Total
A	x_{11}	x_{12}	...	x_{1i}	$x_{1.}$
B	x_{21}	x_{2i}	...	x_{2i}	$x_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Z	x_{i1}	x_{i2}	...	x_{ii}	$x_{i.}$
Total	$x_{.1}$	$x_{.2}$	...	$x_{.i}$	n

Expressão C.4: Cálculo da variância do índice Kappa

$$\widehat{Var(k)} = \frac{1}{n} \left[\frac{\hat{\theta}_1(1 - \hat{\theta}_1)}{(1 - \hat{\theta}_2)^2} + \frac{2(1 - \hat{\theta}_1)(2\hat{\theta}_1\hat{\theta}_2 - \hat{\theta}_3)}{(1 - \hat{\theta}_2)^3} + \frac{(1 - \hat{\theta}_1)^2(\hat{\theta}_4 - 4\hat{\theta}_2^2)}{(1 - \hat{\theta}_2)^4} \right]$$

$$\text{onde, } \hat{\theta}_3 = \frac{\sum x_{ii}(x_{i.} + x_{.i})}{n^2}, \hat{\theta}_4 = \frac{\sum_i \sum_j x_{ij}(x_{j.} + x_{.i})}{n^3}$$

Expressão C.5: Cálculo do intervalo de confiança de 95% para o índice Kappa

$$IC(95\%) = [\hat{k} - 1,96 * \sqrt{\widehat{Var(k)}}; \hat{k} + 1,96 * \sqrt{\widehat{Var(k)}}]$$

APÊNDICE D
METODOLOGIA

Metodologia D.1: Regressão Logística.

A Regressão Logística é uma técnica estatística utilizada quando a variável resposta de interesse é uma variável binária (sucesso ou fracasso). Assim, tenta-se modelar a probabilidade de no evento de interesse ocorrer sucesso, dada uma série de variáveis explicativas.

Metodologia D.2: Modelo logístico proposto:

y_i : Diagnóstico segundo o método ($y_i = 0$: obeso; $y_i = 1$: super obeso) com a seguinte Suposição: $\sum y_i \sim B(n, p_i)$, em que:

n : total de pacientes (192) e

p_i : probabilidade do i-ésimo paciente ser diagnosticado como super obeso pelo método

$$\log \left\{ \frac{p_i}{1-p_i} \right\} = \alpha + \beta_1 * \text{número de doenças}_i + \beta_2 * MG(\%)_i,$$

com $i = 1, 2, \dots, 192$.

APÊNDICE E
SUPOSIÇÕES VERIFICADAS

Ajuste do modelo

Para verificarmos se os modelos estão bem ajustados, foram feitas algumas análises para validar as suposições impostas inicialmente.

Método CDC

Pelo Gráfico B.40, modelo proposto para o método CDC, pode-se verificar que todos os pontos estão dentro da banda ajustada, indicando que a suposição da distribuição binomial ($\sum y_i \sim B(n, p_i)$) é satisfatória.

Através do Gráfico B.41, podemos detectar quais pontos (pacientes) poderiam influenciar a análise, destoando dos resultados obtidos. Pode-se verificar que há 5 pontos aparentemente mais influentes na amostra. Por exemplo: (i) o paciente 1 não possui doenças tem 35 de percentual de massa de gordura e foi considerado como super obeso; (ii) o paciente 22 possui 2 doenças, tem 59 de percentual de massa de gordura e foi classificado como obeso; (iii) o paciente 26 possui 2 doenças, 41 de percentual de massa de gordura e foi classificado como super obeso; (iv) o paciente 175 possui 3 doenças, 39 de percentual de massa de gordura e foi classificado como super obeso (v) o paciente 281 tem 2 doenças, possui 24 % de massa de gordura e foi diagnosticado como super obeso. Podemos observar que os pacientes 1 e 281, seriam diagnosticados como obesos pelo modelo por apresentarem um baixo número de doenças e percentual de gordura abaixo da média, mas o método CDC os classificou como super obesos. O paciente 22 foi classificado como obeso pelo método, mas o modelo o classificaria como super obeso, pois este apresenta duas doenças e alto percentual de massa de gordura. Estes pacientes apareceram como destoantes, pois não possuem o perfil mais esperado daqueles que foram diagnosticados como obesos (menor número de doenças e percentual de massa de gordura), ou super obesos (maior número de doenças ou percentual de massa de gordura).

Para verificar se estes 5 indivíduos poderiam influenciar no resultado do modelo proposto, retiramos estes pontos da amostra e ajustamos um novo modelo logístico. As estimativas dos parâmetros, que podem ser vistas na Tabela A.14, continuaram marginalmente significantes ao nível de significância

de 5% e não apresentaram muitas alterações, permanecendo com as mesmas interpretações, ou seja, os pontos influentes não afetam o desempenho do modelo.

Retirando os 5 pontos influentes a nova razão de chances para o número de doenças será de 1,69 com IC(95%) igual a [1,15; 2,45].

Para o percentual de massa de gordura a nova razão de chances será igual a 1,16 com IC(85%) de [1,09;1,23].

Método WHO

Pelo Gráfico B.42, modelo proposto para o método WHO, pode-se verificar que todos os pontos estão dentro da banda ajustada, indicando que a suposição da distribuição binomial ($\sum y_i \sim B(n, p_i)$) é satisfatória.

Através do Gráfico B.43, podemos detectar quais pontos (pacientes) poderiam influenciar a análise, destoando dos resultados obtidos. Pode-se verificar que há 5 pontos aparentemente mais influentes na amostra, sendo que, os pacientes 1, 22 e 26 já foram detectados e caracterizados como pontos influentes pelo método CDC. Os dois pacientes que ainda aparecem como discrepantes são: (i) o paciente 6 possui 3 doenças, tem 34 de percentual de massa de gordura e foi considerado como obeso e; (ii) o paciente 43 possui 1 doença, tem 21 de percentual de massa de gordura e foi classificado como super obeso. Podemos observar que os pacientes 1 e 43, seriam diagnosticados como obesos pelo modelo por apresentarem baixo número de doenças e percentual de gordura abaixo da média, mas o método WHO os diagnosticou como super obesos. Os pacientes 6 e 22 foram classificados como obeso pelo método, mas o modelo os classificaria como super obesos, pois este apresentam acima de duas doenças e alto percentual de massa de gordura. Estes pacientes apareceram como destoantes, pois não possuem o perfil mais esperado daqueles que foram diagnosticados como obesos (menor número de doenças e/ou percentual de massa de gordura), ou super obesos (maior número de doenças e/ou percentual de massa de gordura).

Para verificar se estes 5 indivíduos poderiam influenciar no resultado do modelo proposto, retiramos estes pontos da amostra e ajustamos um novo

modelo logístico. As estimativas dos parâmetros, que podem ser vistas na Tabela A.15 continuaram marginalmente significantes ao nível de significância de 5% e não apresentaram muitas alterações, permanecendo com as mesmas interpretações, ou seja, os pontos influentes não afetam o desempenho do modelo.

Retirando os 5 pontos influentes a nova razão de chances para o número de doenças será de 1,51 com IC(95%) igual a [1,05; 2,17].

Para o percentual de massa de gordura a nova razão de chances será igual a 1,14 com IC(85%) de [1,07;1,21].

Método Conde & Monteiro

Pelo Gráfico B.44, modelo proposto para o método C&M, pode-se verificar que todos os pontos estão dentro da banda ajustada, indicando que a suposição da distribuição binomial ($\sum y_i \sim B(n, p_i)$) é satisfatória.

Através do Gráfico B.44, podemos detectar quais pontos (pacientes) poderiam influenciar a análise destoando dos resultados obtidos. Pode-se verificar que há 5 pontos aparentemente mais influentes na amostra, sendo que, os pacientes 1, 22, 43 e 281 já foram detectados e caracterizados como pontos influentes pelos métodos CDC e WHO. O outro ponto que aparece como discrepante é o paciente 2 que possui 3 doenças, tem 37 de percentual de massa de gordura e foi considerado como super obeso. Podemos observar que os pacientes 1 e 43 e 281, seriam diagnosticados como obesos pelo modelo por apresentarem baixo número de doenças e percentual de gordura abaixo da média, mas o método C&M os diagnosticou como super obesos. O paciente 22 foi classificado como obeso pelo método, mas o modelo o classificaria como super obeso, pois este apresenta duas doenças e alto percentual de massa de gordura. Estes pacientes apareceram como destoantes, pois não possuem o perfil mais esperado daqueles que foram diagnosticados como obesos (menor número de doenças e/ou percentual de massa de gordura), ou super obesos (maior número de doenças e/ou percentual de massa de gordura).

Para verificar se estes 5 indivíduos poderiam influenciar no resultado do modelo proposto, retiramos estes pontos da amostra e ajustamos um novo modelo logístico. As estimativas dos parâmetros, que podem ser vistas na Tabela A.16, continuaram marginalmente significantes ao nível de significância de 5% e não apresentaram muitas alterações, permanecendo com as mesmas interpretações, ou seja, os pontos influentes não afetam o desempenho do modelo.

Retirando os 5 pontos influentes a nova razão de chances para o número de doenças será de 1,73 com IC(95%) igual a [1,18; 2,52].

Para o percentual de massa de gordura a nova razão de chances será igual a 1,17 com IC(85%) de [1,10;1,25].

APÊNDICE F
DISCRIMINAÇÃO DO MODELO

6.3 Discriminação pelo modelo

Para verificarmos a discriminação do modelo, a probabilidade estimada do modelo foi calculada da seguinte forma:

$$\hat{p} = \frac{e^{(\alpha + \beta_1 * \text{número de doenças}_i + \beta_2 * MG(\%)_i)}}{1 + e^{(\alpha + \beta_1 * \text{número de doenças}_i + \beta_2 * MG(\%)_i)}}$$

com $i = 1, 2, \dots, 192$.

em que $\hat{p}$ é a probabilidade estimada de um paciente ser classificado como super obeso dado o número de doenças e o percentual de massa de gordura.

6.3.1 Método CDC

Foi utilizado um critério de classificação para testar a qualidade de acerto do modelo. Dado que o percentual de super obeso na amostra foi de 38%, este valor foi definido como ponto de corte para a classificação, ou seja, se a probabilidade estimada do paciente for maior que 0,38, diremos que o modelo o classificou como super obeso e se for menor ou igual a 0,38, foi considerado como obeso.

Assim, foi mensurado o número de indivíduos que foram corretamente classificados e os que foram classificados erroneamente, que são os valores que podem ser observados na Tabela A.11. Através desses números, foi possível calcular a especificidade (69,2%), que é o percentual de pacientes classificados pelo modelo como obesos e que de fato foram considerados como obesos pelo método CDC, e a sensibilidade (72,2%), que é o percentual de pacientes classificados como super obesos e que realmente foram considerados super obesos pelo método CDC. Com acerto total de 70,3%, o modelo classificou corretamente um percentual razoável da amostra total.

6.3.2 Método WHO

Foi feito um critério de classificação para testar quão bom está o acerto do modelo. Dado que o percentual de super obeso na amostra foi de 64%, este foi

definido como ponto de corte para a classificação, ou seja, se a probabilidade estimada do paciente for maior que 0,64, diremos que o modelo o classificou como super obeso e se for menor ou igual a 0,64, foi considerado como obeso.

Assim, foi mensurado o número de indivíduos que foram corretamente classificados e os que foram classificados erroneamente, esses valores podem ser observados na Tabela A.12. Através desses números, foi possível calcular a especificidade (66,2%), que é o percentual de pacientes classificados pelo modelo como obesos e que de fato foram considerados como obesos pelo método CDC, e a sensibilidade (63,7%), que é o percentual de pacientes classificados como super obesos e que realmente foram considerados super obesos pelo método CDC. Com acerto total de 64,6%, o modelo classificou corretamente um percentual razoável da amostra total.

6.3.3 Método Conde & Monteiro

Foi feito um critério de classificação para testar quão bom está o acerto do modelo. Dado que o percentual de super obesos na amostra foi de 47%, este foi definido como ponto de corte para a classificação, ou seja, se a probabilidade estimada do paciente for maior que 0,47, diremos que o modelo o classificou como super obeso e se for menor ou igual a 0,47, foi classificado como obeso.

Assim, foi mensurado o número de indivíduos que foram corretamente classificados e os que foram classificados erroneamente, que são os valores que podem ser observados na Tabela A.13. Através desses números, foi possível calcular a especificidade (76,8%), que é o percentual de pacientes classificados pelo modelo como obesos e que de fato foram considerados como obesos pelo método CDC, e a sensibilidade (66,3%), que é o percentual de pacientes classificados como super obesos e que realmente foram considerados super obesos pelo método CDC. Com acerto total de 72,4%, o modelo classificou corretamente um percentual razoável da amostra total.

APÊNDICE A
TABELAS

Tabela A.1: Proporção de pacientes por gênero.

Gênero	Número	%
Masculino	121	57%
Feminino	160	43%

Tabela A.2: Proporção de pacientes para os diferentes tipos de cor.

COR	Número	%
Branco	210	75%
Parda	2	1%
Negro	19	7%
Amarelo	4	1%
Mulato	46	16%

Tabela A.3: Concordância entre o método CDC e WHO.

WHO	CDC		Total
	obeso	super obeso	
obeso	101	1	102
super obeso	74	104	178
total	175	105	280

Tabela A.4: Concordância entre os métodos CDC e C&M.

C&M	CDC		Total
	obeso	super obeso	
obeso	155	7	162
super obeso	20	98	118
total	175	105	280

Tabela A.5: Concordância entre o método WHO e C&M.

C&M	WHO		Total
	obeso	super obeso	
obeso	98	64	162
super obeso	4	114	118
total	102	178	280

Tabela A.6: Soma das doenças para as diferentes classificações segundo os 3 métodos

Número de doenças	CDC		WHO		C&M	
	obesos	super obesos	obesos	super obesos	obesos	super obesos
0	20	9	14	15	19	10
1	47	15	26	36	45	17
2	49	33	31	51	46	36
3	15	19	7	27	14	20
4	2	0	1	1	1	1
Total	133	76	79	130	125	84

Tabela A.7: Porcentagem para soma das doenças para as diferentes classificações segundo os 3 métodos

Número de doenças	CDC		WHO		C&M	
	obesos	super obesos	obesos	super obesos	obesos	super obesos
0	15	12	18	12	15	12
1	35	20	33	28	36	20
2	37	43	39	39	37	43
3	11	25	9	21	11	24
4	2	0	1	1	1	1

Tabela A.8: Resultados do ajuste do modelo logístico proposto para o método CDC com todas as observações.

	Estimativa	Erro Padrão	Valor p
Constante	-5,903	1,142	<0,001
nº de doenças	0,490	0,184	0,008
MG(%)	0,115	0,027	<0,001

Tabela A.9: Resultados do ajuste do modelo logístico proposto para o método WHO com todas as observações.

	Estimativa	Erro Padrão	Valor p
Constante	-3,948	1,056	<0,001
nº de doenças	0,346	0,176	0,049
MG(%)	0,102	0,027	<0,001

Tabela A.10: Resultados do ajuste do modelo logístico proposto para o método C&M com todas as observações.

	Estimativa	Erro Padrão	Valor p
Constante	-5,568	1,111	<0,001
nº de doenças	0,497	0,180	0,006
MG(%)	0,112	0,027	<0,001

Tabela A.11: Discriminação do modelo logístico proposto para o método CDC.

