



UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ZOOTECNIA E
ENGENHARIA DE ALIMENTOS
DEPARTAMENTO DE CIÊNCIAS BÁSICAS

**ESTUDO DE FERRAMENTAS COMPUTACIONAIS PARA ANÁLISE
ESTATÍSTICA E ELABORAÇÃO DE RELATÓRIOS TÉCNICOS**

Aluna: Ana Flávia Giacondino Soligo Lezcano Tatis
<http://lattes.cnpq.br/7975610974331143>

Orientadora: Profa. Dra. Giovana Fumes Ghantous
<http://lattes.cnpq.br/6884787667549337>

Relatório parcial referente ao Projeto FAPESP nº 2020/03228-6, bolsista de Treinamento Técnico Nível I, desenvolvido no período de abril/2020 a janeiro/2022.

Sumário

1	Introdução	3
2	Análise estatística em dados reais	4
2.1	Bancos de dados	4
2.1.1	Estudo nutricional em adultos da cidade do Rio de Janeiro, RJ, Brasil .	4
2.1.2	Estudo nutricional em idosos da cidade de Botucatu, SP, Brasil	5
2.2	Métodos	6
2.3	Resultados e Discussão	8
2.3.1	Descrição dos dados referentes ao estudo realizado em adultos na cidade do Rio de Janeiro, RJ, Brasil	8
2.3.2	Descrição dos dados referentes ao estudo realizado em idosos na cidade de Botucatu, SP, Brasil	16
3	Implementação computacional	32
4	Produção Científica	35
5	Considerações parciais	36

1 Introdução

A realização da descrição de um banco de dados é de extrema importância para iniciar a análise de uma pesquisa científica, pois é o primeiro passo para a organização e transformação dos dados brutos em informações que servirão de base para o estudo inferencial a ser desenvolvido. A estatística descritiva é feita por meio de técnicas gráficas (por exemplo, histogramas, *box-plot*) e numéricas (como as medidas de posição e de dispersão), que resumem os dados de maneira simples e podem ser aplicadas sobre cada variável de interesse separadamente, ou em pares de variáveis, quando se quer verificar uma possível associação entre elas (MORETTIN; BUSSAB, 2010).

É necessário que um pesquisador ou estudante saiba descrever um pequeno banco de dados (*Small Data*), que tenha conhecimento de como realizar cálculos das principais medidas resumo manualmente e também, que saiba como esboçar os gráficos, dado que, é desta maneira que ele aprende os conceitos, sua aplicação prática, e principalmente a interpretação de seus resultados. Porém, quando se trabalha com um conjunto real de dados, este, é geralmente numeroso, o que muitas vezes impossibilita cálculos manuais, como no caso dos chamados *Big Data*; assim, é necessário fazer uso de *softwares*, que não só calculam as medidas rapidamente, como permitem seu armazenamento, visualização e exploração; esses programas estão em constante aprimoramento e portanto seu estudo é contínuo (MORETTIN; SINGER, 2019).

Se, por um lado, programas estatísticos são de grande valia, dado que seu uso permite tanto que se calcule rapidamente todas as medidas necessárias de um extenso banco de dados, quanto que se gere gráficos para cada uma dessas variáveis, por outro lado, é necessário que o usuário do programa tenha aptidão para lidar com o mesmo, possuindo conhecimento suficiente sobre a linguagem de programação utilizada e sobre os métodos estatísticos de interesse. Tais ferramentas, permitem que se faça uma interpretação correta dos resultados que foram fornecidos, sendo imprescindível o auxílio de um estatístico (MORETTIN; SINGER, 2019).

Um aspecto interessante a ser considerado, é o uso de *softwares* estatísticos livres, como por exemplo, o *software R*, nos quais qualquer pessoa pode colaborar com o programa, através da introdução e publicação de pacotes que contém novas funções. Dessa forma, o desenvolvimento destes *softwares* ocorre devido a um esforço conjunto com contribuições provenientes do mundo todo. Entretanto, a programação de um algoritmo não é uma tarefa simples, pois para tal atividade, se faz necessária a criação de um encadeamento lógico de passos, executados a fim de se obter a solução de um problema. Com isso, uma sequência de etapas devem ser seguidas, passando pelas fases de planejamento, análise, projeto, implementação, teste e manutenção (ALMEIDA, 2008). Assim, é fundamental que os envolvidos tenham conhecimento suficiente para produzir funções que retornem resultados

corretos, sendo necessários estudos de simulação na fase de teste das funções (GARCIA, 2011).

Este projeto tem como objetivo a realização e interpretação de análises estatísticas em diferentes programas computacionais, bem como a aprendizagem de uma ferramenta diferenciada para elaboração de relatórios técnicos. Para atingir este objetivo, os *softwares R* (versão 4.0.3) e *SAS University Edition* foram utilizados para análises estatísticas (descritivas e inferenciais) de dados reais e implementação computacional, e a ferramenta *TeXStudio* (*MikTeX* versão 2.9.7351) foi estudada para elaboração de textos científicos.

Este trabalho está organizado em 5 seções. A seção 2 apresenta detalhes sobre os bancos de dados do projeto FAPESP nº 2019/02231-6; as medidas descritivas são exibidas em forma de tabelas e gráficos e os resultados são discutidos, além de um estudo inferencial sobre um dos bancos em questão. Na seção 3 são apresentados os estudos sobre a elaboração de pacotes no *software R* e também as contribuições feitas para o desenvolvimento dos algoritmos pertencentes ao projeto FAPESP nº 2019/02231-6. Na seção 4 são apresentadas as publicações derivadas do presente estudo. Por fim, na seção 5, estão expostas as considerações parciais sobre os objetivos da pesquisa alcançados até o momento presente.

2 Análise estatística em dados reais

2.1 Bancos de dados

Esta seção apresenta informações sobre os bancos de dados de duas cidades brasileiras, as respectivas técnicas de amostragem empregadas para sua obtenção e análises estatísticas descritivas e inferenciais. Os bancos citados foram utilizados no projeto FAPESP 2019/02231-6, intitulado “Detalhes e implementação dos modelos da classe Box-Cox simétrica e aplicações a dados de consumo alimentar”, desenvolvido pela Profa. Dra. Giovana Fumes Ghantous e colaboradores.

2.1.1 Estudo nutricional em adultos da cidade do Rio de Janeiro, RJ, Brasil

O primeiro banco de dados refere-se a uma investigação nutricional de 302 indivíduos, habitantes da cidade do Rio de Janeiro, cedido pelo professor Dr. Eliseu Verley Junior, do Departamento de Epidemiologia do Instituto de Medicina Social da UERJ – Rio de Janeiro. O objetivo do seu estudo foi a avaliação do consumo alimentar de micro e macronutrientes, além disso, coletaram-se dados demográficos e antropométricos.

Para a coleta dos dados fez-se uma amostragem do tipo bola de neve, na qual primeiramente, foram eleitos como entrevistadores, vinte e três graduandos do curso de nutrição, treinados para selecionar indivíduos dispostos a responder detalhes sobre o seu consumo alimentar durante 20 dias não consecutivos, resultando em 20 recordatórios por pessoa, em

períodos de 24 horas (R24h). Por meio do R24h, cada entrevistado relatou o seu consumo de alimentos nas 24 horas anteriores ao horário da entrevista.

Essa técnica de amostragem foi escolhida levando em consideração sua logística e seu custo para aplicação, visando manter a qualidade e a taxa de participação dos envolvidos, para isso, os selecionados preferencialmente eram de uma mesma rede social, sendo que cada entrevistador tinha a sua própria rede apesar de cursarem a mesma universidade, garantindo assim que os entrevistados estivessem espalhados por toda a cidade.

Os dados de consumo foram recolhidos entre março de 2013 e abril de 2014, cobrindo tanto dias de semana como finais de semana. Cada indivíduo foi acompanhado por aproximadamente 3 meses, seguindo a recomendação de que o R24h não deveria ser administrado nos mesmos dias da semana para o mesmo entrevistado.

As variáveis coletadas a partir deste estudo foram: gênero, nível de escolaridade, faixa etária, peso, altura, Índice de Massa Corporal (IMC) e quantidades ingeridas de quilocalorias, proteínas, lipídeos, carboidratos, fibras, cálcio, magnésio, fósforo, ferro, sódio, potássio, cobre, zinco, equivalente de folato dietético (DFE), gordura saturada e vitaminas (A, B1, B2, B6, B12, C e niacina).

2.1.2 Estudo nutricional em idosos da cidade de Botucatu, SP, Brasil

O segundo banco de dados pertence a um estudo epidemiológico realizado no ano de 2003 no município de Botucatu, localizado no interior do estado de São Paulo, o qual teve como objetivo avaliar a adequação do consumo alimentar de 368 idosos. A seleção do grupo de participantes foi feita através de uma amostragem aleatória, que abrangeu um quarto dos domicílios residenciais de cada setor censitário urbano da cidade de Botucatu.

Os moradores das residências sorteadas foram cadastrados e catalogados. A partir daí, novamente por meio de uma amostragem aleatória dentro de cada setor censitário, foram selecionadas pessoas com idade igual ou superior a 60 anos. Pesquisadores de campo treinados visitaram cada residência com o objetivo de verificar a presença dos idosos, assim como sua disposição em participar da pesquisa; aqueles que consentiram em fazer parte do estudo e que se encontravam em sua casa em até quatro visitas integraram o grupo.

Para estimar o tamanho amostral mínimo necessário utilizou-se a equação (VECCHIA et al., 2005).

$$n = \frac{z_{\frac{\alpha}{2}}^2 p(1-p)}{\epsilon^2},$$

em que n representa o número mínimo de indivíduos; z é igual a 1,96 (valor tabelado para um intervalo de confiança bilateral de 95% de uma distribuição normal); ϵ é o erro amostral, fixado em 0,05 e p é uma prevalência da característica de interesse, igual a 0,5. Desprezou-se o fator de correção da redução de heterogeneidade associada ao desenho de conglomerado,

e assim obteve-se o valor de 384 pessoas.

Para o presente estudo, os idosos foram contactados no ano de 2010 para a aplicação dos questionários de qualidade de vida e capacidade funcional. Aqueles que não concordaram em participar ou que já tinham falecido foram substituídos por um novo sorteio no banco de dados. Nessa época foram aplicados questionários sociodemográficos e de morbididades, a escala de Atividades de Vida Diária (AVD), escala de Atividades Instrumentais de Vida Diária (AIVD) e a escala de qualidade de vida de Flanagan (EQVF). Essa amostra foi coletada em 2011.

O instrumento novamente utilizado para a coleta de dados foi o recordatório 24 horas (R24h), considerando que o consumo é variado em diferentes dias da semana, aplicaram-se três recordatórios de 24h para cada indivíduo, sendo necessariamente um dos recordatórios aplicado na segunda-feira para registrar o consumo alimentar referente ao final de semana, e nenhum deles foi aplicado em dias sucessivos.

As variáveis coletadas foram: gênero, características socioeconômicas (trabalha/aposentado), morbididades (hipertensão, diabetes mellitus, doença cardíaca, disfunção na tireoide, colesterol alto, osteoporose, alzheimer, depressão) e prótese, que são variáveis qualitativas dicotômicas; nível de escolaridade e estado civil, que são variáveis qualitativas não dicotômicas e, por fim, idade, peso, altura, IMC, renda, quilocalorias, consumo total em gramas, proteína, proteína animal, proteína vegetal, carboidratos, fibras, cálcio, magnésio, fósforo, ferro, sódio, potássio, cobre, zinco, manganês, selênio, vitaminas (A, B1, B2, B3, ácido pantotênico, B6, B12, folato, C, D, E e K) gordura total, gordura saturada, gordura monossaturada, gordura polissaturada e colesterol, que são variáveis quantitativas contínuas. Adicionalmente, algumas proporções (proteína/carboidrato/gordura) e as respostas de questionários sobre capacidade funcional e qualidade de vida dos idosos foram coletadas. Os questionários aplicados continham perguntas referentes a: atividades da vida diária (AVD), atividades instrumentais da vida diária (AIVD) e a escala Flanagan de qualidade de vida, dados por meio de escalas numéricas. As perguntas de cada questionário estão descritas no anexo I.

2.2 Métodos

As mesmas técnicas estatísticas foram utilizadas para fazer a descrição dos dois bancos de dados. A fim de descrever as variáveis qualitativas foram construídas tabelas de frequências. Com o intuito de descrever as quantitativas contínuas, calcularam-se algumas medidas de posição (mínimo, média, mediana e máximo) e de dispersão (variância e desvio padrão), e construíram-se gráficos de perfil e *box-plot* ajustados (HUBERT; VANDERVIEREN, 2008).

Dentre as variáveis coletadas do primeiro banco de dados, as analisadas no estudo presente foram: nível de escolaridade e gênero, ambas qualitativas; quilocalorias, proteínas, lipídeos, carboidratos, fibras, cálcio, magnésio, fósforo, ferro, sódio, potássio, cobre, zinco,

vitamina A, vitamina B1, vitamina B2, niacina, vitamina B6, vitamina B12, vitamina C e gordura saturada, todas quantitativas contínuas.

Já para o segundo banco, dentre as variáveis coletadas, as analisadas no presente estudo foram: gênero, características socioeconômicas e morbidades (qualitativas dicotômicas) e nível de escolaridade (qualitativa não dicotômica); quilocalorias, proteína, proteína animal, proteína vegetal, carboidratos, fibras, cálcio, magnésio, fósforo, ferro, sódio, potássio, cobre, zinco, manganês, selênio, vitamina A, vitamina B1, vitamina B2, vitamina B3, ácido pantotênico, vitamina B6, vitamina B12, folato, vitamina C, vitamina D, vitamina E, vitamina K, gordura total, gordura saturada, gordura monosaturada, gordura polisaturada, colesterol, consumo total (todas quantitativas contínuas) e proporções de proteína, carboidrato e gordura.

Para avaliação sobre a capacidade funcional dos idosos, utilizou-se a escala de Atividades de Vida Diária (AVD), que contém seis questões e as respostas possíveis eram “sim” ou “não”; à cada resposta afirmativa se atribuiu um ponto, respostas negativas não receberam pontos, e utilizou-se a classificação: idosos que fizeram mais que 4 pontos foram considerados independentes, enquanto aqueles com 4 ou menos pontos foram classificados como dependentes (KATZ et al., 1963).

Outra escala usada para medir a capacidade funcional dos idosos foi a escala de Atividades Instrumentais de Vida Diária (AIVD), que possui nove questões e as opções de respostas eram: “não consegue”, “consegue com ajuda” e “consegue sem ajuda”; à cada pergunta respondida com “não consegue”, atribuiu-se 1 ponto, “consegue com ajuda”, 2 pontos e “consegue sem ajuda”, 3 pontos. Para a classificação foi feita uma soma dos pontos de cada indivíduo e utilizou-se a classificação: idosos que obtiveram mais que 17 pontos foram considerados independentes, enquanto aqueles com 17 ou menos pontos foram classificados como dependente (LAWTON; BRODY, 1969).

Para quantificar a qualidade de vida dos idosos a escala Flanagan foi escolhida, a qual contém quinze questões em que cada questão é avaliada através de uma escala *Likert* 7 pontos, variando de 1 (“muito insatisfeito”) até 7 (“muito satisfeito”). Como não existe ponto de corte específico para a classificação da qualidade de vida, para o presente estudo utilizou-se o valor da mediana, classificando valores acima da mediana como boa qualidade de vida, e abaixo como qualidade de vida ruim.

A cada variável resposta codificada dos questionários sobre capacidade funcional e qualidade de vida supracitados foi utilizada para verificar se tinha uma associação significativa com as variáveis: gênero, faixa etária, estado civil, renda, escolaridade, IMC, diabetes, hipertensão e doença cardíaca. Para realização deste estudo, algumas destas variáveis foram categorizadas ou recategorizadas, como: faixa etária - até 75 anos e acima de 75 anos; estado civil - casados e não casados (solteiro, viúvos, divorciados e separados); renda - até um salário mínimo e acima de um salário mínimo; escolaridade - até o ensino fundamental (en-

sino fundamental e analfabetos) e acima do ensino fundamental (ensino médio e superior); e IMC - normal (≤ 28) e excesso de peso (> 28). Para verificar a associação entre as respostas dos questionários e as variáveis descritas anteriormente um teste qui-quadrado foi utilizado (MORETTIN; BUSSAB, 2010). O nível de significância considerado para os testes foi de 5%.

Os gráficos e as medidas resumo foram realizados por meio do *software R* versão 3.6.2. Para construção dos gráficos de perfil, utilizou-se a função *xyplot()*, que pertence a biblioteca *lattice*; para a construção dos *box-plot* ajustados (HUBERT; VANDERVIEREN, 2008) foi necessária a instalação do pacote *robustbase*, que possui a função *adjbox()*. Para os testes de associação dos estudos sobre a qualidade de vida e a capacidade funcional dos idosos de Botucatu utilizou-se dois *softwares*, no *R* versão 3.6.2., a função *table* para criação de tabelas de contingência e a função *summary* para realização do teste qui-quadrado, no *SAS University Edition* utilizou-se a função *proc freq* associada aos comandos *tables* para criação de tabelas de contingência e *chisq* para realização dos testes qui-quadrado. A editoração deste relatório foi executada por meio do *software TeXStudio* (MikTeX versão 2.9.7351), um programa de edição de textos.

Foi criado um repositório na plataforma *GitHub* para armazenagem dos bancos de dados juntamente com os códigos elaborados nas linguagens *R* e *SAS* utilizados para análise descritiva dos mesmos e algumas análises inferenciais realizadas no segundo banco de dados. O repositório pode ser acessado através do endereço:

<https://github.com/AnaFTatis/Analise-descritiva-e-inferencial-dos-dados>.