Predito	Observado		
	Obeso	super obeso	Total
Obeso	83	20	103
super obeso	37	52	89
Total	120	72	192

Tabela A.12: Discriminação do modelo logístico proposto para o método WHO.

Predito	Observado		
	obeso	super obeso	Total
obeso	47	44	91
super obeso	24	77	101
total	71	121	192

Tabela A.13: Discriminação do modelo logístico proposto para o método C&M.

Predito	Observado		
	Obeso	super obeso	Total
obeso	86	27	113
super obeso	26	53	79
total	112	80	192

Tabela A.14: Resultados do ajuste do modelo logístico proposto para o método CDC sem as cinco observações discrepantes.

	Estimativa	Erro Padrão	Valor p
Constante	-7,207	1,283	<0,001
nº de doenças	0,516	0,194	0,008
MG(%)	0,145	0,030	<0,001

Tabela A.15: Resultados do ajuste do modelo logístico proposto para o método WHO sem as cinco observações discrepantes.

	Estimativa	Erro Padrão	Valor p
Constante	-5,038	1,174	<0,001
nº de doenças	0,413	0,184	0,025
MG(%)	0,128	0,030	<0,001

Tabela A.16: Resultados do ajuste do modelo logístico proposto para o método C&M sem as cinco observações discrepantes.

	Estimativa	Erro Padrão	Valor p
Constante	-7,587	1,322	<0,001
nº de doenças	0,546	0,194	0,005
MG(%)	0,159	0,032	<0,001

APÊNDICE B

GRÁFICOS

Gráfico B.1: Distribuição dos pacientes segundo gênero.

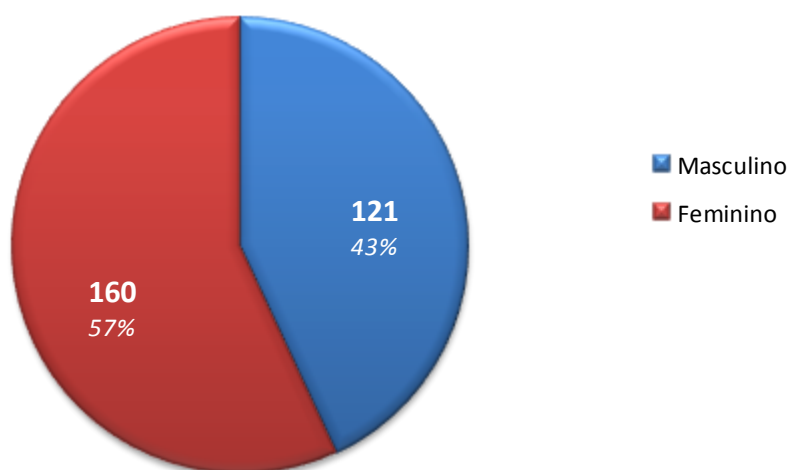


Gráfico B.2: Distribuição dos pacientes segundo grupos étnicos.

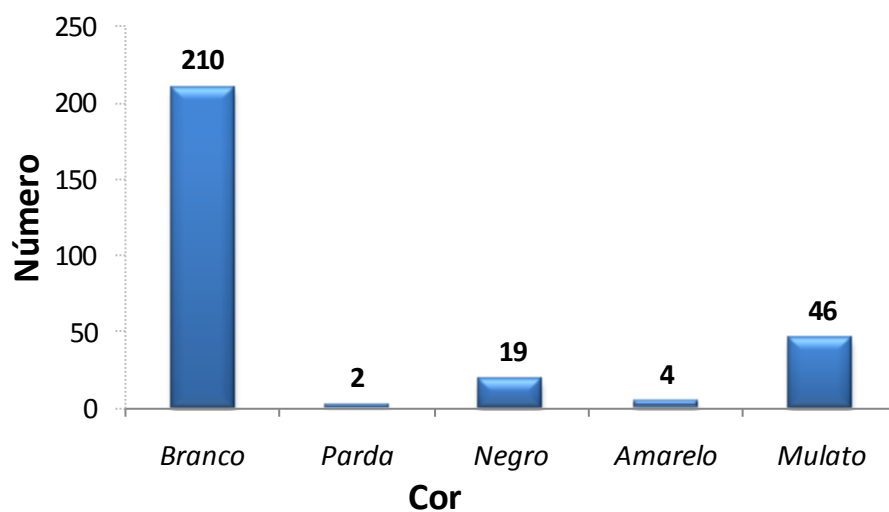


Gráfico B.3: Distribuição dos pacientes segundo grupos étnicos (agrupados)

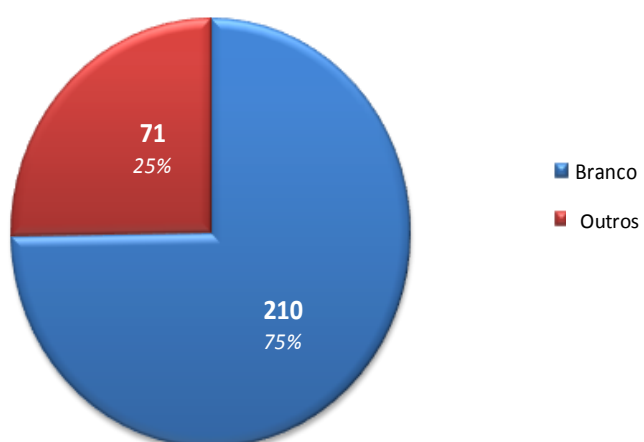


Gráfico B.4: *Box-plot* para a Idade.

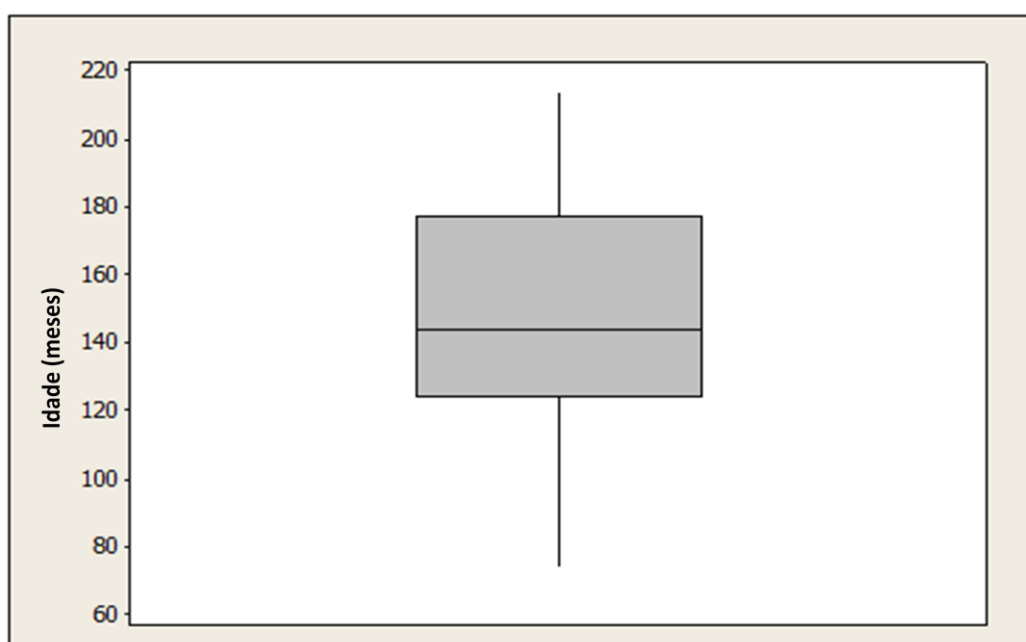


Gráfico B.5: *Box-plot* para a Idade segundo gênero.

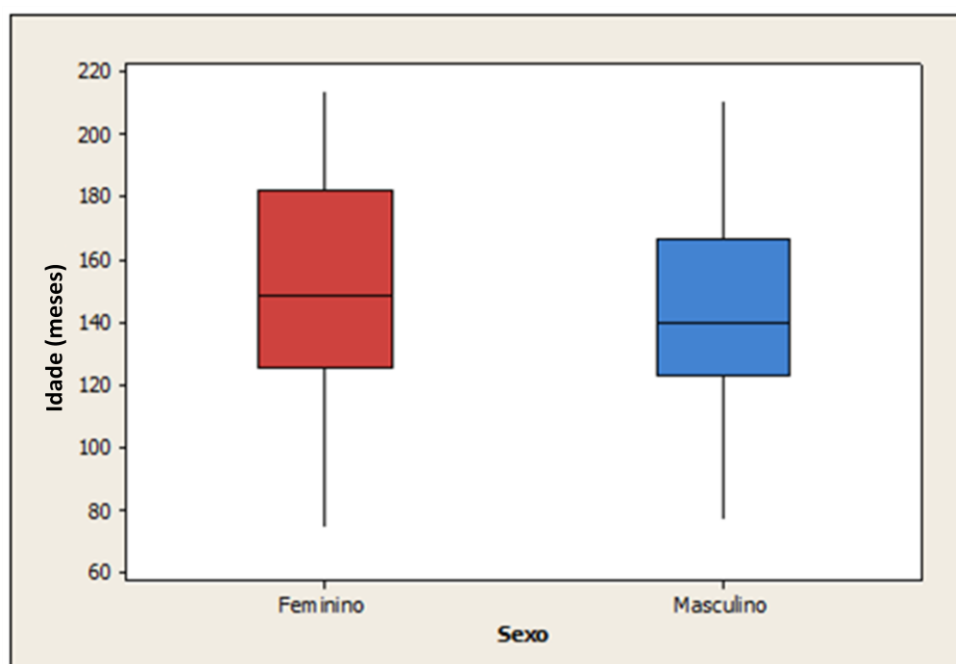


Gráfico B.6: *Box-plot* para a Idade segundo Cor.

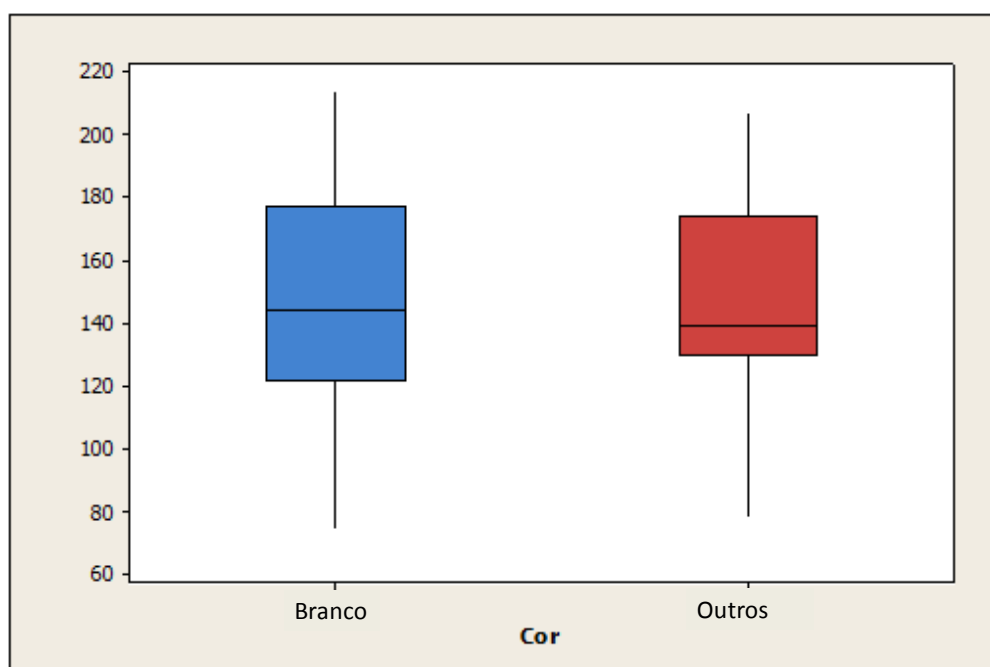


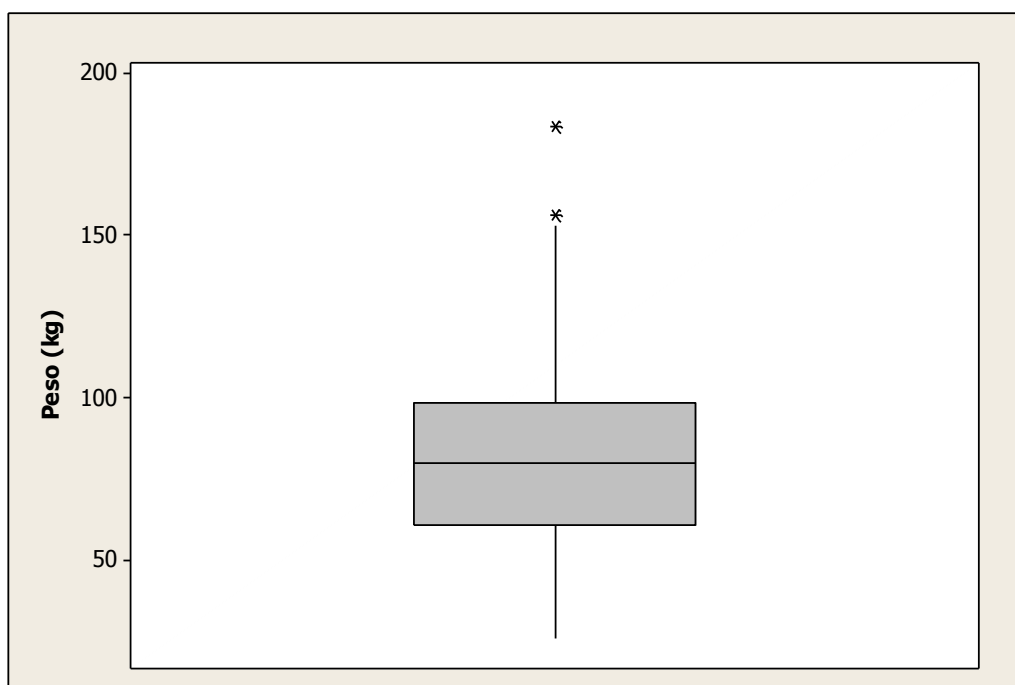
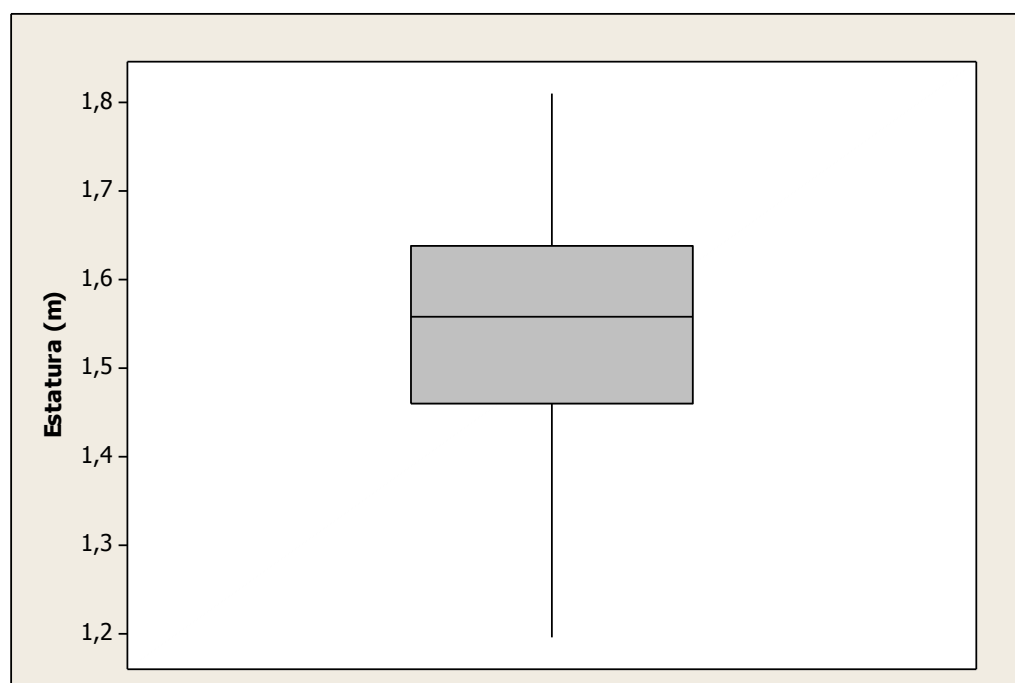
Gráfico B.7: *Box-plot* para o Peso.**Gráfico B.8:** *Box-plot* para a estatura.