2.3 Resultados e Discussão

Esta seção apresenta os resultados sobre a análise estatística realizada referente aos bancos de dados apresentados na seção 2.1, com ênfase nas medidas descritivas.

2.3.1 Descrição dos dados referentes ao estudo realizado em adultos na cidade do Rio de Janeiro, RJ, Brasil

O banco apresenta dados referentes a uma amostra composta por 302 adultos, moradores da cidade do Rio de Janeiro, RJ, Brasil. A tabela 1 apresenta a escolaridade dos indivíduos participantes do estudo.

Tabela 1: Nível de escolaridade de adultos, Rio de Janeiro, 2014

Escolaridade	n	%
Ensino fundamental	22	7,3
Ensino médio	74	24,5
Ensino superior	175	57,9
Não identificado	14	4,6
Dados faltantes	17	5,6
Total	302	100

Fonte: Autoria própria

Na tabela 1, constatou-se que a maioria dos participantes finalizaram ou estavam cursando o ensino superior, abrangendo 175 de 302 entrevistados, ou seja, 57,9%. Conforme descrito na seção 2.1, os entrevistadores eram graduandos do curso de nutrição, que selecionaram indivíduos dispostos a participar do ensaio, desse modo, era esperado que houvesse um grande número de pessoas com o mesmo grau de escolaridade dos entrevistadores.

As tabelas 2 e 3 apresentam medidas descritivas sobre o consumo de macro e micronutrientes dos indivíduos participantes do estudo da cidade do Rio de Janeiro. Primeiramente, a descrição do banco foi realizada com as 6040 observações coletadas (302 indivíduos com 20 recordatórios 24 horas cada). Para um refinamento sobre a coleta, os valores em relação a variável consumo de energia (kcal) foram examinados, encontrando-se 25 observações muito discrepantes. Após uma consulta com uma nutricionista, tais valores foram excluídos, por tratarem-se de valores não condizentes com a realidade esperada para valores de consumo energético. Após a retirada dessas 25 observações, um estudo comparativo foi realizado, no qual percebeu-se que as medidas descritivas não sofreram grandes modificações com a ausência das observações discrepantes. Isso pode ser observado, por exemplo, pelas medidas de posição central. Antes da retirada dos pontos, os valores médio e mediano do consumo de energia foram de 1963,6 kcal e 1803,4 kcal, respectivamente, e com a retirada dos pontos, passaram a ser de 1953,1 kcal e 1802,5 kcal; sendo assim concluiu-se que a retirada dos pontos não influenciou de forma significativa a distribuição geral dos dados. Este fato se deve especialmente ao tamanho amostral, o qual foi consideravelmente alto, assim a retirada de algumas observações, não ocasionou grandes mudanças nas medidas descritivas.

Tabela 2: Medidas descritivas do consumo de macronutrientes, Rio de Janeiro, 2014

Macronutrientes	mín	média	mediana	DP	máx
proteína (g)	0,7	89,7	77,9	52,4	561,6
energia (kcal)	329,5	1953,1	1802,5	807,7	5996,5
fibra total (g)	0,0	31,1	16,4	81,8	2671,5
carboidrato (g)	5,9	246,0	226,2	115,4	1399,3
gordura saturada (g)	0,0	24,8	21,5	15,2	180,9
lipídeo (g)	0,1	70,9	63,5	38,3	389,0

Fonte: Autoria própria

Tabela 3: Medidas descritivas do consumo de micronutrientes, Rio de Janeiro, 2014

Micronutrientes	mín	média	mediana	DP	máx
vitamina A (mcg)	0,0	642,8	396,0	1355,1	26140,6
vitamina B1 (mg)	0,00	1,29	1,14	0,77	17,10
Vitamina B2 (mg)	0,00	1,94	1,61	1,63	42,55
niacina (mg)	0,0	20,0	15,6	16,0	196,6
vitamina B6 (mg)	0,00	1,73	1,45	2,78	188,84
vitamina B12 (mcg)	0,0	9,0	3,9	23,5	570,1
vitamina C (mg)	0,0	120,2	33,4	578,1	15513,4
cálcio (mg)	14,8	714,1	620,8	478,8	5563,4
magnésio (mg)	0,0	238,4	215,0	127,6	1601,8
ferro (mg)	0,1	53,3	11,0	254,4	6852,6
zinco (mg)	0,2	15,1	10,4	69,0	4970,0
cobre (mcg)	0,00	1,35	0,99	2,09	39,62
fósforo (mg)	81,0	1184,5	1035,9	659,8	7123,3
sódio (mg)	29,4	2130,4	1750,4	1595,0	15297,0
potássio (mg)	19,1	2268,2	2093,1	1114,8	14141,4

Fonte: Autoria própria

Nas tabelas 2 e 3 foram observados valores distantes de média e mediana, o que aponta uma distribuição assimétrica dos dados. Esse comportamento é evidente em relação ao consumo de fibras na tabela 2 e de vitamina A e sódio na tabela 3. Em média, a energia e o carboidrato foram os macronutrientes mais consumidos pela população estudada, já os micronutrientes com maior ingestão foram sódio e potássio. Em geral, todos os micro e macronutrientes apresentaram valores altos de desvio padrão, e ao observar o coeficiente de variação¹ (CV) das variáveis, destacaram-se a fibra com maior variabilidade na tabela 2 e a vitamina C, o zinco e o ferro na tabela 3.

¹O coeficiente de variação em porcentagem é calculado através da seguinte equação: $CV = \frac{\text{desvio padrão}}{\text{média}} \times 100$.

As tabelas 4 e 5 apresentam as principais medidas descritivas referentes ao consumo de micro e macronutrientes por parte dos entrevistados de acordo com seu gênero. O banco foi composto majoritariamente por mulheres, sendo 203 contra apenas 82 homens e 17 observações faltantes. A separação por gênero é imprescindível, pois as recomendações nutricionais para pessoas do sexo feminino e masculino são distintas, logo, devem ser analisadas para cada grupo (ASSUMPÇÃO et al., 2017).

Tabela 4: Medidas descritivas do consumo de macronutrientes por gênero, Rio de Janeiro, 2014

Macronutrientes	Homens					Mulheres				
	mín	média	mediana	DP	máx	mín	média	mediana	DP	máx
proteína (g)	0,7	109,3	96,4	52,4	548,5	3,6	81,6	72,4	44,4	561,6
energia (kcal)	354	2355	2210	944	5832	329,5	1794	1681	670	59971
fibra total (g)	0,0	41,7	18,9	113,4	2671,5	0,6	26,5	15,5	64,1	1029,8
carboidrato (g)	17,0	300,2	276,5	145,1	1399,3	5,9	224,5	209,6	95,2	862,0
gordura saturada (g)	0,0	29,4	25,4	17,7	180,9	0,0	22,9	20,1	13,8	153,8
lipídeos (g)	0,2	84,5	75,5	44,5	389,0	0,1	65,6	58,9	34,5	351,5

Fonte: Autoria própria

Tabela 5: Medidas descritivas do consumo de micronutrientes por gênero, Rio de Janeiro, 2014

Micronutrientes	Homens					Mulheres				
	mín	média	mediana	DP	máx	mín	média	mediana	DP	máx
vitamina A (mcg)	0,0	689,7	406,3	1674,8	23732,7	0,0	617,7	386,8	1219,0	26140,6
vitamina B1 (mg)	0,00	1,53	1,36	0,94	17,10	0,07	1,19	1,06	0,65	12,95
Vitamina B2 (mg)	0,00	2,26	1,88	1,67	19,79	0,09	1,78	1,51	1,44	30,90
vitamina B6 (mg)	0,00	2,21	1,76	4,92	188,84	0,04	1,53	1,35	1,19	38,08
vitamina B12 (mcg)	0,0	11,1	4,4	26,1	276,8	0,0	7,8	3,6	19,2	546,5
vitamina C (mg)	0,0	161,9	37,4	751,9	15513,4	0,0	101,2	30,9	490,8	11648,2
niacina (mg)	0,0	24,9	18,5	21,5	196,6	0,2	18,0	14,6	12,8	161,8
cálcio (mg)	19,1	793,7	677,5	520,5	5085,3	14,8	675,4	587,1	458,7	5563,4
magnésio (mg)	0,0	275,0	250,3	139,3	1075,4	17,1	223,1	202,7	121,2	1601,8
ferro (mg)	1,1	75,9	13,6	326,3	6852,6	0,1	44,1	10,2	215,3	4421,4
zinco (mg)	0,2	16,2	12,2	25,1	726,7	0,5	14,5	9,8	82,4	4969,9
cobre (mcg)	0,00	1,63	1,16	2,67	31,46	0,01	1,24	0,93	1,81	39,62
fósforo (mg)	95,7	1367,3	1212,3	729,1	5299,4	81,0	1101,6	968,9	615,5	7123,3
sódio (mg)	38,3	2489,9	2093,7	1735,6	14624,6	29,4	1976,3	1611,7	1520,0	15297,0
potássio (mg)	19,1	2594,6	2400,2	1224,4	8669,9	190,8	2125,3	1966,4	1050,8	14141,4

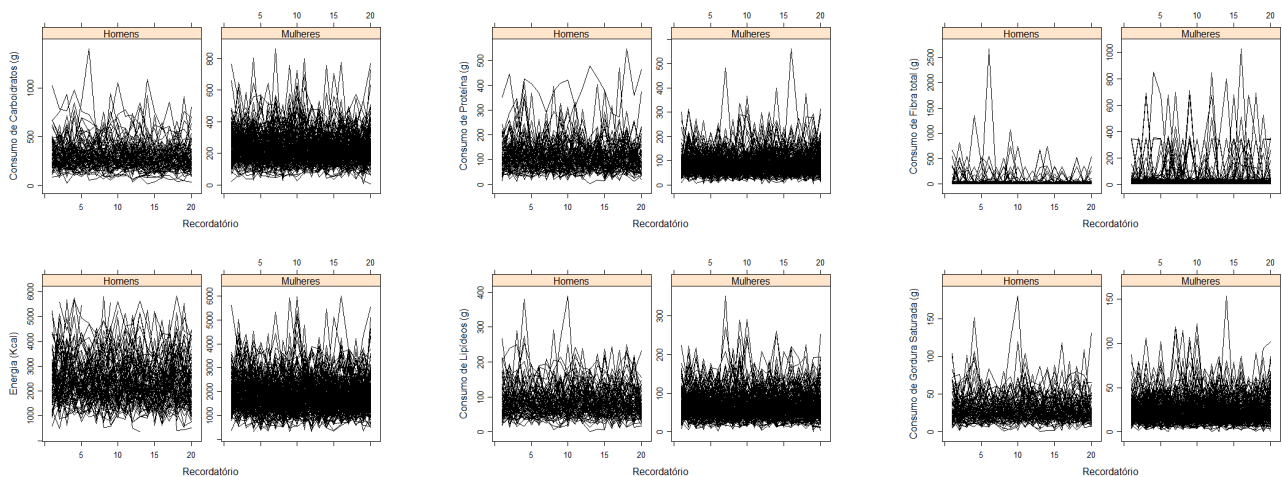
Fonte: Autoria própria

Analisando as tabelas 4 e 5, notou-se que as médias e medianas calculadas, relativas ao consumo dos nutrientes apresentados, resultaram em valores mais elevados para indivíduos do sexo masculino; para ambos, a maior amplitude de consumo foi na quantidade diária de quilocalorias. Interpretando os resultados encontrados para o desvio padrão na tabela 4, percebeu-se que essa medida também apresentou valores maiores para homens, entretanto, calculando os coeficientes de variação, foram encontrados valores próximos para os dois

gêneros, indicando uma variabilidade de consumo de macronutrientes em relação a média semelhante para homens e mulheres. Já na tabela 5 observou-se que a maior parte dos valores médios encontravam-se distantes dos valores medianos, destacando novamente as vitaminas A e C, o sódio e o potássio, como na tabela 3, sinalizando a existência de pontos aberrantes e de uma distribuição assimétrica. Tanto para os homens quanto para as mulheres, os valores mais altos de desvio padrão encontrados foram referentes ao consumo de sódio, porém, a partir do cálculo do coeficiente de variação, constatou-se que a dispersão dos dados em relação a média de consumo foi maior para a vitamina C ($CV = 464\%$) para entrevistados do sexo masculino e para o zinco ($CV = 567\%$), em relação aos entrevistados do sexo feminino.

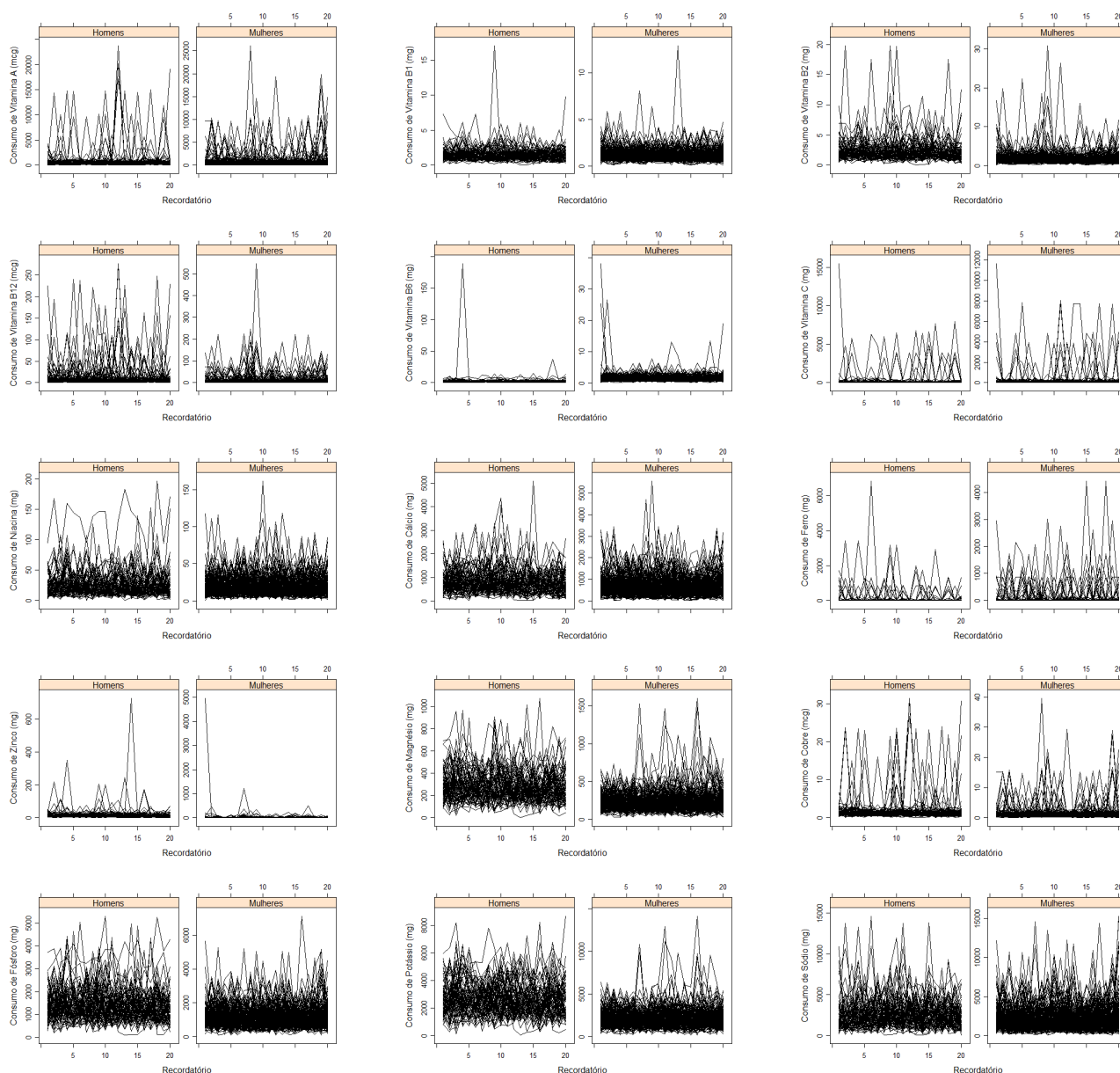
As figuras 1 e 2 apresentam gráficos de perfil construídos para os dois gêneros (feminino e masculino), que ilustram os dados de consumo de macronutrientes e micronutrientes ao longo dos 20 recordatórios respondidos pelos indivíduos do banco descrito no item 2.1.1.

Figura 1: Perfil de consumo de macronutrientes ao longo dos recordatórios por gênero, Rio de Janeiro, 2014



Fonte: Autoria própria

Figura 2: Perfil de consumo de micronutrientes ao longo dos recordatórios por gênero, Rio de Janeiro, 2014



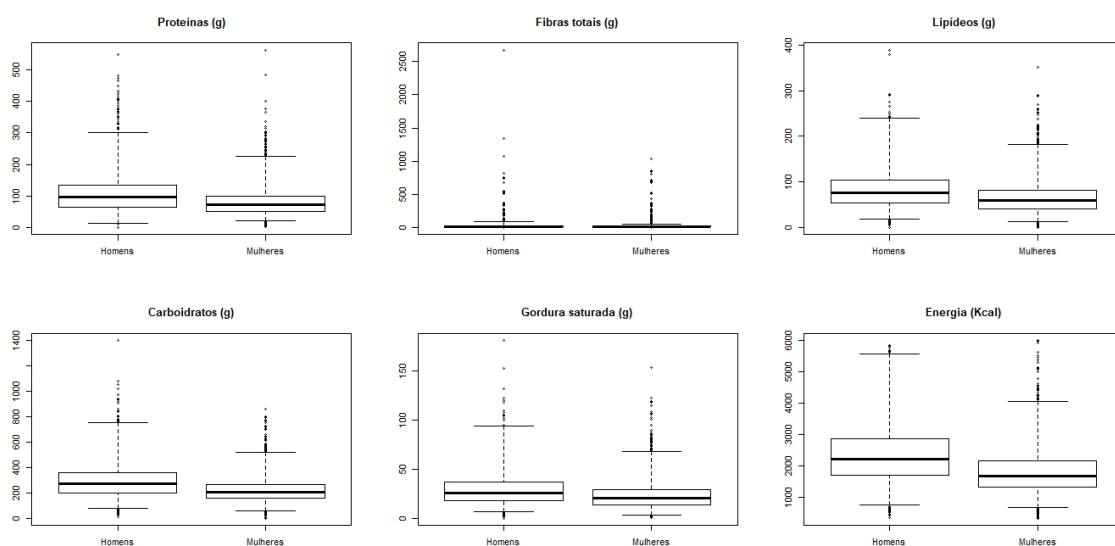
Fonte: Autoria própria

Nas figuras 1 e 2 foram apresentados os perfis de consumo ao longo dos recordatórios de macro e micronutrientes, respectivamente. Em geral, observou-se similaridade entre os perfis construídos para homens e mulheres e percebeu-se que ambos consumiram os nutrientes apresentados de maneira irregular durante os dias da semana. Na figura 1, os gráficos que ilustram o consumo total de fibras durante os recordatórios para homens e mulheres se mostraram os mais desiguais em relação a maneira como esses entrevistados ingerem os nutrientes.

Na figura 2, os menos ingeridos foram a vitamina B1 para ambos os sexos, com consumo médio entre 0 e 5 miligramas e máximo de 15 miligramas e a vitamina B12, que apresentou inconstância e concentração dos valores na parte inferior do gráfico, sendo 500 mcg (0,50 mg) e 250 mcg (0,25 mg) os valores máximos de consumo para mulheres e homens, respectivamente. Além disso, destacaram-se as baixas ingestões de cobre e zinco por parte dos entrevistados; no caso do zinco ocorreram picos de consumo no primeiro e no décimo quinto dia para mulheres e homens, respectivamente.

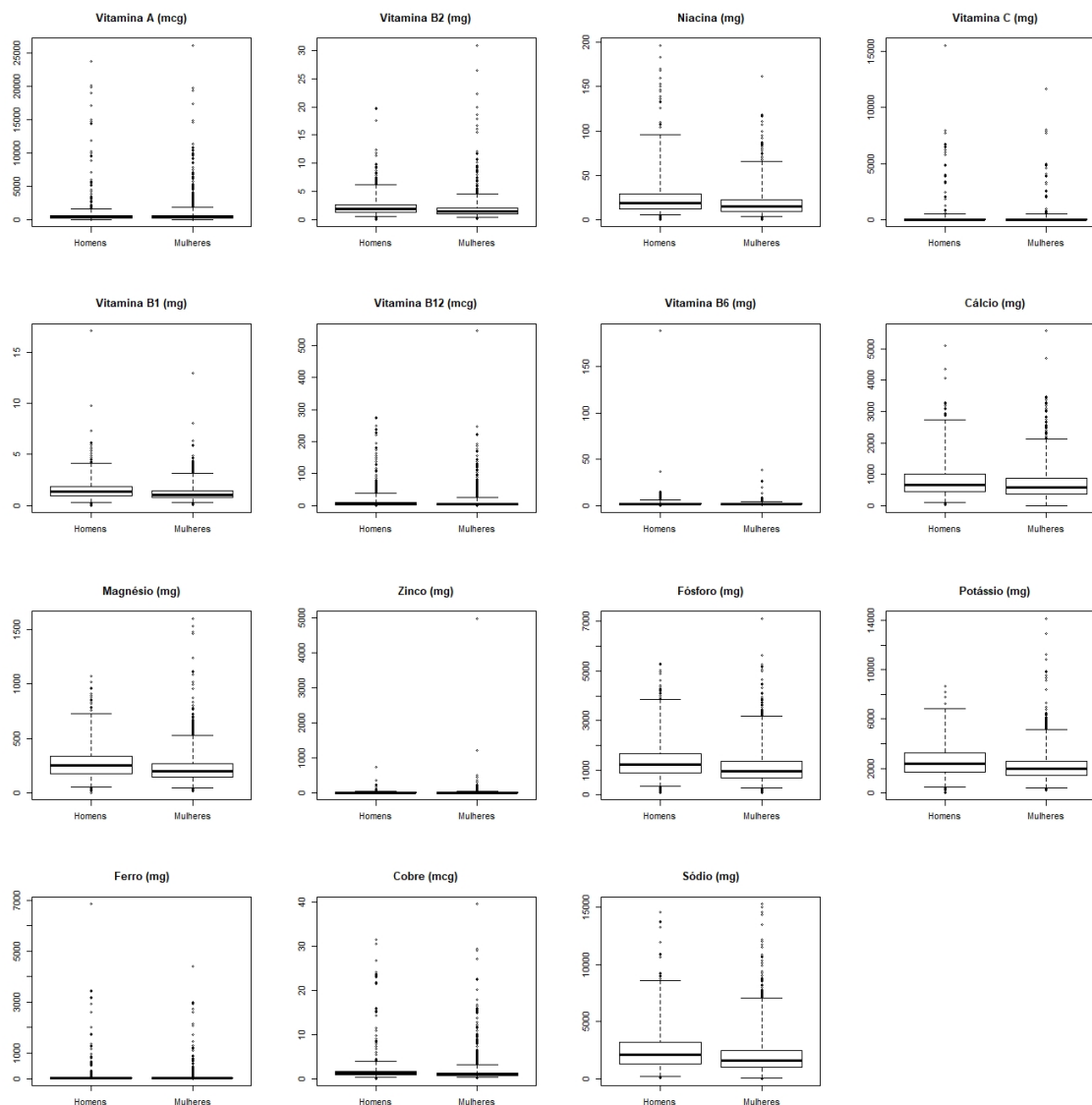
Nas figuras 3 e 4 são apresentados os *box-plot* ajustados (HUBERT; VANDERVIEREN, 2008), cujo objetivo de aplicação é melhorar a percepção dos pontos discrepantes. Os gráficos foram produzidos separadamente para homens e mulheres.

Figura 3: *Box-plot* ajustados referentes ao consumo de macronutrientes por gênero, Rio de Janeiro, 2014



Fonte: Autoria própria

Figura 4: *Box-plot* ajustados do consumo de micronutrientes por gênero, Rio de Janeiro, 2014



Fonte: Autoria própria

As figuras 3 e 4 destacaram a assimetria das distribuições referentes ao consumo de macronutrientes e micronutrientes, respectivamente. Evidenciou-se a presença de diversos pontos aberrantes como, por exemplo, para o consumo de fibras por parte dos entrevistados do sexo masculino na figura 3, para o qual o valor máximo foi de 2671,5 g, enquanto o valor médio (mediana) foi de apenas 18,9 g; e para o consumo de vitamina A por parte das mulheres da amostra na figura 4, para o qual o valor máximo foi de 26140,6 mcg, enquanto o valor da mediana foi de apenas 386,8 mcg. Além disso, observou-se a existência de assimetrias acentuadas à direita nas distribuições de consumo dos nutrientes.

2.3.2 Descrição dos dados referentes ao estudo realizado em idosos na cidade de Botucatu, SP, Brasil

O banco apresenta dados referentes a uma amostra composta por 368 idosos, moradores da cidade de Botucatu, SP, Brasil. A tabela 6 apresenta a escolaridade, a tabela 7 as características socioeconômicas, e a tabela 8 mostra as morbidades mais frequentes relatadas pelos idosos entrevistados. A idade dos entrevistados variou de 52 a 92 anos e o detalhamento sobre o banco se encontra descrito no item 2.1.2.

Tabela 6: Frequências de escolaridade de idosos, Botucatu, 2011

Escolaridade	n	%
Analfabeto	56	15,2
Ensino fundamental	228	62,0
Ensino médio	39	10,6
Ensino superior	45	12,2
Total	368	100

Fonte: Autoria própria

Na tabela 6 nota-se que diferentemente do que foi observado no primeiro banco, em que a maior frequência absoluta se encontrava na categoria ensino superior, neste caso, a maior parte dos entrevistados possui ensino fundamental, sendo 228 de 368, ou seja, 62%. Conforme visto anteriormente, a faixa etária dos entrevistados variou de 52-92 anos, portanto, essas frequências fazem sentido, pois foi apenas na década de 1940 que o Brasil começou a mencionar a educação primária obrigatória, como consta no Decreto-Lei n. 8529, de 2 de janeiro de 1946, e até hoje, apesar das melhorias no ensino, o país continua com uma profunda desigualdade social, o que impede que milhões de brasileiros tenham acesso a educação (RODRIGUES, 2016).

Tabela 7: Frequências das características socioeconômicas de idosos, Botucatu, 2011

Característica		n	%
Trabalha	Sim	39	10,6
	Não	329	89,4
Aposentado	Sim	320	87
	Não	48	13
Total		368	100

Fonte: Autoria própria

Na tabela 7 pôde-se observar que a maioria dos entrevistados é aposentado, sendo 87% do total, e como era de se esperar, levando em consideração a idade dos indivíduos que participaram deste estudo, além disso, a maioria não trabalha, abrangendo 89,4% do total.

Tabela 8: Frequências das morbidades relacionadas pelos idosos, Botucatu, 2011

Morbidades		n	%
Alzheimer	Sim	7	1,9
	Não	361	98,1
Colesterol alto	Sim	56	15,2
	Não	312	84,8
Depressão	Sim	7	1,9
	Não	361	98,1
Diabetes mellitus	Sim	105	28,5
	Não	263	71,5
Disfunção da tireoide	Sim	26	7,1
	Não	342	92,9
Doença cardíaca	Sim	27	7,3
	Não	341	92,7
Hipertensão	Sim	206	56
	Não	162	44
Osteoporose	Sim	33	9
	Não	335	91
Total		368	100

Fonte: Autoria própria

Na tabela 8 percebe-se que as morbidades mais frequentes na população de idosos examinada neste estudo foram a hipertensão, que afeta 56% da população, e a diabetes mellitus, que afeta 28,5%; as menos frequentes foram alzheimer e depressão, atingindo apenas 7 de 368 entrevistados (1,9%) cada.

As tabelas 9 e 10 apresentam as principais medidas descritivas do consumo de macro e micronutrientes dos entrevistados.

Tabela 9: Medidas descritivas do consumo de macronutrientes, Botucatu, 2011

Macronutrientes	mín	média	mediana	DP	máx
proteína (g)	11,4	86,8	80,1	43,8	473,5
proteína animal (g)	0,0	55,7	48,4	35,2	243,7
proteína vegetal (g)	2,5	31,1	27,3	20,9	381,4
fibras (g)	1,2	20,2	17,5	20,5	502,0
energia (kcal)	298,8	1929,0	1822,1	875,0	9547,7
carboidrato (g)	18,8	235,7	215,2	118,8	1797,7
gordura total (g)	7,7	70,8	62,3	42,3	402,6
gordura saturada (g)	1,3	23,0	19,8	15,0	121,4
gordura monosaturada (g)	1,6	24,0	20,8	15,7	163,3
gordura polisaturada (g)	1,5	17,7	14,8	13,0	161,8
colesterol (mg)	0,0	255,2	213,6	187,8	1636,9

Fonte: Autoria própria

Tabela 10: Medidas descritivas do consumo de micronutrientes, Botucatu, 2011

Micronutrientes	mín	média	mediana	DP	máx
vitamina A (mcg)	0,0	1178,0	736,0	2360,1	41372,0
vitamina C (mg)	0,0	144,0	54,6	567,0	10566,0
vitamina D (mcg)	0,0	4,1	3,5	3,1	27,4
Vitamina E (mg)	0,7	6,7	5,7	8,2	227,1
vitamina K (mcg)	3,6	212,8	102,5	1358,3	31036,0
vitamina B1 (mg)	0,2	1,6	1,4	1,0	16,9
vitamina B2 (mg)	0,2	1,6	1,5	0,8	10,8
vitamina B3 (mg)	1,8	21,5	18,5	14,5	258,8
ácido pantotênico (mg)	0,7	5,0	4,6	2,5	42,5
vitamina B6 (mg)	0,2	1,8	1,6	1,4	34,0
folato (mcg)	49,8	403,1	354,1	310,5	6518,1
vitamina B12 (mcg)	0,0	7,1	3,2	15,8	191,5
cálcio (mg)	39,3	680,2	589,9	444,1	6488,5
magnésio (mg)	46,0	270,9	245,5	203,4	4763,4
ferro (mg)	1,4	15,1	13,0	10,3	132,2
zinco (mg)	1,8	11,4	10,3	6,2	85,2
cobre (mg)	0,2	1,3	1,1	1,5	30,0
fósforo (mg)	184,0	1155,2	1056,9	645,3	11916,2
sódio (mg)	215,3	5767,2	3218,2	33173,1	237803,7
potássio (mg)	389,3	2775,5	2480,5	3451,4	99448,1
manganês (mg)	0,6	8,2	2,7	29,0	554,7
selênio (mcg)	18,9	132,4	112,2	150,2	3858,3

Fonte: Autoria própria

Nas tabelas 9 e 10, a partir da leitura dos valores médios notou-se que a energia e o carboidrato foram os macronutrientes mais consumidos pela população estudada, já os micronutrientes com maior consumo foram sódio e potássio. Em geral, todos os nutrientes apresentaram valores altos de desvios padrões, a ingestão de fibra apresentou grande variabilidade em torno da média ($CV = 102\%$), assim como as vitaminas C ($CV = 394\%$) e K ($CV = 638\%$). Como pôde-se observar na tabela 10, o sódio, o manganês e a vitamina A apresentaram valores de média distantes dos valores de mediana, em especial o sódio, com uma diferença de 2549,0 mg entre tais medidas de posição central.

Nas tabelas 11 e 12, as principais medidas descritivas de posição e dispersão (mínimo, média, mediana, máximo e desvio padrão) são apresentadas separadamente para homens e mulheres.

Tabela 11: Medidas descritivas do consumo de macronutrientes por gênero, Botucatu, 2011

Macronutrientes	Homens					Mulheres				
	mín	média	mediana	DP	máx	mín	média	mediana	DP	máx
proteína (g)	19,7	99,4	94,2	51,2	473,5	11,4	79,3	73,0	36,8	275,9
proteína animal (g)	1,5	63,4	57,0	39,3	243,7	0,0	51,1	45,4	31,6	225,6
proteína vegetal (g)	3,2	36,0	32,8	28,4	381,4	2,5	28,2	26,0	13,9	125,3
fibras (g)	1,2	22,6	18,8	30,7	502,0	2,2	18,7	16,6	10,3	89,1
energia (kcal)	542,1	2173,2	2044,3	1015,3	9547,7	298,8	1783,5	1663,5	743,1	8369,9
carboidrato (g)	62,0	257,2	238,5	148,4	1797,7	18,8	222,9	203,1	94,9	1021,6
gordura total (g)	15,5	79,9	71,7	45,4	311,6	7,7	65,5	57,0	39,4	402,6
gordura saturada (g)	3,0	25,8	22,8	16,3	121,4	1,3	21,3	18,5	13,9	108,4
gordura monosaturada (g)	4,2	27,5	25,1	17,5	163,3	1,6	21,9	18,8	14,1	125,0
gordura polissaturada (g)	1,5	19,6	17,2	12,9	113,6	2,3	16,6	13,8	13,0	161,8
colesterol (mg)	12,4	291,1	233,2	204,5	1330,6	0,0	233,8	201,2	173,7	1636,9

Fonte: Autoria própria

Tabela 12: Medidas descritivas do consumo de micronutrientes por gênero, Botucatu, 2011

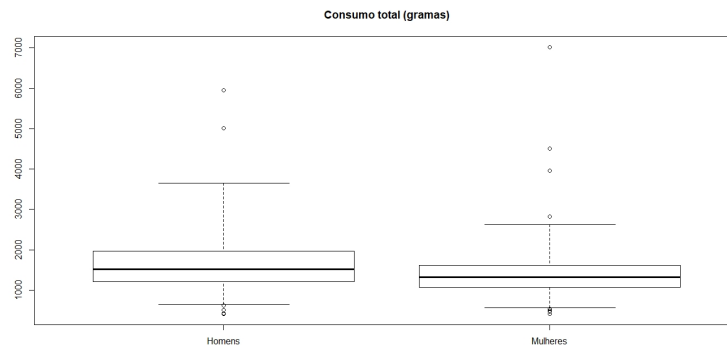
Micronutrientes	Homens					Mulheres				
	mín	média	mediana	DP	máx	mín	média	mediana	DP	máx
vitamina A (mcg)	1,9	1249,7	713,2	3163,4	41372,0	0,0	1135,3	756,9	1712,6	21011,8
vitamina C (mg)	0,0	209,4	54,7	856,6	10566,0	0,0	105,0	54,1	269,3	3931,5
vitamina D (mcg)	0,0	4,3	3,6	3,4	27,4	0,0	3,9	3,4	2,8	25,4
Vitamina E (mg)	0,7	7,5	6,1	12,6	227,1	0,9	6,1	5,5	3,6	37,0
vitamina K (mcg)	3,6	241,1	103,4	1656,9	31036,0	6,6	196,0	101,8	1145,1	23211,8
vitamina B1 (mg)	0,4	1,8	1,5	1,2	16,9	0,2	1,5	1,3	0,7	9,5
vitamina B2 (mg)	0,3	1,7	1,6	0,9	10,8	0,2	1,5	1,4	0,7	8,9
vitamina B3 (mg)	1,8	24,2	20,6	18,2	258,8	3,4	19,8	17,5	11,4	83,8
ácido pantotênico (mg)	1,0	5,4	5,0	3,1	42,5	0,7	4,7	4,4	2,0	18,9
vitamina B6 (mg)	0,3	2,1	1,8	2,1	34,0	0,2	1,7	1,5	0,8	6,5
folato (mcg)	76,1	462,9	399,8	427,5	6518,1	49,8	367,4	336,4	204,3	2685,5
vitamina B12 (mcg)	0,1	6,3	3,5	10,7	87,7	0,0	7,5	2,9	18,2	191,5
cálcio (mg)	59,3	738,4	637,7	555,7	6488,5	39,3	645,5	560,0	357,5	2408,8
magnésio (mg)	51,5	310,6	276,6	292,4	4763,4	46,0	247,2	227,2	116,9	942,2
ferro (mg)	1,4	17,1	15,1	12,4	132,2	2,1	13,9	12,0	8,7	97,3
zinco (mg)	2,3	13,1	12,0	7,4	85,2	1,8	10,4	9,6	5,1	44,8
cobre (mg)	0,2	1,3	1,2	1,5	25,0	0,2	1,2	1,0	1,6	30,0
fósforo (mg)	251,7	1295,1	1174,1	836,9	11916,2	184,0	1071,9	1011,8	478,9	3742,7
sódio (mg)	215,3	6243,0	3604,3	19880,4	237803,7	457,6	5483,6	3004,9	39011,6	156490,5
potássio (mg)	389,3	3249,4	2763,8	5444,2	99448,1	424,0	2493,0	2345,0	1079,4	9374
manganês (mg)	0,7	5,6	3,1	15,1	187,8	0,6	9,8	2,6	34,6	554,7
selênio (mcg)	24,5	141,9	126,0	77,9	602,4	18,9	126,8	102,6	179,8	3858,3

Fonte: Autoria própria

De modo geral, nas tabela 11 e 12 notaram-se valores mais altos de consumo de macro e micronutrientes por indivíduos do sexo masculino. Além disso, na tabela 11 para ambos os sexos, a maior amplitude de consumo foi na quantidade diária de quilocalorias. Ao observar os valores de desvio padrão, percebeu-se que essa medida também apresentou valores mais elevados para entrevistados do sexo masculino, porém, para os coeficientes de variação foram encontrados valores próximos para os dois gêneros, o que indica uma variabilidade de consumo de macro e micronutrientes em relação a média semelhante para homens e mulheres.