Gráfico B.9: *Box-plot* para o percentil da estatura.

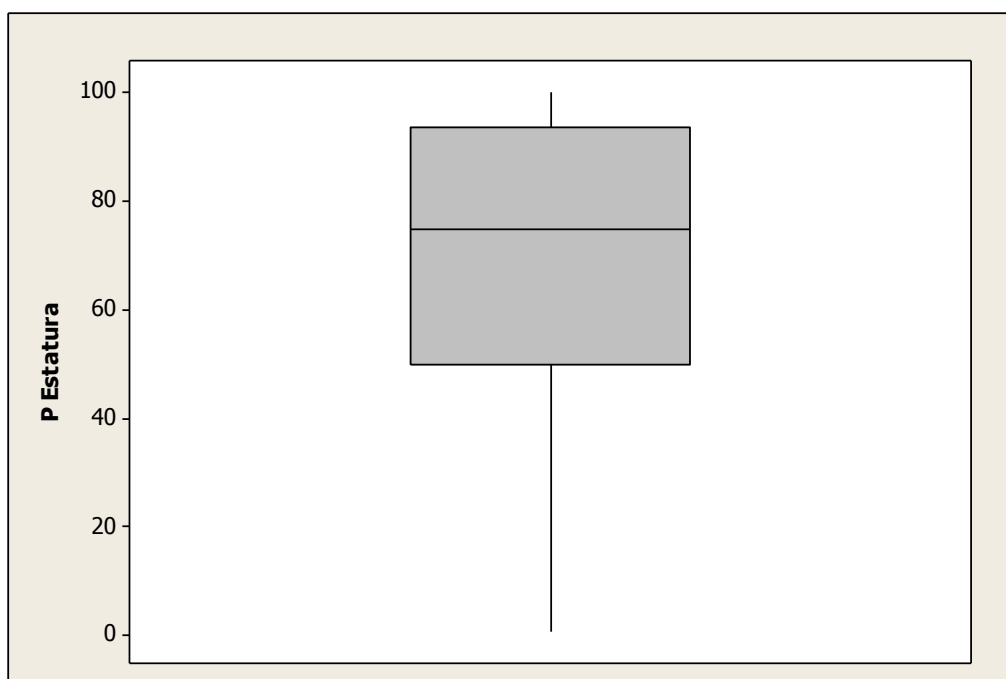


Gráfico B.10: *Box-plot* para o IMC.

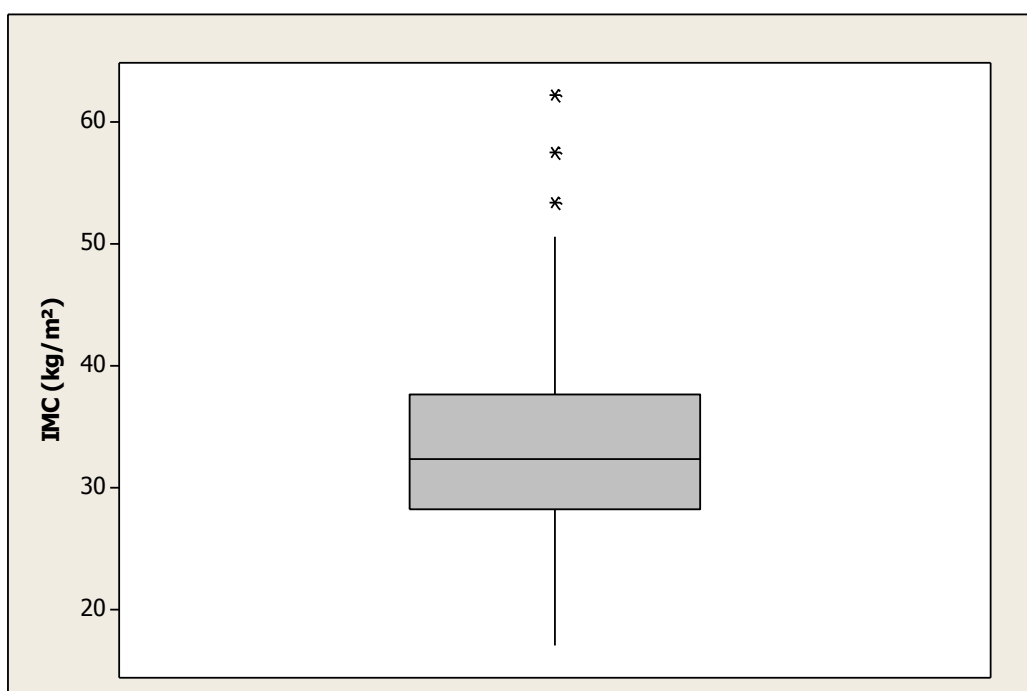


Gráfico B.11: Histograma para o os valores de ZIMC pelo método CDC.

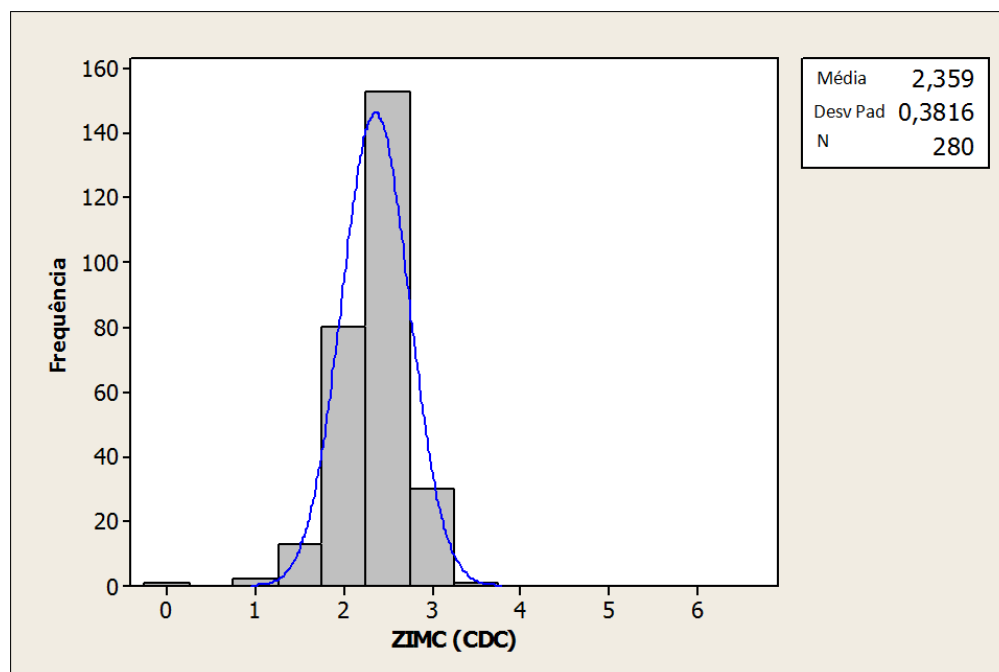


Gráfico B.12: *Box-plot* para o IMC padronizado pelo método CDC.

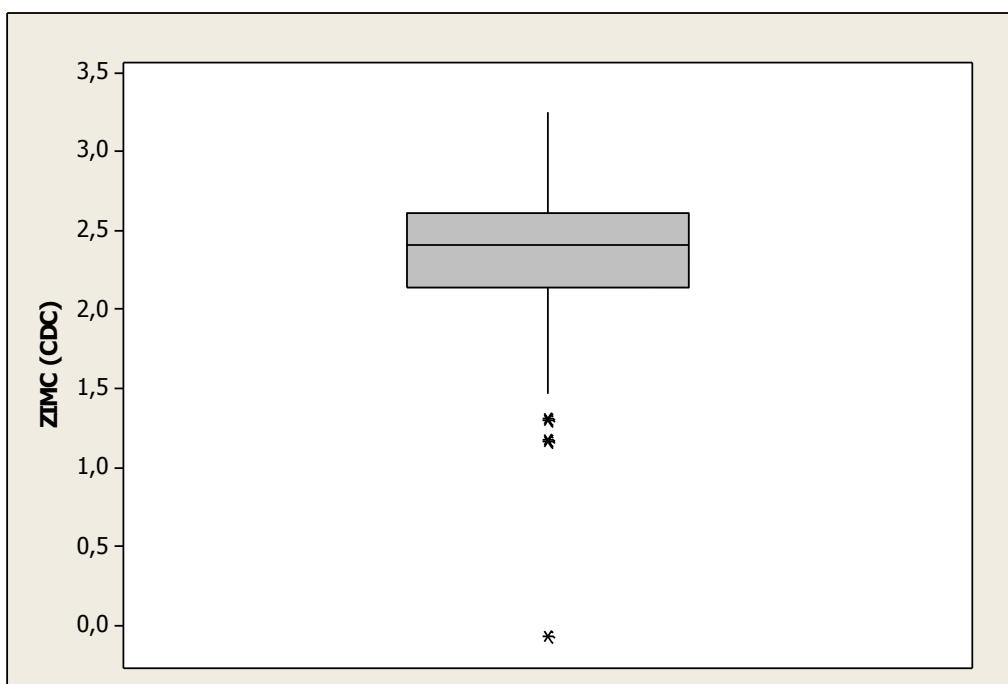


Gráfico B.13: *Box-plot* para o IMC padronizado pelo método CDC segundo cor.

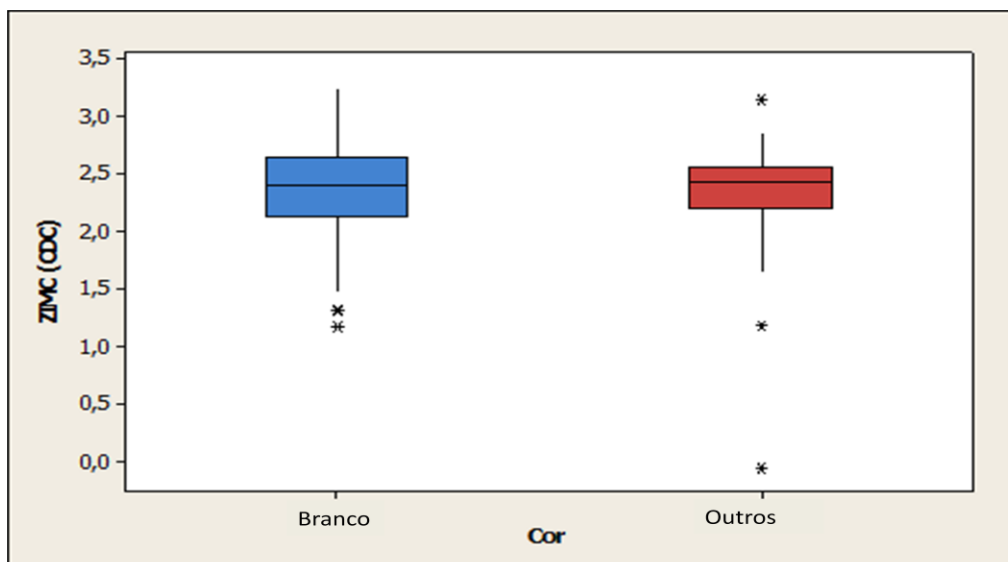


Gráfico B.14: Histograma para os valores de ZIMC pelo método WHO.

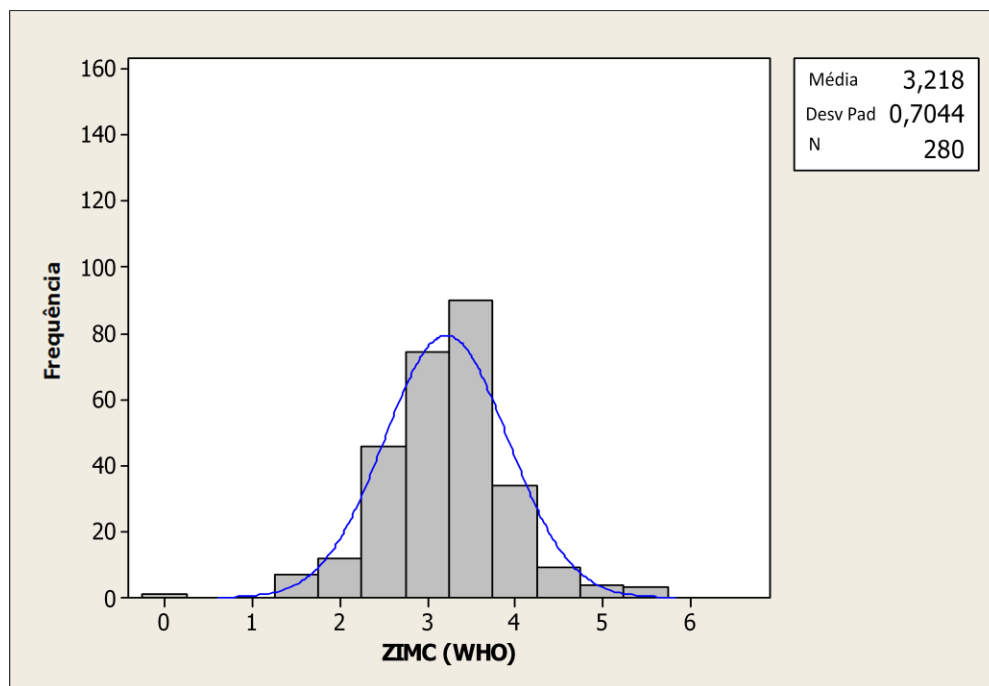


Gráfico B.15: *Box-plot* para o IMC padronizado pelo método WHO.