Para este estudo, foi realizada uma contagem do consumo total em gramas de cada entrevistado, e para análise dessa variável foram construídos os gráficos de perfil e *box-plot* ajustado (HUBERT; VANDERVIEREN, 2008), separados por gênero, apresentados nas figuras 5 e 6.

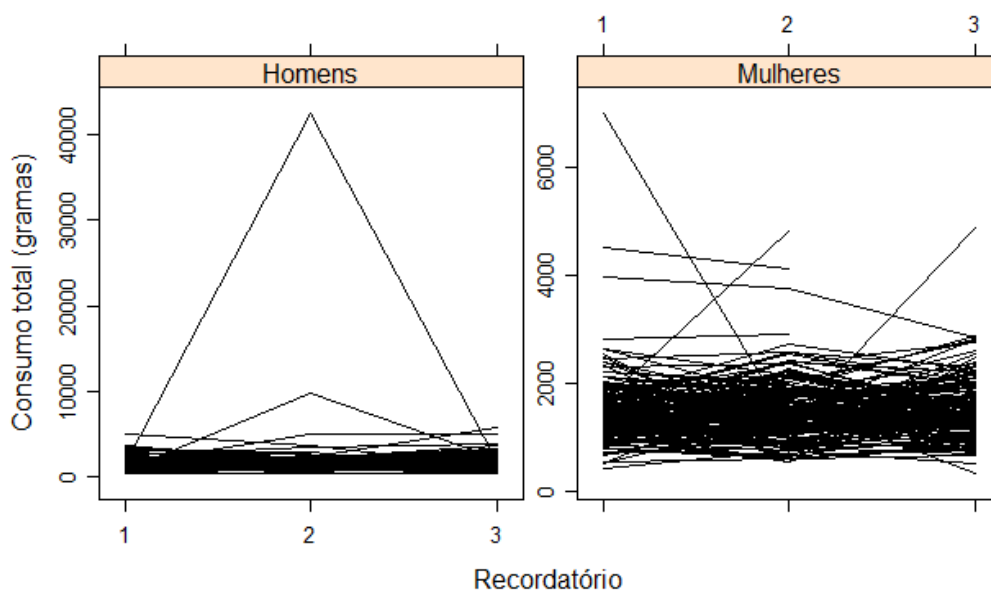
Figura 5: *Box-plot* ajustado do consumo total por gênero, Botucatu, 2011



Fonte: Autoria própria

Como observado nas tabelas anteriores, em geral, homens consumiram mais macro e micronutrientes que as mulheres, e ao observar os *box-plot* ajustados, ilustrado na figura 5, foi possível perceber que o consumo mediano dos homens foi superior ao das mulheres. Constatou-se a presença de pontos aberrantes; em destaque, para as mulheres, o valor máximo foi de 7018 g, enquanto o valor da mediana foi de apenas 1330 g.

Figura 6: Gráfico de perfil do consumo total por gênero, Botucatu, 2011

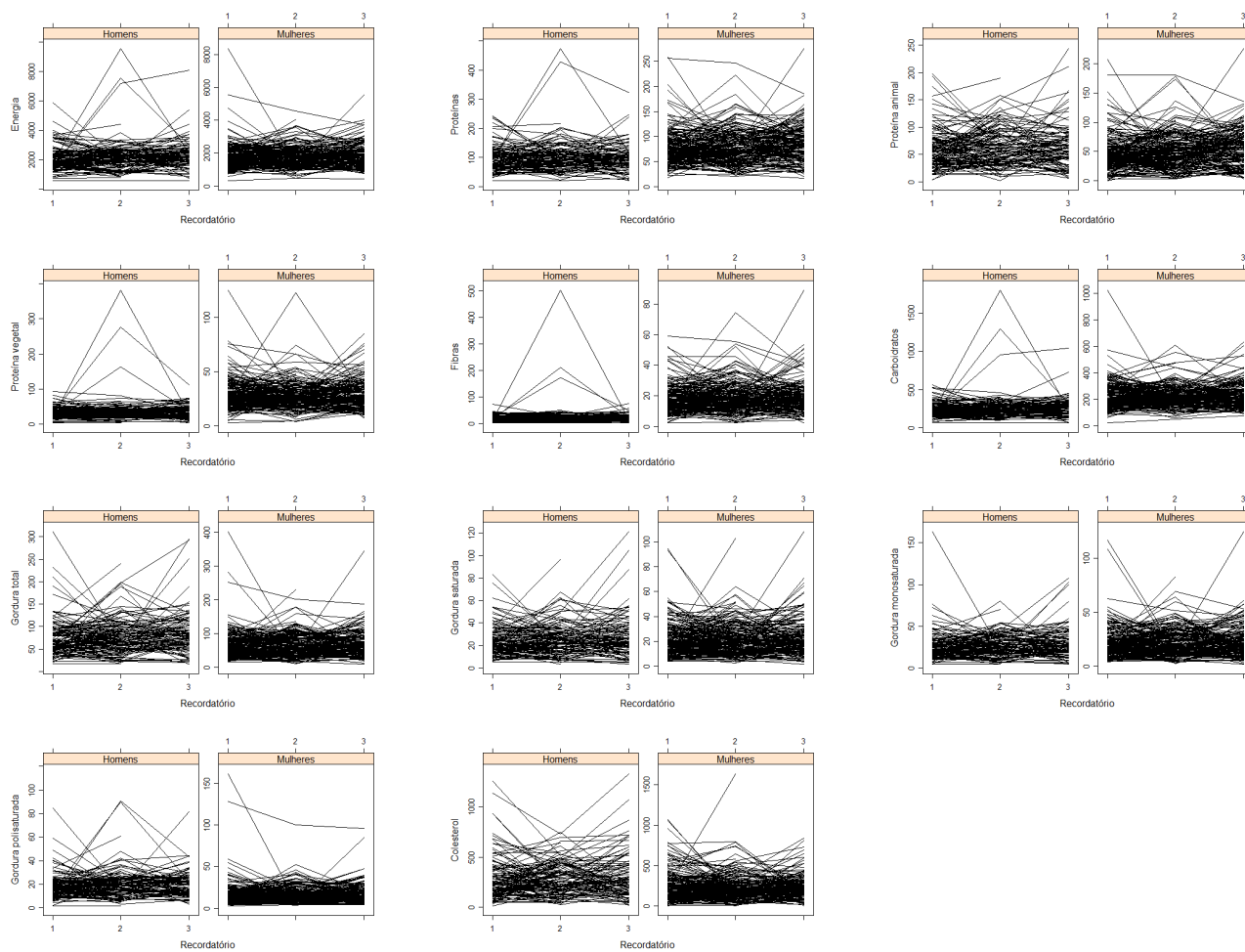


Fonte: Autoria própria

Na figura 6 também é possível perceber a existência de pontos discrepantes (*outliers*). Verifica-se que o dia em que foi aplicado o R24h influenciou nesse consumo total, sendo que para os homens o segundo dia foi o de maior consumo, o qual foi aplicado obrigatoriamente na segunda-feira, para que no estudo estivessem incluídos dados de consumo referentes ao final de semana. No caso das mulheres identificou-se um consumo, ao que parece, mais inconstante durante os dias.

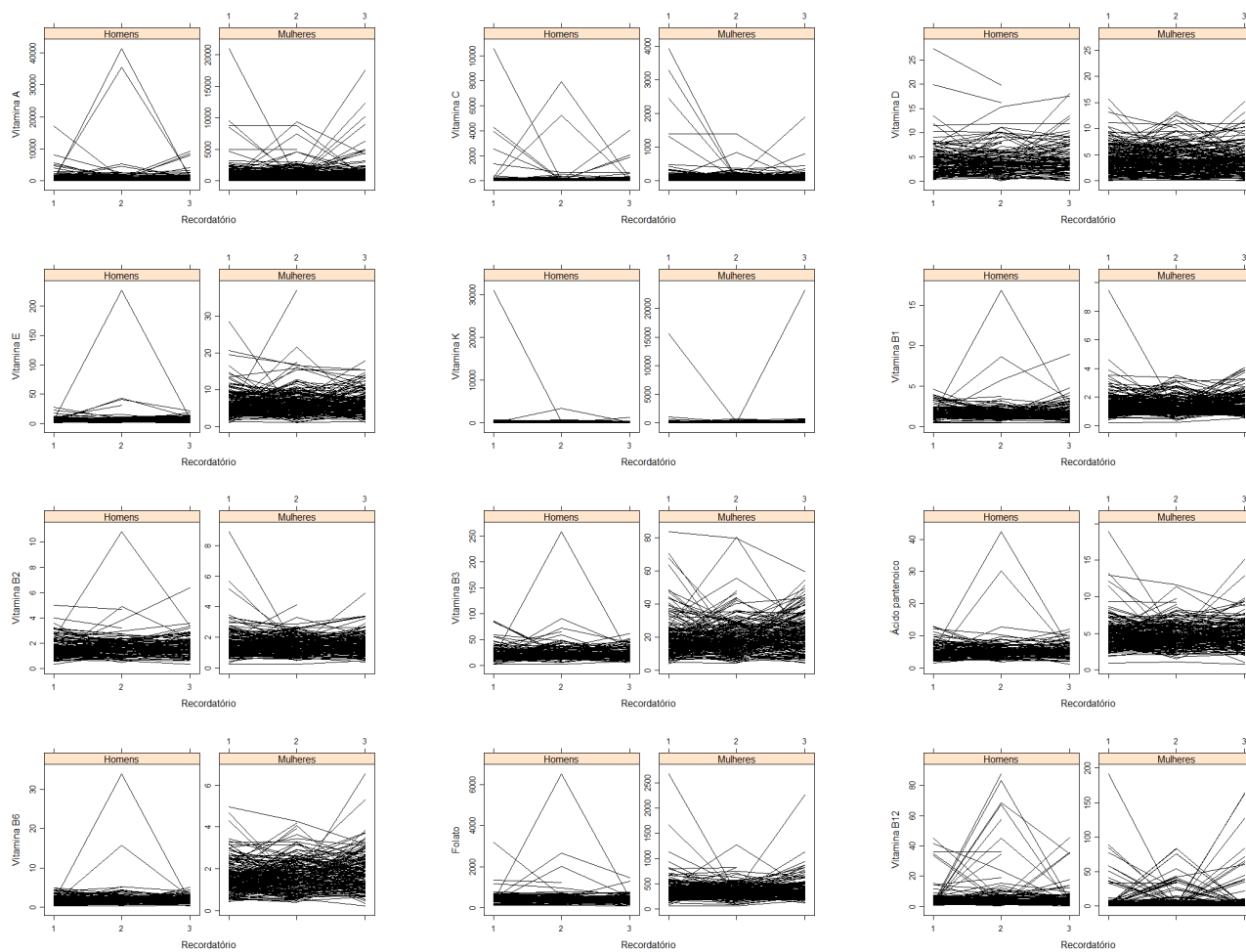
Na figura 7 foram apresentados os gráficos de perfil de consumo dos entrevistados ao longo dos recordatórios para macronutrientes, e nas figuras 8 e 9, apresentam-se os perfis de consumo para micronutrientes (vitaminas e outros) por parte da população estudada.

Figura 7: Perfil de consumo de macronutrientes por gênero, Botucatu, 2011



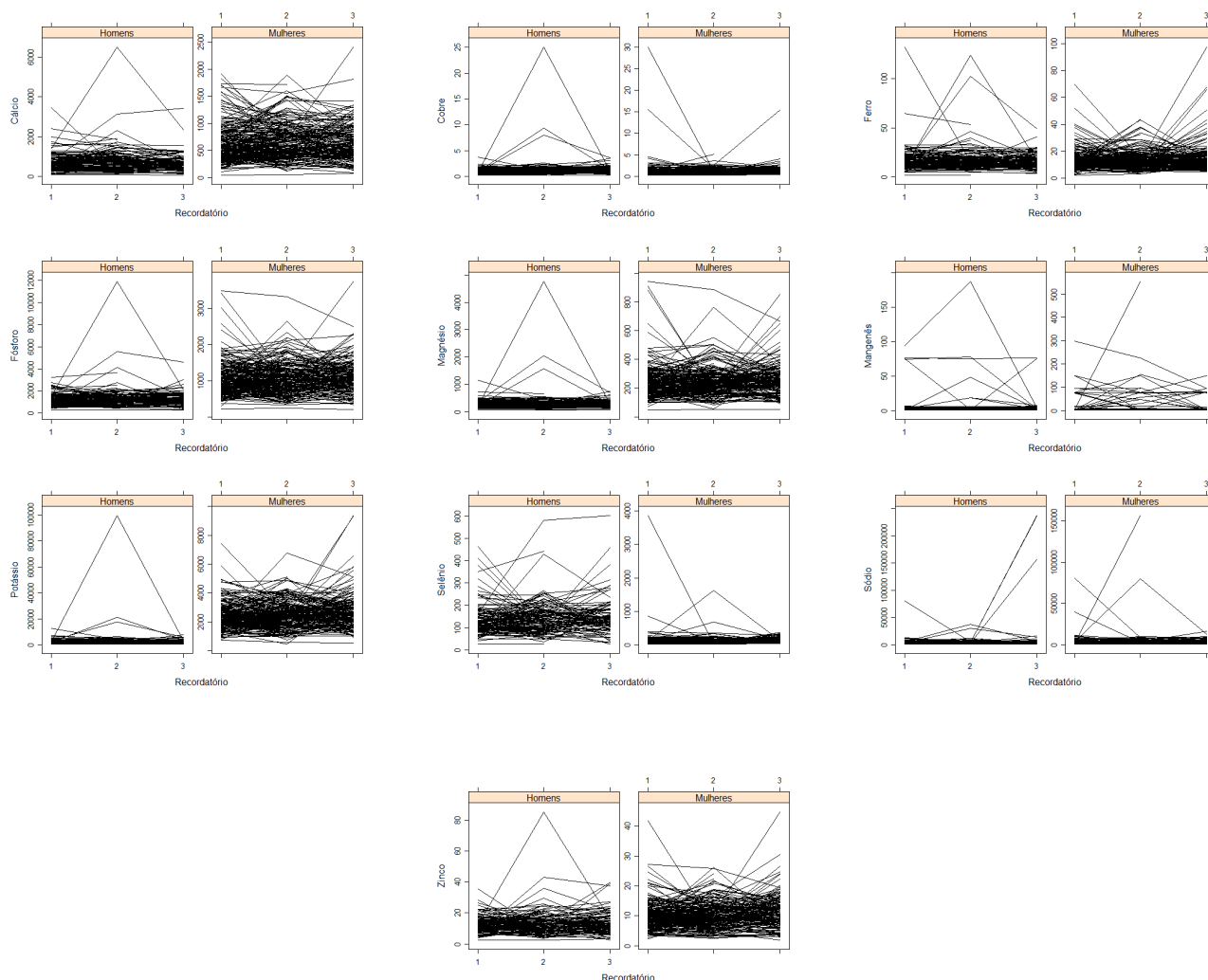
Fonte: Autoria própria

Figura 8: Perfil de consumo de vitaminas por gênero, Botucatu, 2011



Fonte: Autoria própria

Figura 9: Perfil de consumo de micronutrientes por gênero, Botucatu, 2011



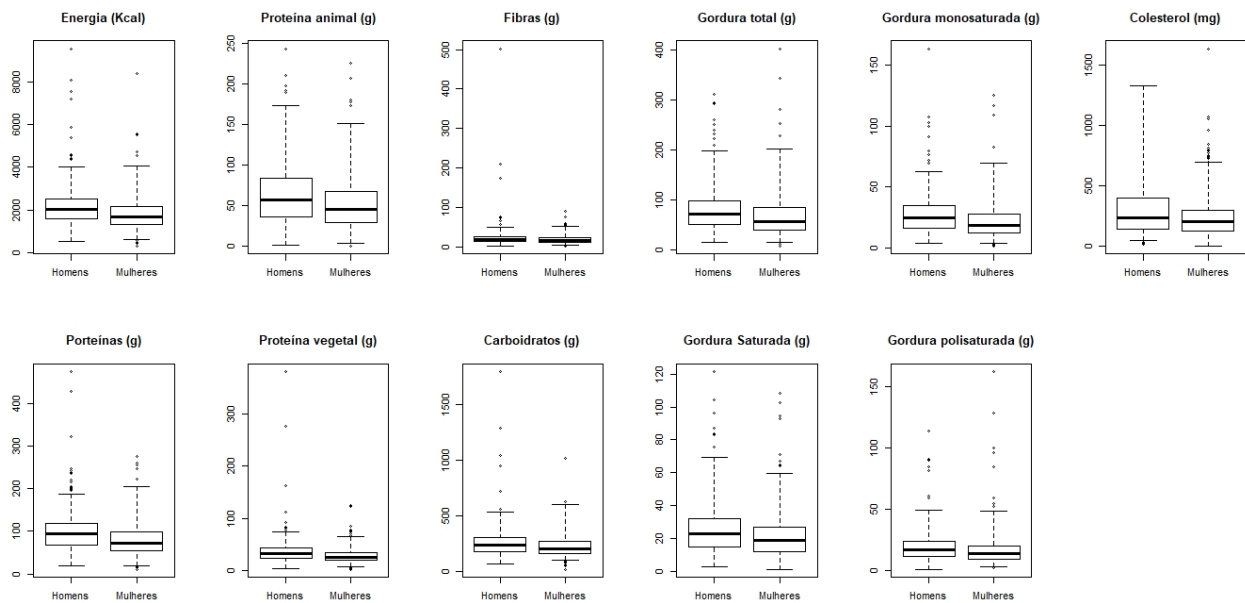
Fonte: Autoria própria

Nas figuras 7, 8 e 9 são apresentados os gráficos referentes ao consumo de macronutrientes, vitaminas e micronutrientes, respectivamente. Nessas figuras observou-se que, em geral, para o sexo masculino, no recordatório dois, o consumo foi mais elevado, sendo baixo nos outros dias, ou seja, no final de semana esses idosos ingerem mais alimentos. Já para as mulheres, não notou-se um padrão tão diferente de consumo de acordo com o dia de coleta dos dados, que aparentou ser mais irregular para todos os macronutrientes, com pontos discrepantes em todos os dias em que foram aplicados os recordatórios.

Com base na figura 8, aponta-se que o consumo de vitamina K para homens e mulheres parece ser o mais baixo, com maior número de dados em torno de zero. Para todos os micronutrientes pôde-se perceber valores discrepantes presentes nos gráficos construídos tanto para os homens, quanto para as mulheres.

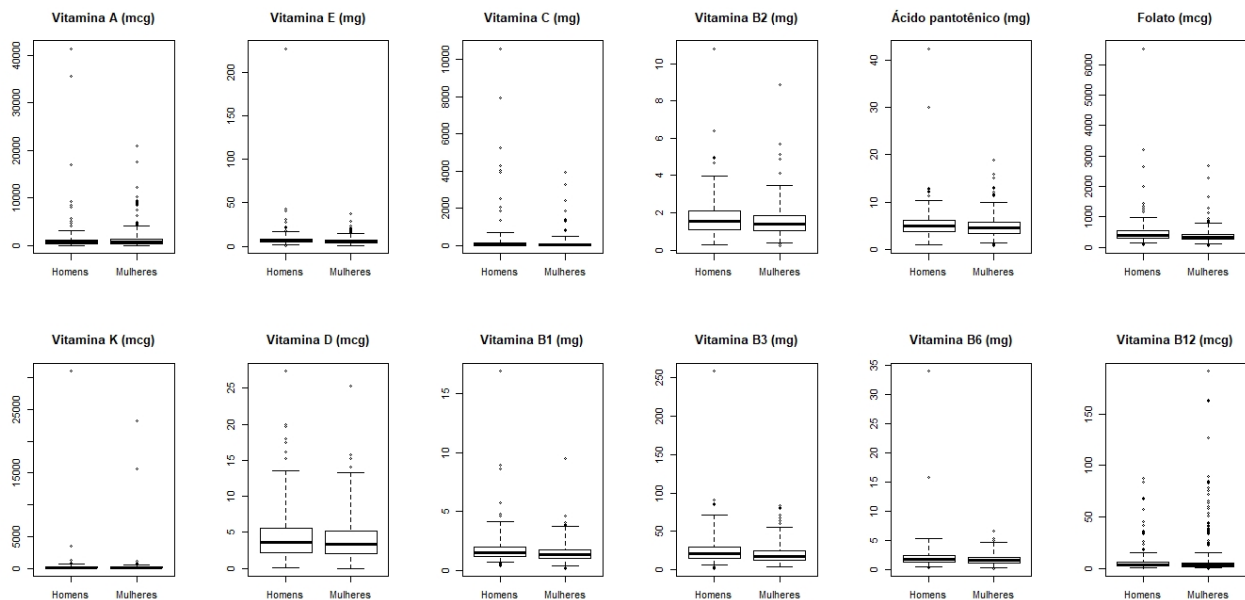
Na figura 10 são apresentados os *box-plot* ajustados (HUBERT; VANDERVIEREN, 2008) do consumo de macronutrientes; nas figuras 11 e 12 são apresentados os mesmos gráficos referentes ao consumo de micronutrientes.

Figura 10: *Box-plot* ajustados do consumo de macronutrientes por gênero, Botucatu, 2011



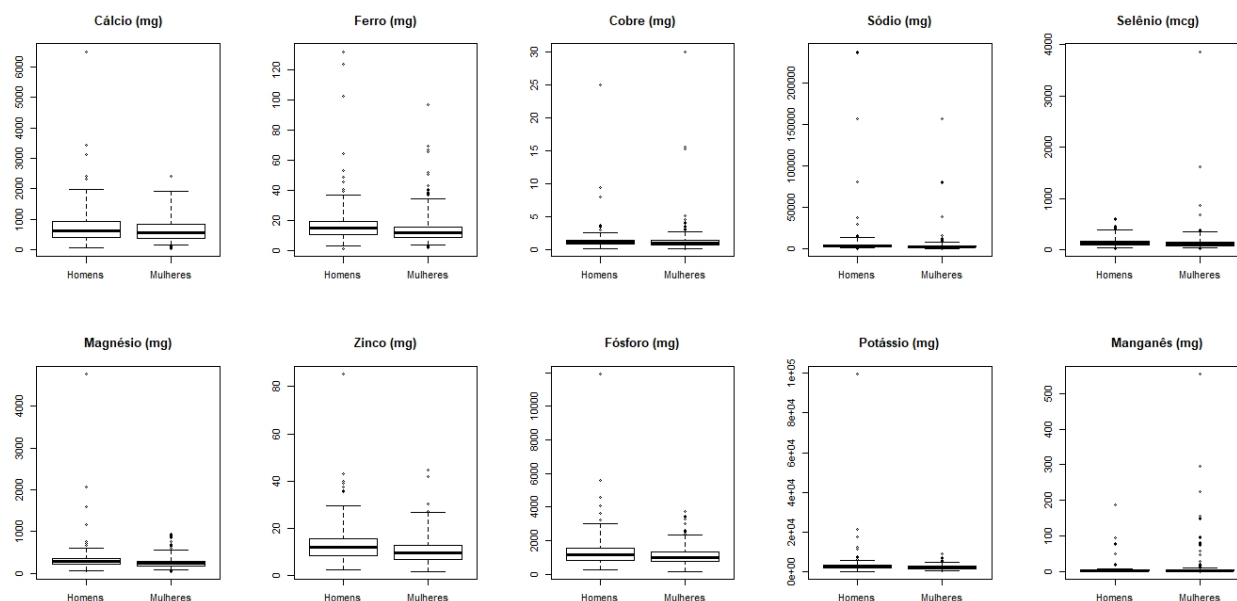
Fonte: Autoria própria

Figura 11: *Box-plot* ajustados do consumo de vitaminas por gênero, Botucatu, 2011



Fonte: Autoria própria

Figura 12: *Box-plot* ajustados do consumo de micronutrientes por gênero, Botucatu, 2011



Fonte: Autoria própria

Nas figuras 10, 11 e 12 observou-se uma assimetria das distribuições referentes ao consumo de macronutrientes, vitaminas e micronutrientes, respectivamente. Evidenciou-se a presença de diversos pontos aberrantes como, por exemplo, para o consumo de quilocalorias por parte dos homens na figura 10, para o qual o valor máximo foi de 9547,7 g, enquanto o valor da mediana foi de somente 2044,3 g, e para o consumo de vitamina A, novamente por entrevistados do sexo masculino, em que o valor máximo foi de 41372,0 mcg, enquanto o valor da mediana foi de apenas 713,2 mcg. Além disso, observou-se a existência de assimetrias acentuadas à direita nas distribuições de consumo desses nutrientes, sendo mais pronunciadas nos gráficos construídos para o consumo de fibras, vitaminas C e K, manganês, sódio e potássio.

Ademais, foi feita a descritiva para valores de proporção de gorduras, carboidratos e proteínas, conforme a tabela 13. De acordo com a literatura, as gorduras totais não devem exceder 30% do valor calórico, ficando entre 20% a 30%, as proteínas devem integrar de 10% a 15% desse valor e, os carboidratos de 50% a 60% (CARDOSO, 2004).

Tabela 13: Medidas de posição referentes a proporção de carboidratos, proteínas e gorduras, Botucatu, 2011

	Homens			Mulheres		
	mín	média	máx	mín	média	máx
carboidratos	16%	47%	76%	22%	50%	74%
proteínas	5%	19%	43%	7%	18%	45%
gorduras	9%	32%	58%	10%	31%	60%

Fonte: Autoria própria

A descrição foi feita separadamente para cada gênero como visto na tabela 13. Para os nutrientes em questão, observou-se que apesar de existirem valores mais altos e mais baixos do que os recomendados, em média, tanto homens quanto mulheres consumiram mais gorduras e proteínas que o indicado e, portanto, menos carboidratos (CARDOSO, 2004).

Questionários sobre Capacidade Funcional e Qualidade de Vida

Estudos sobre capacidade funcional e qualidade de vida dos idosos foram realizados, por meio de questionários de Atividades de Vida Diária (AVD) e Atividades Instrumentais de Vida Diária (AIVD) e a Escala de Qualidade de Vida de Flanagan (EQFV), que tem como objetivo mensurar a qualidade de vida dos idosos a partir de tópicos que abordam sua percepção sobre sua própria saúde e relacionamento com amigos e parentes.

Dos 368 participantes do estudo epidemiológico, 38 não responderam aos questionários sobre capacidade funcional e 34 não responderam ao questionário sobre qualidade de vida. Tem-se ainda que, dos 334 idosos que responderam ao questionário sobre satisfação com a vida, 167 tiveram sua qualidade de vida identificada como “boa” e os demais como “ruim”.

As tabelas 14 e 15 apresentam os resultados obtidos de testes qui-quadrado aplicados para verificar a associação entre as atividades de vida diária (AVD) e as atividades instrumentais de vida diária, respectivamente, às variáveis gênero, faixa etária, estado civil, renda, escolaridade, IMC, diabetes, hipertensão e doença cardíaca. Já na tabela 16, os resultados do estudo de associação entre a escala de qualidade de vida de Flanagan (EQFV) e às covariáveis anteriormente explicitadas são apresentados.

Tabela 14: Associação entre variáveis socioeconômicas, algumas morbidades e as atividades de vida diária, Botucatu, 2011

Variáveis	Categorias	Classificação AVD n (%)		p-valor
		Dependentes	Independentes	
Gênero	Masculino	4 (21,05)	121 (38,91)	0,1194
	Feminino	15 (78,95)	190 (61,09)	
Faixa etária	Até 75 anos	8 (42,11)	222 (71,38)	0,0070
	Acima de 75 anos	11 (57,89)	89 (28,62)	
Estado civil	Casados	4 (21,05)	188 (60,45)	0,0007
	Não Casados	15 (78,95)	123 (39,55)	
Renda	Até um salário mínimo	9 (64,29)	99 (36,13)	0,0338
	Acima de um salário mínimo	5 (35,71)	175 (63,87)	
Escolaridade	Até Ensino fundamental	17 (89,47)	237 (76,21)	0,2634
	Acima Ensino fundamental	2 (10,53)	74 (23,79)	
IMC	Normal (IMC \leq 28)	13 (68,42)	170 (54,66)	0,2414
	Excesso de peso (IMC $>$ 28)	6 (31,58)	141 (45,34)	
Diabetes	Sim	7 (36,84)	86 (27,65)	0,3874
	Não	12 (63,16)	225 (72,35)	
Hipertensão	Sim	14 (73,68)	168 (54,02)	0,0943
	Não	5 (26,32)	143 (45,98)	
Doença cardíaca	Sim	3 (15,79)	22 (7,07)	0,1656
	Não	16 (84,21)	289 (92,93)	

Fonte: Autoria própria

Tabela 15: Associação entre variáveis socioeconômicas, algumas morbidades e as atividades instrumentais de vida diária, Botucatu, 2011

Variáveis	Categorias	Classificação AIVD n (%)		p-valor
		Dependentes	Independentes	
Gênero	Masculino	9 (34,62)	116 (38,16)	0,7208
	Feminino	17 (65,38)	188 (61,84)	
Faixa etária	Até 75 anos	10 (38,46)	220 (72,37)	0,0003
	Acima de 75 anos	16 (61,54)	84 (27,63)	
Estado civil	Casados	9 (34,62)	183 (60,20)	0,0111
	Não Casados	17 (65,38)	121 (39,80)	
Renda	Até um salário mínimo	14 (66,67)	94 (35,21)	0,0041
	Acima de um salário mínimo	7 (33,33)	173 (64,79)	
Escolaridade	Até Ensino fundamental	24 (92,31)	230 (75,66)	0,0529
	Acima Ensino fundamental	2 (7,69)	74 (24,34)	
IMC	Normal (IMC \leq 28)	21 (80,77)	162 (53,29)	0,0068
	Excesso de peso (IMC $>$ 28)	5 (19,23)	142 (46,71)	
Diabetes	Sim	11 (42,31)	82 (26,97)	0,0953
	Não	15 (57,69)	222 (73,03)	
Hipertensão	Sim	15 (57,69)	167 (54,93)	0,7861
	Não	11 (42,31)	137 (45,07)	
Doença cardíaca	Sim	7 (26,92)	18 (5,92)	0,0001
	Não	19 (73,08)	286 (94,08)	

Fonte: Autoria própria

Tabela 16: Associação entre variáveis socioeconômicas, algumas morbidades e a qualidade de vida dos idosos, Botucatu, 2011

Variáveis	Categorias	Classificação Flanagan n (%)		p-valor
		Boa	Ruim	
Gênero	Masculino	73 (43,71)	54 (32,34)	0,0322
	Feminino	94 (56,29)	113 (67,66)	
Faixa etária	Até 75 anos	124 (74,25)	105 (62,87)	0,0251
	Acima de 75 anos	43 (25,75)	62 (37,13)	
Estado civil	Casados	122 (73,05)	72 (43,11)	<0.0001
	Não Casados	45 (26,95)	95 (56,89)	
Renda	Até um salário mínimo	38 (26,95)	68 (44,74)	0,0015
	Acima de um salário mínimo	103 (73,05)	84 (55,26)	
Escolaridade	Até Ensino fundamental	118 (70,66)	135 (80,84)	0,0300
	Acima Ensino fundamental	49 (29,34)	32 (19,16)	
IMC	Normal (IMC \leq 28)	96 (57,49)	90 (53,89)	0,5087
	Excesso de peso (IMC $>$ 28)	71 (42,51)	77 (46,11)	
Diabetes	Sim	41 (24,55)	54 (32,34)	0,1149
	Não	126 (75,45)	113 (67,66)	
Hipertensão	Sim	81 (48,50)	104 (62,28)	0,0113
	Não	86 (51,50)	63 (37,72)	
Doença cardíaca	Sim	8 (4,79)	16 (9,48)	0,0901
	Não	159 (95,21)	151 (90,42)	

Fonte: Autoria própria

Observa-se nas tabelas 14 e 15 que a renda, a faixa etária e o estado civil associaram-se de maneira significativa à classificação da capacidade funcional avaliada tanto com as atividades de vida diária, quanto com as atividades instrumentais de vida diária ($p < 0,05$). Estudos sugerem que os homens mais jovens são mais independentes em relação às mulheres, que são impostas a realizar as tarefas de casa, entretanto, na velhice constata-se uma perda de poder para os homens, refletindo na sua capacidade funcional, o que ocorre de maneira contrária para as mulheres, pois ao passar dos anos, aparentemente conquistam certa liberdade e independência (FIGUEIREDO et al., 2007).