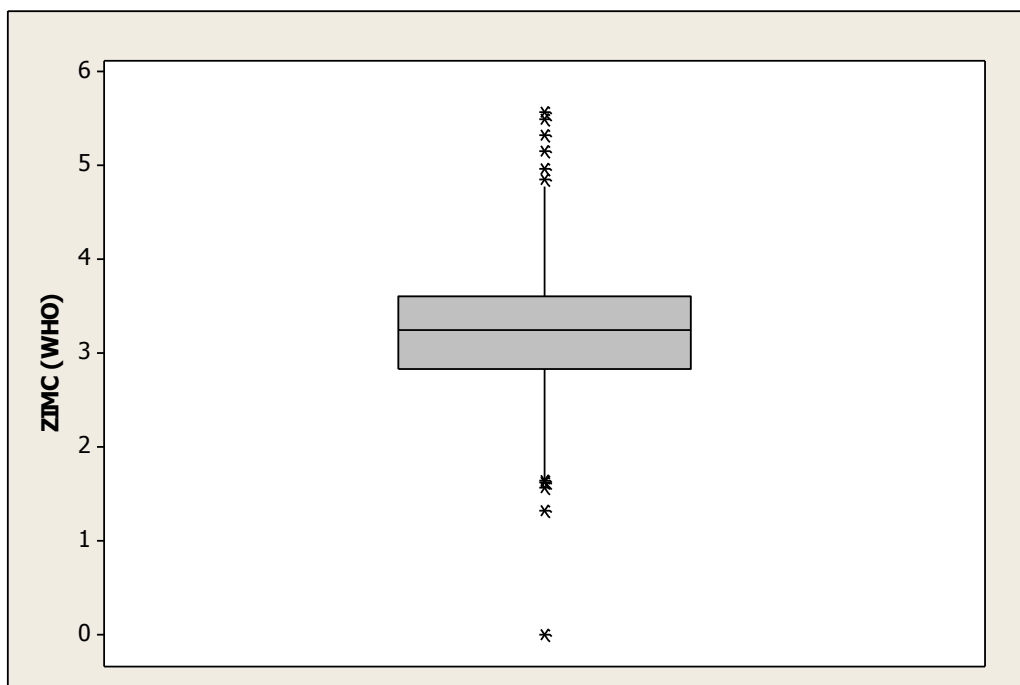


Gráfico B.16: *Box-plot* para o IMC padronizado pelo método WHO segundo cor.

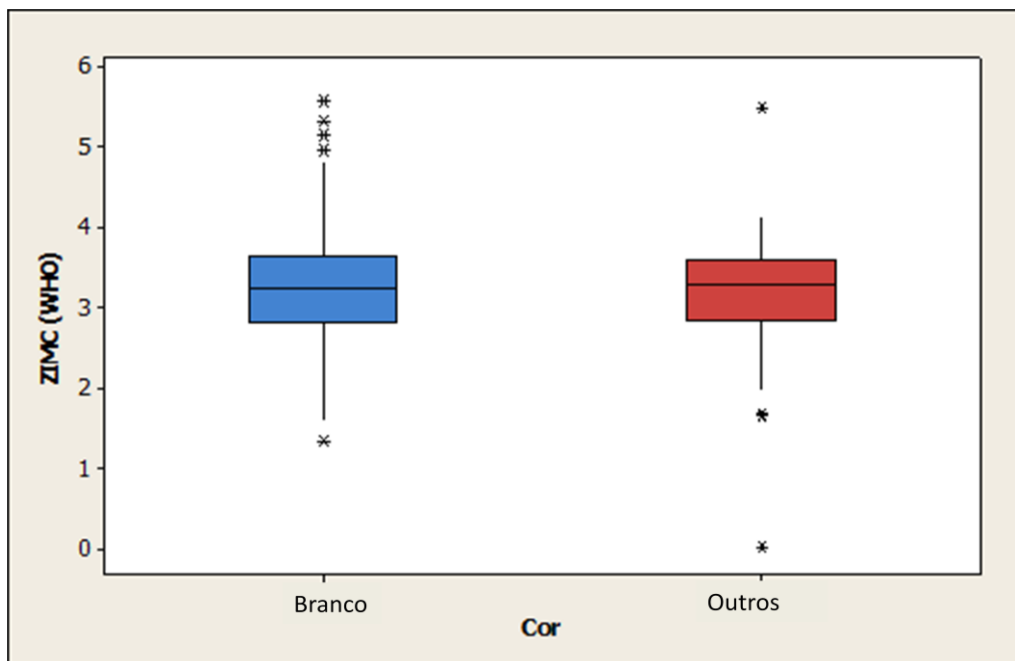


Gráfico B.17: Histograma para o os valores de ZIMC pelo método C&M.

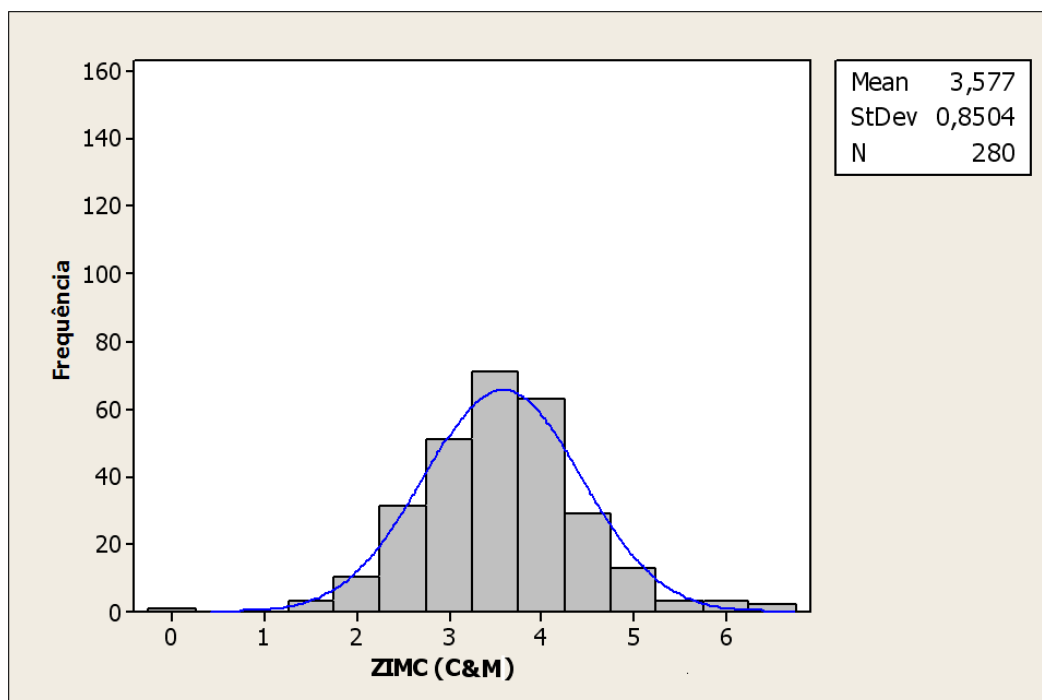


Gráfico B.18: *Box-plot* para o IMC padronizado pelo método C&M.

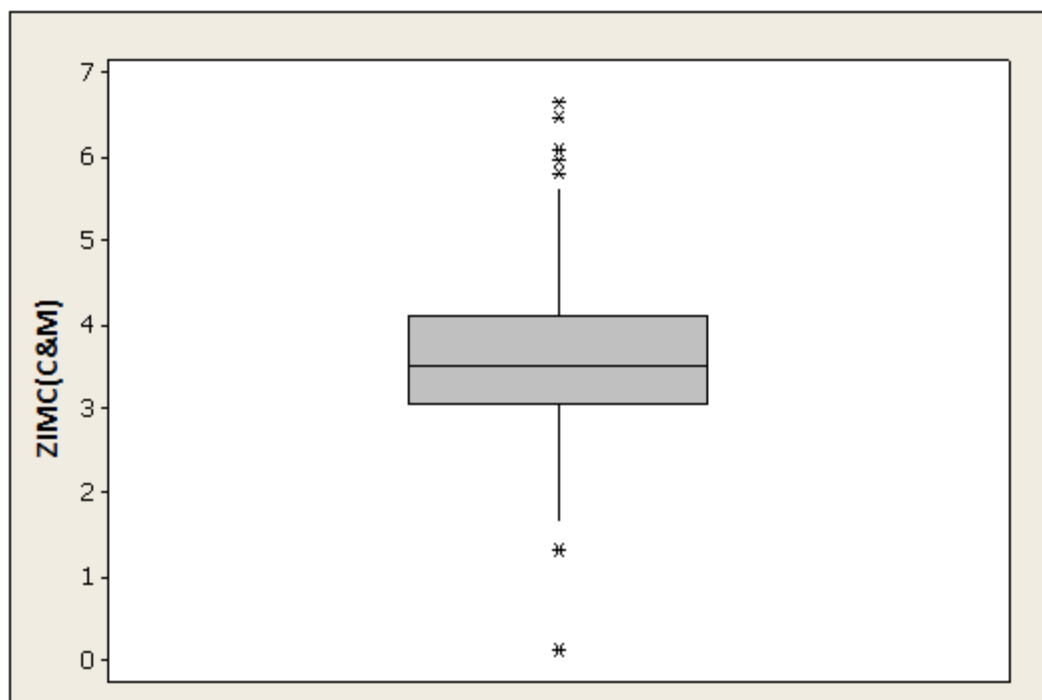


Gráfico B.19: *Box-plot* para o IMC padronizado pelo método C&M segundo cor.

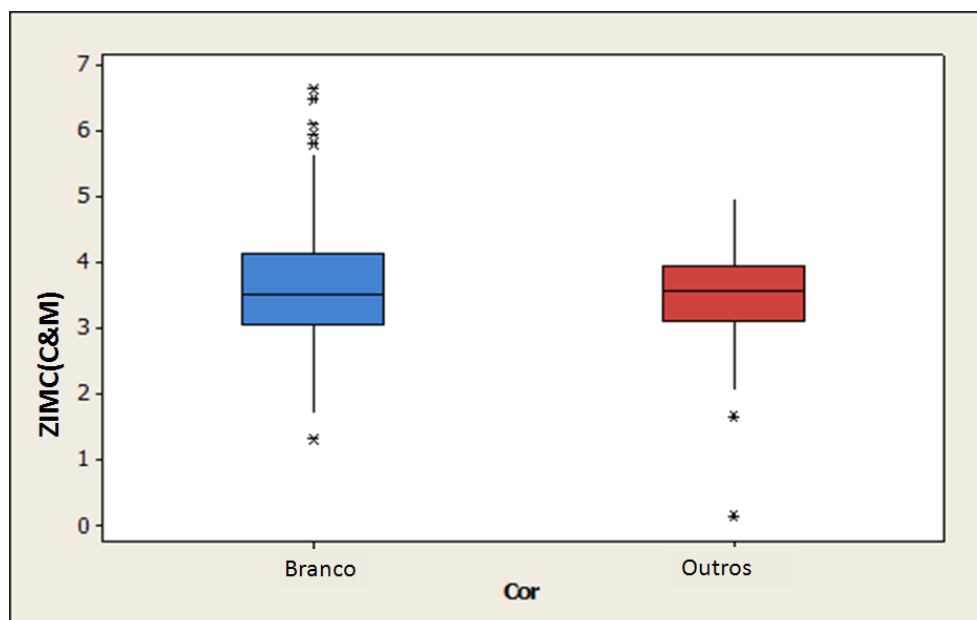


Gráfico B.20: *Box-plot* para a circunferência abdominal calculada no ponto médio.

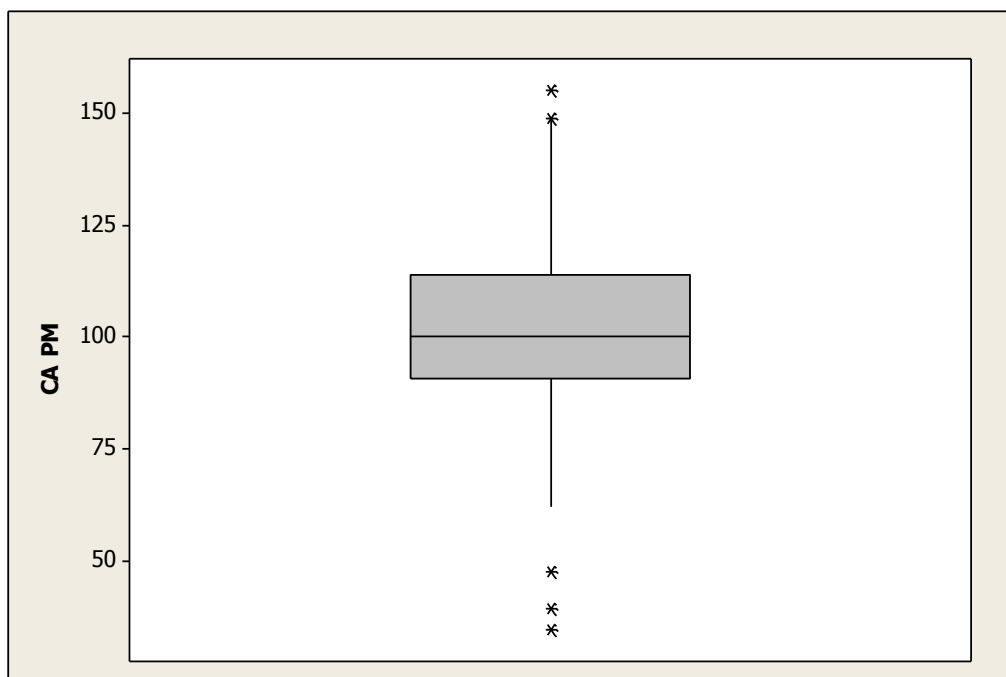


Gráfico B.21: *Box-plot* para a frequência cardíaca.

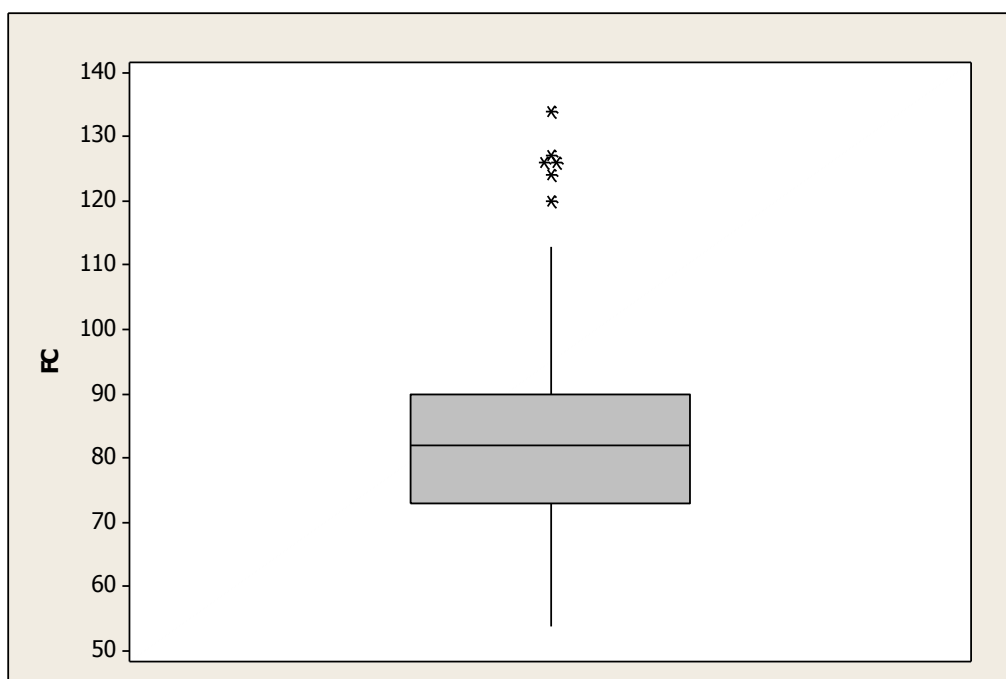


Gráfico B.22: *Box-plot* para a porcentagem de massa de gordura.