A faixa etária também apresentou associação com as classificações para a escala de qualidade de vida de Flanagan. Diversos fatores contribuíram para que a expectativa de vida de um brasileiro aumentasse de 42,7 anos, no ano de 1940, para 75,7 anos em 2016 (PEREIRA, 2019); segundo o IBGE, o número de idosos no país ultrapassa 28 milhões de indivíduos, podendo dobrar nas próximas décadas. Quanto a renda, pessoas que recebem até um salário mínimo sofrem maior exposição a fatores de risco, e dispõem de menos acesso a serviços de saúde que aquelas que ganham mais e, portanto, isso pode ter causado essa diferença em relação a sua autonomia ao realizar as tarefas de auto-cuidado (SILVA; TOMAZ, 2017).

Conclui-se ainda que o estado civil dos idosos está associado com a sua capacidade de realização das atividades de vida diária e das atividades instrumentais de vida diária. Acredita-se que o suporte associado às relações conjugais, possa favorecer um estilo de vida

mais saudável, melhorando sua capacidade funcional e consequentemente sua qualidade de vida (PEREIRA, 2019)

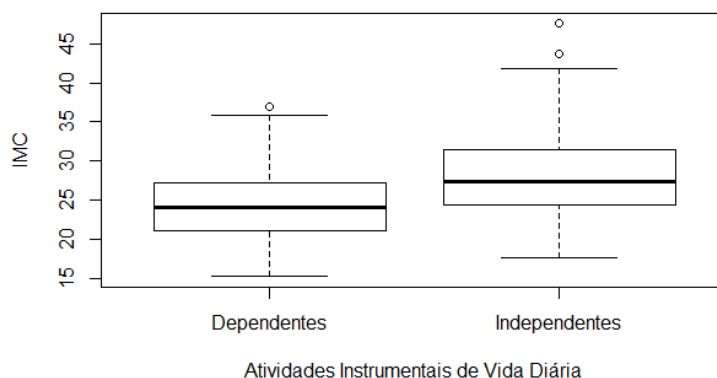
Adicionalmente, as variáveis escolaridade, IMC, gênero, diabetes, hipertensão e doença cardíaca não mostraram associação com a autonomia dos indivíduos (vide tabela 14), e as variáveis, gênero, diabetes e hipertensão não apresentaram associação com as atividades do cotidiano realizadas por estes idosos (conforme tabela 15).

Conforme visto na tabela 8, a hipertensão e a diabetes mellitus foram as doenças mais frequentes nessa população. Embora a hipertensão arterial seja uma das mais importantes causas de óbito prematuro e aposentadoria precoce, não encontrou-se relação entre essa morbidade e a realização de atividades de vida diária (PEREIRA, 2019). O mesmo foi observado para a diabetes mellitus, que também não apresentou associação com as classificações para qualidade de vida, porém, esse resultado difere do que é encontrado em estudos similares. Segundo Schmidt et al. (2009), no Brasil tem-se uma fração não diagnosticada de casos de diabetes e hipertensão, pois tais condições clínicas podem ser assintomáticas e, além disso, o acesso à serviços de saúde e a disponibilidade de exames é diferente para pessoas de classes sociais distintas.

A hipertensão não apontou associação com a capacidade funcional dos idosos, porém, apresentou impacto na qualidade de vida desses indivíduos, como foi dito previamente, essa é uma morbidade comum no Brasil, responsável por uma parte relevante da mortalidade prematura, aposentadoria precoce e faltas no trabalho. Além disso, pesquisas demonstraram que outras variáveis abordadas neste estudo, como faixa etária e gênero, por exemplo, são fatores de risco associados à hipertensão (AMADO; ARRUDA, 2004).

O resultado do p-valor para associação entre as atividades instrumentais de vida diária e o nível de escolaridade foi de 0,0529, um pouco acima de 5%, o que aponta uma certa associação. Alguns estudos propõe que a baixa escolaridade determina condições desfavoráveis de trabalho, menores rendas e dificuldade no acesso à saúde, impactando negativamente a capacidade funcional dos idosos e como consequência determinando uma piora na qualidade de vida (MORAIS, 2019)

Figura 13: *Box-plot* referente a relação entre as variáveis AIVD e IMC, Botucatu, 2011



Fonte: Autoria própria

O Índice de Massa Corporal (IMC) também foi considerado significativo ($p\text{-valor} < 0,05$), ou seja, existe uma relação entre a autonomia dos idosos segundo as AIVD e estar acima do peso ou estar no peso “normal”. Diversos estudos citam o excesso de peso como um fator de risco para a população de idosos, porém, como visto na figura 13, indivíduos dessa amostra que se encontravam nessa condição, mostraram ser mais independentes (PEREIRA, 2019). Os malefícios do excesso de peso são conhecidos e amplamente explorados pela mídia, a fim de conscientizar a população, entretanto existe um termo chamado “paradoxo da obesidade”, que está relacionado com os benefícios e fatores de proteção associados a essa condição.

Segundo um estudo de 2019, idosos com maiores valores de IMC e circunferência abdominal apresentaram maior desempenho muscular, porém, nesse mesmo estudo valores elevados de IMC resultaram em um desempenho negativo das atividades instrumentais de vida diária (MORAIS, 2019). Esse é um assunto controverso e que merece pesquisas mais aprofundadas, todavia, uma possível teoria para explicação do resultado deste estudo é de que os riscos estão relacionados apenas a valores extremos de IMC, de 35 a 45, que foram minoria nessa amostra (SANTOS, 2017). Na tabela 16 estar acima do peso ou não, não apresentou associação com a qualidade de vida dos idosos.

O gênero apresentou significância em relação a escala de Flanagan ($p\text{-valor} = 0,0322$). Alguns autores sugerem que homens e mulheres não são igualmente vulneráveis a situações experimentadas por idosos no país, como problemas de saúde, solidão, transtornos emocionais, situação sócio-econômica desvantajosa e falta de perspectiva profissional (LIMA; BUENO, 2009). Segundo Moraes (2019), as mulheres vivem mais que os homens, porém, estão mais expostas à violência doméstica física e psicológica e à discriminação no acesso à saúde, educação, condições de emprego favorável, cargos políticos e rendas elevadas, dessa

maneira, sua qualidade de vida é severamente prejudicada.

3 Implementação computacional

Para o auxílio na conferência da implementação computacional feita no *software R* do projeto “Detalhes e implementação dos modelos da classe Box-Cox simétrica e aplicações a dados de consumo alimentar” (FAPESP nº2019/02231-6), foi elaborado um pacote de teste com duas funções simples.

Esse pacote foi feito no *software R* versão 3.6.2 utilizando a coleção de ferramentas de desenvolvimento de pacotes *devtools* e foi denominado “marte3”. A primeira função do pacote (Função 1), estima valores de conjuntos discretos, ou seja, realiza interpolações. Essa função foi denominada “interpolação” e possui os argumentos: x_1 (primeiro valor de x), x_2 (segundo valor de x), x_k (valor ao qual se deseja descobrir o correspondente y_k), y_1 (primeiro valor de y) e y_2 (segundo valor de y). A segunda função (Função 2), realiza o cálculo do seno de um número em graus e em radianos e retorna o gráfico da função seno com o valor do seno no eixo y , fixo no intervalo entre -1 e 1, e o valor do ângulo de fase no eixo x , fixo no intervalo entre $-\pi$ e π . Essa função foi denominada “sen0” e depende apenas do argumento t (valor ao qual se quer descobrir o seno correspondente).

O pacote foi inserido na plataforma de hospedagem de códigos fonte *GitHub*. Para utilizar o pacote no *software R*, é necessário requerer a biblioteca *devtools* (`library(devtools)`), e depois instalar o pacote (`install_github("AnaFTatis/Marte3")`), por fim deve-se chamar a biblioteca “marte3” (`library("marte3")`). As funções podem ser acessadas através do endereço:

<https://github.com/AnaFTatis/Marte3>.

O projeto FAPESP nº2019/02231-6 propõe a criação de um pacote para *software R* que abrange novas distribuições para variáveis aleatórias assimétricas positivas, pertencentes à classe Box-Cox simétricas (BCS). A classe proposta é interessante para modelagem de regressão pois permite a interpretação dos parâmetros em termos de quantis, dispersão relativa e assimetria. Pelo fato de incluírem distribuições de cauda pesada, as distribuições Box-Cox simétricas truncadas são úteis para a modelagem de dados positivos assimétricos na presença de observações aberrantes.

O pacote engloba funções para as seguintes BCS's: Box-Cox Normal (BCN), Box-Cox Exponencial Potência (BCPE), Box-Cox Slash (BCSlash), Box-Cox t (BCT), Box-Cox Exponencial Dupla (BCDE), Box-Cox Slash Canônica (BCCSlash), Box-Cox Cauchy (BCCauchy) e Box-Cox Tipo II Logística (BC2Log). Para cada uma das distribuições foram elaboradas funções de densidade de probabilidade (d), de distribuição (p), de quantil (q) e de geração de números aleatórios (r).

Todas as funções supracitadas apresentam como argumentos os parâmetros de: escala relacionada à mediana μ ($\mu > 0$), dispersão relativa associada ao coeficiente de variação

aproximado, baseado nos percentis, σ ($\sigma > 0$) e assimetria, dada pela transformação potência para simetria ν ($-\infty < \lambda < \infty$). Para as funções densidade de probabilidade e distribuição existe ainda o argumento y , que representa o vetor de quantis (sendo $y > 0$), já para a função de quantil deve-se definir p , o vetor de probabilidades ($0 < p < 1$), e para a geração de números aleatórios, deve ser estabelecido a quantidade de números que se deseja gerar, n ($n > 1$).

Ademais, para as quatro funções (d , p , q e r) referentes às distribuições Box-Cox t e Box-Cox Exponencial potência o parâmetro τ é utilizado, associado à curtose da distribuição, o qual controla o decaimento da cauda, a fim de acomodar os pontos discrepantes, sendo $\tau > 0$. Na função Box-Cox Slash este parâmetro é denotado por q .

Em (FUMES, 2015) a classe Box-Cox simétricas foi detalhada, as distribuições BCT e BCSlash apresentam caudas mais pesadas do que as demais supracitadas neste estudo. Nas análises de simulação para verificação das funções elaboradas foram utilizadas em comparativo as funções provenientes do pacote *gamlss* (*Generalized Additive Models for Location Scale and Shape*).

Outra contribuição para a elaboração das funções desenvolvidas no *software R* do projeto FAPESP n°2019/02231-6, foi a descrição da documentação das funções supracitadas, detalhando o título, o uso, os parâmetros, os detalhes, as fórmulas, o alcance e as particularidades de cada uma. Esses documentos foram escritos em inglês, diretamente no *software R*.

Ademais, para assessorar no desenvolvimento do pacote foram feitas as derivadas de primeira e segunda ordem e derivadas mistas do logaritmo da função de verossimilhança em função dos parâmetros μ , σ , λ e τ . A função de verossimilhança pode ser entendida como uma medida de compatibilidade do modelo com o conjunto de dados estudado. Modelos estatísticos baseados em ajustes por máxima verossimilhança apresentam vantagens em relação a outros métodos utilizados tradicionalmente para estimação de parâmetros por conta da sua alta eficiência (NESI et al., 2013). O estimador de máxima verossimilhança é não enviesado, normalmente distribuído, consistente e flexível para diferentes transformações, essas propriedades assintóticas são convenientes para modelagem estatística em geral (WOJCIECHOWSKI et al., 2015). A solução dos estimadores aproximados de verossimilhança é obtida por meio da resolução do sistema de equações desenvolvido quando cada vetor escore é igualado a zero (FUMES, 2015).

O logaritmo da função de verossimilhança é usualmente mais utilizado do que a função de verossimilhança, isso se dá ao fato de que o logaritmo da função é naturalmente mais simples do que a própria função. A curva da a log-verossimilhança indica o valor máximo para a função, ou seja, o ponto mais provável de compatibilidade do modelo com o ajuste de dados. Para uma única observação de y pertencente a uma distribuição BCS (Box-Cox Simétrica) o logaritmo da função de verossimilhança é dado pela equação 1 (FUMES, 2015).

$$\ell(\mu, \sigma, \lambda) = (\lambda - 1) \log y - \lambda \log \mu - \log \sigma + \log r(z^2) - \log R\left(\frac{1}{\sigma|\lambda|}\right) \quad (1)$$

Em que $z = h(y; \mu, \sigma, \lambda)$ com $h(y; \mu, \sigma, \lambda)$ dado na equação 2 Quando $\lambda = 0$ o último termo de ℓ é anulado.

$$Z \equiv h(y; \mu, \sigma, \lambda) = \begin{cases} \frac{1}{\sigma\lambda} \left[\left(\frac{y}{\mu} \right)^\lambda - 1 \right], & \text{if } \lambda \neq 0 \\ \frac{1}{\sigma} \log \left(\frac{y}{\mu} \right), & \text{if } \lambda = 0 \end{cases} \quad (2)$$

Um estudo de 2007 que teve como objetivo detalhar as propriedades assintóticas dos estimadores dos parâmetros, utilizando um modelo de regressão com uma covariável para um processo de renovação (processo nos quais um item ou unidade, após a ocorrência de uma falha, é recolocado sob mesma condição), constatou que os estimadores de máxima verossimilhança foram satisfatórios na maioria dos casos (FOGO; NETO; DEMÉTRIO, 2007).

Segundo o princípio da verossimilhança, é preciso escolher o valor do parâmetro desconhecido que torna aquela amostra a mais provável, o uso desse princípio nos leva a um método de estimação pelo qual se adquirem os estimadores de máxima verossimilhança, que são em geral, muito precisos. Para verificação dos estimadores do pacote proposto pelo projeto FAPESP nº2019/02231-6, foram feitas simulações no *software R*, em que foram calculados os valores de viés e de erro quadrático médio.

Diz-se que um estimador é não-viesado ou não-viciado quando seu valor esperado é o próprio parâmetro que se pretende estudar, ou seja, para um estimador T e um parâmetro θ , $E(T) = \theta$. Quando o parâmetro é viesado mede-se o viés a partir da expressão: $V(T) = E(T) - \theta$, quanto mais próximo de zero for o valor viés, mais seu estimador se aproxima do parâmetro desejado (GUJARATI, 2019).

Em relação a aplicação a dados de consumo alimentar, foram feitos testes no *software R* para todos os modelos da classe BCS, utilizando a função densidade de probabilidade de cada modelo proposto, para todos os nutrientes dos bancos de dados neste estudo apresentados. Os resultados obtidos dos parâmetros estimados foram comparados com aqueles gerados pelo pacote *gamlss*.

Além disso foram calculados os valores de Akaike para comparação dos modelos, o Critério de Informação de Akaike (AIC) auxilia na seleção do modelo mais adequado dentre aqueles que foram ajustados, baseando-se na combinação das teorias da informação e da função de máxima verossimilhança com o conceito de entropia da informação. Assumindo as suposições de regressão não-linear, o AIC é definido pela equação 3, em que N é número de dados, K é a quantidade de parâmetros a serem estimados no modelo e SS é a soma de quadrados da distância vertical dos pontos da curva (MOTULSKY; CHRISTOPOULOS, 2004).

$$AIC = N * \ln \left(\frac{SS}{N} \right) + 2K \quad (3)$$

Por fim, foi produzido um tutorial simples de como criar e publicar um pacote desenvolvido no *software R*, deixando-o disponível para todos os usuários do programa. Este vídeo encontra-se no repositório da autora no *GitHub* e pode ser acessado através do endereço: <https://github.com/AnaFTatis/CriandoPacotesR>. Os passos a serem seguidos são: 1. Carregamento dos pacotes exidos, 2. Criação do pacote, 3. Exclusão de arquivos, 4. Inserção das funções no pacote, 5. Alteração da descrição do pacote, 6. Documentação do pacote, 7. Checagem do pacote, 8. Publicação do pacote via *GitHub*, 9. Instalação e carregamento do pacote e 10. Teste das funções.

4 Produção Científica

Quatro publicações originaram-se deste projeto, a pesquisa intitulada “Fatores Associados à Qualidade de Vida em Idosos de uma Cidade de Médio Porte do Interior de São Paulo”, apresentada na vigésima oitava edição do Simpósio Internacional de Iniciação Científica e Tecnológica da USP (SIICUSP), no dia 26 de novembro de 2020, o artigo nomeado *Factors related to functional capacity and quality of life for elderly people in a middle-sized city of São Paulo, Brazil*, publicado na revista *Health*, em dezembro de 2020, o artigo intitulado “Análise exploratória gráfica para dados assimétricos com presença de pontos discrepantes”, que no momento aguarda submissão na Revista Brasileira de Iniciação Científica (RBIC) e a pesquisa nomeada “Gráficos exploratórios de dados de consumo alimentar com distribuições assimétricas e *outliers*”, apresentada na vigésima nona edição do Simpósio Internacional de Iniciação Científica e Tecnológica da USP (SIICUSP), no dia 26 de outubro de 2021. As publicações foram feitas a partir da análise descritiva e inferencial dos bancos de dados supracitados.

Os resultados referentes sobre os fatores que influenciam na qualidade de vida dos idosos do estudo supracitado foram apresentados na vigésima oitava edição do SIICUSP, bem como tabelas com a caracterização da amostra (análise descritiva). Para o simpósio apresentado no ano de 2020, um vídeo foi gravado e pode ser acessado no endereço: <https://github.com/AnaFTatis/SIICUSP>. Para a elaboração dos *slides* foi feito o emprego do *software TeXStudio*, utilizando a classe de documento *beamer*. O mesmo programa serviu para escrita dos resumos em português e inglês, que se encontram no anexo II deste relatório.

As questões sobre qualidade de vida e capacidade funcional foram detalhadas com o objetivo de avaliar os fatores que influenciam na perda de capacidade funcional e piora na qualidade de vida da amostra de idosos pertencente ao estudo epidemiológico realizado no município de Botucatu (seção 2.1.2). Os resultados encontram-se no artigo supracitado publicado recentemente na revista *Health*. Na publicação, as classificações para o escore da escala de qualidade de vida de Flanagan e de outras duas escalas sobre capacidade funcional foram estudadas como variáveis respostas e os demais fatores como covariáveis.

Regressões logísticas modelaram os fatores que influenciam na perda da capacidade funcional e na piora da qualidade de vida dos idosos. Esse artigo contou com a colaboração da Profa. Dra. Giovana Fumes Ghantous, docente do departamento de Ciências Básicas da USP e do Prof. Dr. José Eduardo Corrente, membro do Escritório Central de Apoio à Pesquisa da Faculdade de Medicina de Botucatu da UNESP. O artigo pode ser acessado em <https://www.scirp.org/journal/paperabs.aspx?paperid=105949>.

O artigo intitulado "Análise exploratória gráfica para dados assimétricos com presença de pontos discrepantes", teve como objetivo a avaliação das ferramentas gráficas para análise exploratória de dados com distribuições assimétricas e pontos discrepantes, a fim de caracterizar a forma das distribuições para um conjunto de dados referente a um estudo nutricional em adultos da cidade do Rio de Janeiro, RJ, Brasil (seção 2.1.1). O artigo supracitado aguarda submissão na Revista Brasileira de Iniciação Científica (RBIC). Nesse artigo foram construídos histogramas, *box-plots* e *box-plots* ajustados. As distribuições de consumo alimentar apresentaram assimetrias acentuadas à direita com presença de pontos discrepantes, e os gráficos de *box-plot* e *box-plot* ajustados classificaram tais pontos de formas distintas. Esse artigo também contou com a colaboração da Profa. Dra. Giovana Fumes Ghantous, docente do departamento de Ciências Básicas da USP e do Prof. Dr. José Eduardo Corrente, membro do Escritório Central de Apoio à Pesquisa da Faculdade de Medicina de Botucatu da UNESP.

A pesquisa nomeada "Gráficos exploratórios de dados de consumo alimentar com distribuições assimétricas e *outliers*", apresentada na vigésima nona edição do Simpósio Internacional de Iniciação Científica e Tecnológica da USP (SIICUSP), ocorreu online de forma síncrona. Essa pesquisa teve como objetivo a avaliação de ferramentas gráficas para análise exploratória de dados de consumo alimentar, os quais, em geral, são assimétricos com presença de *outliers*. Do mesmo modo que na edição anterior, utilizou-se para a elaboração dos *slides* o *software* *TeXStudio*, utilizando a classe de documento *beamer*. Assim como o mesmo programa foi empregado para a escrita dos resumos em português e inglês, que se encontram no anexo II deste relatório.

5 Considerações parciais

Os objetivos do presente estudo incluem a aprendizagem, utilização e interpretação dos resultados obtidos mediante o uso de ferramentas computacionais, amplamente empregadas na área de estatística. Através da análise descritiva dos bancos de dados que compõe este estudo, realizada a fim de caracterizar as populações propostas, foi possível a familiarização com a linguagem de programação dos *softwares* R versão 3.6.2 e *SAS University Edition* e aplicação de técnicas estatísticas básicas.

Ambos os *softwares* supracitados são livres e possuem diversos pacotes, funções e/ou

procedimentos, os mesmos foram utilizados para aprendizagem de importação de arquivos externos, organização e transformação de dados brutos, utilização de pacotes, desempenho de análises descritivas, geração de diferentes gráficos e tabelas, elaboração de funções entre outros.

Em relação ao auxílio na implementação da programação de algoritmos e estudos de simulação, fez-se a elaboração de um pacote teste de funções, que foi importante para o aprendizado do mecanismo dessa prática no *software R*. Ademais, a produção da documentação das funções e a realização dos estudos de simulação demandaram muita pesquisa acerca das ferramentas computacionais, assim como da classe *Box-Cox Simétrica*.

Para editoração deste estudo foi feito uso do *software* livre *TeXStudio* (*MikTeX* versão 2.9.7351), para que isso fosse possível o conhecimento da linguagem *Tex* se fez necessário. Além da escrita de relatórios, o programa de edição de textos mencionado foi empregado para criação de apresentações de *slides*.

Também consta como objetivo do presente estudo o auxílio na elaboração de artigos científicos. A colaboração no artigo intitulado “Factors related to functional capacity and quality of life for elderly people in a middle-sized city of São Paulo, Brazil” publicado na revista *Health* no ano de 2020 foi o primeiro contato da autora com a elaboração deste tipo de publicação. Também foi elaborado outro artigo, intitulado “Análise exploratória gráfica para dados assimétricos com presença de pontos discrepantes”, que aguarda submissão na Revista Brasileira de Iniciação Científica (RBIC). Ambas as publicações foram relevantes na formação da estudante.

A partir deste projeto realizaram-se duas pesquisas apresentadas no Simpósio Internacional de Iniciação Científica e Tecnológica da USP nos anos de 2020 e 2021, e para a realização das mesmas foi necessária a aprendizagem do uso do *software OBS Studio*, esse programa é livre e de código aberto para gravação de vídeos e transmissões em tempo real. Essa experiência foi de grande valia, pois exigiu responsabilidade, dedicação e desenvoltura da autora.

O aprendizado de diferentes programas computacionais, a participação na escrita de artigos e a realização de apresentação em conferência foram importantes não apenas para execução deste trabalho, mas para o desenvolvimento profissional e técnico da autora.

Pirassununga, 30 de janeiro de 2021.



Aluna: Ana Flávia G. S. L. Tatis



Orientadora: Profa. Dra. Giovana Fumes Ghantous

Referências

- ALMEIDA, M. *Curso Essencial de Lógica de Programação*. São Paulo: Digerati Books, 2008.
- AMADO, T. C. F.; ARRUDA, I. K. G. de. Hipertensão arterial no idoso e fatores de risco associados. *Rev. Brasileira de Nutrição Clínica*, São Paulo, v. 19, n. 2, p. 94–99, 2004.
- ASSUMPÇÃO, D. de et al. Diferenças entre homens e mulheres na qualidade da dieta: estudo de base populacional em campinas, são paulo. *Ciência e Saúde Coletiva*, Rio de Janeiro, v. 22, n. 2, p. 347–358, 2017.
- CARDOSO, M. R. V. *Alimentação e Estado Nutricional de Idosos Residentes em Instituições Asilares de Dois Municípios do Sul de Minas Gerais*. 113 p. Dissertação (Mestrado) — UFLA, Lavras, 2004.
- DANTAS, R. A. S.; GOIS, C. F. L.; SILVA, L. M. da. Utilização da versão adaptada da escala de qualidade de vida de flanagan em pacientes cardíacos. *Rev. Latino-Am. Enfermagem*, Ribeirão Preto, v. 13, n. 1, p. 15–20, 2005.
- FIGUEIREDO, M. do L. F. et al. As diferenças de gênero na velhice. *Rev. bras. enferm.*, Brasília, v. 60, n. 4, p. 422–427, 2007.
- FOGO, J. C.; NETO, F. L.; DEMÉTRIO, C. G. B. Propriedades assintóticas dos estimadores de máxima verossimilhança em um processo de renovação com distribuição weibull. *Tend. Mat. Apl. Comput.*, v. 8, n. 3, p. 401–412, 2007.
- FUMES, G. *Modelos Box-Cox simétricos e aplicações a dados nutricionais*. 99 p. Tese (Doutorado) — USP, São Paulo, 2015.
- GARCIA, N. *Lógica de Programação de Computadores*. 1. ed. Joinville: Clube de Autores, 2011.
- GUJARATI, D. *Econometria: Princípios, teorias e aplicações práticas*. [S.l.]: Saraiva, 2019.
- HUBERT, M.; VANDERVIEREN, E. An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis*, v. 12, n. 52, p. 5186–5201, 2008.
- KATZ, S. et al. Studies of illness in the aged. the index of adl: a standardized measure of biological and psychosocial function. *JAMA*, v. 185, n. 12, p. 914–919, 1963.
- LAWTON, M.; BRODY, E. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*, v. 9, n. 3, p. 179–186, 1969.
- LIMA, L. C. V. de; BUENO, C. M. L. B. Envelhecimento e gênero: A vulnerabilidade de idosas no brasil. *Saúde e Pesquisa*, v. 2, n. 2, p. 273–280, 2009.
- MORAIS, K. C. C. de Almeida Antunes de. *Relação Entre Indicadores de Obesidade, Desempenho Muscular e Capacidade Funcional de Idosas Comunitárias*. 66 p. Dissertação (Mestrado) — UFMG, Belo Horizonte, 2019.
- MORETTIN, P. A.; BUSSAB, W. de O. *Estatística Básica*. 6. ed. São Paulo: Saraiva, 2010.

- MORETTIN, P. A.; SINGER, J. M. *Introdução à Ciência de Dados: Fundamentos e aplicações*. Versão parcial preliminar. São Paulo: Departamento de Estatística, USP, 2019.
- MOTULSKY, H.; CHRISTOPOULOS, A. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford, New York: Oxford University Press, 2004.
- NESI, C. N. et al. Verossimilhança na seleção de modelos para predição espacial. *Rev. Bras. Ciênc. Solo*, v. 37, n. 2, p. 352–358, 2013.
- PEREIRA, I. F. da S. *Expectativa de vida livre de fatores de risco relacionados ao estilo de vida na população brasileira*. 138 p. Tese (Doutorado) — UFRN, Natal, 2019.
- RODRIGUES, R. As políticas linguísticas educacionais e a problemática do paradoxo do acesso às práticas letradas e da inclusão social: Interpretações à luz dos estudos da linguística aplicada. *Revista de linguística*, Florianópolis, v. 13, n. 3, p. 1444–1452, 2016.
- SANTOS, C. M. Prevalência da obesidade, obesidade sarcopênica e fatores associados em idosos comunitários: um estudo da rede fibra. *Fisioter. mov.*, Curitiba, v. 30, n. 1, p. 161–169, 2017.
- SCHMIDT, M. I. et al. Prevalência de diabetes e hipertensão no Brasil baseada em inquérito de morbidade auto-referida. *Rev. Saúde Pública*, São Paulo, v. 43, n. 2, p. 74–82, 2009.
- SILVA, M. V. M. da; TOMAZ, A. F. Análise da qualidade de vida e capacidade funcional de idosos. *Revista Tema*, v. 18, n. 28/29, p. 3–18, 2017.
- VECCHIA, R. D. et al. Qualidade de vida na terceira idade: um conceito subjetivo. *Rev. bras. epidemiol.*, São Paulo, v. 8, n. 3, p. 246–252, 2005.
- WOJCIECHOWSKI, J. C. et al. Log-verossimilhança combinada para comparação de estruturas de continuidade espacial em floresta estacional decidual. *Rev. Floresta*, v. 45, n. 4, p. 797–808, 2015.

Anexos

Anexo I - Questionários

Atividades de vida diária (AVD)

O questionário que avalia as “Atividades de Vida Diária” é composto pelas seguintes questões:

1. O(a) senhor(a) não recebe ajuda ou somente recebe ajuda para lavar uma parte do corpo?
2. O(a) senhor(a) pega as roupas e se veste sem qualquer ajuda, exceto para amarrar os sapatos?
3. O(a) senhor(a) vai ao banheiro, usa o banheiro, se veste e retorna sem qualquer ajuda (andador/bengala)?
4. O(a) senhor(a) consegue deitar na cama, sentar-se na cadeira e levantar sem ajuda (andador ou bengala)?
5. O(a) senhor(a) controla completamente urina e fezes?
6. O(a) senhor(a) come sem ajuda (exceto para cortar carne e passar manteiga no pão)?

Atividades Instrumentais de Vida Diária (AIVD)

O questionário que avalia as “Atividades Instrumentais de Vida Diária (AIVD)” é composto pelas seguintes questões:

1. O(a) senhor(a) consegue usar o telefone?
2. O(a) senhor(a) consegue ir a lugares distantes, usando algum transporte, sem necessidade de planejamento especial?
3. O(a) senhor(a) consegue fazer compras?
4. O(a) senhor(a) consegue preparar suas próprias refeições?
5. O(a) senhor(a) consegue arrumar a casa?
6. O(a) senhor(a) consegue fazer trabalhos domésticos como pequenos reparos?
7. O(a) senhor(a) consegue lavar e passar a própria roupa?
8. O(a) senhor(a) consegue tomar seus remédios na dose e no horário corretos?
9. O(a) senhor(a) consegue cuidar de suas finanças?

Escala Flanagan de Qualidade de Vida

Segundo Dantas, Gois e Silva (2005), essa escala foi desenvolvida na década de 1970 pelo psicólogo americano John Flanagan e dispõe dos seguintes tópicos:

1. Objetos que trazem conforto em casa, alimentos, conveniências, segurança financeira.
2. Saúde - estar fisicamente apto e vigoroso.
3. Relacionamento com pais, irmãos e outros parentes - comunicação, visitas, ajuda.
4. Ter e criar filhos.
5. Relacionamento próximo com o cônjuge ou outro significativo.
6. Amigos próximos.
7. Ajudar e encorajar outras pessoas, oferecer-se como voluntário, dar conselhos.
8. Participar de organizações e assuntos públicos.
9. Aprendizagem - frequentar a escola, melhorar sua compreensão e obter conhecimentos adicionais.
10. Compreendendo a si mesmo - conhecer seus recursos e limitações - saber do que se trata a vida.
11. Trabalho - emprego ou em casa.
12. Expressar-se de forma criativa.
13. Socialização - conhecer pessoas, fazer coisas, ir a festas, etc.
14. Ler, escutar músicas, ou entretenimentos visuais.
15. Participar de atividades recreativas.

Anexo II - SIICUSP

Resumo



FATORES ASSOCIADOS À QUALIDADE DE VIDA EM IDOSOS DE UMA CIDADE DE MÉDIO PORTE DO INTERIOR DE SÃO PAULO

Ana Flávia Giacondino Soligo Lezcano Tatis

Giovana Fumes Ghantous

Faculdade de Zootecnia e Eng. de Alimentos/Universidade de São Paulo

anatatis@usp.br

Objetivo

O objetivo deste estudo foi verificar a associação entre a qualidade de vida de idosos avaliada através da escala Flanagan, variáveis demográficas e algumas morbididades referidas pelos entrevistados.

Métodos e Procedimentos

Foram utilizadas informações referentes a 334 idosos provenientes de um estudo epidemiológico realizado na cidade de Botucatu, São Paulo, Brasil, no ano de 2011. O questionário da escala de qualidade de vida de Flanagan (EQVF) e as variáveis: gênero, faixa etária, estado civil, renda, escolaridade, Índice de Massa Corpórea (IMC), hipertensão, diabetes e doença cardíaca foram selecionados para estudo. Para a EQVF, o ponto de corte definido para classificação de qualidade de vida “boa” ou “ruim” foi a mediana da soma das pontuações relativas ao questionário. Para estudar os fatores associados à qualidade de vida foram realizados testes qui-quadrado, o nível de significância adotado foi de 5%. Os *softwares* R versão 3.6.2. e *SAS University Edition* foram usados para realização das análises estatísticas.

O Comitê de Ética em Pesquisa da Faculdade de Medicina de Botucatu da Universidade Estadual Paulista aprovou a realização desta pesquisa, protocolo nº 3111/2009.

Resultados

A tabela 1 apresenta as variáveis que obtiveram associação com a EQVF dos entrevistados.

Tabela 1: Associação entre variáveis socioeconômicas, algumas morbididades e a qualidade de vida dos idosos. Botucatu, 2011.

Variáveis	Categorias	Classificação EQVF		p-valor
		QV Boa n (%)	QV Ruim n (%)	
Gênero	Masculino	73 (43,71)	54 (32,34)	0,0322
	Feminino	94 (56,29)	113 (67,66)	
Faixa etária	Até 75 anos	124 (74,25)	105 (62,87)	0,0251
	Acima de 75 anos	43 (25,75)	62 (37,13)	
Estado civil	Casados	122 (73,05)	72 (43,11)	<0,0001
	Não Casados	45 (26,95)	95 (56,89)	
Renda	Até um SM*	38 (26,95)	68 (44,74)	0,0015
	Acima de um SM*	103 (73,05)	84 (55,26)	
Escolaridade	Até Ensino fund.	118 (70,66)	135 (80,84)	0,0300
	Acima Ensino fund.	49 (29,34)	32 (19,16)	
Hipertensão	Sim	81 (48,50)	104 (62,28)	0,0113
	Não	86 (51,50)	63 (37,72)	

*SM - Salário Mínimo

Observou-se que a qualidade de vida dos idosos está associada ao seu gênero, faixa etária, estado civil, renda, escolaridade e hipertensão. Diabetes, IMC e doença cardíaca não apresentaram associação com a EQVF.

Conclusões

Os resultados encontrados ressaltam pontos problemáticos que persistem na nossa sociedade. Espera-se que este estudo contribua para maior diligência às questões dos idosos e problemas sociais relacionados.

Referências Bibliográficas

- DANTAS, R. A. S.; GÓIS, C. F. L.; SILVA, L. M. da. Utilização da versão adaptada da escala de qualidade de vida de Flanagan em pacientes cardíacos. *Rev. Latino-Am. Enfermagem*, Ribeirão Preto, v.13, n.1, p.15–20, 2005.
- MORETTIN, P. A.; BUSSAB, W. de O. *Estatística Básica*. 6. ed. São Paulo: Saraiva, 2010.