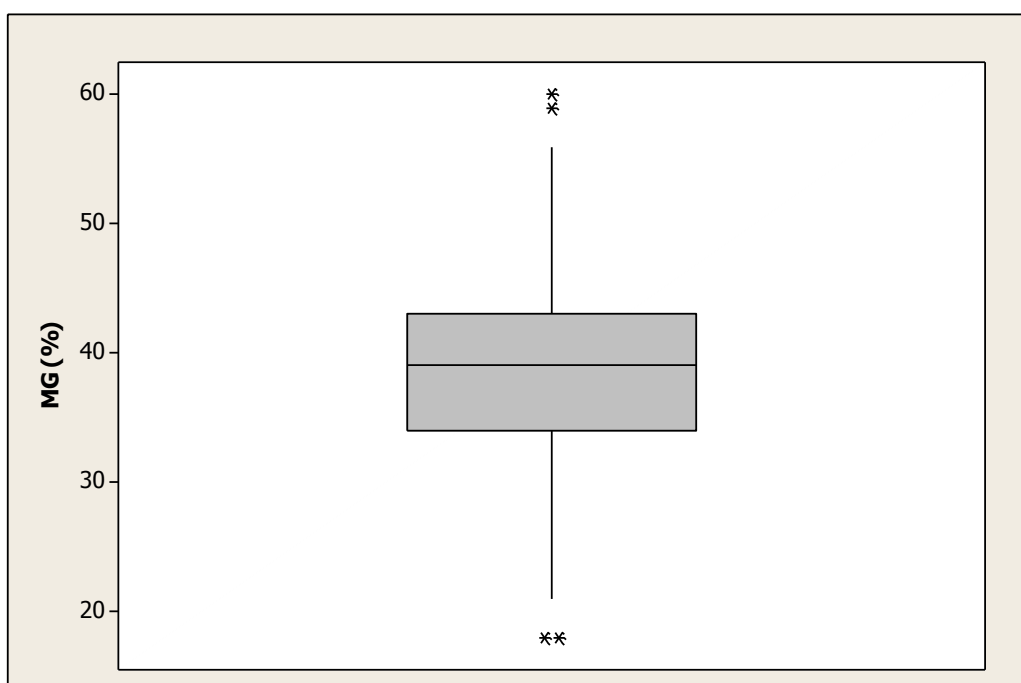


Gráfico B.23: *Box-plot* para a porcentagem de massa de gordura segundo cor.

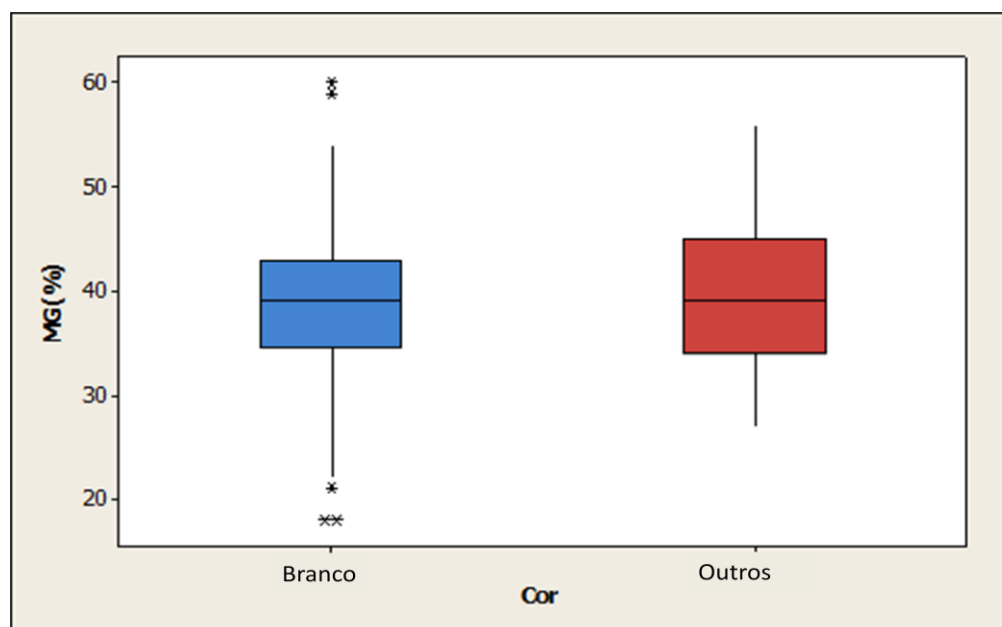


Gráfico B.24: *Box-plot* para o colesterol total.

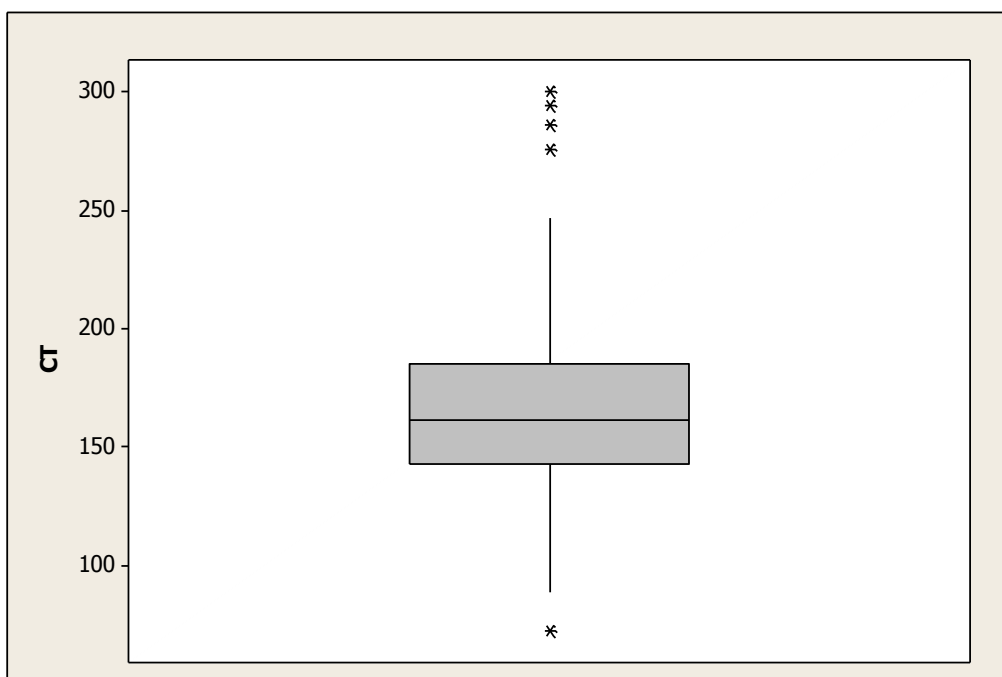


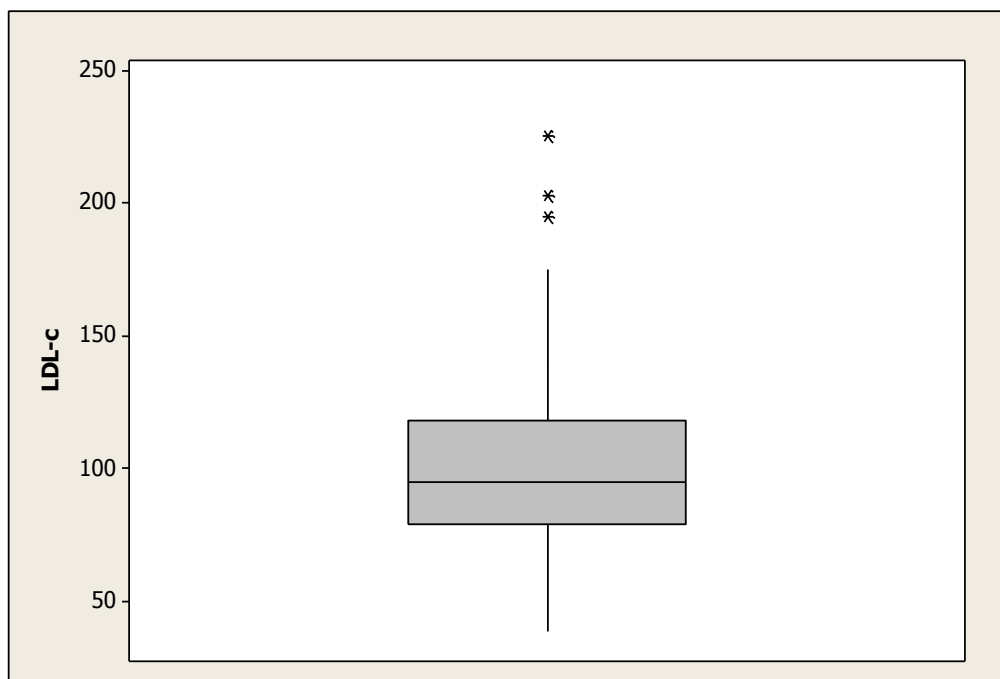
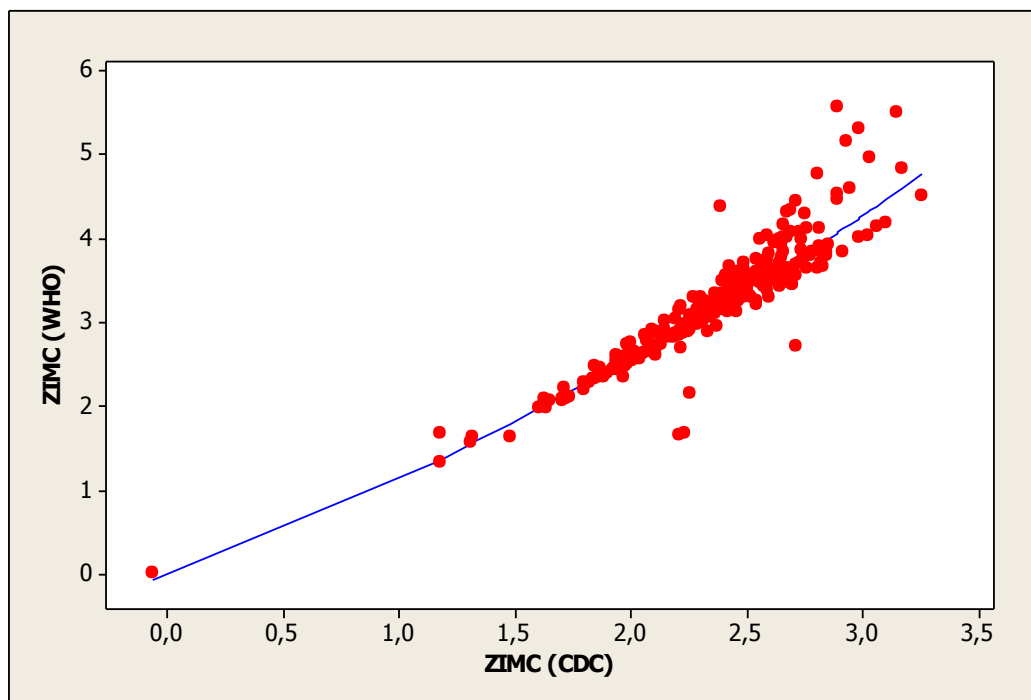
Gráfico B.25: *Box-plot* para o LDL-c.**Gráfico B.26:** Gráfico de dispersão para os valores de ZIMC pelo método de WHO segundo CDC.

Gráfico B.27: Gráfico de dispersão para os valores de ZIMC pelo método C&M segundo CDC.

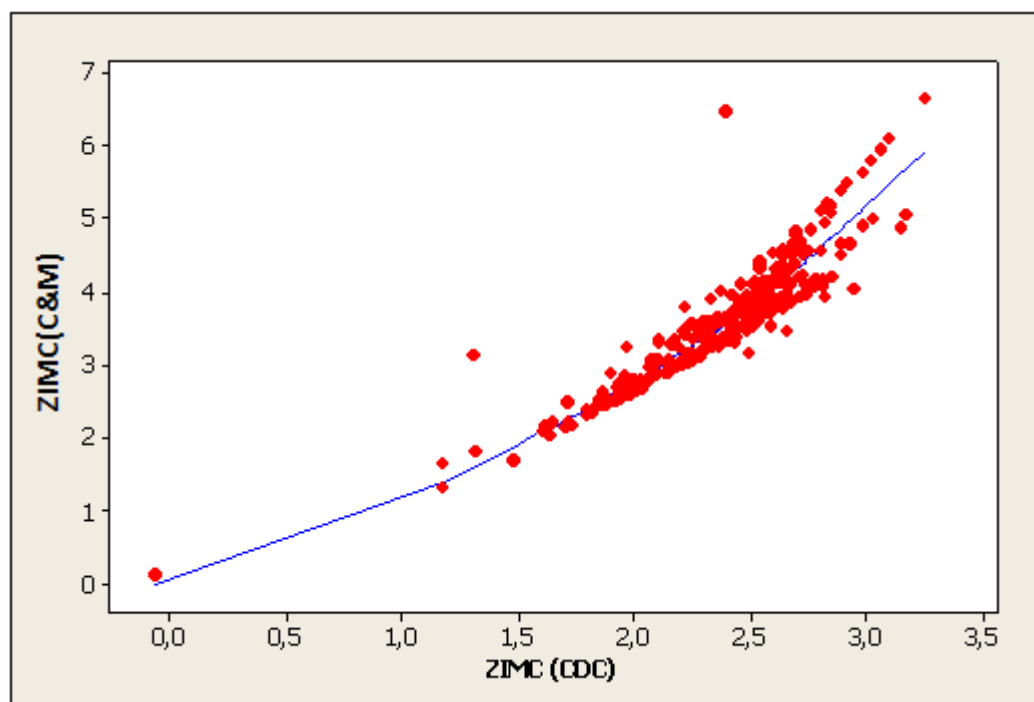


Gráfico B.28: Gráfico de dispersão para os valores de ZIMC pelo método C&M segundo WHO.