FACTORS ASSOCIATED WITH THE QUALITY OF LIFE IN ELDERLY PEOPLE FROM A MID SIZED CITY IN THE INTERIOR OF SÃO PAULO

Ana Flávia Giacondino Soligo Lezcano Tatis

Giovana Fumes Gbantous

Faculty of Animal Science and Food Engineering/ University of São Paulo

anatatis@usp.br

Objective

The aim of this research was to verify an association between the quality of life from elderly people assessed by the Flanagan scale, demographic variables and some morbidities related by the interviewees.

Materials and Methods

This study used information about 334 elderly people of one epidemiological research realized on Botucatu city, São Paulo, Brazil, in the year of 2011. The questionnaire about the Flanagan quality of life scale (FQOLS) and the variables: gender, age group, civil status, income, schooling, Body Mass Index (BMI), hypertension, diabetes and heart disease were selected for study. To FQOLS, the break point defined for the classification of life quality in "good" or "poor" was the median from the sum of the scores concerning to the questionnaire. To study the factors associated to quality of life, chi-square tests were performed and the significance level set was 5%. The softwares R version 3.6.2. and SAS University Edition were used to accomplish the statistical analysis.

This research was approved by the Research Ethics Committee of the Botucatu School of Medicine of Paulista State University, protocol number 3111/2009.

Results

The table 1 presents the variables that revealed association with the interviewees FQOLS.

Table 1: Association between socioeconomic variables, some morbidities and the quality of life from elderly people. Botucatu, 2011.

Variables	Categories	Classification FQOLS		p-value
		Good QL n (%)	Poor QL n (%)	
Gender	Male	73 (43.71)	54 (32.34)	0.0322
	Female	94 (56.29)	113 (67.66)	
Age group	Up to 75 years old	124 (74.25)	105 (62.87)	0.0251
	Over 75 years old	43 (25.75)	62 (37.13)	
Civil status	Married	122 (73.05)	72 (43.11)	<0.0001
	Not married	45 (26.95)	95 (56.89)	
Income	Up to one MW(BR)	38 (26.95)	68 (44.74)	0.0015
	Over one MW(BR)	103 (73.05)	84 (55.26)	
Schooling	Up to elementary sch.	118 (70.66)	135 (80.84)	0.0300
	Over elementary sch.	49 (29.34)	32 (19.16)	
Hypertension	Yes	81 (48.50)	104 (62.28)	0.0113
	No	86 (51.50)	63 (37.72)	

*MW(BR) - Brazilian Minimum Wage

It was observed that the quality of life from elderly people is associated with their gender, age group, civil status, income, schooling and hypertension. Diabetes, BMI and heart disease were not associated with FQOLS.

Conclusions

The results found highlight problematic points that persist in our society. It is expected that this research contribute to greater diligence on the issues of the elderly people and the related social problems.

References

- DANTAS, R. A. S.; GÓIS, C. F. L.; SILVA, L. M. da. Utilização da versão adaptada da escala de qualidade de vida de Flanagan em pacientes cardíacos. *Rev. Latino-Am. Enfermagem*, 2005, v.13, n.1, p.15-20.
- MORETTIN, P. A.; BUSSAB, W. de O. *Estatística Básica*. 6th. ed. São Paulo: Saraiva, 2010.

GRÁFICOS EXPLORATÓRIOS DE DADOS DE CONSUMO ALIMENTAR COM DISTRIBUIÇÕES ASSIMÉTRICAS E *OUTLIERS*

Ana Flávia Giacondino Soligo Lezcano Tatis

Giovana Fumes Ghantous

Faculdade de Zootecnia e Eng. de Alimentos/Universidade de São Paulo

anatatis@usp.br

Objetivo

Este estudo teve como objetivo a avaliação de ferramentas gráficas para análise exploratória de dados de consumo alimentar, os quais, em geral, são assimétricos com presença de *outliers*.

Métodos e Procedimentos

Foram utilizadas informações referentes a 368 idosos provenientes de um estudo epidemiológico realizado na cidade de Botucatu, São Paulo, Brasil, no ano de 2011. O instrumento utilizado para a coleta de dados foi o primeiro recordatório 24 horas (R24h). As variáveis referentes ao consumo de: carboidratos, zinco, ferro, niacina (vitamina B3) e piridoxina (vitamina B6) foram selecionadas. Para comparação, na análise exploratória dos dados utilizou-se o histograma, o gráfico de caixa (*box-plot*) e o gráfico de caixas (*box-plot*) ajustado. De acordo com a literatura, as recomendações sobre adequação de consumo alimentar são distintas para homens e mulheres, por este motivo as análises foram realizadas por gênero. Além da análise exploratória gráfica, foi feita uma análise descritiva das variáveis supracitadas. Os critérios de comparação na área nutricional foram realizados mediante a consulta de tabelas com valores de referências de *EAR* (*Estimated Average Requirement*), a média da distribuição referente à necessidade estimada de um nutriente. O *software* utilizado para obtenção dos gráficos e das medidas resumo foi o *R*, versão 4.1.0. O Comitê de Ética em Pesquisa da Faculdade de Medicina de Botucatu da Universidade Estadual Paulista aprovou a realização desta pesquisa, protocolo nº 3111/2009.

Resultados

A Tabela 1 apresenta as medidas descritivas referentes ao consumo dos nutrientes selecionados para este estudo. Nota-se que as médias e medianas calculadas, resultaram em valores mais elevados para indivíduos do sexo masculino; para ambos, a maior amplitude de consumo foi na quantidade diária de carboidratos. Além disso, observou-se que a maior parte dos valores médios se encontravam distantes dos valores medianos, o que sugere a existência de pontos aberrantes e de distribuições assimétricas. Nas Figuras 1 e 2 foram apresentadas as ferramentas gráficas utilizadas para representar as distribuições referentes ao consumo de nutrientes pela população de idosos. Como esperado, as caixas desenhadas para os dois tipos de diagramas de caixas, assim como as medidas de posição central, foram exatamente iguais nos dois gráficos, as diferenças encontradas foram nos tamanhos dos “bigodes” dos gráficos de caixa, os quais foram distintos devido a quantidade de pontos aberrantes determinada por cada uma das ferramentas gráficas utilizadas. Nas Figuras 1 e 2 também percebeu-se a existência de assimetrias acentuadas à direita nas distribuições de consumo dos nutrientes analisados. Mediante a análise dos “bigodes”, nota-se que os gráficos de caixa ajustados representaram a cauda esquerda de maneira mais acurada. Os consumos medianos de carboidratos, zinco, ferro, niacina e piridoxina dos idosos, foram superiores aos valores recomendados pelas *EARs*. Neste estudo apenas foram feitas comparações diretas com os valores encontrados para o consumo mediano dos nutrientes e seus respectivos valores de *EAR*, com a finalidade de complementação da interpretação dos resultados obtidos por meio do uso das ferramentas gráficas.

Tabela 1: Medidas descritivas do consumo de nutrientes por gênero, Botucatu, 2011.

Nutrientes	Homens							Mulheres						
	mín	média	mediana	DP	CV	máx	EAR	mín	média	mediana	DP	CV	máx	EAR
carboidrato (g)	62,0	257,2	238,5	148,4	58	1797,7	100	18,8	222,9	203,1	94,9	43	1021,6	100
zinco (mg)	2,3	13,1	12,0	7,4	57	85,2	9,4	1,8	10,4	9,6	5,1	49	44,8	6,8
ferro (mg)	1,4	17,1	15,1	12,4	73	132,2	6	2,1	13,9	12,0	8,7	63	97,3	5
niacina (vitamina B3) (mg)	1,8	24,2	20,6	18,2	75	258,8	12	3,4	19,8	17,5	11,4	57	83,8	11
piridoxina (vitamina B6) (mg)	0,3	2,1	1,8	2,1	100	34,0	1,4	0,2	1,7	1,5	0,8	49	6,5	1,3

Mín = mínimo, DP = desvio padrão, Máx = máximo, CV (%) = coeficiente de variação em porcentagem.

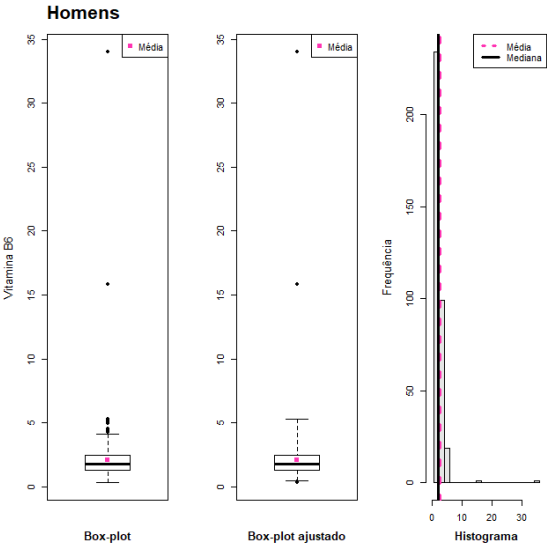


Figura 1: Gráficos referentes ao consumo de piridoxina por homens. Botucatu, 2011.

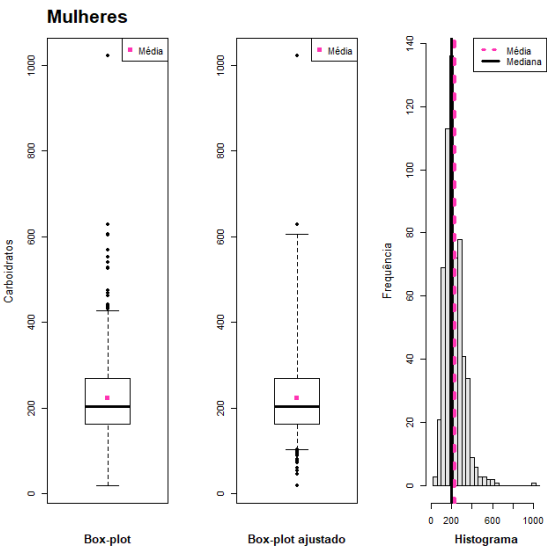


Figura 2: Gráficos referentes ao consumo de carboidratos por mulheres. Botucatu, 2011.

Conclusões

Verificou-se que as ferramentas gráficas são indispensáveis para análise descritiva de um conjunto de dados. Especialmente quando se tratam de distribuições assimétricas, os *box-plot* ajustados podem expressar com mais precisão a presença de pontos discrepantes, sendo desse modo, uma alternativa interessante para casos nos quais se deseja detectar pontos que estão muito distantes do conjunto de dados e que são dignos de melhor investigação.

Referências Bibliográficas

HUBERT, M.; VANDERVIERENB, E. An adjusted box-plot for skewed distributions. **Computational Statistics and Data Analysis**, v. 52, n. 12, p. 5186-5201, Agosto 2008. ISSN 0167-9473.

MORETTIN, P. A.; BUSSAB, W. de O. **Estatística Básica**. 6. ed. São Paulo: Saraiva, 2010.

PADOVANI, R. M. et al. Dietary reference intakes: application of tables in nutritional studies. **Rev. Nutr.**, Campinas, v. 19, n. 6, p. 741-760, Nov/Dez 2006. ISSN 1678-9865.



EXPLORATORY GRAPHS OF FOOD CONSUMPTION DATA WITH ASYMMETRIC DISTRIBUTIONS AND OUTLIERS

Ana Flávia Giacondino Soligo Lezcano Tatis

Giovana Fumes Ghantous

Faculty of Animal Science and Food Engineering/University of São Paulo

anatatis@usp.br

Objective

This study aimed to evaluate graphical tools for exploratory analysis of food consumption data, which, in general, are asymmetrical with presence of outliers.

Materials and Methods

Information referring to 368 older adults from an epidemiological study conducted in the city of Botucatu, São Paulo, Brazil, in 2011, was used. The instrument adopted in the data collection was the first 24-hour recall (24hR). The following variables were selected: carbohydrates, zinc, iron, niacin (vitamin B3) and pyridoxine (vitamin B6). For comparison, the histogram, the box-plot and the adjusted box-plot were used in the exploratory data analysis. According to the literature, the recommendations on adequate food intake are different for men and women, for this reason the analyses were produced by gender. In addition to the graphical exploratory analysis, a descriptive analysis of the above-mentioned variables was performed. The comparison criteria in the nutritional area were made by consulting tables with reference values of EAR (Estimated Average Requirement), the mean of the distribution referring to the estimated need of a nutrient. The software used to obtain the graphs and summary measures was R, version 4.1.0. This research was approved by Research Ethics Committee of the Botucatu Medical School of São Paulo State University - Brazil, protocol n° 3111/2009.

Results

The Table 1 presents the descriptive measures regarding the intake of the nutrients selected for this study. It should be noted that the calculated means and medians resulted in higher values for male individuals; for both, the greatest range of consumption was in the daily amount of carbohydrates. Furthermore, it was observed that most of the mean values were far from the median values, which suggests the existence of outliers and asymmetric distributions. Figures 1 and 2 presented the graphic tools used to represent the distributions of nutritional intake by the elderly population. As expected, the boxes drawn for the two types of *box-plots*, as well as the measures of central tendency, were exactly the same in the two graphs, the differences found were in the sizes of the whiskers of the box-plots, which were distinct due to the amount of outliers determined by each of the graphic tools used. In the Figures 1 and 2 it was also noticed the existence of sharp asymmetries to the right in the distributions of the intake of nutrients analyzed. Upon analysis of the whiskers, it is noted that the adjusted box-plots represented the left tail more accurately. The median intake of carbohydrates, zinc, iron, niacin and pyridoxine of the elderly were higher than the values recommended by EARs. In this study, only direct comparisons were made with the values found for the median consumption of nutrients and their respective values of EAR, in order to complement the interpretation of the results obtained through the use of graphic tools.

Table 1: Descriptive measures of nutrients intake by gender, Botucatu, 2011.

Nutrients	Men							Women						
	min	mean	median	SD	CV	max	EAR	min	mean	median	SD	CV	max	EAR
carbohydrate (g)	62.0	257.2	238.5	148.4	58	1797.7	100	18.8	222.9	203.1	94.9	43	1021.6	100
zinc (mg)	2.3	13.1	12.0	7.4	57	85.2	9.4	1.8	10.4	9.6	5.1	49	44.8	6.8
iron (mg)	1.4	17.1	15.1	12.4	73	132.2	6	2.1	13.9	12.0	8.7	63	97.3	5
niacin (vitamin B3) (mg)	1.8	24.2	20.6	18.2	75	258.8	12	3.4	19.8	17.5	11.4	57	83.8	11
pyridoxine (vitamin B6) (mg)	0.3	2.1	1.8	2.1	100	34.0	1.4	0.2	1.7	1.5	0.8	49	6.5	1.3

Min = minimum, SD = standard deviation, Max = maximum, CV (%) = coefficient of variation.

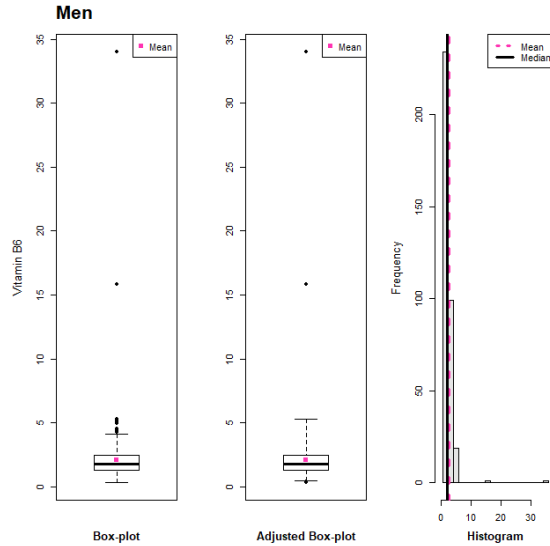


Figure 1: Graphs referring to pyridoxine consumption by men. Botucatu, 2011.

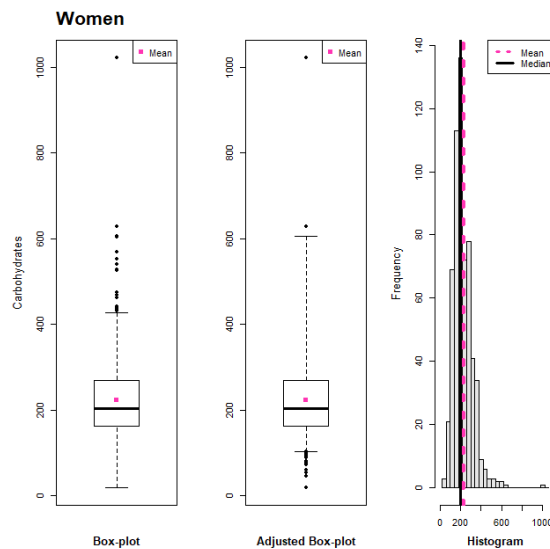


Figure 2: Graphs referring to carbohydrates consumption by women. Botucatu, 2011.

Conclusions

It was found that graphic tools are essential for descriptive analysis of a data set. Especially when it refers to asymmetric distributions, adjusted box-plot can express more accurately the presence of outliers, to cases that want to detect points that are very far from the data set and which are worthy of better investigation.

References

- HUBERT, M.; VANDERVIERENB, E. An adjusted box-plot for skewed distributions. *Computational Statistics and Data Analysis*, v. 52, n. 12, p. 5186-5201, August 2008. ISSN 0167-9473.
- MORETTIN, P. A.; BUSSAB, W. de O. *Estatística Básica*. 6. ed. São Paulo: Saraiva, 2010.
- PADOVANI, R. M. et al. Dietary reference intakes: application of tables in nutritional studies. *Rev. Nutr.*, Campinas, v. 19, n. 6, p. 741-760, Nov/Dec 2006. ISSN 1678-9865.