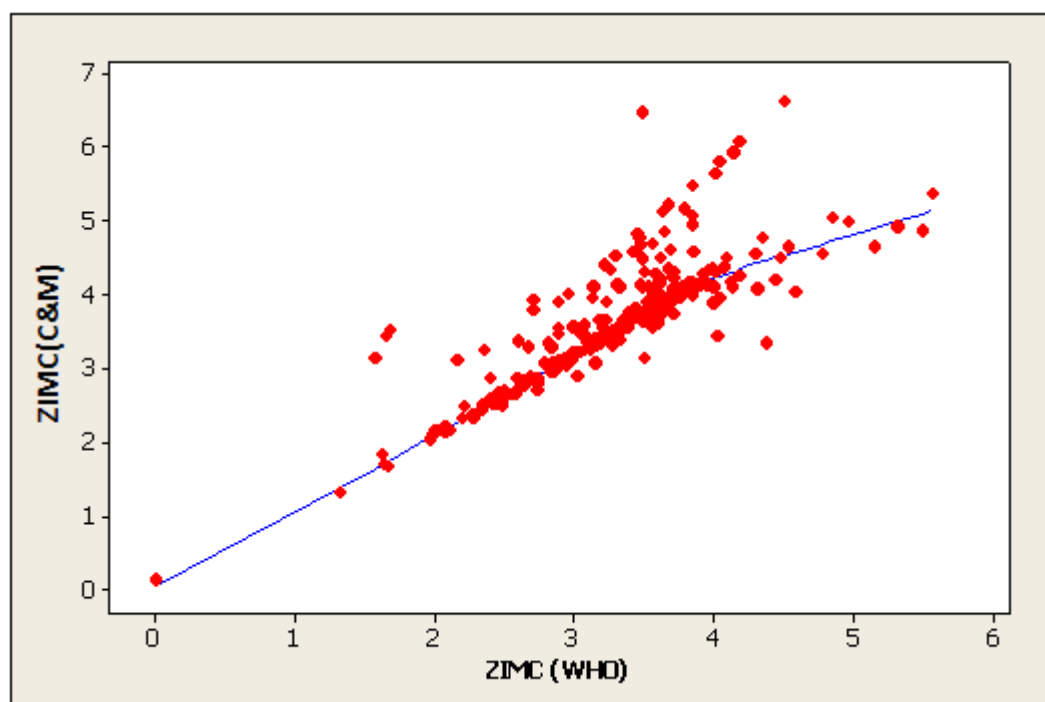


Gráfico B.29: Distribuição de pacientes segundo ocorrência de hipertensão arterial.

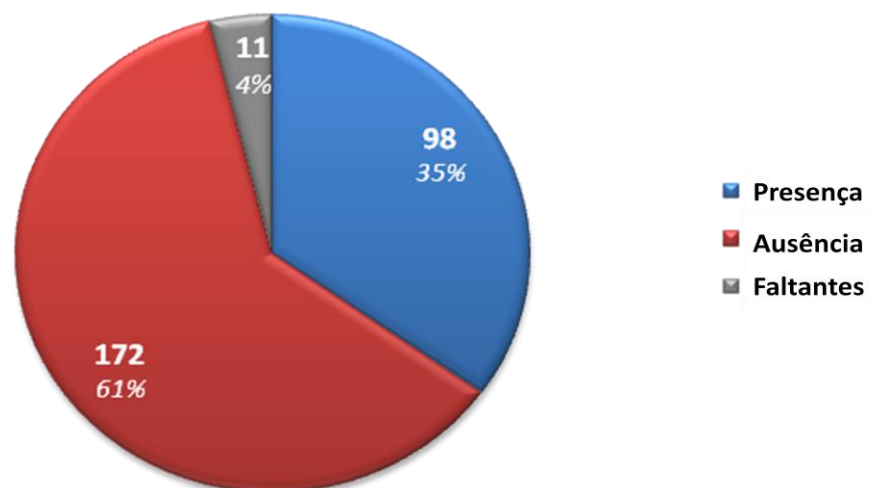


Gráfico B.30: Distribuição de pacientes segundo ocorrência de glicemia alterada.

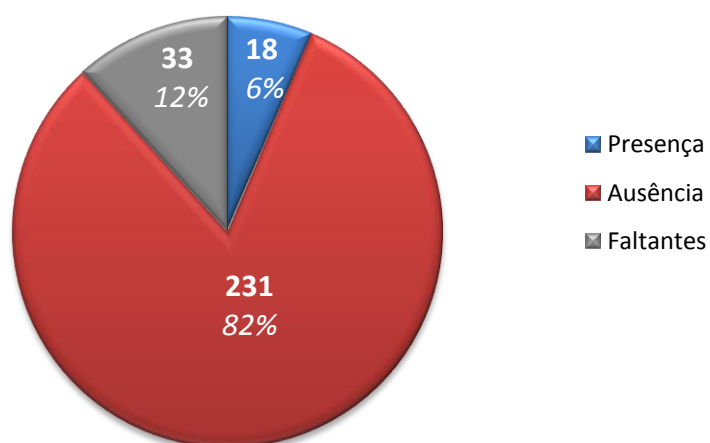


Gráfico B.31: Distribuição de pacientes segundo ocorrência de dislipidemia.

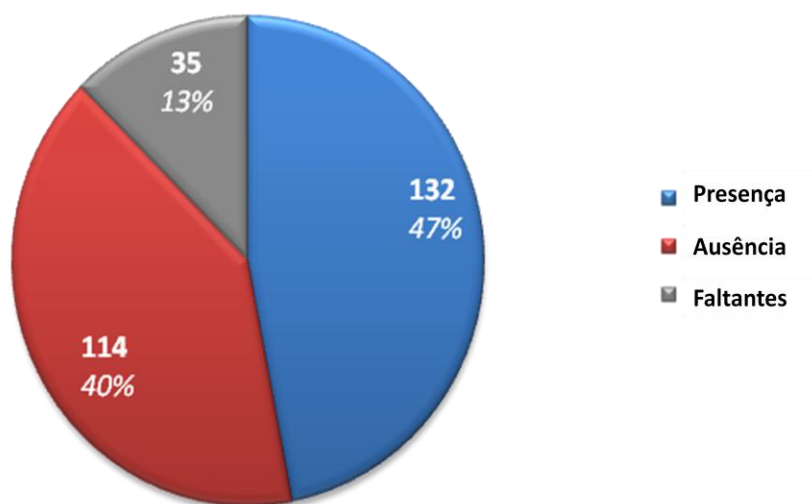


Gráfico B.32: Distribuição de pacientes segundo ocorrência de resistência à insulina.

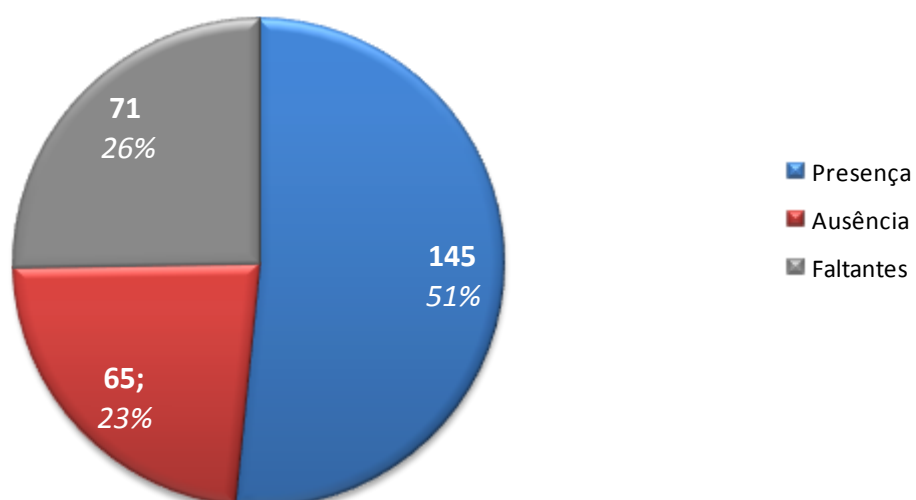


Gráfico B.33: Distribuição dos pacientes segundo número de doenças diagnosticadas

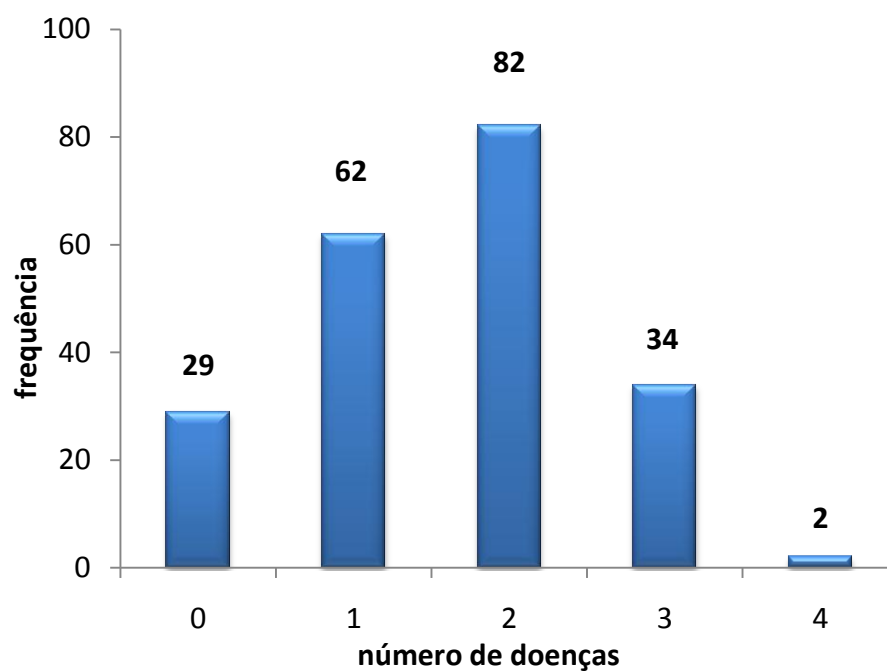


Gráfico B.34: Box-plot para ZIMC pelo método de CDC segundo número de doenças.

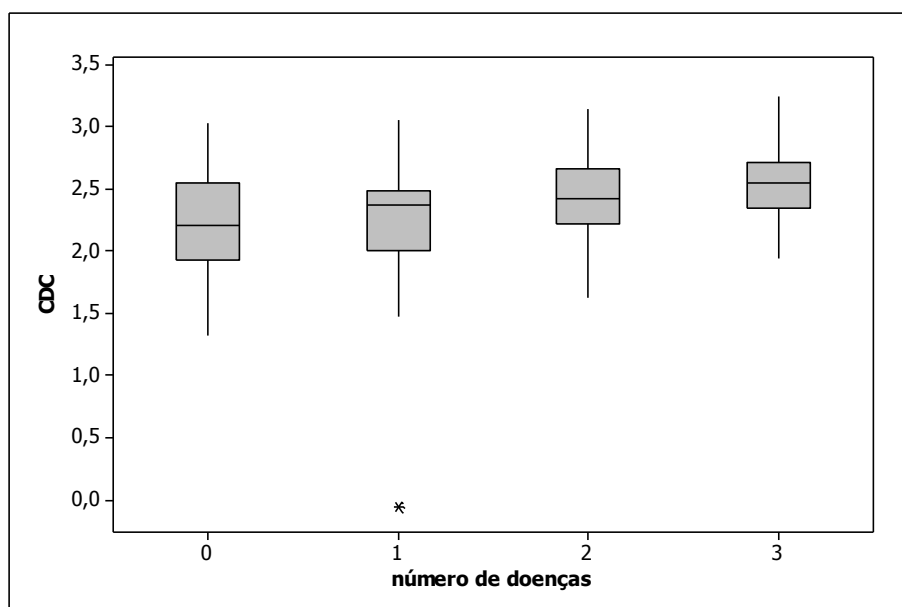


Gráfico B.35: *Box-plot* para ZIMC pelo método de WHO segundo número das doenças.

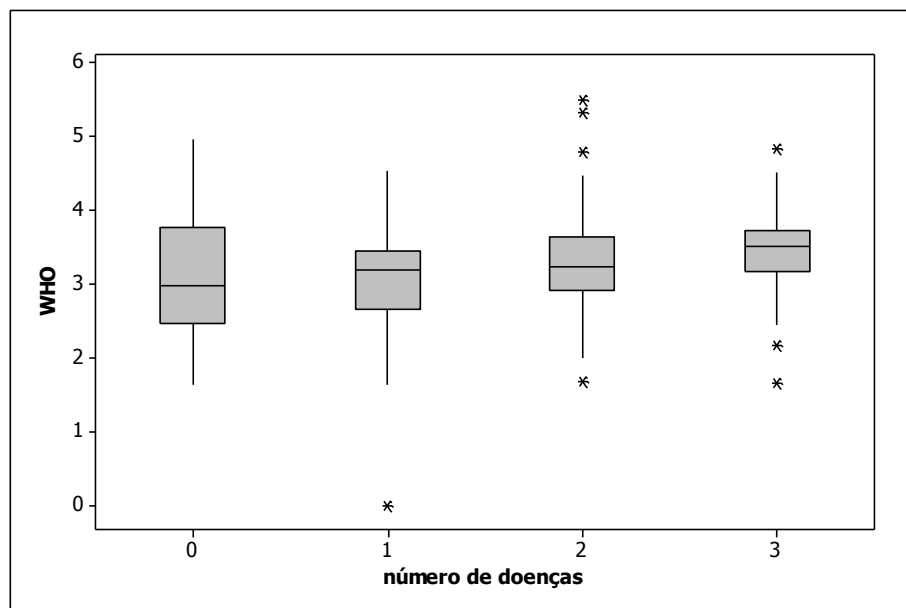


Gráfico B.36: *Box-plot* para ZIMC pelo método C&M segundo número das doenças.

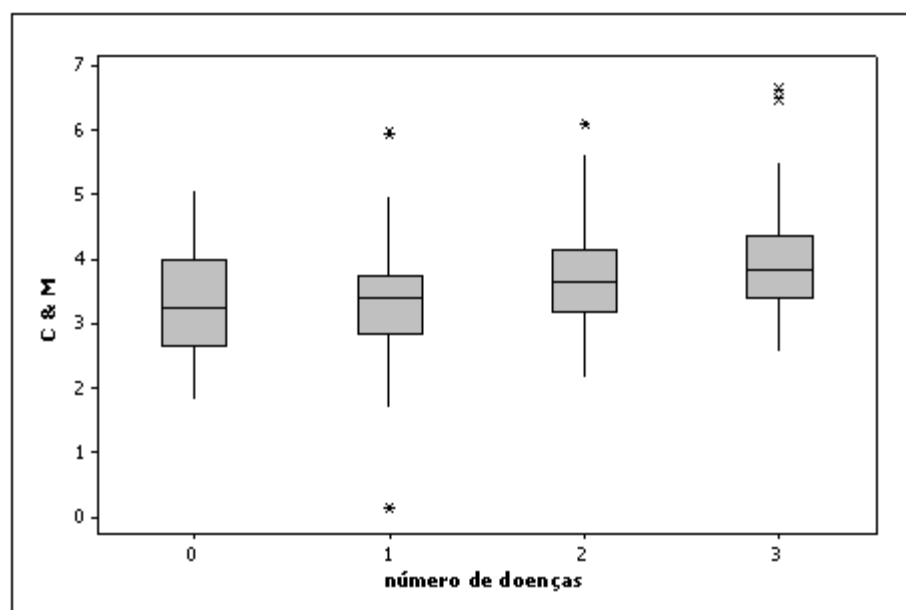


Gráfico B.37: Gráfico da probabilidade estimada segundo o número de doenças e percentual de massa de gordura para o método CDC.

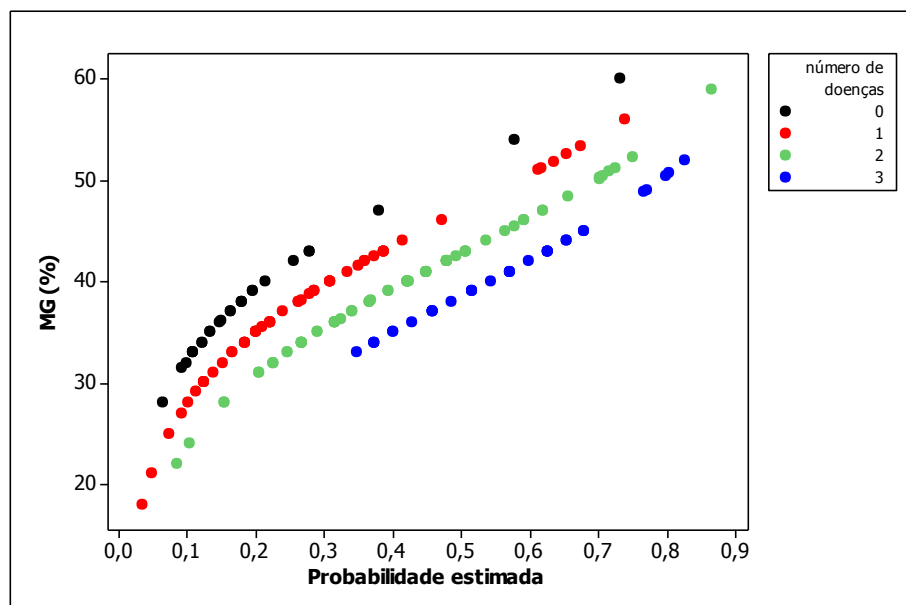


Gráfico B.38: Gráfico da probabilidade estimada segundo o número de doenças e percentual de massa de gordura para o método WHO.

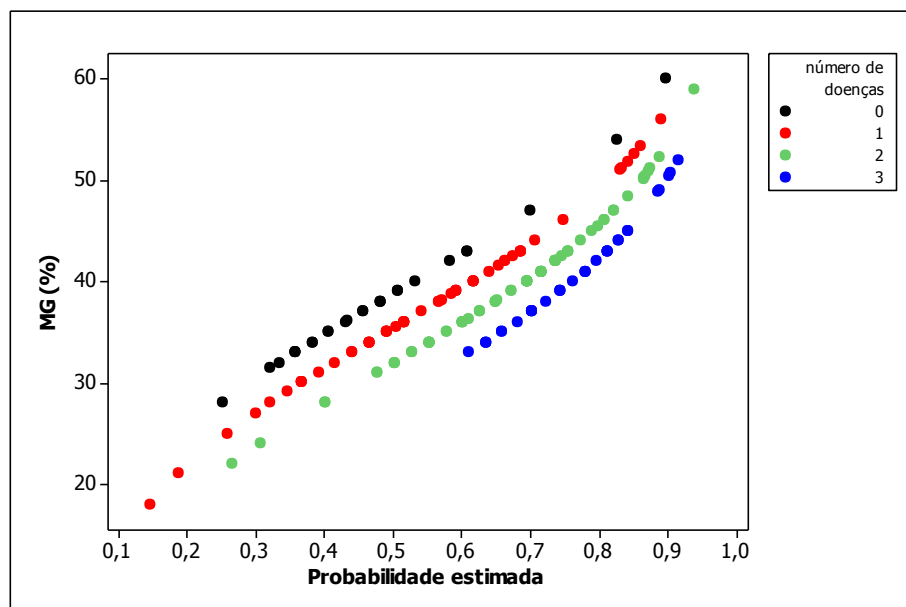


Gráfico B.39: Gráfico da probabilidade estimada segundo o número de doenças e percentual de massa de gordura para o método C&M.

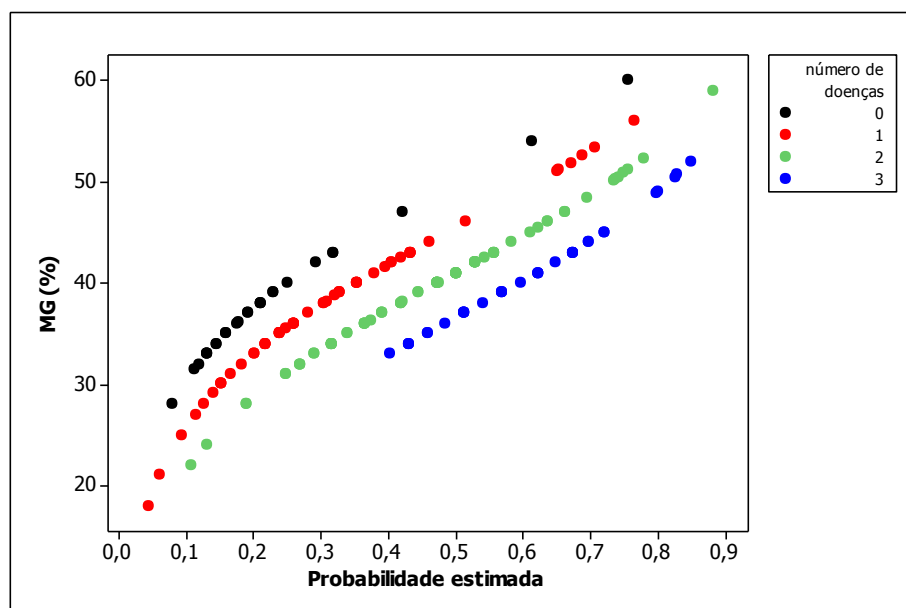


Gráfico B.40: Gráfico normal de probabilidades com banda empírica de 95% de confiança para os resíduos do modelo logístico proposto para o método CDC

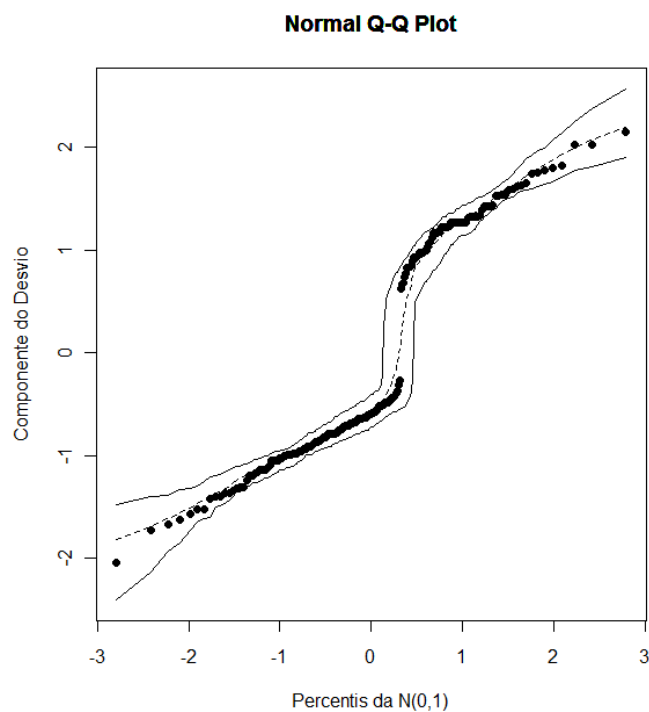


Gráfico B.41: Gráfico dos pontos influentes do modelo logístico proposto para o método CDC

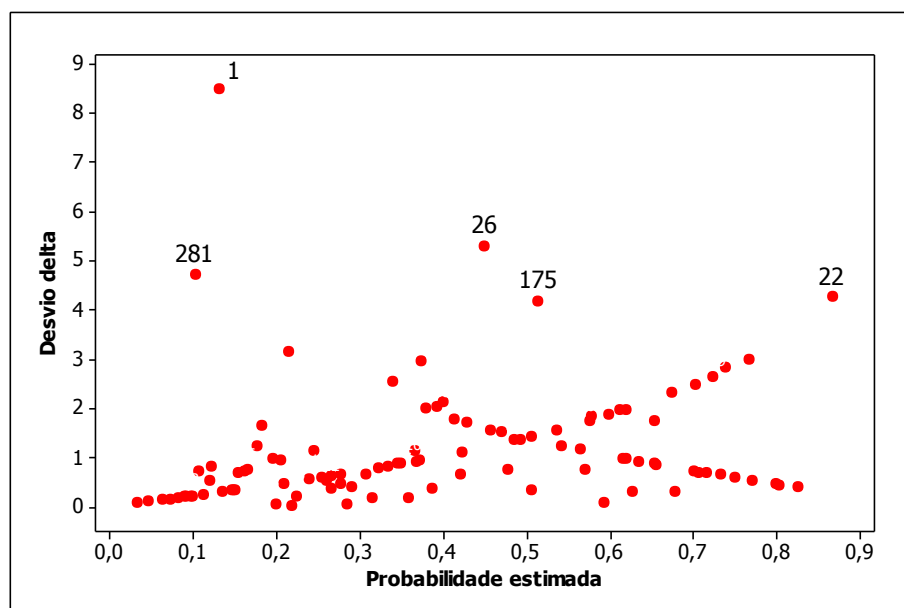


Gráfico B.42: Gráfico normal de probabilidades com banda empírica de 95% de confiança para os resíduos do modelo logístico proposto para o método WHO

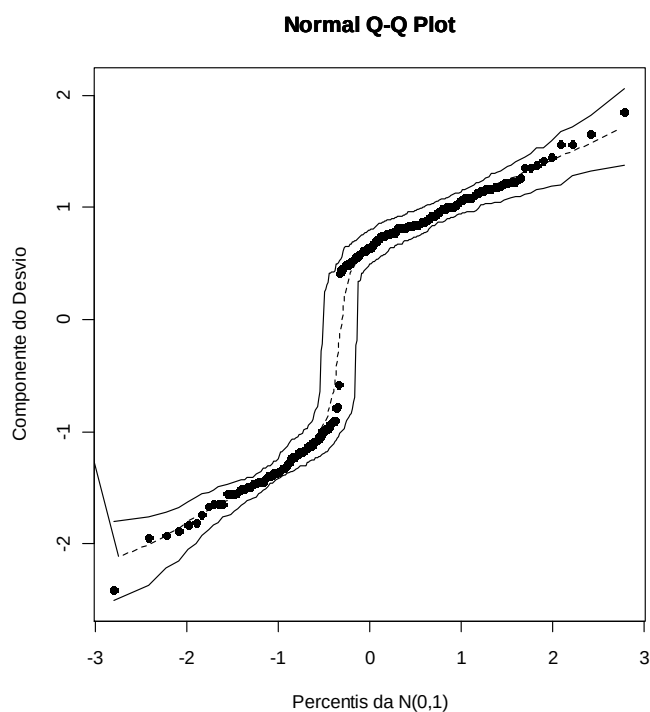


Gráfico B.43: Gráfico dos pontos influentes do modelo logístico proposto para o método WHO

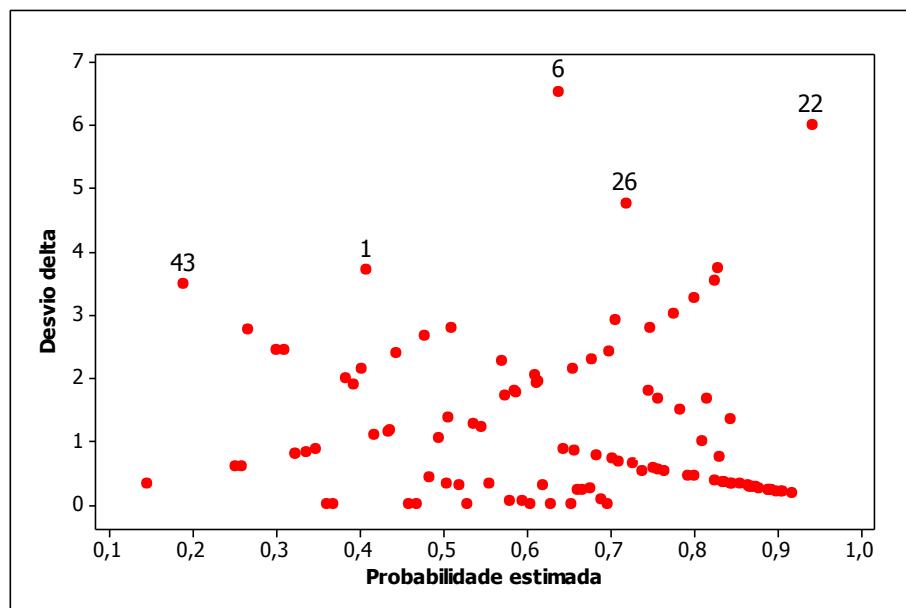


Gráfico B.44: Gráfico normal de probabilidades com banda empírica de 95% de confiança para os resíduos do modelo logístico proposto para o método C&M

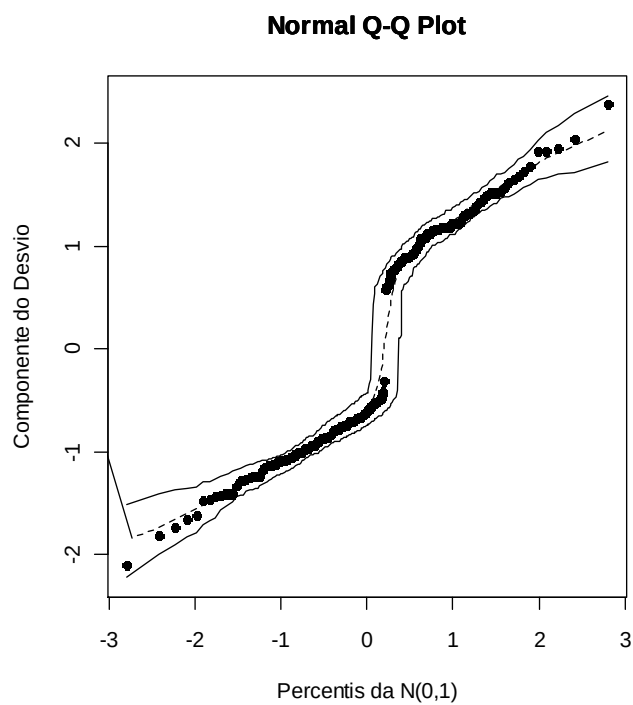
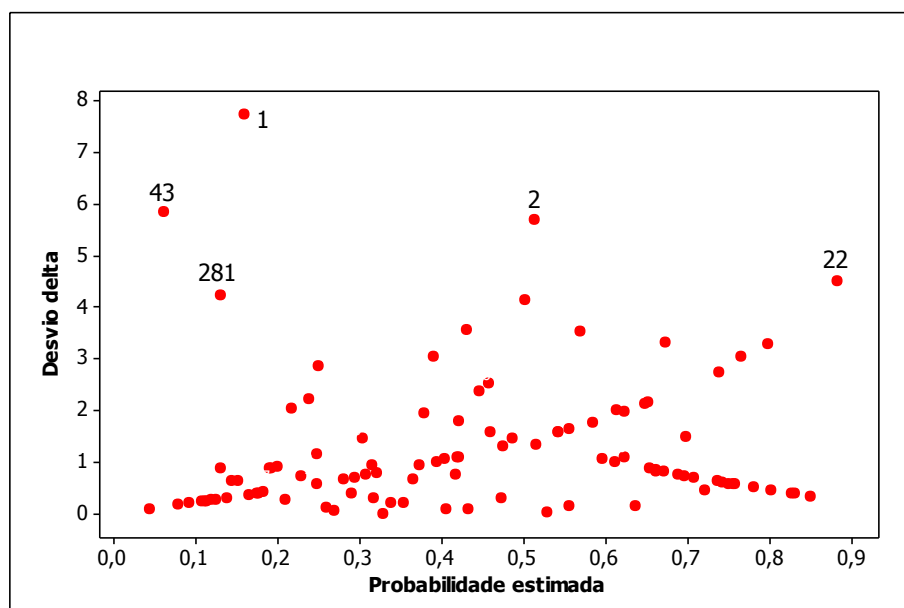


Gráfico B.44: Gráfico dos pontos influentes do modelo logístico proposto para o método C&M



APÊNDICE C
EXPRESSÕES

Expressão C.1: Cálculo do Z pelo método LMS:

O método LMS assume que, para dados independentes com valores positivos, a transformação Box-Cox idade-específica pode ser empregada para torná-los normalmente distribuídos; A amostra foi separada por gênero e para cada gênero foi separada em faixas etárias, cada uma de 3 meses, denominados estratos. Em cada estrato, foram calculados os parâmetros L, M e S. O parâmetro M expressa mediana do IMC (kg/m²) em cada estrato; S representa o coeficiente de variação de cada estrato e o parâmetro L, o coeficiente (Box-Cox) empregado para a transformação matemática dos valores do IMC com o objetivo de obter distribuição normal em cada estrato. O valor selecionado para o coeficiente L é aquele cuja transformação produza a menor soma dos quadrados dos desvios da variável.

$$Z = \frac{\left[\left(\frac{IMC}{M}\right)^L - 1\right]}{(LS)} \text{ para } L > 0$$

$$Z = \frac{1}{S} \ln\left(\frac{IMC}{M}\right) \text{ para } L = 0$$

Expressão C.2: Equação de regressão:

A regressão foi construída para se obter um valor de corte para classificação pelo método de Conde & Monteiro (C&M), já que este valor não existe na literatura.

$$ZIMC(C\&M) = 1,22 + 0,996 ZIMC(CDC)$$

Pela análise de resíduos, todas as suposições foram satisfeitas. Como o valor de corte do método CDC é 2,5, valor fornecido pela pesquisadora, teremos a seguinte equação:

$$ZIMC(C\&M) = 1,22 + 0,996 * 2,5 = 3,71$$

Portanto 3,71 será a nota de corte estabelecida para o método de C&M.

Expressão C.3: Cálculo do índice Kappa

$$\hat{k} = \frac{\hat{\theta}_1 - \hat{\theta}_2}{1 - \hat{\theta}_2}$$

$$\text{onde, } \hat{\theta}_1 = \frac{\sum x_{ii}}{n}, \hat{\theta}_2 = \frac{\sum x_{i.} + x_{.i}}{n^2}$$

Com x_{ii} , $x_{i.}$, $x_{.i}$, $x_{..}$ e n definidos na tabela abaixo:

Classificação B	Classificação A				
	A	B	...	z	Total
A	x_{11}	x_{12}	...	x_{1i}	$x_{1.}$
B	x_{21}	x_{2i}	...	x_{2i}	$x_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Z	x_{i1}	x_{i2}	...	x_{ii}	$x_{i.}$
Total	$x_{.1}$	$x_{.2}$	...	$x_{.i}$	n

Expressão C.4: Cálculo da variância do índice Kappa

$$\widehat{Var(k)} = \frac{1}{n} \left[\frac{\hat{\theta}_1(1 - \hat{\theta}_1)}{(1 - \hat{\theta}_2)^2} + \frac{2(1 - \hat{\theta}_1)(2\hat{\theta}_1\hat{\theta}_2 - \hat{\theta}_3)}{(1 - \hat{\theta}_2)^3} + \frac{(1 - \hat{\theta}_1)^2(\hat{\theta}_4 - 4\hat{\theta}_2^2)}{(1 - \hat{\theta}_2)^4} \right]$$

$$\text{onde, } \hat{\theta}_3 = \frac{\sum x_{ii}(x_{i.} + x_{.i})}{n^2}, \hat{\theta}_4 = \frac{\sum_i \sum_j x_{ij}(x_{j.} + x_{.i})}{n^3}$$

Expressão C.5: Cálculo do intervalo de confiança de 95% para o índice Kappa

$$IC(95\%) = [\hat{k} - 1,96 * \sqrt{\widehat{Var(k)}}; \hat{k} + 1,96 * \sqrt{\widehat{Var(k)}}]$$

APÊNDICE D
METODOLOGIA

Metodologia D.1: Regressão Logística.

A Regressão Logística é uma técnica estatística utilizada quando a variável resposta de interesse é uma variável binária (sucesso ou fracasso). Assim, tenta-se modelar a probabilidade de no evento de interesse ocorrer sucesso, dada uma série de variáveis explicativas.

Metodologia D.2: Modelo logístico proposto:

y_i : Diagnóstico segundo o método ($y_i = 0$: obeso; $y_i = 1$: super obeso) com a seguinte Suposição: $\sum y_i \sim B(n, p_i)$, em que:

n : total de pacientes (192) e

p_i : probabilidade do i-ésimo paciente ser diagnosticado como super obeso pelo método

$$\log \left\{ \frac{p_i}{1-p_i} \right\} = \alpha + \beta_1 * \text{número de doenças}_i + \beta_2 * MG(\%)_i,$$

com $i = 1, 2, \dots, 192$.

APÊNDICE E
SUPOSIÇÕES VERIFICADAS

Ajuste do modelo

Para verificarmos se os modelos estão bem ajustados, foram feitas algumas análises para validar as suposições impostas inicialmente.

Método CDC

Pelo Gráfico B.40, modelo proposto para o método CDC, pode-se verificar que todos os pontos estão dentro da banda ajustada, indicando que a suposição da distribuição binomial ($\sum y_i \sim B(n, p_i)$) é satisfatória.

Através do Gráfico B.41, podemos detectar quais pontos (pacientes) poderiam influenciar a análise, destoando dos resultados obtidos. Pode-se verificar que há 5 pontos aparentemente mais influentes na amostra. Por exemplo: (i) o paciente 1 não possui doenças tem 35 de percentual de massa de gordura e foi considerado como super obeso; (ii) o paciente 22 possui 2 doenças, tem 59 de percentual de massa de gordura e foi classificado como obeso; (iii) o paciente 26 possui 2 doenças, 41 de percentual de massa de gordura e foi classificado como super obeso; (iv) o paciente 175 possui 3 doenças, 39 de percentual de massa de gordura e foi classificado como super obeso (v) o paciente 281 tem 2 doenças, possui 24 % de massa de gordura e foi diagnosticado como super obeso. Podemos observar que os pacientes 1 e 281, seriam diagnosticados como obesos pelo modelo por apresentarem um baixo número de doenças e percentual de gordura abaixo da média, mas o método CDC os classificou como super obesos. O paciente 22 foi classificado como obeso pelo método, mas o modelo o classificaria como super obeso, pois este apresenta duas doenças e alto percentual de massa de gordura. Estes pacientes apareceram como destoantes, pois não possuem o perfil mais esperado daqueles que foram diagnosticados como obesos (menor número de doenças e percentual de massa de gordura), ou super obesos (maior número de doenças ou percentual de massa de gordura).

Para verificar se estes 5 indivíduos poderiam influenciar no resultado do modelo proposto, retiramos estes pontos da amostra e ajustamos um novo modelo logístico. As estimativas dos parâmetros, que podem ser vistas na Tabela A.14, continuaram marginalmente significantes ao nível de significância

de 5% e não apresentaram muitas alterações, permanecendo com as mesmas interpretações, ou seja, os pontos influentes não afetam o desempenho do modelo.

Retirando os 5 pontos influentes a nova razão de chances para o número de doenças será de 1,69 com IC(95%) igual a [1,15; 2,45].

Para o percentual de massa de gordura a nova razão de chances será igual a 1,16 com IC(85%) de [1,09;1,23].

Método WHO

Pelo Gráfico B.42, modelo proposto para o método WHO, pode-se verificar que todos os pontos estão dentro da banda ajustada, indicando que a suposição da distribuição binomial ($\sum y_i \sim B(n, p_i)$) é satisfatória.

Através do Gráfico B.43, podemos detectar quais pontos (pacientes) poderiam influenciar a análise, destoando dos resultados obtidos. Pode-se verificar que há 5 pontos aparentemente mais influentes na amostra, sendo que, os pacientes 1, 22 e 26 já foram detectados e caracterizados como pontos influentes pelo método CDC. Os dois pacientes que ainda aparecem como discrepantes são: (i) o paciente 6 possui 3 doenças, tem 34 de percentual de massa de gordura e foi considerado como obeso e; (ii) o paciente 43 possui 1 doença, tem 21 de percentual de massa de gordura e foi classificado como super obeso. Podemos observar que os pacientes 1 e 43, seriam diagnosticados como obesos pelo modelo por apresentarem baixo número de doenças e percentual de gordura abaixo da média, mas o método WHO os diagnosticou como super obesos. Os pacientes 6 e 22 foram classificados como obeso pelo método, mas o modelo os classificaria como super obesos, pois este apresentam acima de duas doenças e alto percentual de massa de gordura. Estes pacientes apareceram como destoantes, pois não possuem o perfil mais esperado daqueles que foram diagnosticados como obesos (menor número de doenças e/ou percentual de massa de gordura), ou super obesos (maior número de doenças e/ou percentual de massa de gordura).

Para verificar se estes 5 indivíduos poderiam influenciar no resultado do modelo proposto, retiramos estes pontos da amostra e ajustamos um novo

modelo logístico. As estimativas dos parâmetros, que podem ser vistas na Tabela A.15 continuaram marginalmente significantes ao nível de significância de 5% e não apresentaram muitas alterações, permanecendo com as mesmas interpretações, ou seja, os pontos influentes não afetam o desempenho do modelo.

Retirando os 5 pontos influentes a nova razão de chances para o número de doenças será de 1,51 com IC(95%) igual a [1,05; 2,17].

Para o percentual de massa de gordura a nova razão de chances será igual a 1,14 com IC(85%) de [1,07;1,21].

Método Conde & Monteiro

Pelo Gráfico B.44, modelo proposto para o método C&M, pode-se verificar que todos os pontos estão dentro da banda ajustada, indicando que a suposição da distribuição binomial ($\sum y_i \sim B(n, p_i)$) é satisfatória.

Através do Gráfico B.44, podemos detectar quais pontos (pacientes) poderiam influenciar a análise destoando dos resultados obtidos. Pode-se verificar que há 5 pontos aparentemente mais influentes na amostra, sendo que, os pacientes 1, 22, 43 e 281 já foram detectados e caracterizados como pontos influentes pelos métodos CDC e WHO. O outro ponto que aparece como discrepante é o paciente 2 que possui 3 doenças, tem 37 de percentual de massa de gordura e foi considerado como super obeso. Podemos observar que os pacientes 1 e 43 e 281, seriam diagnosticados como obesos pelo modelo por apresentarem baixo número de doenças e percentual de gordura abaixo da média, mas o método C&M os diagnosticou como super obesos. O paciente 22 foi classificado como obeso pelo método, mas o modelo o classificaria como super obeso, pois este apresenta duas doenças e alto percentual de massa de gordura. Estes pacientes apareceram como destoantes, pois não possuem o perfil mais esperado daqueles que foram diagnosticados como obesos (menor número de doenças e/ou percentual de massa de gordura), ou super obesos (maior número de doenças e/ou percentual de massa de gordura).

Para verificar se estes 5 indivíduos poderiam influenciar no resultado do modelo proposto, retiramos estes pontos da amostra e ajustamos um novo modelo logístico. As estimativas dos parâmetros, que podem ser vistas na Tabela A.16, continuaram marginalmente significantes ao nível de significância de 5% e não apresentaram muitas alterações, permanecendo com as mesmas interpretações, ou seja, os pontos influentes não afetam o desempenho do modelo.

Retirando os 5 pontos influentes a nova razão de chances para o número de doenças será de 1,73 com IC(95%) igual a [1,18; 2,52].

Para o percentual de massa de gordura a nova razão de chances será igual a 1,17 com IC(85%) de [1,10;1,25].

APÊNDICE F
DISCRIMINAÇÃO DO MODELO

6.1 Discriminação pelo modelo

Para verificarmos a discriminação do modelo, a probabilidade estimada do modelo foi calculada da seguinte forma:

$$\hat{p} = \frac{e^{(\alpha + \beta_1 * \text{número de doenças}_i + \beta_2 * MG(\%)_i)}}{1 + e^{(\alpha + \beta_1 * \text{número de doenças}_i + \beta_2 * MG(\%)_i)}}$$

com $i = 1, 2, \dots, 192$.

em que $\hat{p}$ é a probabilidade estimada de um paciente ser classificado como super obeso dado o número de doenças e o percentual de massa de gordura.

6.1.1 Método CDC

Foi utilizado um critério de classificação para testar a qualidade de acerto do modelo. Dado que o percentual de super obeso na amostra foi de 38%, este valor foi definido como ponto de corte para a classificação, ou seja, se a probabilidade estimada do paciente for maior que 0,38, diremos que o modelo o classificou como super obeso e se for menor ou igual a 0,38, foi considerado como obeso.

Assim, foi mensurado o número de indivíduos que foram corretamente classificados e os que foram classificados erroneamente, que são os valores que podem ser observados na Tabela A.11. Através desses números, foi possível calcular a especificidade (69,2%), que é o percentual de pacientes classificados pelo modelo como obesos e que de fato foram considerados como obesos pelo método CDC, e a sensibilidade (72,2%), que é o percentual de pacientes classificados como super obesos e que realmente foram considerados super obesos pelo método CDC. Com acerto total de 70,3%, o modelo classificou corretamente um percentual razoável da amostra total.

6.1.2 Método WHO

Foi feito um critério de classificação para testar quão bom está o acerto do modelo. Dado que o percentual de super obeso na amostra foi de 64%, este foi

definido como ponto de corte para a classificação, ou seja, se a probabilidade estimada do paciente for maior que 0,64, diremos que o modelo o classificou como super obeso e se for menor ou igual a 0,64, foi considerado como obeso.

Assim, foi mensurado o número de indivíduos que foram corretamente classificados e os que foram classificados erroneamente, esses valores podem ser observados na Tabela A.12. Através desses números, foi possível calcular a especificidade (66,2%), que é o percentual de pacientes classificados pelo modelo como obesos e que de fato foram considerados como obesos pelo método CDC, e a sensibilidade (63,7%), que é o percentual de pacientes classificados como super obesos e que realmente foram considerados super obesos pelo método CDC. Com acerto total de 64,6%, o modelo classificou corretamente um percentual razoável da amostra total.

6.1.3 Método Conde & Monteiro

Foi feito um critério de classificação para testar quão bom está o acerto do modelo. Dado que o percentual de super obesos na amostra foi de 47%, este foi definido como ponto de corte para a classificação, ou seja, se a probabilidade estimada do paciente for maior que 0,47, diremos que o modelo o classificou como super obeso e se for menor ou igual a 0,47, foi classificado como obeso.

Assim, foi mensurado o número de indivíduos que foram corretamente classificados e os que foram classificados erroneamente, que são os valores que podem ser observados na Tabela A.13. Através desses números, foi possível calcular a especificidade (76,8%), que é o percentual de pacientes classificados pelo modelo como obesos e que de fato foram considerados como obesos pelo método CDC, e a sensibilidade (66,3%), que é o percentual de pacientes classificados como super obesos e que realmente foram considerados super obesos pelo método CDC. Com acerto total de 72,4%, o modelo classificou corretamente um percentual razoável da amostra total.